UNIVERSIDAD DE DEUSTO

MULTI VARIABLE ANALYSIS OF GAZE AND INTERACTION FOR REAL TIME USER PROFILING IN COGNITIVE INTELLIGENT THERAPIES BASED ON SERIOUS GAMES

Tesis doctoral por

MAITE FRUTOS PASCUAL

Directora

BEGOÑA GARCÍA-ZAPIRAIN SOTO

Bilbao, August 2016

UNIVERSIDAD DE DEUSTO

MULTI VARIABLE ANALYSIS OF GAZE AND INTERACTION FOR REAL TIME USER PROFILING IN COGNITIVE INTELLIGENT THERAPIES BASED ON SERIOUS GAMES

Tesis doctoral por MAITE FRUTOS PASCUAL
dentro del Programa de Doctorado en Ing. Informática y Telecomunicación
Directora BEGOÑA GARCÍA-ZAPIRAIN SOTO

La doctoranda La directora

Bilbao, August 2016

Para aita, ama y Jess

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements.

MAITE FRUTOS PASCUAL
August 2016

Agradecimientos

Gracias a *Deiker*, por darme la oportunidad de completar mis estudios de doctorado en la Universidad de Deusto, y gracias a mi directora, *Begoña Garcia-Zapirain* por abrirme las puertas al mundo de la investigación, sin tu primera oportunidad nada hubiese sido posible. Gracias también a todas las familias que de manera desinteresada accedieron a formar parte de esta tesis a través de los estudios aquí incluidos, y mi especial agradecimiento a todos los niños que participaron en ellos. Fue un placer compartir un ratito con cada uno de vosotros. Y por supuesto mi eterna gratitud a *Mari Jose Pecharromán* y *Patricia Clemente* del Colegio Vizcaya y *Angela Magaz* del grupo Albor Cohs, de nuevo mil gracias por vuestra disposición y ayuda desinteresada. Gracias también a todos los que me acogieron y me ayudaron de una manera u otra durante mi estancia en la Universidad de Wolverhampton (Reino Unido).

Gracias a mis ex-compañeros de laboratorio, los que están y los que estuvieron (*Álvaro*, *Fernando*, *Nuria*, *Yoli*, *Zelai*, *Gonzalo*, *Jose*, *Josetxu*, *Alain*, *Kattalin*, *Maria*). sin duda por años que pasen y laboratorios que visite no podré nunca encontrar mejores. Estos años de txiki-vueltas han hecho mucho más llevadero el proceso… *Evidas*, gracias por vuestro apoyo.

Gracias a mis amigos, los de siempre, por todos los momentos ajenos al papel, el ordenador y los papers, ¡qué sería de la vida sin ellos!. Gracias *Maite*, *Gemma*, *Cynthia*, *Patricia*, *Amaia* (y ahora también *Laura*) por seguir creciendo conmigo, por todos los ratitos juntas y por seguir sacandome esas risas tan necesarias que le hacen a uno olvidarse de todo. Gracias Ana por tantos paseos por Deusto y por escucharme en cada paso. Gracias Xabi y Maite por estar siempre ahí aunque sea en la distancia.

Y por último y más importante, gracias a mi familia, mi roca, por estar siempre ahí, confiar y creer en mí más que yo misma. Gracias *aita* por tantas conversaciones (y las que nos quedan), por ser siempre honesto conmigo, por saber enseñarme a intentar sacar lo mejor de mi. Gracias *ama* por estar siempre siempre orgullosa. Gracias *Jess* por todos nuestros ratitos y sobre todo, por los domingos por la tarde. Gracias *aitite* y *amama* por siempre presumir de nietos. Y por supuesto, gracias *René* por ser mi otra roca, por cambiarme la visión de la vida y enseñarme que si hay ganas uno es capaz de cualquier cosa (¡y también por leerte la tesis!). Al final me convertí en pato de Central Park.

Gracias en definitiva a todas esas personas que han formado parte de mi vida, de una manera u otra, durante estos años.

Abstract

Popularity of video games is by no means a new trend in European households. The health of the video and computer games industry, together with the variety of genres and technologies available, mean that videogame concepts and programmes are being applied in numerous different disciplines. This growth in games as a mainstream entertainment has raised the question of how to take advantage of this digital games trend for educational purposes.

Generating games that are flexible and customizable enough to adapt themselves autonomously to different users' needs is a key advance in the development of game-based learning experiences. These adaptive digital educational games able to customize user interaction according to the individual performance of the player can enhance gameplay experience, making it more unique and personal. New game developments inside this trend provide systems with an efficient way of learning based

on the users themselves, personalizing their experience with the system, potentially increasing their effects.

Traditionally, to get that adaptivity in serious games and simulations, developers tend to use the approach of balancing players ability and skills with the game challenge level. This may be further extended to focus on other features such as player's learning mental process, or the use of objective biofeedback recordings while players are interacting with the system.

One of this potential solutions for objective evaluation of players' state is the use of gaze interaction and visual attention. Even though the use of eyetracking data in the serious games field has increased recently, it is mainly used as an additional input system for the game. This dissertation is focused on the analysis of gaze patterns in junction with the user interaction with the system in order to generate intelligent profiles of user behaviours enabling the development of adaptive algorithms and game engines that allow the complete customization of the gameplay. This user profiles will give a new insight into the attention patterns of users, allowing the adaptation of game contents to their visual attention patterns and interaction techniques.

This dissertation is presented as a PhD by publication contribution. It is presented as a collection of four articles exploring different aspects of intelligent serious games. It explores the effects of gaze pattern behaviours and player interaction while using a serious games approach. This project was undertaken to design a profiling framework for serious games development.

Table of contents

List of figures

List of tables

“Begin at the beginning, - the King said gravely, - and go on till you come to the end: then stop."

Alice in Wonderland and through the Looking Glass

Lewis Carroll

1 Introduction

Popularity of video games is by no means a new trend in European households. Nearly 80% of children aged between 6 to 12 years old play videogames in Spain, according to the report published by the *Interactive Software Federation of Europe* [isf 15]. This growth in games as a mainstream entertainment has raised the question of how to take advantage of this digital games trend for educational purposes [Plass 15].

Several studies suggest that the future of pedagogy will inevitably be linked to the proposal of combined play and learning so as to promote creativity in future generations [Samuelsson 08]. The main assumption under this idea is that digital games enable learners to learn in an engaging, motivating and pleasant way [Law 12]. One of the promising ways that games can be used for fostering education is by adapting themselves to each child individually [Andersen 12]. Players have different learning abilities and training needs, however, serious games do not usually take player individuality into account. This may lead to the generation of stereotyped training conditions, that affect the replay value of this games [Lopes 11].

Generating games that are flexible and customizable enough to adapt themselves autonomously to different users' needs is a key advance in the development of game-based learning experiences [Göbel 10]. These adaptive digital educational games able to customize user interaction according to the individual performance of the player can enhance gameplay experience, making it more unique and personal [Law 12, Lopes 11]. New game developments inside this trend provide systems with an efficient way of learning based on the users themselves, customizing and personalizing their experience with the system, which may increase their potential effects [Tobail 11].

Adaptivity and adaptability are terms frequently used in the literature. Adaptability is the personal choice of users during the learning experience, where they can modify certain parameters to adapt the environment to their needs, while the adaptivity refers to the capacity of the system to modify the environment based on certain features [Burgos 07, Akbulut 12].

Traditionally to get that adaptivity in serious games and simulations, developers tend to use the approach of balancing players ability and skills with the game challenge level, which may be further extended to focus on other features such as player's learning mental process [Lopes 11]. This tailoring of learning experiences and e-systems to final users' needs can be accomplished based on several distinct criteria: human factors [Brusilovsky 04, Paterno 99], cognitive styles [Calcaterra 05, Chen 08], learning styles, prior knowledge [Greene 10], anxiety measurements, achievement motivation or self-efficacy, among others [Park 03, Mampadi 11].

It is necessary to track and to evaluate the learner's knowledge state to get the basis for the decisions to be made. This dissertation is focused on performing that evaluation based not only in user interaction but also in their visual attention performance. Research using eye tracking provides an opportunity to test theories about multimedia learning, information processing and visual attention[Mayer 10]. Visual attention is the control mechanism that selects meaningful inputs and suppresses those with lower importance [Sundstedt 13].

Inferring relevant feedback from gaze is done by capturing eye movements with proactive systems that may help to analyse user behaviour [Muir 12]. Moreover, following the eye-mind hypothesis put forth by Carpenter in 1980, there is a close link between the direction of the human gaze and the focus of attention [Just 80], provided that the visual environment in front of the eyes is pertinent to the task that we want to study [Hyönä 10]. Eye tracking sensors collect information about the location and duration of an eye fixation within a specific area on a computer monitor. The study of gaze data can be suitable for providing an insight into gaze behaviour of players, thus assisting them in obtaining tailored software developments based on their visual attention patterns.

The use of eye tracking can change the gaming experience for all players, benefiting themselves from the incorporation of a new interactive, non-intrusive and attention-aware tool for customizing gaming experiences [Isokoski 09]. The analysis of this gaze related data can help developers and designers identifying problems with gameplay, due to a misguided visual perception of the game environment [Sundstedt 13]. Moreover, the information extracted from players visual attention patterns can be useful in determining what is noticed and what is not noticed for incrementing the challenge level of the game.

In computer games, users attention is generally focused on the current task, while task-irrelevant details remain unnoticed, in the process known as in-attentional blindness [Mack 98]. Computer games usually impose an intensive task, so players become blind for all elements in the game environment that are not relevant or do not contribute significantly to the current task [Bernhard 11], the use of eye-trackers may help in understanding where players focus their attention during game play [Sundstedt 08].

Recently, researchers began to introduce eye-trackers devices and gaze data processing techniques in serious games and computer games [Deng 14, Almeida 11, Isokoski 09]. Studies throughout the literature, such as those conducted by Nacke *et al.*, evaluated the use of gaze data as an alternative way of controlling interaction with games. They obtained favourable outcomes where this challenge results in positive affection and feelings of flow and immersion [Ekman 08].

Even though the use of eyetracking data in the serious games field has increased recently, it is mainly used as an additional input system for the game. This dissertation is focused on the analysis of gaze patterns in junction with the user interaction with the system in order to generate intelligent profiles of user behaviours enabling the development of adaptive algorithms and game engines that allow the complete customization of the gameplay. This user profiling framework will give a new insight into users' attention behaviours, allowing the adaptation of game contents to their visual attention patterns and interaction techniques.

1.1 Research hypothesis and objectives

The analysis of the problem introduced throughout the previous section lead to the envisage of the following hypothesis.

Adaptive multi-variable algorithms based on gaze behaviour and user interaction for tailoring intelligent systems inside the serious games field can be used to profile users' activity patterns within these systems, enabling the development of intelligent frameworks for the real time customization of the game-play.

Based on the aforementioned hypothesis, this dissertation aims to profile the visual attention patterns and interaction performance of children while using intelligent cognitive therapies based on serious games. In order to reach the main aim of this dissertation, the following specific objectives need to be fulfilled.

– SO1: Define the current state of intelligent serious games. This objective is fulfilled by the comprehensive review of the existing techniques for creating intelligent serious games inside the literature. This objective aims to identify the current techniques employed in the tailoring and adaptation of current development of serious games, as well as to establish an actual panorama of the state of the art of these developments.

– SO2: Select the different questions that are going to be tackled during the research process. The achievement of this objective is supported by the study of the state of the art related not only to intelligent serious games, but to the development of different adaptive systems based on the use of eye-tracking. This objective need to be accomplished before the start of the implementation of the final technique, since this will be intimately linked to the elected problems for the testing phase.

- SO3: Design and implementation of the technique. This objective must follow the philosophy established in the objectives SO1 and SO2. This technique will take into account different data analysis algorithms for gaze data retrieval and interpretation and artificial intelligent techniques for the implementation of intelligent serious games. These two fields together will lead to the creation of a personalized solution that fulfils the main research problem exposed in the hypothesis.

- SO4: Configuration of the test-bed environment for the exposed technique. This objective is composed by the design and development of a specific methodology that can be used inside the environment described, as well as by the performance of te appropriate set of tests. This work will be composed by the performance of different set of tests that will help to obtain a better degree of understanding of the research question exposed. After the performance of the different studies that compose this dissertation, the parametrization of the analysis of the results need to be performed.

– SO5: Analysis and evaluation of the results. All the results obtained from the different pilot studies and the final tests will be thoroughly analysed. This analysis will be part of the main contribution of this research work. For this evaluation different statistical methods will be used along with decision making and machine learning techniques. The main goal of this section will be the objective evaluation of users' visual patterns for determining different user profiles that allow the creation of a framework for promoting cognitive development activities based on serious games adapt themselves according to the users' needs.

In addition to the specific objectives mentioned above, the following aims have been imposed as a way to contribute to the research community. These two objectives have been established based on the nature of the topic of this dissertation.

– Maximize the scientific contribution of this dissertation with the publication of several articles in scientific renamed conferences and journals relevant to the proposed topic. These scientific contributions are part of the dissertation, and they have been included throughout the different chapters of this document.

– Maximize the clarity and reproducibility of the different methodologies and data processing algorithms employed.

This will allow future developers and researchers to implement, improve and/or replicate all the different research questions tackled throughout this document.

1.2 Scientific and social impact and contribution

Once this dissertation framework has been introduced, this section aims to highlight the social and scientific impact of the presented research work.

The use of biofeedback techniques, such as eye-tracking for visual attention evaluation, is believed to improve the adaptability and immersion of users into therapies and e-learning developments. These measurements are implemented by the use of different sensors and biofeedback techniques that will help to allow real time adaptation of systems, helping to boost user engagement. Using these real time adaptation of therapies with children with and without learning disabilities could be a powerful way to ensure that children are always experiencing the most suitable challenging level for their current development state, which may lead in the improvement of their knowledge acquaintance strategies and academic curricula.

The other main field of this dissertation is the use of the techniques, methodologies and technologies involved in the serious games field. The literature analysis performed about its use manifest that combination of serious games together with biofeedback could be a step forward in the customization of game dynamics, objectives and rules taking into consideration users' needs at any time. This may lead to a more accessible experience, potentially boosting the efficiency of the developments involved.

The research opportunity that this novel approach gives is the analysis of visual pattern behaviours in users, in combination with their interaction performance, to create adaptive algorithms able to perform real time adaptation of game based learning interventions for children with and without learning disabilities. The design and implementation of these algorithms will lead to the acquaintance of the most suitable level of challenge for every user, generating intelligent user profiles that goes one step further of the "one size fits all" approaches, easy to find in current game based learning literature.

Furthermore, the chance to create a network of adaptive contents or completely tailored educational experiences. This can be implemented according to each user specific needs, leading to an advance in current educational systems, boosting their portability and autonomy. The continuous monitoring of biological signals together with the interaction results obtained from users give a new dimension in the research of learning therapies based on serious games. The modular development and design of these systems enable the addition of new variables in the future, that will help in the complete objective monitoring of the user, logging its progression.

The scientific-technical interest of the proposed topic lays on the framework design for the development of the multi-variable algorithms capable to adapt interfaces, work schemes and intensity levels of exercises based on serious games. This framework will profile users according to their interaction and visual attention patterns in order to obtain a personalized, adaptive and accessible intervention based on each users' needs.

The objective biological variables involved in this development will be acquainted by the use of eye tracking sensors recording user visual interaction during the testing procedure. This data will be analysed together with the interaction data recorded during the test. These parameters will then be used for the classification of the user state and the generation of different user profiles that will help in the decision making process of the system.

The scientific impact of this research topic is to join the use of biological variables with the features recorded during gameplay (scores, times, user history). These mix of features will be the input of the classification multi-variable algorithm for processing raw data recorded, giving a customization suggestion to the system based on the processed information. This suggestion can be used to adapt the intensity of the game or to switch between different games. The use of this customization strategies could help in the development of intelligent serious games that are always customized for final users' needs.

The scientific impact of this dissertation has been validated by the research community with the publication of different articles in various journals with JCR. This dissertation is comprised by a set of 4 different articles, 3 already published and one under review in international journals with impact factor.

1.3 Research methodology

This section explains the methodology followed by the research work of this dissertation. An experimental methodology approach was used due to the nature of the study. This dissertation has been submitted on a PhD by publication basis, so the dissemination of scientific publications is presented throughout the whole research process, as it is displayed in Figure 1.1.

The philosophy behind this methodology is to, not only contribute to the scientific community with the different findings obtained during the process, but to create an iterative methodology that allow the refinement of the final system. This is done through different development iterations that are based on the pilot studies conducted prior to the final analysis and evaluation of the system, as it is displayed in Figure 1.1.

Figure 1.1 shows the different stages that were implemented during the research process. These steps are synthesised in the following paragraphs:

– *Revision of the Literature*: The main objective of this step is to analyse, comprehend and understand the current state of the art of all the technologies and techniques involved in

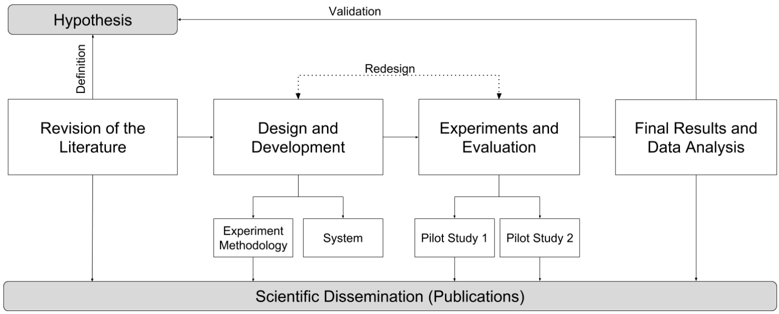


Fig. 1.1 Research methodology used in this dissertation

this work. In order to perform this task, the most relevant literature will be selected among the available publications in the scientific community. National and international journal articles, conference communications and proceedings and books will be reviewed. The knowledge obtained during this stage will lead to the formulation of the hypothesis and the generation of a review article with the available knowledge and future recommendations.

– *Design and Development*: After the literature analysis and the processing of the knowledge acquired in the previous step this stage will lead to the definition, design and development of the different iterations of the system. The completion of this stage will lead to the final assessment of the established hypothesis.

- *Experiments and Evaluation*: This stage aims to test the different iterations obtained after the design and development step. This section will lead to the completion of three different experiments: two initial pilot studies that will test a set of research questions that will help to the performance of the final research study. These data analysis and quantification of users' responses during the pilot phase will be used in the redesign of the overall developed system. All the knowledge used during these steps will be backed up by the concepts acquired during the first stages of the proposed methodology.

- *Final Results and Data Analysis*: This stage aims to compare the results and obtained parameters during the performance of the different user tests. These comparisons will be performed in order to assess the consistency of user profiling solutions using serious games and eye-tracking. This step will also explore the available options for the automatic generation of profiles. This stage will lead to the final evaluation of the hypothesis, testing whether it could be validated or not under the selected scenario.

1.4 Summary and organization of the study

This section outlines the structure and content of the different chapters that are part of this dissertation. This dissertation is presented as a PhD by publication so part of the chapters described below are included in the shape of published articles in peer reviewed journals with impact factor.

- **Chapter 1 - Introduction:** This first chapter has introduced the overall design, methodology and concept of the outlined research study. Its main objective is to set the main concepts and the research scenario. It will leave the reader with an overall idea of the conducted research study and system development. On top of this, the hypothesis and specific objectives introduced in this section will be key factors in the overall development of the study.

- **Chapter 2 - Related Work:** The second chapter in this dissertation introduces a general vision of the current state of development of the main areas involved in this research work. First, it gives an overall analysis of the use of eye-tracking for evaluation visual attention interaction in the development of different technological solutions, focusing on serious games-based developments, which is the main area of interest of this dissertation. The second part of this state of the art chapter is about the evaluation of the use of different artificial intelligence techniques in the development of serious games. The main goal of this part is to collect all the relevant articles published during the last decade and create a trend analysis about the use of certain artificial intelligence algorithms related to decision making and learning in the field of intelligent serious games. The analysis of the literature involved in these two sections will relevant in the final system design and in the performance of the research study.

– **Chapter 3 - System Design:** The third chapter is about the system design and development of the proposed framework. After introducing the most relevant research studies and developments in the area of serious games and eye-tracking, the system design of the current work is exposed. This design is backed up by the performance of two different pilot studies, that helped to create the final version of this development. This chapter will present the system design, experiment methodology and final development of the system.

– **Chapter 4 - Methodology:** This chapter details the research methodology followed in each of the two pilot studies and the final test performed during in this dissertation. All the relevant information regarding participants' sample and the devices and technologies involved will be detailed throughout this section.

- **Chapter 5 - Results:** The fifth chapter covers the description of the model presented, and it will lead the reader through the different results. This chapter reflects about the use of eyetracking and visual attention in combination with game interaction for user profiling. It discusses the relation between visual attention and user performance during the process of this work. This chapter is formed by two published articles and one under review contribution. The scientific impact of the research work proposed will be covered in this section.

- **Chapter 6 - Conclusion:** The sixth and final chapter of this dissertation introduces the different thoughts and conclusions extracted from the final evaluation of the research work presented. This chapter will cover the specific objectives introduced in section 1.1, and whether they were successfully met during the process. Future lines will be discussed in this section.

"Where should I go?" - Alice.
"That depends on where you want to end up." - The Cheshire Cat.
Alice in Wonderland and through the Looking Glass

Lewis Carroll

2

Related Work

This chapter reviews the most relevant literature with regards to the three main pillars that sustain this dissertation: Visual attention, the use of eye-tracking and the use of artificial intelligent techniques in the serious games field. The analysis of these fields will try to leave the reader with certain knowledge of the current panorama in the creation of intelligent serious games and current role visual attention plays in their design and development.

2.1 Visual attention

This section introduces visual attention, defining and introducing the different concepts and theories associated with it. After this introduction, the current panorama of the use of visual attention together with eye tracking in the serious games field is analysed.

2.1.1 Visual attention and perception

Human brain is limited with regard of its information-processing capabilities, implying that visual information need to be filtered so as to process the most important sensory inputs. Visual attention is defined then as the filtering process that select the most meaningful information from the visual stimulus input, suppressing those with a lower visual importance [Bernhard 10]. Thus, one of the most widespread definitions of attention is selectivity in perception [Bundesen 05], influencing perception by enhancing the contrast sensitivity and spatial resolution of visual inputs [Carrasco 11].

The common model for visual attention is divided into bottom-up processes, which reacts to low level features, and top-down processes, which select the target of interest based on users' goals and tasks on performance [James 13].

2.1.2 Bottom up control of visual attention

Bottom up control of visual attention is commonly defined as a goal-driven attentional process [Corbetta 02, Theeuwes 10]. Bottom-up processes are thought to operate on raw sensory inputs [Connor 04], rapidly and involuntarily capturing attention driven by properties inherent in the stimuli, that is, by salient visual features (i.e. flashing lights in a new environment) [Desimone 95, Ungerleider 00].

2.1.3 Top down control of attention

Top down control of attention is an essential cognitive ability that allows the brain to prioritize and process complex natural environments [Al-Aidroos 12], selecting visual stimulus that are relevant to the performance of a certain task (i.e. looking for something in a scene) [Buschman 07]. This process modulation underlies on the ability to focus attention on task-relevant stimuli, ignoring irrelevant distractions [Gazzaley 12]. Several approaches and methodologies have been proposed in the literature, showing that the human brain acts by enhancing activity in sensory regions for items that are relevant and suppressing the activity for items that are irrelevant to task goals. This top down control of attention is the cognitive process that take part during the player interaction with the proposed system.

2.2 Gaze interaction

This section introduces the features involved in gaze interaction and the related work generated by the scientific community in determining the different algorithms, paradigms and features involved in linking gaze interaction and visual attention behaviours.

2.2.1 Eye movements

Gaze interaction may provide a natural addition to the use of adaptive software, giving an objective insight of user visual behaviour patterns and visual attention. Gaze interaction is formed by a set of different eye movements, such as fixations, saccadic movements and smooth pursuits, which are outlined throughout this section.

Fixations

Fixations are the period of time when the eyes remain fairly still and new information is acquired from the visual array [Rayner 09]. They are catalogued as eye movements that stabilize the retina over a specific object of interest [Duchowski 07]. Fixations are characterized by miniature eye movements such as tremor, drift and microsaccades. Microsaccades are spatially random eye movements with an arc of 1 to 2 minutes [Duchowski 07]. These micro movements of the eye that form part of the fixation need to be taken into account by data processing algorithms, since they can be misclassified as noise due to their variation.

Fixation duration has been considered as a good indicator for the estimation of different cognitive functions such as object identification, memory, and monitoring of task-relevant objects [Bernhard 10].

Saccades

Saccadic movements are rapid changes in position of the eyeballs. They are typically found between fixational pauses during reading [Westheimer 54]. The saccade is generally considered as a ballistic eye movement whose trajectory, once begun, cannot be influenced [Becker 69]. They are defined as rapid eye movements used in repositioning the fovea to a new location in the visual environment [Duchowski 07]. Saccadic movements range in duration between 10 ms to 100 ms and the executor is effectively blind during this transition [Shebilske 83].

Smooth pursuits

Smooth pursuit movements are involved when visually following a moving target. In this specific movement, the eyes are capable of matching its velocity with the target in question [Carpenter 77, Duchowski 07]. They contrast to saccadic movements in that they occur when the eyes move rapidly from one target to the other or when an object is suddenly displaced [Vidal 12].

In the literature, different approaches have been reported for filtering this movements: Kalman filters for processing smooth pursuits [Abd-Almageed 02], dispersion and velocity analysis for classifying their nature [Liang 08, San Agustin 10, Koh 10], or labelling small pursuits as what falls in between the threshold between fixations and saccades [Grindinger 06, Komogortsev 07].

Nystagmus

Nystagmus are eye movements characterized as smooth pursuit movement interspersed with saccades and invoked to compensate for the retinal movement of the target [Duchowski 07].

2.2.2 Eye movement analysis

The goal of eye movement measurement and analysis is to objectively determine visual attention behaviour. Eye signals can be approximated by linear filters [Duchowski 07], which is an operational simplification of the underlying non-linear process [Carpenter 77]. From a signal processing standpoint, linear filtering analysis is enough for the localization of the distinct features that conform the eye movement signal, establishing a useful approximation in the sense of pattern recognition.

One of the main strengths of eye movement signals analysis is to characterize the signal in terms of salient eye movements (i.e., fixations or saccades). Majority of filter algorithms used in gaze analysis are focused on allocating where the eye signal changes abruptly, determining the end of a fixation and the beginning of a saccade, and the other way round for the end of the saccade and the beginning of the fixation.

There are three main automatic approaches for processing raw gaze data in 2D space [Duchowski 07]:

– Average: This method is based on averaging temporal signal over time. If there is little or no variance between different samples the set of data is deemed to be a fixation.

– Differentiation: This method assumes a uniform sampling rate in where successive samples are subtracted to estimate the eye movement velocity. Below a certain threshold the set of data is considered a fixation while the ones with a higher speed are classified as saccades. This method is considered more suitable for real time gaze data analysis.

– Area: This method is based on the selection of certain areas of interest. Inside these areas fixations are identified for detecting area impact into visual attention.

Previous to any of these algorithms a noise filtering pre-processing step need to be done. This is to eliminate the excessive noise that may be produced due to the inherent instability of the eye and blinks.

Dispersion-based fixation detection algorithms

This approach, also known as Dwell-Time Fixation detection algorithm [Holmqvist 11], is based on averaging the gaze coordinates for measuring the distance between them . A low distance in between a selected amount of raw data potentially signifies a fixation while wider distances may be related to a saccade [Duchowski 07].

(a) **Dispersion-Threshold Identification (IDT)**

This algorithm works with the concept that fixation points, since they have a lower velocity, tend to be clustered together in the spatial dimension [Salvucci 00]. Fixations are identified as a group of consecutive points within a certain dispersion (maximum separation) and a minimum duration threshold [Stark 81, Widdel 84].

I-DT algorithms use a moving window that spans a consecutive number data points, determined by the given duration threshold and sampling frequency, looking for potential fixations. This algorithm checks the dispersion between *(x, y)* maximum values [Salvucci 00].

(b) **MST Identification (I-MST)**

This algorithm is based on minimum spanning tree (MST) algorithms, thus, a connected undirected graph algorithm whose principle is connecting all vertices together with the minimal weighting for its edges [Camerini 88]. The use of this algorithm can provide a flexible and controllable representation for dispersion-based fixation identification [Goldberg 95].

Velocity-based fixation detection algorithms

In the velocity-based approaches the velocity of the signal is calculated within a sample window and then compared to a velocity threshold [Duchowski 07]. They all follow the approach that if the velocity is smaller than the given threshold then the data is classified as a fixation, while if the velocity is higher, then it is a saccadic movement. The advantage of this algorithms is that a short-term differential filter can be often used to detect the saccadic onset, decreasing the window size.

(a) **Velocity-Threhold Identification (I-VT)**

This algorithm separates fixation and saccades based on their point-to-point velocities [Erkelens 95, Sen 84]. Velocity profiles hugely varies between saccades and fixations, being low for fixations ($<$100 deg/sec) and higher for saccades ($>$300 deg/sec) allowing a robust separation process between them [Salvucci 00].

(b) **HMM Identification (I-HMM)**

Hidden Markov Model Fixation identification uses probabilistic analysis to determine fixations given a protocol [Salvucci 98]. I-HMM is represented by a two state model which represent velocity distributions for saccade and fixation points [Salvucci 00].

Area-based fixation detection algorithms

Area fixation identification identifies fixations that occur within specified target areas [Den Buurman 81]. The target areas are defined as rectangular regions of interest that represent units of information in the visual field.These algorithms identify fixations occurring close to relevant targets [Salvucci 00].

2.2.3 Eye tracking and visual attention

Scientific research on eye movements began at the end of the 19th century when the first reliable methods for the measurement of eye position were developed [Huey 98, Buswell 35, Schütz 11]. Even though eyetrackers were intrusive and cumbersome at the beginning, the recent advances on eyetracking technology allow the use of this devices effortlessly and almost in a complete transparent way to final users, avoiding distractions [Sundstedt 13].

Eye-tracking studies are thought to be a reasonable method for inferring how visual attention behaves under different stimuli and tasks. The observation of eye-movements is not a new area of research within psychology-related fields, it have been studied in depth over the last decades [Rayner 78, Rayner 98, Rayner 09].

Both eyes have only a small area, the *fovea*, that can see accurately. The *fovea* is circular and cover about 1-5 degrees of the visual field [Duchowski 07]. When perceiving a scene as part of a visual stimulus, humans move their eyes to direct the light from different parts of the scene, this movement is known as a saccadic movement.

Eye-tracking allow researchers to have an insight into the different main features that can be extracted from gaze interaction. Eye-trackers can be used to estimate position of the foveal focus, fixation duration and counts, amplitudes of the saccades and an estimation of the pupil size and diameter for further analysis [Duchowski 07].

Even though gaze data information is reliable and available for researchers through the use of eyetracking sensors, up to now there is little success in the prediction of a sequence of fixation movements for a human observer, when he/she is looking at an arbitrary scene [Schütz 11]. Different paradigms and factors have been established for the analysis of visual patterns and the features that drive users' fixations.

Salience

This paradigm established that the salient parts of a scene are the ones that attracts visual attention first. In recent years, salience map models have been vastly analysed throughout the literature on visual search [Purcell 12, Siebold 13, Ramirez-Moreno 13, Anderson 15, Pantelis 16].

Different implementations can be found for salience model definitions. Itti & Koch presented an approach where the image was linearly filtered and differences were computed for three different features: intensity, colour and orientation [Itti 00, Itti 01]. This is based on the internal work done by neurons on early stages of visual processing. Kienzle *et al.* collected a large number of fixations on a series of pre-calibrated images. Then they applied machine learning techniques for determining different patches where people performed the highest number of fixations [Kienzle 09]. In summary, there is some evidence of the role of main stimulus in the target selected for fixations, however the results up to date with video sequences showed a large degree of variability [Schütz 11].

Although it has been largely studied and new developments are in process which could improve the success rate of fixation detection, salience models by themselves have a discrete effect on guiding gaze.

Object recognition

Salience approaches work on individual features and have no knowledge about the different objects that compound the scene under observation. This object recognition approaches are based on the concept that saccadic movement and fixations are driven by an object recognition process rather than a feature-based approach.

This approach is not as popular as the salience map model, but several studies found that observers tend to fixate faces and animals in scenes even when no specifically instructed task for searching them has been set [Thorpe 96, Crouzet 10, Drewes 11]. Other studies found that the preferred fixation point for objects was close to its centre, supporting the role of object-based selection [Nuthmann 10]. Although this is not a common approach for fixation estimation, objects and living beings play an important role in visual target selection processes.

Task-based

This approach is focused on how the execution of an active task influences the visual pattern behaviour. Several articles have analysed gaze patterns while users under study are performing specific domestic everyday tasks [Land 99, Hayhoe 00], sports-related tasks [Bahill 84, Land 00, Hayhoe 05] or doing different general interest and more abstract tasks interacting with objects [Ballard 95, Epelboim 98, Johansson 01, Herst 01, Brouwer 09]. Other studies were focused on the coordination of hand, eye and body movements [Land 09]. However, the main task studied in gaze interaction is reading behaviours, where a vast amount of articles have been published assessing gaze interaction in a variety of reading related tasks, such as mathematics, science problems, tests and subtitles in films [Witzel 12, Tsai 12, Lai 13, Kunze 13a, Kunze 13b, Bax 13, Kruger 14, Bisson 14, Andrá 15, Rodrigue 15, Giovinco 15]. Majority of studies concluded that, while subjects are performing a certain task, they are mostly fixated on task-relevant features in the scene. Task demands are clearly controlling eye movements when subjects are pursuing an specific goal.

2.2.4 The use of eye tracking in games and serious games

Research using eye tracking sensors affords a unique opportunity to test aspects of theories about multimedia learning concerning processing during learning [Mayer 10]. Moreover, the use of this approach may help to understand where players focus their attention during gameplay [Sundstedt 08], as well as how they confront unfamiliar games and software [Pretorius 10].

However, it was not until recently that researchers began to analyse and introduce eye tracking sensors and techniques in serious games and games [Deng 14, Almeida 11, Isokoski 09]. Games that can be controlled solely through eye movement would be accessible to persons with decreased mobility or control. Moreover, the use of eye tracking data can change interaction with games, producing new input experiences based on visual attention [Isokoski 09].

Eye tracking devices have been used in the design of educational games, in terms of assessing usability based on user gaze-behaviours when interacting with the game [Zain 11, Kiili 14].El-Nasr and Yan used eye tracker sensors to analyse attention patterns within an interactive 3D game environment, so as to improve game level design and graphics [El-Nasr 06].

Kickmeier-Rust *et al.* focused on assesing the effectiveness and efficiency of Serious Games [Kickmeier Rust 11]. For this purpose, they assessed these variables with gaze data and gaze paths, in order to obtain interaction strategies in specific game situations . Sennersten and Lindley also evaluated the effectiveness of virtual environments in games through the analysis of visual attention using eye tracking data [Sennersten 10]. Johansen *et al.* discussed the efficiency of eye tracker sensors in assessing users' behaviour during gameplay [Johansen 08].

Józsa and Hamornik used recorded eye tracking data to evaluate learning curves in university students while using a 7 hidden differences puzzle game. They used this data to assess similarities and differences in information acquisition strategies considering gender and education dependent characteristics [Józsa 11]. Dorr *et al.* conducted a similar study which concluded that expert and novice players use different eye movement strategies [Dorr 07]. Muir *et al.* used eye-tracking data to capture user attention patterns and present results on how those patterns were affected by existing user knowledge, attitude towards getting help and performance while using the educational game Prime Club [Muir 12].

Radoslaw *et al.* [Mantiuk 12] used eye tracker sensors assessing render quality in games . They argued that gaze-dependant rendering was specially important when immersing in Serious Games where immersing in virtual environments played a primary role [Turner 14]. Smith and Graham and Hillaire *et al.*

concluded that use of an eye tracker increases video game immersion, altering the gameplay experience [Smith 06, Hillaire 08].

Chang *et al.* developed the game *WAYLA*, as a mean to evaluate the potential to offer new interaction experiences based on eye tracking and visual attention. These authors took advantage of the popularity and arrival of more affordable eye tracker sensors [Chang 13].

Li and Zhang used eye-movement analysis to assess patients' mental engagement in a rehabilitation game. Therapists use this feedback to adjust rehabilitation exercises to users' needs [Li 14]. Continuing with the health-related field, Lin *et al.* developed an eye-tracking system for eye motion disability rehabilitation as a joystick-controlled game [Lin 04]. Vickers *et al.* developed a framework which integrated automatic modification of game tasks, interaction techniques and input devices according to a user ability profile [Vickers 13]

Walber *et al.* presented *EyeGrab*, a game for image classification controlled by the players' gaze. The main purpose of this game was to collect eye tracking data to enrich image context information [Walber 12].

Other studies such as those conducted by Nacke *et al.* evaluated the use of eye tracker sensors as an alternative way of controlling interaction with games. They obtained positive results where this challenge was related to positive affection and feelings of flow and immersion [Nacke 10]. Ekman *et al.* goes one step further, discussing the limitations of using pupil-based interaction and providing suggestions for using pupil size as an input modality [Ekman 08].

2.3 ARTICLE I: Intelligent serious ga-mes

This section introduces the first of the four articles that are part of this dissertation. It presents a panorama about the use of artificial intelligent techniques within the serious games field. The development of intelligent serious games lead to the creation of new software developments able to adapt themselves to players' final needs. This is one of the core objectives of this dissertation.

The video-games market has become an established and ever-growing global industry. The health of the video and computer games industry, together with the variety of genres and technologies available, mean that videogame concepts and programmes are being applied in numerous different disciplines. One of these is the field known as serious games. The main goal of this article is to collect all the relevant articles published during the last decade and create a trend analysis about the use of certain artificial intelligence algorithms related to decision making and learning in the field of serious games. A categorization framework was designed and outlined to classify the 129 papers that met the inclusion criteria. The authors made use of this categorization framework for drawing some conclusions regarding the actual use of intelligent serious games. The authors consider that over recent years enough knowledge has been gathered to create new intelligent serious games to consider not only the final aim but also the technologies and techniques used to provide players with the most realistic experience. However, researchers may need to improve their testing methodology for developed serious games, so as to ensure they meet their final purposes.

This paper is available under the early access section of the journal *IEEE Transactions on Computational Intelligence and AI in Games (IF: 1.481 [Q1])* [Frutos-Pascual 15b], it was accepted on the 25th of December 2015. This article is included on its original full version, being the first page of the manuscript included in the one-column preliminary version that is currently available online. However, the full article has been included in the two-column format of the journal. This format is not available in press yet.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCIAIG.2015.2512592, IEEE Transactions on Computational Intelligence and AI in Games

Review of the Use of AI Techniques in Serious Games: Decision-Making and Machine Learning

Maite Frutos-Pascual Deustotech Life [eVIDA]
University of Deusto
Avda Universidades 24, Bilbao, Spain
Email: maitefrutos@deusto.es Begoña García Zapirain Deustotech Life [eVIDA]
University of Deusto
Avda Universidades 24, Bilbao, Spain
Email: mbgarciazapi@deusto.es

Abstract

The video-games market has become an established and ever-growing global industry. The health of the video and computer games industry, together with the variety of genres and technologies available, mean that videogame concepts and programmes are being applied in numerous different disciplines. One of these is the field known as serious games. The main goal of this article is to collect all the relevant articles published during the last decade and create a trend analysis about the use of certain artificial intelligence algorithms related to decision making and learning in the field of serious games. A categorization framework was designed and outlined to classify the 129 papers that met the inclusion criteria. The authors made use of this categorization framework for drawing some conclusions regarding the actual use of intelligent serious games. The authors consider that over recent years enough knowledge has been gathered to create new intelligent serious games to consider not only the final aim but also the technologies and techniques used to provide players with a nearly real experience. However, researchers may need to improve their testing methodology for developed serious games, so as to ensure they meet their final purposes.

I. Introduction

Since the beginning of the twenty-first century, the video-games market has become an established and ever-growing global industry. The health of the video and computer games industry, together with the variety of genres and technologies available, mean that videogame concepts and programmes are being applied in numerous different disciplines. One of these is the field known as serious games.

The term serious games was coined by North American researcher Clark Abt in his book "Serious Games" in the 70s (1). Although the spirit of this trend has been maintained over the last decades, the technologies, applications and scope have changed significantly.

Review of the Use of AI Techniques in Serious Games: Decision-Making and Machine Learning

Maite Frutos-Pascual
Deustotech Life [eVIDA]
University of Deusto
Avda Universidades 24, Bilbao, Spain
Email: maitefrutos@deusto.es

Begonya García-Zapirain
Deustotech Life [eVIDA]
University of Deusto
Avda Universidades 24, Bilbao, Spain
Email: mbgarciazapi@deusto.es

***Abstract*—The video-games market has become an established and ever-growing global industry. The health of the video and computer games industry, together with the variety of genres and technologies available, mean that videogame concepts and programmes are being applied in numerous different disciplines. One of these is the field known as serious games. The main goal of this article is to collect all the relevant articles published during the last decade and create a trend analysis about the use of certain artificial intelligence algorithms related to decision making and learning in the field of serious games. A categorization framework was designed and outlined to classify the 129 papers that met the inclusion criteria. The authors made use of this categorization framework for drawing some conclusions regarding the actual use of intelligent serious games. The authors consider that over recent years enough knowledge has been gathered to create new intelligent serious games to consider not only the final aim but also the technologies and techniques used to provide players with a nearly real experience. However, researchers may need to improve their testing methodology for developed serious games, so as to ensure they meet their final purposes.**

I. Introduction

Since the beginning of the twenty-first century, the video-games market has become an established and ever-growing global industry. The health of the video and computer games industry, together with the variety of genres and technologies available, mean that videogame concepts and programmes are being applied in numerous different disciplines. One of these is the field known as serious games.

The term serious games was coined by North American researcher Clark Abt in his book “Serious Games” in the 70s (1). Although the spirit of this trend has been maintained over the last decades, the technologies, applications and scope have changed significantly.

Serious games represent a genre designed to be more than “just” fun (2). Moreover, the educational value associated to serious games goes beyond academic purposes, locating their focus on skill practice and entertainment value during exposure (3; 4; 5; 6).

The main purpose of a serious game is both to be fun and entertaining, and educational. A serious game is thus designed both to be attractive and appealing to a broad target audience, and to meet specific educational goals (7). They are designed to foster knowledge, skills or routine habits in the player.

Serious games span a broad range of fields and areas of expertise. In the literature, serious games were divided into several categories based on different classification schemes. These models can be divided into two main categories: market-based and purpose-based classifications (8).

Several authors established different categories of **market-based** classification. This segmentation is based primarily on the different “markets” or fields the *serious games* are developed for. The different segments identified in the literature are:

- *Military games, government games, educational games, corporate games, health-care games, political, religious and art games* (9).
- *Health, public policy, strategic communication, human performance engineering, training and simulation, education, game evaluation* (10).
- *Educational, social change, military, occupation and marketing* (11).
- *Defense, teaching and training, advertising, information and communications, health, culture and activism* (12).
- *K-12 edutainment, higher education, health-care, corporate, military, non-government and other (13).*

Different authors also provide different categories for **purpose-based** classifications, or intention they were intended to satisfy.

- *Advergames, activism games, training and simulation games, edugames, newsgames and edumarket games*. (14).
- *Business games, health and medicine, news, activism, advergames and political games*. (15).

Other classification approaches use alternatives to the market/purpose distinction, proposing labels or tags as means of classification. (16). Meanwhile, the G/P/S classification model considers a gameplay, purpose and scope trio (17). Finally, classification can also be conducted according to learning principles, target age group or game platform (12).

The main goal of this article is to collect all the relevant articles published during recent years and create a trend analysis about the use of certain artificial intelligence algorithms related to decision making and machine learning in the field of serious games. A categorization framework was designed and outlined to classify available articles in the literature. The authors made use of this categorization framework for performing an analysis of the actual use of intelligent serious games.

This article consist of an analysis of serious games, offering a literary review of their use combined with certain artificial intelligence techniques in the area of *decision making* and *machine learning*. Other areas such as *path finding* were initially considered for study but were not sized enough for further analysis, so they were excluded from this review. The article is divided into several sections. First of all, a complete methodology is introduced presenting the form and function of data collected from the literature review. Then, each of the following subsections presents a contextualization and classification of available articles. Finally, the article ends with the discussion and conclusion section.

II. Method

This article is the result of a systematic search of the serious game term combined with AI-related parameters. Although the serious games trend as we know it arose during the beginning of the 21st century, due to the amount of studies released in this field authors will limit the scope of this review to the last decade, specially between 2005 and 2014.

A. Data collection

1) Databases searched: This review was carried out mainly using the search engine provided by the *Web of Knowledge*. However, searches were also carried out in other electronic databases for this review, for instance databases particularly related to education, computer science, information technology, social sciences and health: *ACM* (Association for Computing Machinery), *IEEE* (Institute of Electrical and Electronics Engineers), *BioMed Central*, *Science Direct*, *EBSCO*, Emerald, and PsycINFO.

2) Search terms: Search terms for games in conjunction with terms for possible outcomes, impacts of effect games, as well as with related AI parameters. The authors performed a combined search of the terms *game, serious game, games, serious games, play, playful, game based learning* in combination with *intelligent, artificial, intelligence, adaptive, decision trees, fuzzy logic, markov system, goal orientation, GOAP, finite state machines, näive bayes, artificial neural network, case-based reasoning, support vector machines, adaptive hypermedia, dynamic difficulty adjustment, machine learning, decision tree, genetic algorithms* and *reinforcement learning*. A third parameter was introduced to narrow down the search, using words for possible outcomes such as *learning, education, health, training, motivation, behaviour, skills, attitude, military*, among others (18)

3) Selection of papers for inclusion: Abstracts were selected for the retrieval of the paper if they were judged to include data about the implementation of *serious games* including at least one of the displayed AI techniques. In order to achieve this purpose, papers had to (a) include specifically the purpose of the serious game, and its target audience and definition (b)describe the implementation of the AI technique used and its purpose inside the game and (c) they have been published between January 2005 and September 2014. Using these three conditions 129 out of more than 300 papers met the inclusion criteria and were identified as relevant for the review. Figure 1 displays inclusion rate per year. Abstracts that clearly implemented a serious game and which seem to have some AI implementation algorithm although its usage were not defined in the abstract were also retrieved for further analysis.

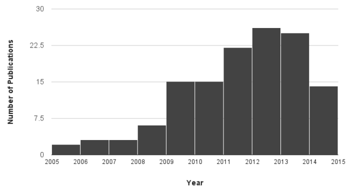


Fig. 1. Articles included in this review per year

4) Selection of papers for exclusion: Papers that did not clearly specify the technique used, the final aim of the *serious game* or did not satisfy the conditions outlined in the previous section were not selected for inclusion in the current article.

B. Data analysis

1) Coding of papers: The 129 papers meeting the inclusion criteria were coded using a data extraction form that was developed by considering several features related to the outcomes and impacts of the game in several salient dimensions.

2) Information extraction: Information from papers was extracted and coded within the following categories:

1) *Year and Country* - Year of publication of the article and first author's main country of affiliation.
2) *AI Flagship* - This AI flagship classification is based on the work published by Georgios N. Yannanakis, from the *Center of Computer Games Research* in Copenhagen. His work presents four key game AI research areas that represent new perspectives (19). For this review, the following flagships were considered:
    a) *Player Experience Modeling (PEM)* - In this category, AI techniques are used for constructing new model of player experience. This can stem from different types of data collected from the players. These PEM approaches are:
        i) Objective - Collecting data by monitoring changes in players' physiology. These bodily alterations may assist in adapting games to the users state and behaviours.
        ii) Subjective - Subjective player experience modelling based on self-reporting. This can guide machine learning algorithms to capture different aspects of player experience.
        iii) Gameplay - Any element derived from the interaction between the player and the game.

   b) *Procedural Content Generation* - This flagship refers to the development of algorithms that generate game content automatically. These elements may affect player experience.
   c) *Game Data Mining* - Games where users' data is collected during gameplay with the aim of addressing specific questions.
3) *AI Usage* - This category classifies the final purpose of the AI implementation for each of the included articles (for example for classifying users, altering NPC behaviors, adapting gameflow, among others)
4) *Market/Purpose Classification* -
   a) *Serious Games Market* - Classify the articles based on the different "markets" or fields the serious games were developed for.
   b) *Serious Games Purpose* - Classify the references based on the intention the serious games were developed for.
5) *Platform/Delivery* - Platform for delivery of the game were classified as PC, mobile/tablet, online game or others.
6) *Type* - This category refers to the project type. This is divided into: new developments of serious games, new frameworks for building serious games and other (research studies or game engines, among others).
7) *Testing Procedure* - For studies tested with real users, further information regarding the sample size, gender balance and age range was analysed and coded. For articles with different testing procedures, their study design and data analysis were reviewed and coded.
8) *Publication Channel* - Publishing medium was also analysed and coded. Articles were divided into: *(a)* Type I - International journals with impact factor, *(b)* Type II - International journals, *(c)* Type III - International conferences, workshops and symposiums and *(d)* Type IV - Book chapters.

C. Article outline

The authors will discuss all these categories in two main fields of AI Section III - Decision making and Section IV - Machine Learning. Each section reviews several methods and algorithms related to those areas. Authors have selected those most implemented in the field of serious games for their inclusion in this article. Each of the sections outlines the different trends in the design and development of AI in serious games. These trends are based on the different coded categories included in the methodology section. Finally, global tendencies are analysed in the discussion and conclusion section.

III. Decision Making

This section deals with one of the most popular techniques inside the storyline of serious games, the decision making field.

These algorithms and techniques are used for logical and rational decision-making processes based on the information compiled by the system. The input of the decision- making systems is the previously collected knowledge, and the output is an action request (20).

These algorithms and techniques are very popular in the design and development of serious games, i.e Sordoni *et al.*. used decision-making agents in a game that was oriented towards raising people's awareness about the importance of making rational decisions concerning natural resources management in protected environments (21). Cantwell *et al.* developed a collaborative exergame for the elderly, in which game speed and difficulty were adjusted in the light of user profile and performance (22). The subsections below provide information about other developments classified by specific AI techniques within the field of decision making.

A. Decision Trees

The main objective of decision trees is to create a prediction model on the basis of a set of decision rules obtained from compiled data during system performance. These decision trees have been used throughout the literature by several authors. This section categorizes the latest and most relevant publications using this technique applied to the serious games field.

Twenty seven articles described research that employed decision trees and could be further analysed and coded following the categories outlined in the methods section. Please refer to Table I for a complete list of coded articles in this section.

1) AI flagship: The vast majority of the coded articles (19 out of 27) fall into the category of *PEM - Gameplay based*. For example, Costa *et al* were focused on the enhancement of executive memory and attention. The internal mechanisms of this serious game were based on a decision tree that considered players' errors and progresses to create a customized gameplay experience (23)

Four studies made use of *Game Data Mining*. Keshtkar *et al.* presented a framework based on data mining techniques and language processing approaches for evaluating players' personality and behaviour while using the educational game *Land Science* (24). The remaining four did not correspond to any AI flagship.

2) AI usage: Decision trees were implemented for modelling the game flow in 14 out of 27 articles, therefore, providing customized experiences to users based on their interaction with the game. This was implemented as a set of decision trees for the internal mechanism of the game (25; 26; 27; 28; 29).

In 8 of the revised articles, the decision trees were used for assessing motivation and users' states while playing, letting the system to obtain and analyse users' behaviour information. (30; 31; 32)

Two articles designed systems based on decision trees to involve players in further analysis or even to let users build them. For example, Haworth *et al.* displayed decision trees during gameplay to assess the users' effectiveness in analytic reasoning while playing in a logic labyrinth-style game (33). M. Good *et al.* designed the scientific discovery game *The*

Cure. This game was centred on the task of gene selection for breast cancer survival. The objective of each level is to choose a set of genes that produces a better decision tree classifier, which is stored for the next time improving the system (34).

3) Categorization of games: Articles were classified according to the following features:

- **Serious games market:** The most common market for serious games employing decision trees was health with 10 papers out of 27. Next was the educational with 8, then business with 4 (24; 35; 36; 37) and culture and activism with 2 (38; 27). Remaining three articles were classified as political (26), military (39) and uncategorised respectively.
- **Serious games purpose:** Regarding their purpose, selected articles were classified as follows: 4 under the category of health (23; 31; 30; 34), 13 under training field, 7 under edugames, 2 under activism-related games (40; 38) and final one was uncategorised.
- **Platform/delivery:** 19 out of the 27 analysed serious games were designed for PC, while 5 were for use online (30; 34; 26; 40; 38) and 3 were specifically developed for mobile devices (41; 35; 37).
- **Project type:** The majority of analysed papers were new developments of serious games, while the remaining 7 were framework proposals and design guidelines combining decision trees and serious games.

4) Sample of participants: Out of the 27 articles included for revision in this section only 14 of them include any information about testing procedures. Eight out of those 14 outline further information about sample's age range. Only 5 included detailed description of the users in terms of gender and age (30; 25; 42; 43; 40).

5) Publication type: Publication channels were well-balanced, with 7 articles published in international journals with impact factor and 8 in international journals without impact factor. Another 8 were presented in international conferences, workshops and symposiums and lastly 4 were published as chapters in books.

B. Fuzzy Logic

Fuzzy logic is based on approximate reasoning. It goes beyond binary approximations, and was designed to cope with gray areas that go beyond true or false (20).

Fuzzy logic is relatively popular in the video-game industry. It is considered to be of great utility when employing probabilistic methods to reflect any kind of uncertainty (50). The most relevant publications are briefly introduced below and classified in the terms outlined in the methods section. Sixteen articles employed fuzzy logic. Please refer to Table II for a complete list of the coded articles in this section.

1) AI flagship: Eleven out of sixteen articles fall into the category of *PEM - Gameplay based*. One example is the Serious Game *Karo* which infers gameplay outcomes based on players decision and actions (51)

Three articles were classified under the *PEM - Objective* flagship as they made use of different movement and biological sensors together with user data for customizing the gameplay (52; 53).

One article was categorized as *Procedural Content Generation*. This serious game was focused on the training of commanders in their strategies after toxic chemical spillages. It combined fuzzy logic with artificial neural networks for modelling NPCs behaviours.

The last article was classified under the *Game Data Mining* label. This game was developed as a multi-agent-based model of team play, for characterizing individual personalities by Myers-Briggs Type Indicator (MBTI) (54).

2) AI usage: Fuzzy logic was implemented for addressing a wide range of possibilities throughout the literature.

The most common usage of fuzzy logic in serious games was as techniques for adapting the gameplay, with 6 references in this section.

Fuzzy systems were also implemented for assessing user state during gameplay; classifying users' state (54; 55) or managing feedback during the game (56).

Fuzzy systems were also considered for monitoring and promoting immersive experiences (53; 57), as well as for the controlling of NPCs (58; 59).

Fuzzy logic was also implemented in serious games for actually teaching fuzzy logic, as means of observing the behaviours of virtual characters (60) or robots (61) guided by a certain set of rules during gameplay.

3) Categorization of games: Articles were classified according to the following features:

- **Serious games market:** Six out of 16 articles in this section belonged to the health category inside the serious games market. The market with the second highest number of articles in this section are education (60; 61; 62), social change (54; 58; 51) and business (63; 64; 65) fields with 3 references each. The military market (59) did also have representation inside the fuzzy logic implementation.
- **Serious games purpose:** The respective purposes of the identified publications in this section were quite balanced: 3 articles belonged to the health category, 6 were related to training and the remaining 7 were classified as edugames.
- **Platform/delivery:** Out of the 16 included articles, 13, were designed for PC while the remaining 3 were available online.
- **Project type:** Thirteen of the reviewed studies were concerned with new serious games software development, while the remaining three were categorized as platform (61), framework (52) and research study (54).

4) Sample of participants: All the outlined articles were focused on internal mechanisms, designs or algorithms involved in the development of serious game, and did not display data about user interaction, performance or game efficiency and efficacy. However, one of the articles describes a testing procedure using NetLogo for the simulation study (54).

5) Publication type: Publication channels were not balanced in this section, with 5 articles published in international

TABLE I
DECISION TREES AND SERIOUS GAMES - CODED ARTICLES

Authors	Country	Year	AI Flagship	AI Usage	Type	Platform	SG Market	SG Purpose	Participants				Publication*	Reference
									Total	Males	Females	Age		
Good *et al.*	US	2014	Massive Scale Game Data Mining	Transparent	Development	Online	Health	Health	1077	–	–	–	Type II	(34)
Castro-Sanchez *et al.*	GB	2014	–	Gameflow	Development	Mobile	Health	Training	–	–	–	–	Type II	(41)
Frutos-Pascual *et al.*	ES	2014	PEM - Gameplay	Assistance	Development	Online	Health	Health	17	10	7	12 - 19	Type I	(30)
Keshtkar *et al.*	US	2014	Massive Scale Game Data Mining	Assessing	Framework	PC	Business	Training	12	–	–	6 - 12	Type IV	(24)
Sabourin *et al.*	US	2013	PEM - Gameplay	Gameflow	Development	PC	Educational	Edugames	260	129	131	13.4 (SD = 0.57)	Type II	(25)
Lin *et al.*	TW	2013	PEM - Gameplay	Suggestion	Development	PC	Educational	Training	92	48	44	18 - 26	Type I	(42)
Moya *et al.*	ES	2013	–	Optimization	Framework	PC	–	–	35	–	–	6 - 70	Type I	(44)
Horan and Gardner	UK	2013	PEM - Gameplay	Gameflow	Development	Online	Political	Edugames	6	–	–	–	Type IV	(26)
Petridis *et al.*	UK	2013	PEM - Gameplay	Gameflow	Development	PC	Culture and activism	Edugames	5	–	–	–	Type II	(27)
Hussaan and Sehaba	FR	2012	PEM - Gameplay	Assessing	Development	PC	Health	Edugames	–	–	–	–	Type IV	(45)
Fu-Hsing *et al.*	TW	2012	Massive Scale Game Data Mining	Assessing	Framework	Online	Educational	Activism	8	4	4	11 - 12	Type II	(40)
Parsons *et al.*	NZ	2012	PEM - Gameplay	Gameflow	Development	Mobile	Business	Training	14	–	–	–	Type III	(35)
Goulding *et al.*	UK	2012	PEM - Gameplay	Evaluation	Development	PC	Educational	Training	–	–	–	–	Type I	(46)
Rijcken *et al.*	NL	2012	PEM - Gameplay	Gameflow	Development	Online	Culture and activism	Activism	–	–	–	–	Type II	(38)
Hussaan *et al.*	FR	2011	PEM - Gameplay	Gameflow	Development	PC	Health	Edugames	–	–	–	–	Type III	(28)
Chiang	TW	2011	Massive Scale Game Data Mining	Assessing	Development	PC	Business	Training	Tested with different games				Type III	(36)
Qin *et al.*	CN	2011	PEM - Gameplay	Assessing	Development	PC	Health	Training	–	–	–	–	Type I	(32)
Van der Spek	NL	2011	PEM - Gameplay	Gameflow	Development	PC	Health	Training	41	36	5	23.3 (SD = 1.89)	Type I	(43)
Santos *et al.*	BR	2011	PEM - Gameplay	Assessing	Development	PC	Health	Health	80	–	–	10 - 17	Type III	(31)
Ghergulescu and Huntean	IE	2011	PEM - Gameplay	Assessing	Framework	PC	Educational	Edugames	–	–	–	–	Type III	(47)
Costa *et al.*	BR	2010	PEM - Gameplay	Gameflow	Development	PC	Health	Health	–	–	–	–	Type III	(23)
Haworth *et al.*	CA	2010	–	Transparent	Development	PC	Educational	Training	–	–	–	–	Type II	(33)
Vidani and Chittaro	IT	2009	PEM - Gameplay	Gameflow	Development	PC	Health	Training	–	–	–	–	Type III	(48)
Bellotti *et al.*	IT	2009	PEM - Gameplay	Gameflow	Framework	PC	Educational	Training	–	–	–	–	Type I	(49)
Lilly and Warnes	UK	2009	–	Gameflow	Framework	Mobile	Business	Edugames	–	–	–	–	Type IV	(37)
Kim *et al.*	US	2009	PEM - Gameplay	Gameflow	Development	PC	Military	Training	31	–	–	–	Type II	(39)
Tang and Hannegham	MY	2008	PEM - Gameplay	Gameflow	Framework	PC	Educational	Training	–	–	–	–	Type III	(29)

** (a) Type I: International Journals with Impact Factor, (b) Type II: International Journals, (c) Type III: International conferences, workshops and symposiums, (d) Type IV: Chapters in Books*

journals with impact factor, 1 in an international journal, 7 in international conferences, workshops and symposiums and 3 published as chapters in books.

C. Markov Systems

Markov systems go a step further than fuzzy logic, assigning a meaning to the values of truth or belonging. Markov systems are memory-less, and they are represented as a set of finite transitions among a determined number of possible states (67).

Reviewing the literature inside the *serious games* field, there are some references to their usage. The authors found ten articles that met all the inclusion criteria outlined in section II-A3. Please refer to Table III for a complete list of coded articles in this section.

1) AI flagship: Six out of ten articles fell into the category of *PEM - Gameplay based*. This is the case of the game *PlayPhysics* which was developed for teaching physics to engineering students and made use of Hidden Markov Models for adapting the gameplay based on players' interaction (68).

One article was classified under the *PEM - Objective* flagship. The serious game *Heap Motiv* implemented Keller ARCS motivational model along with biological variables such as heart rate and EEG to identify users' state and establish different motivational strategies (69).

Two studies were categorized as *Procedural Content Generation*. They were about a framework designed for computer assisted language learning games using trained Markov Decision Processes (70; 71).

Lastly, one article was categorized as *Game Data Mining*, which presented a new game-based approach for conducting cognitive science experiments. They implemented Markov decision processes to infer cognitively relevant information from players actions during gameplay (72).

2) AI usage: As described in earlier sections, more than half of all articles considered, seven out of ten, were focused on adapting the gameplay.

Markov systems were also implemented for collecting information from players (72; 73) and assessing player motivation (69).

3) Categorization of games: Articles were classified according to the following features:

- **Serious games market:** The educational market was the most referenced, with five articles in this category. Health had three articles (74; 73; 75). Military (76) and culture and activism (77) fields were balanced with one article each.
- **Serious games purpose:** Regarding the final purpose of the reviewed studies, the vast majority of them were categorized as edugames. The three remaining articles were considered to have a health-related purpose.
- **Platform/delivery:** All but one of the analysed studies were designed and implemented for PC. The remaining article was developed as an online tool (72).
- **Project type:** Seven out of 10 of the reviewed studies were new serious games software developments, while three of the remaining articles were categorized as frameworks (70; 71; 72). Lastly, one article presented a set of guidelines for designing serious games (74).

4) Sample of participants: Eight out of ten articles were tested in one way or another with real participants. The *Heap Motiv* game gave detailed description about the sample characteristics and the experiment setting (69). Su *et al.* tested their framework proposal with both simulated learners and over 250 real participants (70; 71), but they did not specify gender segmentation, nor age range. The game *PlayPhysics* was tested with 28 university students (68), while the *Solis*

TABLE II
FUZZY LOGIC AND SERIOUS GAMES - CODED ARTICLES

Authors	Country	Year	AI Flagship	AI Usage	Type	Platform	SG Market	SG Purpose	Participants				Publication*	Reference
									Total	Males	Females	Age		
Karime et al.	AE	2014	PEM - Objective	Adapt the gameplay	Framework	PC	Health	Health	–	–	–	–	Type I	(52)
Pirovano et al.	IT	2014	PEM - Gameplay	Teach Fuzzy logic	Development	PC	Educational	Edugame	–	–	–	–	Type I	(60)
Petit dit Dariel et al.	FR	2013	PEM - Gameplay	Adapt the gameplay	Development	PC	Health	Training	–	–	–	–	Type I	(66)
Zaldivar et al.	MX	2013	PEM - Gameplay	Teaching fuzzy logic	Platform	PC	Educational	Edugame	–	–	–	–	Type II	(61)
Oliveira et al.	BR	2013	PEM - Gameplay	Adapt the gameplay	Development	PC	Business	Training	–	–	–	–	Type III	(63)
Farhangian et al.	NZ	2013	Game Data Mining	Classify users	Research	PC	Social Change	Training	Simulation study using NetLogo				Type III	(54)
De Oliveira et al.	BR	2012	PEM - Objective	Assess user state	Development	PC	Health	Health	–	–	–	–	Type III	(55)
Pirovano et al.	IT	2012	PEM - Objective	Monitor exercises execution online	Development	Online - PC	Health	Health	–	–	–	–	Type III	(53)
Saraiva et al.	BR	2011	PEM - Gameplay	Adapt the gameplay	Development	Online	Training	Training	–	–	–	–	Type III	(64)
Brasil et al.	BR	2011	PEM - Gameplay	Actions Gameplay	Development	Online	Training	Edugame	–	–	–	–	Type III	(65)
Imbeault et al.	CA	2011	PEM - Gameplay	Assess user state	Development	PC	Health	Training	20	–	–	–	Type III	(56)
Buditjahjanto and Miyauchi	ID	2011	PEM - Gameplay	Infer NPC behaviour	Development	PC	Social Change	Edugame	–	–	–	–	Type I	(58)
Shen et al.	SG	2010	PEM - Gameplay	Adapt the gameplay	Development	PC	Educational	Edugame	–	–	–	–	Type IV	(62)
Cai et al.	SG	2009	PEM - Gameplay	Foster immersion	Development	PC	Health	Edugame	–	–	–	–	Type I	(57)
Balas et al.	CZ	2008	PEM - Gameplay	Adapt the gameplay	Development	PC	Social change (civics)	Edugame	–	–	–	–	Type III	(51)
Djordjevich et al.	US	2008	Procedural Content Generation	Infer NPC behaviour	Development	PC	Military	Training	–	–	–	–	Type III	(59)

** (a) Type I: International Journals with Impact Factor, (b) Type II: International Journals, (c) Type III: International conferences, workshops and symposiums, (d) Type IV: Chapters in Books*

Curse game was tested by over 1400 real users, though authors did not provide any further analysis of these data (77).

5) Publication type: Publication channels in this section were: two articles published in international journals with impact factor, one study was available in an international journal, and five articles were published in international conferences, workshops and symposiums. The remaining two studies were published as book chapters.

D. Goal Oriented Behavior

Reviewing the literature, goal-oriented behaviour has been used for providing NPCs with a set of behaviours and guidelines for making them more realistic. Goal oriented behaviour deals with a series of techniques that produce action sequences so as to achieve a determined goal (80). *Goal Oriented Action Planning - GOAP* is composed of a set of goals and actions. One of its main goals is determined by the properties of the environment. Actions are described in terms of their effects on the environment and they are formulated with a planner (81).

Only six articles fulfilled the inclusion criteria of the goal oriented behaviour category. Please refer to Table IV for the complete list of coded articles in this section.

1) AI flagship: Two out of the six included articles fall into the category of *PEM - Gameplay based*.

Three studies were categorized as *Procedural Content Generation*. Two of them introduced the framework *cOnciens*. This framework described a set of tools for game AI developers. It was focused on model gaming scenarios using social structures (82; 83).

The last reviewed study in this section was categorized as *Game Data Mining*, which implemented techniques within the videogame industry in a mechanical assembly line (84).

2) AI usage: The majority of articles included in this section, four out of six, implemented GOAP techniques for controlling NPC behaviour (85; 86). Two of these four were additionally focused on modelling scenarios during gameplay (82; 83).

One article implemented GOAP techniques for evaluating users' responses (87).

A final article was focused on gameplay assembly plans. The authors tested feasibility through the use of a game focused on car repair, applying GOAP to the decision process of repairing (84).

3) Categorization of games: Articles were classified according to the following features:

- **Serious games market:** One article was categorized as educational (85), another was labelled under the field of business (84). The remaining four studies were uncategorised as regards of the serious game market. These four articles proposed a customizable framework able to be used in a variety of serious games' markets (82; 83).
- **Serious games purpose:** Considering the purpose of the game, included articles were divided as follows: One article was considered to be under the field of training (84), while another was classified under the edugames field (85). The remaining four were uncategorised as regards their purpose.
- **Platform/delivery:** All the revised studies were designed and/or implemented for PC.
- **Project type:** Five out of the six revised articles were framework proposals for serious games development. The remaining study was a new development for applying videogame techniques in a mechanical assembly line (84).

4) Sample of participants: Only one article was tested with real participants (87). The serious games framework proposed by Alvarez-Napagao *et al.* was tested in various settings and serious games for assessing its usability (82; 83).

5) Publication type: One of the articles was published in an international journal with impact factor (87), while the remaining five were published in international conferences, workshops and symposiums.

E. Rule-Based Systems

Rule-based systems are outlined as a way of storing and handling knowledge to interpret information in an efficient way (88). The most common use of this in the field of *Serious Games* is by using a set of pre-implemented guidelines using boolean operators in an expert system for assisting in decision making and tactical movements during the game.

Regarding the field of Serious Games, ten articles were found implementing these techniques (see Table V for a complete relation of coded articles).

1) AI flagship: The majority of reviewed articles in this category, eight out of ten, fall into the category of *PEM - Gameplay based*. For example, Khayat *et al.* create customized

TABLE III
MARKOV SYSTEMS AND SERIOUS GAMES - CODED ARTICLES

Authors	Country	Year	AI Flagship	AI Usage	Type	Platform	SG Market	SG Purpose	Participants				Publication*	Reference
									Total	Males	Females	Age		
Su et al.	GB	2014	Procedural Content Generation	Adapt gameplay	Framework	PC	Educational	Edugames	Simulated learners + 278 real				Type II	(70)
Navarro-Newball et al.	CO	2014	PEM - Gameplay	Infer information from players	Development	PC	Health	Health	–	–	–	–	Type I	(73)
Sucar *et al.*	MX	2013	PEM - Gameplay	Adapt gameplay	Development	PC	Health	Health	6	Feasibility pilot study with chronic stroke patients			Type I	(75)
Su *et al.*	TW	2013	Procedural Content Generation	Adapt gameplay	Framework	PC	Educational	Edugames	Simulated learners + 278 real				Type III	(70)
Derbali *et al.*	CA	2013	PEM - Objective	Assess players' motivation	Development	PC	Educational	Edugames	41	20	21	23.7 (group 1) 25.3 (group 2)	Type III	(69)
Rafferty *et al.*	US	2012	Game Data Mining	Infer information from players	Framework	Online	Educational	Edugames	225	–	–	–	Type III	(72)
Bouchard *et al.*	CA	2012	PEM - Gameplay	Adapt gameplay	Guidelines	PC	Health	Health	–	–	–	–	Type IV	(74)
Neto *et al.*	PT	2011	PEM - Gameplay	Adapt gameplay	Development	PC	Culture and activism	Edugames	More than 1400 (data not analyzed)				Type III	(77)
Muñoz *et al.*	IE	2010	PEM - Gameplay	Adapt gameplay	Development	PC	Educational	Edugames	28	–	–	University Students	Type IV	(68)
Johnson *et al.*	US	2007	PEM - Gameplay	Adapt gameplay	Development	PC	Military	Edugames	8	–	–	–	Type III Type IV	(78; 79; 76)

** (a) Type I: International Journals with Impact Factor, (b) Type II: International Journals, (c) Type III: International conferences, workshops and symposiums, (d) Type IV: Chapters in Books*

TABLE IV
GOAL ORIENTED BEHAVIOURS AND SERIOUS GAMES - CODED ARTICLES

Authors	Country	Year	AI Flagship	AI Usage	Type	Platform	SG Market	SG Purpose	Participants				Publication*	Reference
									Total	Males	Females	Age		
Qu *et al.*	NL	2013	Procedural Content Generation	Evaluate users' responses	Framework	PC	–	–	40	Real world + Virtual reality			Type I	(87)
Yu *et al.*	UK	2012	Game Data Mining	Generate assembly plans	Development	PC	Business	Training	–	–	–	–	Type III	(84)
Alvarez-Napagao *et al.*	ES	2012	Procedural Content Generation	Model scenarios and NPC behaviors	Framework	PC	–	–	–	–	–	–	Type III	(82)
Alvarez-Napagao *et al.*	ES	2010	Procedural Content Generation	Model scenarios and NPC behaviors	Framework	PC	–	–	–	–	–	–	Type III	(82)
Jepp *et al.*	PT	2010	PEM-Gameplay	NPC behaviors	Framework	PC	–	–	–	–	–	–	Type III	(86)
Rebolled-Mendez *et al.*	UK	2009	PEM - Gameplay	NPC Behaviour	Framework	PC	Educational	Edugames	–	–	–	–	Type III	(85)

** (a) Type I: International Journals with Impact Factor, (b) Type II: International Journals, (c) Type III: International conferences, workshops and symposiums, (d) Type IV: Chapters in Books*

sets of mini-games for children based on their performance in a IQ test and their responses to previous exercises (89).

One article was categorized as *Procedural Content Generation*. This study introduced a framework for creating adaptive educational Serious Games, giving the players the opportunity to create their own play experience (90).

The remaining study was uncategorised as regards of its AI flagship.

2) AI usage: Four out of ten studies were related to the assessment and evaluation of users in one way or another. For example, Arya *et al.* developed a fuzzy rule based system that study facial expressions and social competences in children with ASD to evaluate their emotions (91) while Patton *et al.* developed a Serious Game for guiding students with calligraphy problems. They evaluated user performance using a rule-based system (92).

Two articles made use of AI techniques with the end purpose of guiding users and responding to their utterances (93; 94).

Two studies were focused on modelling NPCs behaviours as a rule based system based on the requirements related to the character-based game development (95) or the iteration between some predefined behaviours of the game (96).

The remaining two articles that used AI techniques were related to the player experience. Players were able to create their own games using a set of rule-based decisions inside a framework for creating adaptive educational games (90). The last study was focused on inferring players' gameflow based on the recorded data and a set of rules (97).

3) Categorization of games: Articles were classified according to the following features:

- **Serious games market:** Two articles were categorized as part of the health market (89; 91), another four were labelled under education (92; 98; 90; 93). Three studies were classified under business (94; 95; 96). The remaining study was labelled under culture and activism (97)
- **Serious games purpose:** Considering the purpose of the game, the articles were divided as follows: Seven articles were considered to be edugames, while the remaining three were labelled under the term of training.
- **Platform/delivery:** While the majority of the studied serious games in this section were developed for PC, seven out of ten, the remaining three were implemented for tablet PC (92), mobile devices (90) and for uses online (89).
- **Project type:** Seven out of the ten consulted articles were new developments of Serious Games. The three remaining studies were framework proposals for building Serious Games.

4) Sample of participants: Four of the articles were tested with real participants.One of which gives accurate data regarding the gender and age of the participants (93). Two of the remaining three clarify that the preliminary tests were performed with undergraduate students. The last does not give any further information about the sample, only on the number of participants.

Two out of the three included frameworks in this section provide data about their performance. Both of them were tested for assessing their strengths (90; 98).

Four articles did not provide any information about participants or experimental procedures.

5) Publication type: Eight of the reviewed articles were published in international conferences, workshops and symposiums. The remaining two articles were published in international journals.

F. Finite-State Machines

Finite state machines constitute a mathematical computational model. They are represented as an abstract system that can be on one of a finite number of states each time. Specific events or conditions are needed to switch between different states, these switches being named as transitions. One specific finite state machine is defined by a list of states and a set of conditions for switching between them (99).

TABLE V
RULE-BASED SYSTEMS AND SERIOUS GAMES - CODED ARTICLES

Authors	Country	Year	AI Flagship	AI Usage	Type	Platform	SG Market	SG Purpose	Participants				Publication*	Reference
									Total	Males	Females	Age		
Sehaba and Hussaan	FR	2013	PEM - Gameplay	Evaluate users' performance	Framework	PC	Educational	Edugame	Comparative study between framework and expert system				Type II	(98)
Haladjan et al.	DE	2012	Procedural Content Generation	Users can create their own games	Framework	Mobile devices	Educational	Edugame	Multiple-case and quasi-experimental study				Type III	(90)
Khayat et al.	EG	2012	PEM - Gameplay	Assess users	Development	Online	Health	Training	–	–	–	–	Type III	(89)
Ward et al.	US	2012	PEM - Gameplay	Guide users	Development	PC	Educational	Edugame	11	7	4	11.1	Type III	(93)
Rojas-Barahona	FR	2012	PEM - Gameplay	Respond to players utterances	Development	PC	Business	Training	40	–	–	–	Type III	(94)
Souza *et al.*	BR	2009	PEM - Gameplay	Model NPC's behavior	Development	PC	Business	Training	–	–	–	–	Type III	(95)
Arya et al.	CA	2009	–	Identify user emotions	Framework	PC	Health	Edugame	50	–	–	University Students	Type II	(91)
Patton *et al.*	US	2008	PEM - Gameplay	Evaluate users' performance	Development	Tablet PC	Educational	Edugame	16	–	–	Senior and Graduate level students	Type III	(92)
Core, Gomboc et al.	US	2006	PEM - Gameplay	Rule based behaviors	Development	PC	Business	Edugame	–	–	–	–	Type III	(?)
El Rhalibi *et al.*	UK	2005	PEM - Gameplay	Gameflow	Development	PC	Culture and activism	Edugame	–	–	–	–	Type III	(97)

* *(a) Type I: International Journals with Impact Factor, (b) Type II: International Journals, (c) Type III: International conferences, workshops and symposiums, (d) Type IV: Chapters in Books*

Reviewing the literature concerning this technique, 19 articles were found. They were categorized as regards of the characteristics outlined in the methods section. Table VI illustrates all the coded studies.

1) AI flagship: The majority of reviewed articles in this category, eleven out of nineteen, fall into the category of *PEM) - Gameplay based*. For example, the Serious Game *Mission - Master your fear* is a therapeutic game for kindergarten children. The different levels and game scenarios were based on children previous interaction and they were implemented as a finite state machine (100).

Three studies were categorized as *PEM - Objective*. One of these articles included voice recognition algorithms for processing players' speech (101) while the other two implemented motion recognition systems based on accelerometers (102; 103). Using these sensors they were able to evaluate hand position, gesture recognition, and velocity among others for difficulty adjustment.

One article was labelled under the *Game Data Mining* category. In the Serious Game *Missing*, the developers applied machine learning techniques to the game log so as to evaluate different strategies and demographic variances between players (104)

The remaining four studies were uncategorised as regards of their AI flagship.

2) AI usage: The purpose of implementing finite state machines varies between a wide range of categories. The most common was for controlling the gameflow. Six articles were found in this field. They used finite state machines for switching between levels (100; 105), controlling the CPR actuation (106) or customizing the gameplay (107), among others.

Three articles were found in each of the following categories: Building the game behaviour (108; 109; 110), evaluating users (104; 111; 102) and controlling NPCs behaviours (112; 113; 114).

Two articles used finite state machines for controlling virtual characters in the game (103; 101).

The remaining article was developed as a game engine where AI techniques were available in a high-level scripting language (115). The remaining article was uncategorised as regards of AI purpose.

3) Categorization of games: Articles were classified according to the following features:

- **Serious games market:** Four articles were categorized as for the health market (101; 104; 106; 103), nine studies were under educational. Two articles were under business (111; 114). Another two were under culture and activism (102; 116). One article was categorized in the military field (115). The remaining article was uncategorised.
- **Serious games purpose:** Considering the purpose of the game, the articles were divided as follows: Eight articles were considered to be on edugames, while other nine were labelled under training and only one was categorized as health-related. The remaining article was uncategorised.
- **Platform/delivery:** The majority of included serious games in this section were developed for PC. The remaining two were implemented for mobile devices (102) and for being use online (114).
- **Project type:** Sixteen out of the nineteen consulted articles were new developments of Serious Games. The three remaining studies were classified as design methodology (112), game engine (115) and authoring environment (107).

4) Sample of participants: Eight of the articles were tested with real participants. Three of which give accurate data regarding the gender and age of the participants (101; 102; 117). Two of the remaining three clarify the age range of the participants. The authors were not able to access detailed information on participants in the remaining article.

The feasability of the Serious Game *Stories from the History of Czechoslovakia* was tested but no further analysis of system usability was outlined in the study (118).

The authoring environment *SeGAE* was applied to the Serious Game *Blossom Flowers* for adapting different aspects of the game and the character, proving its efficiency (107).

The remaining eleven articles did not give any further information whether they have been tested or not with real participants.

5) Publication type: Fifteen articles were published in international conferences, workshops and symposiums. The remaining four articles were made available through international journals with impact factor (102), international journals (117) and book chapters (116; 114).

G. Results in Decision Making

This section reveals the preliminary conclusions in the use of decision making techniques inside the *serious games* field. Firstly, all the included references in this section are outlined in tables I to VI, according to the implemented AI method. They are classified and a brief description of each study is

TABLE VI
FINITE STATE MACHINES AND SERIOUS GAMES - CODED ARTICLES

Authors	Country	Year	AI Flagship	AI Usage	Type	Platform	SG Market	SG Purpose	Participants				Publication*	Reference
									Total	Males	Females	Age		
Tan et al.	AU	2013	PEM - Objective	Control characters	Development	PC	Health	Training	2	1	1	7-10	Type III	(101)
Li et al.	US	2013	Game Data Mining	Evaluate users	Development	PC	Health	Training	67	–	–	28.4 (SD=11.2)	Type III	(104)
Morgan et al.	US	2013	PEM - Gameplay	Evaluate users	Development	PC	Business	Edugame	Tested but not further information				Type III	(111)
Wattanasoontorn et al.	ES	2012	PEM - Gameplay	Gameflow	Development	PC	Health	Training	–	–	–	–	Type III	(106)
Nyoman *et al.*	ID	2012	PEM - Gameplay	Gameflow	Development	PC	Educational	Training	–	–	–	–	Type III	(105)
Rehm and Leichtenstern	DK	2012	PEM - Objective	Evaluate users	Development	Mobile	Culture and Activism	Training	20	15	5	–	Type I	(102)
Sisler et al.	CZ	2012	PEM - Gameplay	Gameflow	Development	PC	Educational	Edugame	Feasability study. Game not tested				Type III	(118)
Ines and Abdelkader	FR	2011	PEM - Objective	Control characters	Development	PC	Health	Health (Rehab)	3	–	–	20-30	Type III	(103)
Szczesna et al.	PL	2011	PEM - Gameplay	Gameflow	Development	PC	Educational	Training	–	–	–	–	Type III	(100)
Wendel et al.	DE	2010	PEM - Gameplay	NPC Behaviours	Development	Online	Business	Edugame	–	–	–	–	Type IV	(114)
Sagae et al.	US	2010	PEM - Gameplay	Game behaviour	Development	PC	Educational	Training	–	–	–	–	Type III	(109)
Irvine and Gongora	GB	2010	–	NPC Behaviours	Methodology	PC	–	–	–	–	–	–	Type III	(112)
Yessad et al.	FR	2010	PEM - Gameplay	Customize gameplay	Authoring Environment	PC	Educational	Edugame	The approach was tested with the Serious Game Blossom Flowers.				Type III	(107)
Rankin and Vargas	AU	2009	–	Build the game behaviour	Development	PC	Educational	Edugame	–	–	–	–	Type III	(108)
Silverman et al.	US	2009	PEM - Gameplay	–	Development	PC	Culture and Activism	Edugame	–	–	–	–	Type IV	(116)
Florez-Puga et al.	ES	2008	PEM - Gameplay	NPC Behaviours	Development	PC	Educational	Edugame	–	–	–	–	Type III	(113)
Binsubaih et al.	UK	2006	PEM - Gameplay	Gameflow	Development	PC	Educational	Training	56	–	–	–	Type II	(117)
Darken et al.	US	2007	–	Code Games	Game Engine	PC	Military	Training	–	–	–	–	Type III	(115)
Duggan et al.	IE	2005	–	Build the game world	Development	PC	Educational	Edugame	–	–	–	–	Type III	(110)

** (a) Type I: International Journals with Impact Factor, (b) Type II: International Journals, (c) Type III: International conferences, workshops and symposiums, (d) Type IV: Chapters in Books*

outlined in previous sections while main statistical outcomes regarding the decision making field are summarized in this section.

1) AI flagship: Table VII displays the number of publications segmented by AI flagship type and decision making method.

TABLE VII
AI FLAGSHIP - DECISION MAKING TRENDS

Technique	N	Player Experience Modeling			Procedural Content Generation	Game Data Mining	Uncat.
		Subjective	Objective	Gameplay			
Decision Trees	27	0	0	19	0	4	4
Fuzzy Logic	16	0	3	11	1	1	0
Markov Systems	10	0	1	6	2	1	0
Goal Oriented Behavior	6	0	0	2	3	1	0
Rule-based Systems	10	0	0	8	1	0	1
Finite-State Machines	19	0	3	11	0	1	4
Total:	88	0	7	57	7	8	9
Percentage:	100	0.00%	7.95%	64.77%	7.95%	9.09%	10.23%

As it is shown in Table VII, most of the articles revised in this section, 64.77%, fell under the category of *PEM - Gameplay* (57 out of 88).

The remaining 31 studies were divided between the *PEM - Objective*, Procedural Content Generation and Game Data Mining categories.

Nine articles remained uncategorised as regards of AI flagship, while no one was labelled under the *PEM - Subjective* field.

2) Categorization of games: Articles were classified according to the following features:

– **Serious games market:** Table VIII displays the number of publications segmented by the serious games market and decision making method. As is displayed in Table VIII, the markets with the highest number of articles in their category are education (34.09%) and health (28.41%). Decision trees algorithms were the most used in health, education and business markets. Fuzzy logic and markov systems were also popular in health, education and social change.

– **Serious games purpose:**Table IX displays the number of publications according to the serious games purpose and decision making method. Edugames and training are the categories with the highest number of labelled articles, with 42.05% and 36.36% of the included articles in each category respectively. Regarding the purpose, several algorithms were balanced used in the edugame segment: decision trees, fuzzy logic, markov systems, rule-based systems and finite state machines. In training, finite state machines and decision trees were the most used. The remaining segments were balanced.

TABLE VIII
SERIOUS GAMES MARKET - DECISION MAKING TRENDS

Technique	N	Health	Education	Business	Culture and Activism	Social Change	Political	Military	Uncat.
Decision Trees	27	10	8	4	2	0	1	1	1
Fuzzy Logic	16	6	3	3	0	3	0	1	0
Markov Systems	10	3	5	0	1	0	0	1	0
Goal Oriented Behavior	6	0	1	1	0	0	0	0	4
Rule-based Systems	10	2	4	3	1	0	0	0	0
Finite-State Machines	19	4	9	2	2	0	0	1	1
Total:	88	25	30	13	6	3	1	4	6
Percentage:	100	28.41%	34.09%	14.77%	6.82%	3.41%	1.14%	4.55%	6.82%

TABLE IX
SERIOUS GAMES PURPOSE - DECISION MAKING TRENDS

Technique	N	Health	Edugame	Training	Activism	Uncat.
Decision Trees	27	4	7	13	2	1
Fuzzy Logic	16	3	7	6	0	0
Markov Systems	10	3	7	0	0	0
Goal Oriented Behavior	6	0	1	1	0	4
Rule-based Systems	10	0	7	3	0	0
Finite-State Machines	19	1	8	9	0	1
Total:	88	11	37	32	2	6
Percentage:	100	12.50%	42.05%	36.36%	2.27%	6.82%

– **Platform/delivery:** Table X displays the number of publications according to platform/delivery and decision making method. The vast majority of the reviewed articles, 80.68%, were developed for PC. The second category with the highest number of references is online serious games, and the last mobile devices.

– **Project type:** Table XI displays the number of publications segmented by platform/delivery and decision making method. Of the 88 available studies, 63 of them were concerned with new developments of serious games,

TABLE X
PLATFORM/DELIVERY - DECISION MAKING TRENDS

Technique	N	PC	Mobile/Tablet	Online	Other
Decision Trees	27	19	3	5	0
Fuzzy Logic	16	13	0	3	0
Markov Systems	10	9	0	1	0
Goal Oriented Behavior	6	6	0	0	0
Rule-based Systems	10	7	2	1	0
Finite-State Machines	19	17	1	1	0
Total:	88	71	6	11	0
Percentage:	100	80.68%	6.82%	12.50%	0.00%

while 19 were classified as frameworks. Only under the category of Goal Oriented Behaviours was the number of articles in frameworks higher than the number on new developments.

TABLE XI
PROJECT TYPE - DECISION MAKING TRENDS

Technique	N	Development	Framework	Other
Decision Trees	27	20	7	0
Fuzzy Logic	16	13	1	2
Markov Systems	10	6	3	1
Goal Oriented Behavior	6	1	5	0
Rule-based Systems	10	7	3	0
Finite-State Machines	19	16	0	3
Total:	88	63	19	6
Percentage:	100	71.59%	21.59%	6.82%

3) Sample of participants: Of the 88 articles screened, 39 were tested in one way or another while 49 remained untested. Thus, 39 out of 88 articles did perform some kind of user test with real participants, or, in the case of frameworks, did tested their approach in a real context. The remaining 49 did not provide further information about any testing procedure, and were focused on the system's architecture, implementation and design.

4) Publication type: Table XII displays the number of publications according to the publishing channel and decision making method. Half of the revised articles were available through international conferences, workshops and symposiums. The remaining articles were more or less balanced in number between the remaining three categories: International journals with impact factor, international journals and as chapters in books.

TABLE XII
PUBLICATION TYPE - DECISION MAKING TRENDS

Technique	N	Type I	Type II	Type III	Type IV
Decision Trees	27	7	8	8	4
Fuzzy Logic	16	5	1	9	1
Markov Systems	10	2	1	5	2
Goal Oriented Behavior	6	1	0	5	0
Rule-based Systems	10	0	2	8	0
Finite-State Machines	19	1	1	15	2
Total:	88	16	13	50	9
Percentage:	100	18.18%	14.77%	56.82%	10.23%

** (a) Type I: International Journals with Impact Factor, (b) Type II: International Journals, (c) Type III: International conferences, workshops and symposiums, (d) Type IV: Chapters in Books*

IV. MACHINE LEARNING

A system learns from an experience when it improves in the carrying out of a determined task, optimizing its performance. In the *serious games* area this learning from experience is applied to the patterns of the users themselves, and also employed in the optimization of the behaviour of Non Player Characters and the agents of play, providing them with greater realism

The intrinsic problem of algorithms of learning in the games field in general and of serious games in particular is the amount of usage time required to generate an effective learning system (119).

Below is a summary of the technical literature dealing with the most used techniques in this application theme.

A. Näive Bayes Classifier

Naive Bayes classifiers constitute a supervised classification and prediction technique based on Bayes theorem. It is a learning and prediction technique based on storing a series of attribute-class pairs in a contingency table. All the attributes that have a class are independent from each other. New attributes are incorporated into one class or other according to the probability they have of belonging to one or other class (20).

Reviewing the literature concerning this technique, thirteen articles were found. They were categorized as regards of the characteristics outlined in the methods section. Please refer to Table XIII for the complete list of coded articles in this section.

1) AI flagship:: Nine out of the thirteen articles included in this section fell into the *PEM - Gameplay based* category. For example, Muñoz *et al* developed a serious game for teaching physics to engineering students. A Näive Bayes classifier was used to adapt the gameplay based on users' performance (68).

Two articles were classified as *PEM - Objective*. An EEG device was used in one of them for the objective measurement of determining the risk of developmental problems among users (120).

The remaining articles were classified under the *Game data mining* label as regards of the AI flagship. The serious game *MyHealthyKids* made use of Näive Bayes to predict, based on the information gathered during the gameplay, whether a child is prone to be obese or not (121).

2) AI usage:: The majority of the included articles in this section used Näive Bayes Classifier to analyse users' gameplay and classify them as regards of their performance.

The remaining four articles used Näive Bayes for assess users (122), adapt the gameflow (25; 68) and analyse players' interaction (18).

3) Categorization of games:: Articles were classified according to the following features:

– **Serious games market:** Reviewed articles were quasi-balanced as regards of the serious games market they belong to. Three articles out of thirteen were labelled under health field, three were both classified under education and under business field. The remaining four were

unclassified as regards of the serious games market they belong to.
– **Serious games purpose:** Included articles in this section were classified as follows: Six articles were under the edugame category, another two were training games. The remaining four were unclassified.
– **Platform/delivery:** Ten out of the thirteen reviewed articles were developed for PC, while one was a web-mobile development and the remaining two were studies with no specified platform/delivery.
– **Project type:** Nine of the reviewed articles were new development of serious games while three were framework designs and one was a study about the implications of motivation and attention during gameplay.

4) Sample of participants: All the included articles in this section were tested in one way or another. Nine articles were tested with real participants. Two other papers were tested over real datasets (123; 124). The remaining two articles were tested upon an existed development, so as to test their approach in real serious games (18; 122).

5) Publication type: Publication channels were segmented as follows: one article was published in an international journal with impact factor, three papers appeared in international journals, while other five were published as book chapters and in international conferences, workshops and symposiums respectively.

B. Artificial Neural Networks

Artificial neural networks constitute a paradigm of learning. ANNs are composed of a set of "neurons" or states which are defined by a set of relevant features. These systems receive a group of inputs and generate a specific output. The connections between neurons are associated with a determined probabilistic weight, which gives knowledge to the system. ANNs emulate human brain behaviour (20).

Twelve articles described research that employed artificial neural networks and could be further analysed and coded following the categories outlined during the methods section. Please refer to Table XIV for the complete list of coded articles in this section.

1) AI flagship:: The majority of reviewed articles in this section, eight out of twelve, were labelled under *PEM - Gameplay* category as regards of their AI flagship.

Three articles took different measurements of players' actions on input devices such as eyetracker, motion capturing sensors or even joysticks while playing (130; 131; 132). They were under the *PEM - Objective* category.

The remaining article was developed as a game data mining system. It was a cognitive diagnostics tool for serious educational games (133).

2) AI usage:: Six articles were focused on altering gameflow based on the result of an implemented artificial neural network.

One framework (132), one study (133) and three developments (134; 135; 131) were focused on user evaluation and classification.

The remaining article made use of artificial neural networks for controlling NPC behaviours (85).

3) Categorization of games:: Articles were classified according to the following features:
– **Serious games market:** Seven articles under educational market. Three were under health field (130; 131; 132) and one belonged to culture and activism (136). The remaining study was uncategorised.
– **Serious games purpose:** Six articles were classified as edugames, three were under training (136; 137; 130), two were health-related (131; 132). The remaining article was uncategorised.
– **Platform/delivery:** All but one of the included studies were developed for PC. The remaining article described a system developed for mobile devices (137).
– **Project type:** Six articles were new developments of serious games, while three studies were framework designs and implementations (138; 85; 132). The remaining three were research studies (139; 140; 133).

4) Sample of participants: Four articles did not give any further information about their testing procedures.

Three studies were pilot-tested, one of them gave the number of participants (136) while another only refers being tested with a small pilot group (135). The last pilot study performed the testing procedure with one single user (131).

Four articles gave age of the participants, but gender were only given in two of them (136; 134).

Two out of the three included studies applied their findings to the analysis of existing serious games (139; 140).

5) Publication type: The majority of articles were published in international conferences, workshops and symposiums.

The remaining three articles were published in an international journal with impact factor (133), in an international journal (134) and as a book chapter (140).

C. Case-Based Reasoning

Case-based reasoning is the technique based on the ability or capacity to solve or determine the outputs of the system based on the solutions adopted before. Systems implementing this technique start with a known training data-set from which a set of knowledge is drawn for future references.

Case-based reasoning techniques are applied in the field of interactive *serious games* mainly when they are related to automatic customizations of its contents based on user interaction. These techniques provide systems with an efficient way of learning from users, providing them with customized and personal experiences (143).

Ten articles were found in this section that fulfilled the inclusion criteria for this review in the category of case based reasoning. Please refer to Table XV for the complete list of coded articles in this section.

1) AI flagship:: All but one of the included references fell into *PEM - Gameplay* category.

The remaining study was oriented towards Game Data Mining. It was an automobile supply chain simulation game, in where users actions and decisions taken were knowledge-inputs to the case-based reasoning system, improving its response for future situations (144).

TABLE XIII
NÄIVE BAYES CLASSIFIER AND SERIOUS GAMES - CODED ARTICLES

Authors	Country	Year	AI Flagship	AI Usage	Type	Platform	SG Market	SG Purpose	Participants				Publication*	Reference
									Total	Males	Females	Age		
Tseng and Wu	TW	2014	PEM - Objective	Classify Users	Development	PC (w/ online server)	Health	Exergame	85	36	49	55 - 85	Type II	(125)
Feldman *et al.*	AR	2014	PEM - Gameplay	Classify users	Framework	PC	–	–	47	–	–	University Students	Type I	(126)
Husain et al	MY	2013	Game Data Mining	Classify Users	Development	Web-Mobile	Health	Edugame	15 (Questionnaires) 30 (Children Persuasive Module)	– –	– –	26 - 48 –	Type III	(121)
Mori *et al.*	IT	2013	PEM - Gameplay	Interaction	Framework	PC	–	–	Developed a game to test the aproach of conversational agents in SG in a real context				Type II	(18)
Sabourin *et al.*	US	2013	PEM - Gameplay	Gameflow	Development	PC	Educational	Edugames	260	129	131	13.4 (SD = 0.57)	Type II	(25)
Keshtkar *et al.*	US	2012	PEM - Gameplay	Classify users	Development	PC	Business	Edugames	Tested over dataset with 83% accuracy				Type III	(123)
Nor Wan *et al.*	MY	2012	PEM - Gameplay	Classify users	Development	PC	Health	Health	31	10	21	over 65	Type III	(127)
Weichselbraun *et al.*	CH	2011	Game Data Mining	Assess users	Framework	–	–	–	Tested upon Sentiment Quiz, a web-based social verification game for sentiment detection.				Type III	(122)
Derbali *et al.*	CA	2011	PEM - Objective	Classify Users	Study	–	–	–	33	22	11	26.7 (SD 4.1)	Type III	(120)
Morgan *et al.*	US	2011	PEM - Gameplay	Classify users	Development	PC	Business	Edugames	21	–	–	High School	Type III	(128)
Muñoz *et al.*	IE	2010	PEM - Gameplay	Adapt gameplay	Development	PC	Educational	Edugames	28	–	–	University Students	Type III	(68)
Vaassen and Daelemans	NL	2010	PEM - Gameplay	Classifiy users	Development	PC	Business	Training	339 Dutch sentences				Type III	(124)
Rowe *et al.*	US	2009	PEM - Gameplay	Classifiy users	Development	PC	Educational	Edugame	59	–	–	Undergrad.	Type III	(129)

** (a) Type I: International Journals with Impact Factor, (b) Type II: International Journals, (c) Type III: International conferences, workshops and symposiums, (d) Type IV: Chapters in Books*

TABLE XIV
ARTIFICIAL NEURAL NETWORKS AND SERIOUS GAMES - CODED ARTICLES

Authors	Country	Year	AI Flagship	AI Usage	Type	Platform	SG Market	SG Purpose	Participants				Publication*	Reference
									Total	Males	Females	Age		
Lamb *et al.*	US	2014	Game Data Mining	Gather user info	Study	PC	Educational	Edugame	–	–	–	–	Type I	(133)
Zhan *et al.*	US	2014	PEM - Objective	Gameflow	Development	PC	Health	Training	7	–	–	13 - 17	Type III	(130)
Gigliotta *et al.*	IT	2014	PEM - Gameplay	Gameflow	Study	PC	Educational	Edugame	Analyse different serious games				Type IV	(140)
Miglino *et al.*	IT	2012	PEM - Gameplay	Gameflow	Study	PC	Educational	Edugame	Analyse different serious games				Type III	(139)
Syufagi *et al.*	ID	2012	PEM - Gameplay	Classify users	Development	PC	Educational	Edugames	33	18	15	16 - 19	Type II	(134; 141; 142)
Otis *et al.*	CA	2012	PEM - Objective	Evaluate users	Development	PC	Health	Health	1	1	0	–	Type III	(131)
Grappiolo *et al.*	DK	2011	PEM - Gameplay	Gameflow	Development	PC	Culture and activism	Training	6 (Pilot study)	1	5	19 - 58	Type III	(136)
Puzenat and Verlut	FR	2010	PEM - Objective	Evaluate users state	Framework	PC	Health	Health	36	–	–	2 - 8	Type III	(132)
Arrabales *et al.*	ES	2009	PEM - Gameplay	Gameflow	Framework	PC + Bots	–	–	–	–	–	–	Type III	(138)
Rebolled-Mendez *et al.*	GB	2009	PEM - Gameplay	NPC Behaviour	Framework	PC	Educational	Edugames	–	–	–	–	Type III	(85)
Hildmann *et al.*	GB	2008	PEM - Gameplay	Gameflow	Development	Mobile	Educational	Training	–	–	–	–	Type III	(137)
Mills and Dalgarno	AU	2007	PEM - Gameplay	Evaluate users behavior	Development	PC	Educational	Edugame	Small pilot-test group				Type III	(135)

** (a) Type I: International Journals with Impact Factor, (b) Type II: International Journals, (c) Type III: International conferences, workshops and symposiums, (d) Type IV: Chapters in Books*

2) AI usage:: Eight articles implemented case-based reasoning techniques for adapting the gameflow, providing final users with personalized gaming experiences.

One of the included references used case-based reasoning for users' responses evaluation (144).

The remaining article was uncategorised as regards of its AI usage.

3) Categorization of games:: Articles were classified according to the following features:

– **Serious games market:** Four articles were labelled under educational field and 3 under business. Remaining three articles were classified as culture and activism (145), social change (146) and uncategorised respectively.
– **Serious games purpose:** Five articles were labelled for having an edugame purpose and other four inside the training category. The remaining article was uncategorised as regards of its purpose.
– **Platform/delivery:** Five of the developments and framework studied were intended for PC while the remaining four were designed for use online.
– **Project type:** Four articles were new serious games developments, while other four were frameworks for the creation of new games. The remaining two articles were: game engine implementation (147) and serious games design model (146).

4) Sample of participants: Three of the articles did test their developments with real participants. The *Rashi Game* was tested in a pilot study with students (148). No further information was given about the number of participants, age of gender. Operation ARA creators referred having tested their development with a group of 81 undergraduate college students (149). The serious game *The Life of Moses* was tested with a small gender-balanced group of undergraduate students (150).

5) Publication type: Seven articles were published in international conferences, workshops and symposiums while the remaining three were published as book chapters.

D. Support Vector Machines

Support vector machines are a set of supervised algorithms focused on data analysis and pattern recognition. Data is clustered using hyperplanes in a multidimensional space that maximizes separation between classes. This clustering is based on decision boundaries.

Six articles were found in this section fulfilling the inclusion criteria of support vector machines category. Please refer to Table XVI for the complete list of coded articles in this section.

1) AI flagship:: Five articles fell into the category of *PEM - Objective* as regards of their AI flagship. They made use of sensors, such as EEG (154; 155; 156) or eyetracker (157; 158), for recording biological data while playing. The remaining article was labelled as *PEM - Gameplay* (159).

2) AI usage:: The included articles were focused in predicting (159), evaluating (155) or classifying (154**?** ; 156) users behaviours while playing. One article used support vector machines for controlling NPC behaviours (160).

3) Categorization of games:: Articles were classified according to the following features:

– **Serious games market:** Three articles were classified as belonging for health field (154; 158; 160) other two were under military field (159; 155). The remaining article was uncategorised.
– **Serious games purpose:** Considering the purpose of the serious game, two articles were labelled both for

TABLE XV
CASE BASED REASONING AND SERIOUS GAMES - CODED ARTICLES

Authors	Country	Year	AI Flagship	AI Usage	Type	Platform	SG Market	SG Purpose	Participants				Publication*	Reference
									Total	Males	Females	Age		
Hulpus *et al.*	IE	2014	PEM - Gameplay	Gameflow	Framework	PC	Educational	Edugame	–	–	–	–	Type IV	(147)
Forsyth *et al.*	US	2013	PEM - Gameplay	Gameflow	Development	PC	Educational	Edugame	81	–	–	Undergraduate college students.	Type IV	(149)
Triki *et al.*	TN	2013	PEM - Gameplay	–	Model	Online	Social Change	Training	–	–	–	–	Type III	(146)
Tobail *et al.*	IE	2012	PEM - Gameplay	Gameflow	Framework	Online	Business	Training	–	–	–	–	Type III	(151)
Floryan & Woolf	US	2011	PEM - Gameplay	Gameflow	Development	PC	Educational	Edugame	Pilot study with students				Type III	(148)
Lopes & Bidarra	NL	2011	PEM - Gameplay	Gameflow	Framework	PC	–	–	–	–	–	–	Type III	(152)
Tobail *et al.*	IE	2011	Game Data Mining	Evaluate users' responses	Framework	Online	Business	Training	–	–	–	–	Type III	(144)
Hulpus *et al.*	IE	2010	PEM - Gameplay	Gameflow	Game Engine	PC	Business	Edugame	–	–	–	–	Type IV	(153)
McKenzie & McCalla	CA	2009	PEM - Gameplay	Gameflow	Development	Online	Culture and activism	Training	–	–	–	–	Type III	(145)
Chiann-Ru	TW	2006	PEM - Gameplay	Gameflow	Development	Online	Educational	Edugame	10	5	5	Undergraduate students	Type III	(150)

* *(a) Type I: International Journals with Impact Factor, (b) Type II: International Journals, (c) Type III: International conferences, workshops and symposiums, (d) Type IV: Chapters in Books*

having health-related purposes (154; 155) and training purposes (160; 158). One article was classified for being an edugame (159),and the last one was uncategorised as regards of its final purpose (156).

– **Platform/delivery:** All the included articles were designed and/or developed for PC.
– **Project type:** Two of the articles were new developments of serious games (158; 159), while other three were new framework designs (155; 156; 160). The remaining article was a game system implementation (154)

4) Sample of participants: All the articles were tested with real participants. Three of them gave further information about gender segmentation (156; 160) and age of the participants (155; 160).

5) Publication type: Three articles were published in international journals with impact factor, while other three were made available through international conferences, workshops and symposiums.

E. Results in Learning

This section exposes preliminary conclusions in the use of machine learning techniques inside the *Serious Games* field. Firstly, all the included references in this section are outlined in tables XIII to XVI, according to the implemented AI method. They are classified and a brief description of each study is outlined in previous sections while main statistical outcomes regarding the decision making field are summarized in this section.

1) AI flagship: Table XVII displays the number of publications segmented by AI flagship type and machine learning method.

TABLE XVII
AI FLAGSHIP - MACHINE LEARNING TRENDS

Technique	N	Player Experience Modeling			Procedural Content Generation	Game Data Mining	Uncat.
		Subjective	Objective	Gameplay			
Näive Bayes Classifier	13	0	2	9	0	2	0
Artificial Neural Networks	12	0	3	8	0	1	0
Case-Based Reasoning	10	0	0	9	0	1	0
Support Vector Machines	6	0	5	1	0	0	0
Total:	41	0	10	27	0	4	0
Percentage:	100	0.00%	24.39%	65.85%	0.00%	9.76%	0.00%

As it is shown in Table XVII, most of the articles revised in this section, 65.85%, fell under the category of PEM - Gameplay (27 out of 41).

The remaining 14 studies were divided between PEM-Objective (10 out of 41) and Game Data Mining (4 out of 41) categories.

None of the reviewed articles in the machine learning section were labeled under *PEM - Subjective* or under *Procedural Content Generation* flagships.

2) Categorization of games: Articles were classified according to the following features:

– **Serious games market:** Table XVIII displays the number of publications segmented by the serious games market and machine learning method. As is displayed in Table XVIII, the market with the highest number of articles in its category is education (34.15%), followed by health market (21.95%) and business (14.63%). The reviewed algorithms are quite balanced in this section, however, artificial neural networks stands out in its use inside the education market.

TABLE XVIII
SERIOUS GAMES MARKET - MACHINE LEARNING TRENDS

Technique	N	Health	Education	Business	Culture and Activism	Social Change	Political	Military	Uncat.
Näive Bayes Classifier	13	3	3	3	0	0	0	0	4
Artificial Neural Networks	12	3	7	0	1	0	0	0	1
Case-Based Reasoning	10	0	4	3	1	1	0	0	1
Support Vector Machines	6	3	0	0	0	0	0	2	1
Total:	41	9	14	6	2	1	0	2	7
Percentage:	100	21.95%	34.15%	14.63%	4.88%	2.44%	0.00%	4.88%	17.07%

– **Serious games purpose:** Table XIX displays the number of publications segmented by the serious games purpose and machine learning method. Edugames and training are the categories with the highest number of labelled articles, with 43.90% and 26.83% of the included articles in each category respectively. In this section the 17.07% of the reviewed serious games were uncategorised as regards of their purpose. None of the articles were labelled under the activism field.

TABLE XIX
SERIOUS GAMES PURPOSE - MACHINE LEARNING TRENDS

Technique	N	Health	Edugame	Training	Activism	Uncat.
Näive Bayes Classifier	13	1	6	2	0	4
Artificial Neural Networks	12	2	6	3	0	1
Case-Based Reasoning	10	0	5	4	0	1
Support Vector Machines	6	2	1	2	0	1
Total:	41	5	18	11	0	7
Percentage:	100%	12.20%	43.90%	26.83%	0.00%	17.07%

– **Platform/delivery:** Table XX displays the number of publications according to platform/delivery and machine learning method. The vast majority of the reviewed

TABLE XVI
SUPPORT VECTOR MACHINES AND SERIOUS GAMES - CODED ARTICLES

Authors	Country	Year	AI Flagship	AI Usage	Type	Platform	SG Market	SG Purpose	Participants				Publication*	Reference
									Total	Males	Females	Age		
Bernardini *et al.*	UK	2013	PEM - Objective	Classify users	Development	PC	Health	Training	19	–	–	–	Type I + Type III	(157; 158)
Berta *et al.*	IT	2013	PEM - Objective	Classify users	Framework	PC	–	–	22	17	5	Engineering, Masters and PhD students.	Type I	(156)
Parsons & Reinebold	US	2012	PEM - Objective	Evaluate users	Framework	PC	Military	Health	337	–	–	21 - 36	Type I	(155)
Ahn & Lee	KR	2011	PEM - Objective	Classify users	Game System	PC	Health	Health	10	–	–	–	Type III	(154)
Raybourn *et al.*	US	2010	PEM - Gameplay	Predict users' behaviours	Development	PC	Military	Edugame	78	–	–	–	Type III	(159)

** (a) Type I: International Journals with Impact Factor, (b) Type II: International Journals, (c) Type III: International conferences, workshops and symposiums, (d) Type IV: Chapters in Books*

articles, the 78.05%, were developed for PC. The second category with the highest number of references is online serious games developments (12.20%), and the last mobile devices or other platforms (4.88% each).

TABLE XX
PLATFORM/DELIVERY - MACHINE LEARNING TRENDS

Technique	N	PC	Mobile / Tablet	Online	Other
Näive Bayes Classifier	13	10	1	0	2
Artificial Neural Networks	12	11	1	0	0
Case-Based Reasoning	10	5	0	5	0
Support Vector Machines	6	6	0	0	0
Total:	41	32	2	5	2
Percentage:	100%	78.05%	4.88%	12.20%	4.88%

– **Project type:** Table XXI displays the number of publications segmented by platform/delivery and decision making method. 21 out of the 41 available articles were concerned with new developments of serious games, while 13 were classified as frameworks. The number of frameworks and developments are very balanced under the categories of case-based reasoning and support vector machines.

TABLE XXI
PROJECT TYPE - MACHINE LEARNING TRENDS

Technique	N	Development	Framework	Other
Näive Bayes Classifier	13	9	3	1
Artificial Neural Networks	12	6	3	3
Case-Based Reasoning	10	4	4	2
Support Vector Machines	6	2	3	1
Total:	41	21	13	7
Percentage:	100	51.22%	31.71%	17.07%

3) Sample of participants: 30 out of 41 articles did perform any kind of user test with real participants, or, in the case of frameworks, did tested their approach in a real context. The remaining 11 did not provide further information about any testing procedure.

4) Publication type: Table XXII displays the number of publications according to the publishing channel and machine learning method. 60.98% of the included articles were made available through International conferences, workshops and symposiums. The 19.51% was published as chapters in books while the 12.20% appeared in international journals with impact factor and the remaining 9.76% was published in international journals.

TABLE XXII
PUBLICATION TYPE - MACHINE LEARNING TRENDS

Technique	N	Type I	Type II	Type III	Type IV
Näive Bayes Classifier	13	1	3	5	4
Artificial Neural Networks	12	1	1	9	1
Case-Based Reasoning	10	0	0	7	3
Support Vector Machines	6	3	0	4	0
Total:	41	5	4	25	8
Percentage:	100%	12.20%	9.76%	60.98%	19.51%

** (a) Type I: International Journals with Impact Factor, (b) Type II: International Journals, (c) Type III: International conferences, workshops and symposiums, (d) Type IV: Chapters in Books*

V. DISCUSSION & CONCLUSION

The number and distribution of papers identified using the search terms listed in this article confirm the large and increasing interest in the serious games field related to artificial intelligent algorithms and techniques. Our inclusion criteria identified 129 papers providing evidence for the development and integration of some specific AI algorithms related to fields of decision making and machine learning.

The papers selected for this review were very diverse in terms of their area of research and application, market and purpose, and project and delivery type, reflecting the varied backgrounds of the researchers and their different interests in serious games. However, all of them have the AI field in common. This, together with the multi-feature analysis performed in this article, helped the authors to create a common analysis framework for organizing and comparing all the included studies.

Results are quite consistent between the two main sections of these articles, which may help to establish some trends regarding the use of decision making and machine learning algorithms in serious games design and development.

a) Algorithms used: Considering the decision making section, the number of articles included under the umbrella of decision trees is considerably higher than the studies found for the rest of algorithms. Decision trees are computationally undemanding, and have been found useful as components of intelligent systems (161).

However, in the machine learning section, the included number of articles is quite balanced between näive bayes classifiers, artificial neural networks and case-based reasoning algorithms. Support vector machine algorithms had the lower inclusion rate in the machine learning section. The vast majority of the SVM studies in this review were focused on classifying, evaluating or predicting users' states and behaviours using biological sensors. SVM are supervised classifiers that have been improved to successfully work with limited quantity and quality of training samples (162), which may help in their implementation with biological data.

b) AI flagship: Some 65% of articles included in both decision making and machine learning sections belongs to the *PEM - Gameplay* flagship. This is an interesting outcome since both of the reviewed sections produced the same result. According to Georgios N. Yannakakis, Gameplay-based PEM is the less intrusive and most computationally efficient approach for games (19), which may be the reason of its high incidence in this review.

PEM-Subjective flagship did not have representation in this review, which may be because authors were more focused in algorithm-centered studies.

c) AI usage: AI techniques were applied with a wide variety of final purposes for each of the articles. The most common implementations were for altering the gameflow or for assessing/classifying users' state and behaviour while playing. The production of intelligent serious games that dynamically adapt themselves to users' needs and performance have been proved to be efficient in terms of improvement comparisons (163).

d) Serious games market: Education and health markets were the most widespread markets considering reviewed algorithms. No recent key figures were found regarding serious games market analysis. However, the market study published by IDATE which ranges from 1952 to 2009 proved that serious games in most cases were designed for education (164).

e) Serious games purpose: Edugame related purposes were the most employed purpose in this review. This may be because the use of serious games for learning purposes is already established (165).

f) Platform/Delivery: The vast majority of the articles included in both decision making and machine learning categories were designed and/or developed for PC. Studies that involved online or mobile serious games were in discreet middle distance. Similar results were found by Connolly *et al.* in his review about computer and serious games (165).

However, in recent years the number of smartphones sold to end users worldwide has increased sharply. In 2013 the number of smartphones sold to customers increased over fifty percent on figure from 2011. This means that almost 20% of world's population owned a smart device. This figure is expected to grow to 34% by 2017 (166). This increase in the use of portable devices connected to the Internet could mark a change in trends regarding serious games delivery platform.

g) Testing procedure: Only about a half of all the articles included in this review, thus 69 out of 129, did perform some kind of user test with real participants, or, in the case of frameworks did test their approach in a real context. The serious games community is very fragmented, an creating a systematic procedure for validating the usability and penetration of serious games may be a challenge (167). However, authors consider that serious games are closely related to final users, so they need to be tested with real participants, so as to ensure their usability and effectiveness. Thus, available studies regarding the use of artificial intelligence in serious games may need to improve their testing procedures. Testing procedures in serious games need to be carefully designed, since multiple variables take part in them. Pre and post test assessment protocols may help in testing efficacy and efficiency in serious games and in their learning outcomes. Serious games also imply that they are, above all, games, so fun and engagement elements must be also ensured and tested (168).

It may also be useful to refine our working classification of outcomes proposing a higher level classification for the testing procedure. Thus, focusing on the appropriateness and adequacy of the employed methodology by analysing and categorizing study design, sampling, and data collection methods.This higher level classification will help to understand the underlying testing procedures, providing the community with a better understanding of the testing protocols used in serious games design. Serious games that use AI techniques should also consider the performance of those algorithms in the final game. Intelligent games should be carefully designed and algorithms need to perfectly suit the final objectives of the game.

h) Publication type: More than half of the articles included in this review were available through international conferences workshops and symposiums. Traditionally, there was a preference for conference publication in the field of computer science (169). Articles included in this review were published in 56 different international conferences, workshops and symposiums. The conference with the higher number of publications in this category was *VS-GAMES*. Only eight out of these 56 conferences were specifically oriented to either games or serious games design and development. Nine out of the 56 conferences were indexed in the computer science conference rankings. Regarding journals, 21 papers were published in international journals with impact factor, these publications were distributed as follows: 8 articles were found in Q1 journals, other 5 were in Q2 journals, other 6 were in Q3 and the remaining 2 were published in Q4 journals. These figures suggest a growing interest for the serious games field that goes beyond specialized conferences.

The current study has extended our understanding in the trends an tendencies of the categorization of AI techniques inside the serious games field. It is important to determine whether our classification is the most useful way of characterizing the different outcomes. Although several aspects related to the serious games field and AI algorithms have been studied, there was some ambiguity about which category an outcome should be code under. For example, in the case of AI flagships, sometimes the *PEM - gameplay* overlaps with *Game Data Mining*, and how they are viewed depends upon the perspective of the reader. However, due to the number of references inside the *PEM - Gameplay* field, this did not affect the final outcome for this trend analysis.

AI techniques offer significant potential in the development of serious games, enhancing the player experience in all the stages of its gaming experience. As it has been reviewed in this article, AI algorithms may help to improve several stages inside the serious games development:

– **Inner working of the game:** Serious games are constantly moving closer to modern games development, following the same pattern with regard to AI techniques. Most modern games addresses three basic game-related needs when implementing intelligent algorithms: move characters, make decisions inside the gameflow, and think tactically (20). These needs have also been found throughout this article, where several articles made use of AI techniques for the inner workings of the game.
– **Personalized Gaming experience:** The design and development of adaptive intelligent serious games with content changes based on user interaction makes player experience, training and education more customized. AI techniques provide systems with an efficient way of learning based on the users themselves, providing them with customized personal experiences, which may increase their potential effects (170).

In summary, AI techniques are not only not limited in the context of serious games but also they have a promising potential in the future of this area. The revised literature highlights the potential of intelligent serious games, and the wide range of possibilities they provide to researchers, professionals, and final users. The future of serious games will probably be closer to modern games development involving more AI algorithms, arts and animations and ending with serious games that resemble more to modern video-games, engaging users and game professionals into their path.

A. Limitation

The current review has a number of limitations. As with all reviews, it was limited by the search terms used, the databases employed and the time period selected. However, this review helps to contextualize the use of some mainstream artificial intelligence algorithms in the field of serious games, giving an overview of the trends and limitations over the last decade.

B. Conclusion

Finally, the authors consider that over recent years enough knowledge has been gathered to create new intelligent serious games to consider not only the final aim but also the technologies and techniques used to provide players with a nearly real experience. This new age of serious games is very close to the world of video-games, and they generate new solutions completely adapted to their target audience. However, researchers may need to improve their methodology for testing developed serious games, so as to ensure they meet their final purposes. Moreover, the authors would like to encourage other researchers to extend this article to other AI specific techniques and/or addressing new AI-related features, to extend this state of the art in the field of serious games, creating a knowledge-hub for researchers in the area.

ACKNOWLEDGMENT

The authors would like to thank DEIKER Research Agency at the University of Deusto, and researcher Philip Heaton at Deustotech-Life for his help with revising this article.

REFERENCES

[1] C. C. Abt, *Serious games*. University Press of America, 1987.
[2] T. Dumbleton and J. Kirriemuir, “13 digital games and education,” *Understanding digital games*, p. 223, 2006.
[3] U. Ritterfeld, M. Cody, and P. Vorderer, *Serious games: Mechanisms and effects*. Routledge, 2009.
[4] C. Girard, J. Ecalle, and A. Magnan, “Serious games as new educational tools: how effective are they? a meta-analysis of recent studies,” *Journal of Computer Assisted Learning*, vol. 29, no. 3, pp. 207–219, 2013.
[5] J. P. Gee, “What video games have to teach us about learning and literacy,” *Computers in Entertainment (CIE)*, vol. 1, no. 1, pp. 20–20, 2003.
[6] F. L. Greitzer, O. A. Kuchar, and K. Huston, “Cognitive science implications for enhancing training effectiveness in a serious gaming context,” *Journal on Educational Resources in Computing (JERIC)*, vol. 7, no. 3, p. 2, 2007.
[7] F. Bellotti, B. Kapralos, K. Lee, P. Moreno-Ger, and R. Berta, “Assessment in and of serious games: an overview,” *Advances in Human-Computer Interaction*, vol. 2013, p. 1, 2013.
[8] B. Sawyer and P. Smith, “Serious games taxonomy,” in *Slides from the Serious Games Summit at the Game Developers Conference*, 2008.
[9] D. R. Michael and S. L. Chen, *Serious games: Games that educate, train, and inform*. Muska & Lipman/Premier-Trade, 2005.
[10] M. Zyda, “From visual simulation to virtual reality to games,” *Computer*, vol. 38, no. 9, pp. 25–32, 2005.
[11] J. Alvarez and L. Michaud, “Serious games. advergaming, edugaming, training and more,” *IDATE, Monpellier*, 2008.
[12] R. Ratan and U. Ritterfeld, “Classifying serious games,” *Serious games: Mechanisms and effects*, pp. 10–24, 2009.
[13] A. J. Stapleton, “Serious games: Serious opportunities,” in *Australian Game Developers? Conference, Academic Summit, Melbourne*, 2004.
[14] L. Michaud and J. Alvarez, “Serious games: Advergaming, edugaming, training,” *Paris: IDATE*, 2008.
[15] B. P. Bergeron, *Developing serious games*. Charles River Media Hingham, 2006.
[16] J. S. Breuer and G. Bente, “Why so serious? on the relation of serious games and learning,” *Eludamos. Journal for Computer Game Culture*, vol. 4, no. 1, pp. 7–24, 2010.
[17] D. Djaouti, J. Alvarez, and J.-P. Jessel, “Classifying serious games: The g/p/s model,” *Handbook of research on improving learning and motivation through educational games: Multidisciplinary approaches*, pp. 118–136, 2011.
[18] D. Mori, R. Berta, A. De Gloria, V. Fiore, and L. Magnani, “An easy to author dialogue management system

for serious games," *Journal on Computing and Cultural Heritage (JOCCH)*, vol. 6, no. 2, p. 10, 2013.

[19] G. N. Yannakakis, "Game ai revisited," in *Proceedings of the 9th conference on Computing Frontiers*. ACM, 2012, pp. 285–292.

[20] I. Millington and J. D. Funge, *Artificial intelligence for games*. Taylor & Francis US, 2009.

[21] A. Sordoni, J.-P. Briot, I. Alvarez, E. Vasconcelos, M. de Azevedo Irving, G. Melo, and V. S. Patto, "Design of a participatory decision making agent architecture based on argumentation and influence function–application to a serious game about biodiversity conservation," *RAIRO-Operations Research*, vol. 44, no. 4, pp. 269–283, 2010.

[22] D. Cantwell, D. O. Broin, R. Palmer, and G. Doyle, "Motivating elderly people to exercise using a social collaborative exergame with adaptive difficulty." 2012.

[23] R. Costa, I. Mendonca, and D. Souza, "Exploring the intelligent agents for controlling user navigation in 3d games for cognitive stimulation," in *8th International Conference on Disability, Virtual Reality and Associated Technologies*, vol. 1, 2010, pp. 1–6.

[24] F. Keshtkar, C. Burkett, H. Li, and A. C. Graesser, "Using data mining techniques to detect the personality of players in an educational game," in *Educational Data Mining*. Springer, 2014, pp. 125–150.

[25] J. L. Sabourin, L. R. Shores, B. W. Mott, and J. C. Lester, "Understanding and predicting student self-regulated learning strategies in game-based learning environments," *International Journal of Artificial Intelligence in Education*, vol. 23, no. 1-4, pp. 94–114, 2013.

[26] B. Horan and M. Gardner, "+ spaces: Serious games for role-playing government policies," in *Understanding Learning in Virtual Worlds*. Springer, 2013, pp. 101–116.

[27] P. Petridis, I. Dunwell, F. Liarokapis, G. Constantinou, S. Arnab, S. de Freitas, and M. Hendrix, "The herbert virtual museum," *Journal of Electrical and Computer Engineering*, vol. 2013, 2013.

[28] A. M. Hussaan, K. Sehaba, and A. Mille, "Helping children with cognitive disabilities through serious games: project cles," in *The proceedings of the 13th international ACM SIGACCESS conference on Computers and accessibility*. ACM, 2011, pp. 251–252.

[29] S. Tang, M. Hanneghan, T. Hughes, C. Dennett, S. Cooper, M. A. Sabri, C. Carter, A. El Rhalibi, M. Merabti, P. Fergus *et al.*, "Towards a domain specific modelling language for serious game design," in *6th International Game Design and Technology Workshop (GDTW'08), Liverpool, UK*, 2008, pp. 43–52.

[30] M. Frutos-Pascual, B. G. Zapirain, and A. M. Zorrilla, "Adaptive tele-therapies based on serious games for health for people with time-management and organisational problems: Preliminary results," *International journal of environmental research and public health*, vol. 11, no. 1, pp. 749–772, 2014.

[31] F. E. Santos, A. P. Bastos, L. C. Andrade, K. Revoredo, and P. Mattos, "Assessment of adhd through a computer game: An experiment with a sample of students," in *Games and Virtual Worlds for Serious Applications (VS-GAMES), 2011 Third International Conference on*. IEEE, 2011, pp. 104–111.

[32] J. Qin, Y.-P. Chui, W.-M. Pang, K.-S. Choi, and P.-A. Heng, "Learning blood management in orthopedic surgery through gameplay," *Computer Graphics and Applications, IEEE*, vol. 30, no. 2, pp. 45–57, 2010.

[33] R. Haworth, S. S. T. Bostani, and K. Sedig, "Visualizing decision trees in games to support children's analytic reasoning: any negative effects on gameplay?" *International Journal of Computer Games Technology*, vol. 2010, p. 3, 2010.

[34] B. M. Good, S. Loguercio, O. L. Griffith, M. Nanis, C. Wu, and A. I. Su, "The cure: Making a game of gene selection for breast cancer survival prediction," *arXiv preprint arXiv:1402.3632*, 2014.

[35] D. Parsons, K. Petrova, and H. Ryu, "Mobile gaming-a serious business!" in *Wireless, Mobile and Ubiquitous Technology in Education (WMUTE), 2012 IEEE Seventh International Conference on*. IEEE, 2012, pp. 17–24.

[36] Z. Chiang, "The pricing strategy in chain store operation simulation system based on decision tree," in *Computational Sciences and Optimization (CSO), 2011 Fourth International Joint Conference on*. IEEE, 2011, pp. 392–395.

[37] J. Lilly and M. Warnes, "Designing mobile games for learning–the mgbl approach," *Serious Games on the Move, O. Petrovic and A. Brand, Eds. Vienna/New York: Springer*, pp. 3–25, 2009.

[38] T. Rijcken, J. Stijnen, and N. Slootjes, "Simdelta: Inquiry into an internet-based interactive model for water infrastructure development in the netherlands," *Water*, vol. 4, no. 2, pp. 295–320, 2012.

[39] J. M. Kim, R. W. Hill, Jr, P. J. Durlach, H. C. Lane, E. Forbell, M. Core, S. Marsella, D. Pynadath, and J. Hart, "Bilat: A game-based environment for practicing negotiation in a cultural context," *International Journal of Artificial Intelligence in Education*, vol. 19, no. 3, pp. 289–308, 2009.

[40] F.-H. Tsai, K.-C. Yu, and H.-S. Hsiao, "Exploring the factors influencing learning effectiveness in digital game-based learning." *Educational Technology & Society*, vol. 15, no. 3, pp. 240–250, 2012.

[41] E. Castro-Sanchez, E. Charani, L. Moore, M. Gharbi, and A. Holmes, "Taking antimicrobial stewardship initiatives to the next level: Development of a serious prescribing game for acute care," in *International Journal of Infectious Diseases*. Elsevier, 2014, pp. 46–47.

[42] C. F. Lin, Y.-c. Yeh, Y. H. Hung, and R. I. Chang, "Data mining for providing a personalized learning path in creativity: An application of decision trees," *Computers & Education*, vol. 68, pp. 199–210, 2013.

[43] E. D. van der Spek, P. Wouters, and H. van Oostendorp, “Code red: Triage or cognition-based design rules enhancing decisionmaking training in a game environment,” *British Journal of Educational Technology*, vol. 42, no. 3, pp. 441–455, 2011.

[44] S. Moya, S. Grau, and D. Tost, “The wise cursor: assisted selection in 3d serious games,” *The Visual Computer*, vol. 29, no. 6-8, pp. 795–804, 2013.

[45] A. M. Hussaan and K. Sehaba, “Generator of adaptive learning scenarios: design and evaluation in the project cles,” in *21st Century Learning for 21st Century Skills*. Springer, 2012, pp. 166–179.

[46] J. Goulding, W. Nadim, P. Petridis, and M. Alshawi, “Construction industry offsite production: A virtual reality interactive training environment prototype,” *Advanced Engineering Informatics*, vol. 26, no. 1, pp. 103–116, 2012.

[47] I. Ghergulescu and C. H. Muntean, “Learner motivation assessment with¡ e-adventure¿ game platform,” in *World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education*, vol. 2011, no. 1, 2011, pp. 1212–1221.

[48] A. C. Vidani and L. Chittaro, “Using a task modeling formalism in the design of serious games for emergency medical procedures,” in *Games and Virtual Worlds for Serious Applications, 2009. VS-GAMES’09. Conference in*. IEEE, 2009, pp. 95–102.

[49] F. Bellotti, R. Berta, A. De Gloria, and L. Primavera, “Adaptive experience engine for serious games,” *Computational Intelligence and AI in Games, IEEE Transactions On*, vol. 1, no. 4, pp. 264–280, 2009.

[50] M. Fazzolari, R. Alcalá, Y. Nojima, H. Ishibuchi, and F. Herrera, “A review of the application of multiobjective evolutionary fuzzy systems: Current status and further directions,” *Fuzzy Systems, IEEE Transactions on*, vol. 21, no. 1, pp. 45–65, 2013.

[51] D. Balas, C. Brom, A. Abonyi, and J. Gemrot, “Hierarchical petri nets for story plots featuring virtual humans.” in *AIIDE*, 2008.

[52] A. Karime, M. Eid, J. M. Alja’am, A. Saddik, and W. Gueaieb, “A fuzzy-based adaptive rehabilitation framework for home-based wrist training,” *IEEE Transactions on Instrumentation and Measurement*, 2014.

[53] M. Pirovano, R. Mainetti, G. Baud-Bovy, P. L. Lanzi, and N. A. Borghese, “Self-adaptive games for rehabilitation at home,” in *Computational Intelligence and Games (CIG), 2012 IEEE Conference on*. IEEE, 2012, pp. 179–186.

[54] M. Farhangian, M. K. Purvis, M. Purvis, and B. T. R. Savarimuthu, “Modelling the effects of personality and temperament in small teams,” in *Coordination, Organizations, Institutions, and Norms in Agent Systems IX*. Springer, 2014, pp. 25–41.

[55] F. de Oliveira, R. Lanzillotti, R. da Costa, R. Gonçalves, P. Ventura, and L. de Carvalho, “Arvet and saptept: A virtual environment and a system supported by fuzzy logic in virtual reality exposure therapy for ptsd patients,” in *12th International Conference on Computational Science and Its Applications (ICCSA)*, 2012.

[56] F. Imbeault, B. Bouchard, and A. Bouzouane, “Serious games in cognitive training for alzheimer’s patients,” in *Serious Games and Applications for Health (SeGAH), 2011 IEEE 1st International Conference on*. IEEE, 2011, pp. 1–8.

[57] Y. Cai, C. Miao, A.-H. Tan, Z. Shen, and B. Li, “Creating an immersive game world with evolutionary fuzzy cognitive maps,” *IEEE computer graphics and applications*, vol. 30, no. 2, pp. 58–70, 2010.

[58] I. A. Buditjahjanto and H. Miyauchi, “An intelligent decision support based on a subtractive clustering and fuzzy inference system for multiobjective optimization problem in serious game,” *International Journal of Information Technology & Decision Making*, vol. 10, no. 05, pp. 793–810, 2011.

[59] D. D. Djordjevich, P. G. Xavier, M. L. Bernard, J. H. Whetzel, M. R. Glickman, and S. J. Verzi, “Preparing for the aftermath: Using emotional agents in game-based training for disaster response,” in *Computational Intelligence and Games, 2008. CIG’08. IEEE Symposium On*. IEEE, 2008, pp. 266–275.

[60] M. Pirovano and P. L. Lanzi, “Fuzzy tactics: A scripting game that leverages fuzzy logic as an engaging game mechanic,” *Expert Systems with Applications*, vol. 41, no. 13, pp. 6029–6038, 2014.

[61] D. Zaldivar, E. Cuevas, M. A. Pérez-Cisneros, J. H. Sossa, J. G. Rodríguez, and E. O. Palafox, “An educational fuzzy-based control platform using lego robots,” *International Journal of Electrical Engineering Education*, vol. 50, no. 2, pp. 157–171, 2013.

[62] Z. Shen, C. Miao *et al.*, “An emotional agent in virtual learning environment,” in *Transactions on edutainment IV*. Springer, 2010, pp. 22–33.

[63] C. Oliveira, M. Cintra, and F. Mendes Neto, “Learning risk management in software projects with a serious game based on intelligent agents and fuzzy systems,” in *8th conference of the European Society for Fuzzy Logic and Technology (EUSFLAT-13)*. Atlantis Press, 2013.

[64] I. S. Brasil, F. M. M. Neto, J. F. S. Chagas, R. M. de Lima, D. F. L. Souza, M. F. Bonates, and A. Dantas, “An inteligent agent-based virtual game for oil drilling operators training,” in *Virtual Reality (SVR), 2011 XIII Symposium on*. IEEE, 2011, pp. 9–17.

[65] I. S. Brasil, F. M. M. Neto, J. F. S. Chagas, R. Monteiro, D. F. L. Souza, M. F. Bonates, and A. Dantas, “An intelligent and persistent browser-based game for oil drilling operators training,” in *Serious Games and Applications for Health (SeGAH), 2011 IEEE 1st International Conference on*. IEEE, 2011, pp. 1–9.

[66] O. J. Petit dit Dariel, T. Raby, F. Ravaut, and M. Rothan-Tondeur, “Developing the serious games potential in nursing education,” *Nurse education today*, 2013.

[67] C. Gardiner, *Stochastic methods*. Springer-Verlag, Berlin–Heidelberg–New York–Tokyo, 1985.

[68] K. Muñoz, P. Mc Kevitt, T. Lunney, J. Noguez, and L. Neri, "Playphysics: an emotional games learning environment for teaching physics," pp. 400–411, 2010.

[69] L. Derbali, R. Ghali, and C. Frasson, "Assessing motivational strategies in serious games using hidden markov models," in *The Twenty-Sixth International FLAIRS Conference*, 2013.

[70] P.-h. Su, Y.-B. Wang, T.-h. Yu, and L.-s. Lee, "A dialogue game framework with personalized training using reinforcement learning for computer-assisted language learning," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*. IEEE, 2013, pp. 8213–8217.

[71] P.-h. Su, C.-h. Wu, and L.-s. Lee, "A recursive dialogue game for personalized computer-aided pronunciation training," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, pp. 127–141, 2014.

[72] A. N. Rafferty, M. Zaharia, and T. L. Griffiths, "Optimally designing games for cognitive science research," in *CogSci2012: The Annual Meeting of the Cognitive Science Society*, 2012.

[73] A. Navarro-Newball, D. Loaiza, C. Oviedo, A. Castillo, A. Portilla, D. Linares, and G. Álvarez, "Talking to teo: Video game supported speech therapy," *Entertainment Computing*, vol. 5, no. 4, pp. 401–412, 2014.

[74] B. Bouchard, F. Imbeault, A. Bouzouane, and B.-A. J. Menelas, "Developing serious games specifically adapted to people suffering from alzheimer," in *Serious Games Development and Applications*. Springer, 2012, pp. 243–254.

[75] L. E. Sucar, F. Orihuela-Espina, R. L. Velazquez, D. J. Reinkensmeyer, R. Leder, and J. Hernández-Franco, "Gesture therapy: An upper limb virtual reality-based motor rehabilitation platform," *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, vol. 22, no. 3, pp. 634–643, 2014.

[76] W. L. Johnson and A. Valente, "Tactical language and culture training systems: Using artificial intelligence to teach foreign languages and cultures." in *AAAI*, 2008, pp. 1632–1639.

[77] J. N. Neto, R. Silva, J. P. Neto, J. M. Pereira, and J. Fernandes, "Solis' curse-a cultural heritage game using voice interaction with a virtual agent," in *Games and Virtual Worlds for Serious Applications (VS-GAMES), 2011 Third International Conference on*. IEEE, 2011, pp. 164–167.

[78] W. L. Johnson, N. Wang, and S. Wu, "Experience with serious games for learning foreign languages and cultures," in *SimTecT Conference*, 2007.

[79] W. L. Johnson, "Serious use of a serious game for language learning," *Frontiers in Artificial Intelligence and Applications*, vol. 158, p. 67, 2007.

[80] J. Orkin, "Goal oriented action planning," *CTAN http://web. media. mit. edu/˜ jorkin/goap. html*, 2010.

[81] ——, "Three states and a plan: the ai of fear," in *Game Developers Conference*, vol. 2006, 2006, p. 4.

[82] S. Álvarez Napagao, I. Gómez Sebastià, J. Vázquez Salceda, F. Koch *et al.*, "conciens: Organizational awareness in real-time strategy games," 2012.

[83] S. Alvarez-Napagao, I. Gómez-Sebastià, S. Panagiotidi, A. Tejeda-Gómez, L. Oliva, and J. Vázquez-Salceda, "Socially-aware emergent narrative," in *Agents for Educational Games and Simulations*. Springer, 2012, pp. 139–150.

[84] H. Yu, T. Lim, J. Ritchie, R. Sung, S. Louchart, I. Stănescu, I. Roceanu, and S. de Freitas, "Exploring the application of computer game theory to automated assembly," *Procedia Computer Science*, vol. 15, pp. 266–273, 2012.

[85] G. Rebolledo-Mendez, S. de Freitas, and A. R. G. Gaona, "A model of motivation based on empathy for ai-driven avatars in virtual worlds," in *Games and Virtual Worlds for Serious Applications, 2009. VS-GAMES'09. Conference in*. IEEE, 2009, pp. 5–11.

[86] P. Jepp, M. Fradinho, and J. M. Pereira, "An agent framework for a modular serious game," in *Games and Virtual Worlds for Serious Applications (VS-GAMES), 2010 Second International Conference on*. IEEE, 2010, pp. 19–26.

[87] C. Qu, W. Brinkman, P. Wiggers, and I. Heynderickx, "The effect of priming pictures and videos on a question–answer dialog scenario in a virtual environment," *Presence*, vol. 22, no. 2, pp. 91–109, 2013.

[88] R. Davis and J. J. King, "The origin of rule-based systems in ai," *Rule-Based Expert Systems: The Mycin Experiments of the Stanford Heuristic Programming Project*, 1984.

[89] G. A. E. Khayat, T. F. Mabrouk, and A. S. Elmaghraby, "Intelligent serious games system for children with learning disabilities," in *Computer Games (CGAMES), 2012 17th International Conference on*. IEEE, 2012, pp. 30–34.

[90] J. Haladjian, D. Ismailović, B. Köhler, and B. Brügge, "A quick prototyping framework for adaptive serious games with 2d physics on mobile touch devices," in *IADIS International Conference Mobile Learning*, 2012, pp. 197–204.

[91] A. Arya, S. DiPaola, and A. Parush, "Perceptually valid facial expressions for character-based applications," *International Journal of Computer Games Technology*, vol. 2009, 2009.

[92] H. W. Patton, "Learning via gaming: An immersive environment for teaching kids handwriting," in *Frontiers In Education Conference - Global Engineering: Knowledge Without Borders, Opportunities Without Passports, 2007. FIE '07. 37th Annual*, 2008.

[93] A. Ward, M. McKeown, C. Utay, O. Medvedeva, and R. Crowley, "Interactive stories and motivation to read in the raft dyslexia fluency tutor," in *Intelligent Virtual Agents*. Springer, 2012, pp. 260–267.

[94] L. M. Rojas-Barahona, A. Lorenzo, C. Gardent *et al.*, "Building and exploiting a corpus of dialog interactions between french speaking virtual and human agents." in *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC-2012)*, 2012, pp. 1428–1435.

[95] A. J. Souza, P. C. Tedesco, D. R. Silva, G. L. Ramalho *et al.*, "An integrated development model for character-based games," in *Games and Digital Entertainment (SBGAMES), 2009 VIII Brazilian Symposium on*. IEEE, 2009, pp. 191–196.

[96] M. G. Core, H. C. Lane, M. Van Lent, D. Gomboc, S. Solomon, and M. Rosenberg, "Building explainable artificial intelligence systems," in *Proceedings of the National Conference on Artificial Intelligence*, vol. 21, no. 2. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 2006, p. 1766.

[97] A. El Rhalibi, N. Baker, and M. Merabti, "Emotional agent model and architecture for npcs group control and interaction to facilitate leadership roles in computer entertainment," in *Proceedings of the 2005 ACM SIGCHI International Conference on Advances in computer entertainment technology*. ACM, 2005, pp. 156–163.

[98] K. Sehaba and A. M. Hussaan, "Goals: generator of adaptive learning scenarios," *International Journal of Learning Technology*, vol. 8, no. 3, pp. 224–245, 2013.

[99] D. Benedettelli, "Finite state machines," *Creating Cool MINDSTORMS® NXT Robots*, pp. 47–58, 2008.

[100] A. Szczesna, J. Grudzinski, T. Grudzinski, R. Mikuszewski, and A. Debowski, "The psychology serious game prototype for preschool children," in *Serious Games and Applications for Health (SeGAH), 2011 IEEE 1st International Conference on*. IEEE, 2011, pp. 1–4.

[101] C. T. Tan, A. Johnston, K. Ballard, S. Ferguson, and D. Perera-Schulz, "speak-man: towards popular gameplay for speech therapy," in *Proceedings of The 9th Australasian Conference on Interactive Entertainment: Matters of Life and Death*. ACM, 2013, p. 28.

[102] M. Rehm and K. Leichtenstern, "Gesture-based mobile training of intercultural behavior," *Multimedia systems*, vol. 18, no. 1, pp. 33–51, 2012.

[103] D. L. Ines and G. Abdelkader, "Mixed reality serious games: The therapist perspective," in *Serious Games and Applications for Health (SeGAH), 2011 IEEE 1st International Conference on*. IEEE, 2011, pp. 1–10.

[104] H. Li, H. Munoz-Avila, L. Ke, C. Symborski, and R. Alonso, "Discovery of player strategies in a serious game," in *First AAAI Conference on Human Computation and Crowdsourcing*, 2013.

[105] I. N. Sukajaya, S. Mardi, I. Purnama, M. Hariadi, M. H. Purnomo, and A. Vitianingsih, "Multi-parameter dynamic difficulty game's scenario using box-muller of gaussian distribution," in *Computer Science & Education (ICCSE), 2012 7th International Conference on*. IEEE, 2012, pp. 1666–1671.

[106] V. Wattanasoontorn, I. Boada, C. Blavi, and M. Sbert, "The framework of a life support simulation application," *Procedia Computer Science*, vol. 15, pp. 293–294, 2012.

[107] A. Yessad, J. Labat, and F. Kermorvant, "Segae: A serious game authoring environment," in *Advanced Learning Technologies (ICALT), 2010 IEEE 10th International Conference on*. IEEE, 2010, pp. 538–540.

[108] J. R. Rankin and S. S. Vargas, "Fps extensions modelling esgs," in *Advances in Computer-Human Interactions, 2009. ACHI'09. Second International Conferences on*. IEEE, 2009, pp. 152–155.

[109] A. Sagae, W. L. Johnson, and R. Row, "Serious game environments for language and culture education," in *Proceedings of the NAACL HLT 2010 Demonstration Session*. Association for Computational Linguistics, 2010, pp. 29–32.

[110] B. Duggan, H. McAtamney, and F. Mtenzi, "Learning games programming with dalek world." in *7th. International Conference on Computer Games: AI, Animation, Mobile, Educational and Serious Games, CNBDI*. Dublin Institute of Technology, 2005.

[111] B. Morgan, F. Kehtkar, A. Graesser, and D. Shaffer, "Automating the mentor in a serious game: A discourse analysis using finite state machines," in *HCI International 2013-Posters? Extended Abstracts*. Springer, 2013, pp. 591–595.

[112] D. Irvine and M. A. Gongora, "Cogbox-combined artificial intelligence methodologies to achieve a semi-realistic agent in serious games," in *Artifical Intelligence and Soft Computing*. Springer, 2010, pp. 492–499.

[113] G. Flórez-Puga, M. Gomez-Martin, B. Diaz-Agudo, and P. Gonzalez-Calero, "Dynamic expansion of behaviour trees," in *Proceedings of Artificial Intelligence and Interactive Digital Entertainment Conference. AAAI Press*, 2008, pp. 36–41.

[114] V. Wendel, M. Babarinow, T. Hörl, S. Kolmogorov, S. Göbel, and R. Steinmetz, "Woodment: web-based collaborative multiplayer serious game," in *Transactions on edutainment IV*. Springer, 2010, pp. 68–78.

[115] P. McDowell, R. Darken, J. Sullivan, and E. Johnson, "Delta3d: a complete open source game and simulation engine for building military training systems," *The Journal of Defense Modeling and Simulation: Applications, Methodology, Technology*, vol. 3, no. 3, pp. 143–154, 2006.

[116] B. G. Silverman, D. Pietrocola, N. Weyer, R. Weaver, N. Esomar, R. Might, and D. Chandrasekaran, "Nonkin village: An embeddable training game generator for learning cultural terrain and sustainable counter-insurgent operations," in *Agents for Games and Simulations*. Springer, 2009, pp. 135–154.

[117] A. Binsubaih, S. Maddock, and D. Romano, "A serious game for traffic accident investigators," *Interactive Technology and Smart Education*, vol. 3, no. 4, pp. 329–346, 2006.

[118] V. Šisler, C. Brom, J. Cuhra, K. Činátl, and J. Gemrot, "Stories from the history of czechoslovakia, a serious game for teaching history of the czech lands in the 20th century–notes on design concepts and design process," in *Entertainment Computing-ICEC 2012*. Springer, 2012, pp. 67–74.
[119] S. Bakkes, C. T. Tan, and Y. Pisan, "Personalised gaming: a motivation and overview of literature," in *Proceedings of the 8th Australasian Conference on Interactive Entertainment: Playing the System*. ACM, 2012, p. 4.
[120] L. Derbali, P. Chalfoun, and C. Frasson, "A theoretical and empirical approach in assessing motivational factors: From serious games to an its," in *Proceedings of the Twenty-Fourth International Florida Artificial Intelligence Research Society Conference*. AAAI.
[121] W. Husain, M. H. M. Adnan, L. K. Ping, J. Poh, and L. K. Meng, "Myhealthykids: Intelligent obesity intervention system for primary school children," in *The Third International Conference on Digital Information Processing and Communications (ICDIPC2013)*. The Society of Digital Information and Wireless Communication, 2013, pp. 627–633.
[122] A. Weichselbraun, S. Gindl, and A. Scharl, "Using games with a purpose and bootstrapping to create domain-specific sentiment lexicons," in *Proceedings of the 20th ACM international conference on Information and knowledge management*. ACM, 2011, pp. 1053–1060.
[123] F. Keshtkar, C. Burkett, A. Graesser, and H. Li, "Detecting players personality behavior with any effort of concealment," in *Computational Linguistics and Intelligent Text Processing*. Springer, 2012, pp. 502–514.
[124] F. Vaassen and W. Daelemans, "Emotion classification in a serious game for training communication skills," in *Computational Linguistics in the Netherlands*, 2010.
[125] K. C. Tseng and C.-C. Wu, "An expert fitness diagnosis system based on elastic cloud computing," *The Scientific World Journal*, vol. 2014, 2014.
[126] J. Feldman, A. Monteserin, and A. Amandi, "Detecting students' perception style by using games," *Computers & Education*, vol. 71, pp. 14–22, 2014.
[127] S. N. W. Shamsuddin, H. Ugail, V. Lesk, and E. Walters, "Vread: a virtual simulation to investigate cognitive function in the elderly," in *Cyberworlds (CW), 2012 International Conference on*. IEEE, 2012, pp. 215–220.
[128] B. Morgan, F. Keshtkar, Y. Duan, P. Nash, and A. Graesser, "Using state transition networks to analyze multi-party conversations in a serious game," in *Intelligent Tutoring Systems*. Springer, 2012, pp. 162–167.
[129] J. Rowe, B. Mott, S. McQuiggan, J. Robison, S. Lee, and J. Lester, "Crystal island: A narrative-centered learning environment for eighth grade microbiology," in *Workshop on Intelligent Educational Games at the 14th International Conference on Artificial Intelligence in Education, Brighton, UK*, 2009, pp. 11–20.
[130] L. Zhang, J. W. Wade, D. Bian, A. Swanson, Z. Warren, and N. Sarkar, "Data fusion for difficulty adjustment in an adaptive virtual reality game system for autism intervention," in *HCI International 2014-Posters? Extended Abstracts*. Springer, 2014, pp. 648–652.
[131] M.-D. Otis and B. Menelas, "Toward an augmented shoe for preventing falls related to physical conditions of the soil," in *Systems, Man, and Cybernetics (SMC), 2012 IEEE International Conference on*. IEEE, 2012, pp. 3281–3285.
[132] D. Puzenat and I. Verlut, "Behavior analysis through games using artificial neural networks," in *Advances in Computer-Human Interactions, 2010. ACHI'10. Third International Conference on*. IEEE, 2010, pp. 134–138.
[133] R. L. Lamb, L. Annetta, D. B. Vallett, and T. D. Sadler, "Cognitive diagnostic like approaches using neural-network analysis of serious educational videogames," *Computers & Education*, vol. 70, pp. 92–104, 2014.
[134] M. A. Syufagi, M. H. Purnomo, and M. Hariadi, "Tendency of players is trial and error: Case study of cognitive classification in the cognitive skill games," *Jurnal Ilmu Komputer dan Informasi*, vol. 5, no. 1, p. 31, 2012.
[135] C. Mills and B. Dalgarno, "A conceptual model for game-based intelligent tutoring systems," *Proceedings of the 2007 Australasian Society for Computers in Learning in Tertiary Education*, 2007.
[136] C. Grappiolo, Y.-G. Cheong, J. Togelius, R. Khaled, and G. N. Yannakakis, "Towards player adaptivity in a serious game for conflict resolution," in *Games and Virtual Worlds for Serious Applications (VS-GAMES), 2011 Third International Conference on*. IEEE, 2011, pp. 192–198.
[137] H. Hildmann, A. Uhlemann, and D. Livingstone, "A mobile phone based virtual pet to teach social norms and behaviour to children," in *Digital Games and Intelligent Toys Based Education, 2008 Second IEEE International Conference on*. IEEE, 2008, pp. 15–17.
[138] R. Arrabales, A. Ledezma, and A. Sanchis, "Towards conscious-like behavior in computer game characters," in *Computational Intelligence and Games, 2009. CIG 2009. IEEE Symposium on*. IEEE, 2009, pp. 217–224.
[139] O. Miglino, O. Gigliotta, M. Schembri, and A. Di, "Collective adaptive agents as techniques to build-up edutainments systems," in *Italian Workshop on Artificial Life and Evolutionary Computation WIVACE*, 2012.
[140] O. Gigliotta, O. Miglino, M. Schembri, and A. Di, "Building up serious games with an artificial life approach: two cases studies," in *Evolution, Complexity and Artificial Life*, S. Cagnoni, M. Mirolli, and M. Villani, Eds. Springer, 2013.
[141] M. Syufagi, P. Hery, and M. Hariadi, "Modeling serious games based on cognitive skill classification using

learning vector quantization with petri net," in *Advanced Computer Science and Information System (ICACSIS), 2011 International Conference on*. IEEE, 2011, pp. 159–164.

[142] M. A. Syufagi, M. Hariadi, and M. H. Purnomo, "Petri net model for serious games based on motivation behavior classification," *International Journal of Computer Games Technology*, vol. 2013, p. 1, 2013.

[143] A. Ram, S. Ontañón, and M. Mehta, "Artificial intelligence for adaptive computer games." in *FLAIRS Conference*, 2007, pp. 22–29.

[144] A. Tobail, J. Crowe, and A. Arisha, "Learning by gaming: supply chain application," in *Simulation Conference (WSC), Proceedings of the 2011 Winter*. IEEE, 2011, pp. 3935–3946.

[145] A. McKenzie and G. McCalla, "Serious games for professional ethics: An architecture to support personalization," in *AIED 2009: 14 th International Conference on Artificial Intelligence in Education Workshops Proceedings*. Citeseer, 2009, p. 69.

[146] S. Triki, N. B. B. Saoud, J. Dugdale, and C. Hanachi, "Coupling case based reasoning and process mining for a web based crisis management decision support system," in *Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE), 2013 IEEE 22nd International Workshop on*. IEEE, 2013, pp. 245–252.

[147] I. Hulpuş, C. Hayes, and M. O. Fradinho, "A framework for personalised learning-plan recommendations in game-based learning," in *Recommender Systems for Technology Enhanced Learning*. Springer, 2014, pp. 99–122.

[148] M. Floryan and B. Woolf, "Rashi game: Towards an effective educational 3d gaming experience," in *Advanced Learning Technologies (ICALT), 2011 11th IEEE International Conference on*. IEEE, 2011, pp. 436–437.

[149] C. Forsyth, A. Graesser, B. Walker, K. Millis, P. I. Pavlik Jr, and D. Halpern, "Didactic galactic: Types of knowledge learned in a serious game," in *Artificial Intelligence in Education*. Springer, 2013, pp. 832–835.

[150] C.-R. Song, "Design and usability ofweb-based serious game for taiwanese efl undergraduate students," in *Networking, International Conference on Systems and International Conference on Mobile Communications and Learning Technologies, 2006. ICN/ICONS/MCL 2006. International Conference on*. IEEE, 2006, pp. 220–220.

[151] A. Tobail, J. Crowe, and A. Arisha, "Serious gaming learning: Supply chain multi-agent web-based simulation game," 2012.

[152] R. Lopes and R. Bidarra, "A semantic generation framework for enabling adaptive game worlds," in *Proceedings of the 8th International Conference on Advances in Computer Entertainment Technology*. ACM, 2011, p. 6.

[153] I. Hulpuş, M. Fradinho, and C. Hayes, "On-the-fly adaptive planning for game-based learning," in *Case-Based Reasoning. Research and Development*. Springer, 2010, pp. 375–389.

[154] J.-S. Ahn and W.-H. Lee, "Using eeg pattern analysis for implementation of game interface," in *Consumer Electronics (ISCE), 2011 IEEE 15th International Symposium on*. IEEE, 2011, pp. 348–351.

[155] T. D. Parsons and J. L. Reinebold, "Adaptive virtual environments for neuropsychological assessment in serious games," *Consumer Electronics, IEEE Transactions on*, vol. 58, no. 2, pp. 197–204, 2012.

[156] R. Berta, F. Bellotti, A. De Gloria, D. Pranantha, and C. Schatten, "Electroencephalogram and physiological signal analysis for assessing flow in games," *Computational Intelligence and AI in Games, IEEE Transactions on*, vol. 5, no. 2, pp. 164–175, 2013.

[157] S. Bernardini, K. Porayska-Pomsta, and H. Sampath, "Designing an intelligent virtual agent for social communication in autism," in *Proceedings of the Ninth Annual AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, 2013.

[158] S. Bernardini, K. Porayska-Pomsta, and T. J. Smith, "Echoes: an intelligent serious game for fostering social communication in children with autism," *Information Sciences*, 2013.

[159] E. M. Raybourn, N. Fabian, E. Tucker, and M. Willis, "Beyond game effectiveness part ii: A qualitative study of multi-role experiential learning," in *The Interservice/Industry Training, Simulation & Education Conference (I/ITSEC)*, vol. 2010, no. 1. NTSA, 2010.

[160] M. Zhang, M. Xu, Y. Liu, G. He, L. Han, P. Lv, and Y. Li, "The framework and implementation of virtual network marathon," in *VR Innovation (ISVRI), 2011 IEEE International Symposium on*. IEEE, 2011, pp. 161–167.

[161] J. R. Quinlan, "Decision trees and decision-making," *Systems, Man and Cybernetics, IEEE Transactions on*, vol. 20, no. 2, pp. 339–346, 1990.

[162] G. Mountrakis, J. Im, and C. Ogole, "Support vector machines in remote sensing: A review," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 66, no. 3, pp. 247–259, 2011.

[163] J. Solana, C. Cáceres, A. García-Molina, P. Chausa, E. Opisso, T. Roig-Rovira, E. Menasalvas, J. M. Tormos-Muñoz, and E. J. Gómez, "Intelligent therapy assistant (ita) for cognitive rehabilitation in patients with acquired brain injury," *BMC medical informatics and decision making*, vol. 14, no. 1, p. 58, 2014.

[164] J. Alvarez, D. Djaouti, and O. Rampnoux, "Serious games market: Some key figures (from 1950's to 2000's)," *Paris: IDATE*, 2009.

[165] T. M. Connolly, E. A. Boyle, E. MacArthur, T. Hainey, and J. M. Boyle, "A systematic literature review of empirical evidence on computer games and serious games," *Computers & Education*, vol. 59, no. 2, pp. 661–686, 2012.

[166] T. S. Portal, "Number of smartphones sold to end users worldwide from 2007 to 2014 (in million units)," http://statista.com/statistics/263437/global-smartphone-sales-to-end-users-since-2007/, [Online; accessed 30-April-2015].

[167] I. Mayer, G. Bekebrede, C. Harteveld, H. Warmelink, Q. Zhou, T. Ruijven, J. Lo, R. Kortmann, and I. Wenzler, "The research and evaluation of serious games: Toward a comprehensive methodology," *British Journal of Educational Technology*, vol. 45, no. 3, pp. 502–527, 2014.

[168] G. Sim, S. MacFarlane, and J. Read, "All work and no play: Measuring fun, usability, and learning in software for children," *Computers & Education*, vol. 46, no. 3, pp. 235–248, 2006.

[169] M. Ernst, "Choosing a venue: conference or journal?" 2008.

[170] M. Muir and C. Conati, "An analysis of attention to student–adaptive hints in an educational game," in *Intelligent Tutoring Systems*. Springer, 2012, pp. 112–122.

"If I had a world of my own, everything would be nonsense. Nothing would be what it is, because everything would be what it isn't. And contrary wise, what is, it wouldn't be. And what it wouldn't be, it would. You see?."

Alice in Wonderland and through the Looking Glass

Lewis Carroll

3 System Design

This chapter contains a detailed description of all the methods employed in the design and development of this dissertation, which will lead to the validation of the hypothesis introduced in section 1.1.

The final aim of this chapter is to create a framework for the analysis of gaze interaction in children while they are performing certain tasks inside a serious games development. The final aim is to evaluate gaze pattern behaviours and interaction measurements for automatically profiling users while using serious games, adapting the software development to final users' needs.

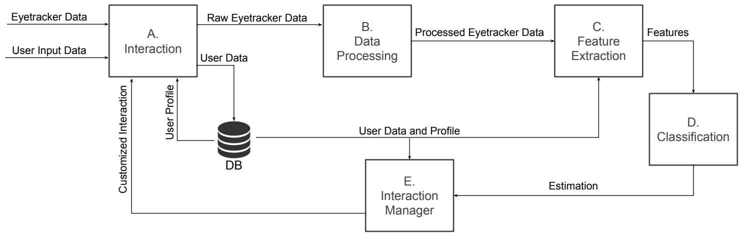
This chapter is divided as follows: Section 3.1 describes, from a high level perspective, the design of the proposed system, including all the different functional blocks that are part of it. Section 3.2 details each of these high level blocks, extracting all their relevant features and explaining their functionality, paradigms and algorithms.

3.1 High level design

Figure 3.1 displays the different blocks that create the high-level architecture of the system. This system has been designed using a modular approach, allowing new steps and modules to be easily added and implemented in every stage. The five high-level design blocks that create the core of this system are: the interaction module, the data processing block, the feature extraction process, the classifier block and the interaction manager.

During the following subsections each of the blocks in Figure 3.1 is presented together with its high-level functionality and the description of its inputs and outputs.

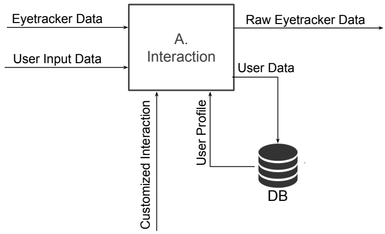
Fig. 3.1 System architecture



3.1.1 A. Interaction

Block A has been named as the interaction block. This module displays the appropriate level and serious games type from the available options. As displayed in Figure 3.2 this module has four different inputs: The raw data coming from the eyetracking source, the interaction data generated in real time by the player, the recommendation from the interaction manager about the appropriate degree of challenge for that particular user and previous history of the user. This module does also record both gaze and interaction data streams, storing them in the appropriate format enabling the analysis module to extract the meaningful information.

Fig. 3.2 High level design - Interaction module

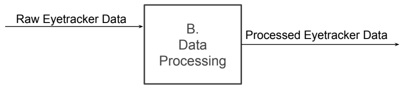


As it is shown in Figure 3.2 this module has two different outputs, one is directly stored in the interaction database for being retrieved and analysed and the other is the raw gaze data recording, stored as an *.xml* file for its further analysis in block B. Please refer to subsection 3.2.1 for a detailed review of the inner workings of the Interaction module.

3.1.2 B. Data processing

The data processing module analyses the raw gaze data recorded in the previous module, giving a meaning to all the different stored eye positions. The input and output of this system are displayed in Figure 3.3. Although the high level representation of this module is rather simple, the inner workings of this section are quite complex. This is the core block for the extraction and classification of the saccadic eye movements and fixations.

Fig. 3.3 High level design - Data processing module

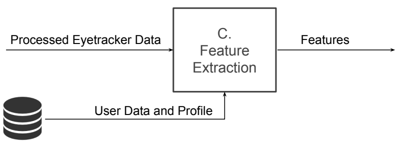


Gaze data cleaning, filtering and clustering is done in this section, and it is explained in further detail in subsection 3.2.2.

3.1.3 C. Feature extraction

This is the feature extractor module for the processed gaze data and the performance recording generated by the player on each interaction with the game. Inputs and outputs of this block are displayed in Figure 3.4.

Fig. 3.4 High level design - Feature extraction module

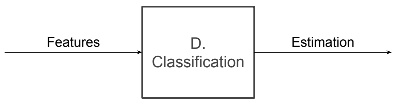


The set of features generated in this module are then further analysed by the classification module. For a detailed review of the inner workings of the feature extraction module please refer to subsection 3.2.3.

3.1.4 D. Classification

This module estimates user performance based on the features it gets from the feature extraction process on every interaction. Based on current performance and gaze patterns, this block estimates the final player score. Figure 3.5 displays this block's main input and output.

Fig. 3.5 High level design - Classification module

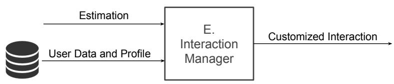


The output estimation uses previous knowledge of the system. It employs machine learning approaches for improving its classification performance with every new iteration. This module gives a correct or incorrect estimation based on previous knowledge, gaze patterns and interaction behaviour of the player. In case of incorrect estimations the system re-do the classification process for learning the new behaviour while the estimation is then labelled as positive or negative, based on the nature of the error. This classification process is explained in detail in subsection 3.2.4.

3.1.5 E. Interaction manager

This module is the last one in the proposed architecture. Since this system has been implemented in a modular and circular fashion, the outputs of this last module are inputs in the first interaction block. An overview of Interaction Manager inputs and output is displayed in Figure 3.6.

Fig. 3.6 High level design - Interaction manager module



The input in this module are the performance estimation given by the classification module together with current player's interaction history and performance. This module generates a unique gameplay for each player based on his or her personal gaming profile and visual attention patterns. A detailed explanation of this module, along with its implemented architectural hierarchy is introduced in subsection 3.2.5.

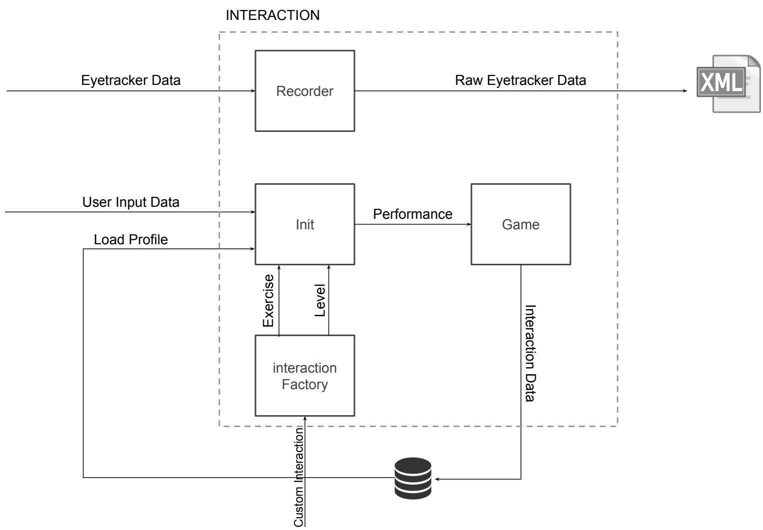
3.2 Low level design

This section gives a detailed overview of all the modules that compound the proposed system architecture. Each subsection in this low level design section will give a detailed summary of all the techniques, algorithms and paradigms used for all the five modules introduced in section 3.1 High Level Design.

3.2.1 A. Interaction

The Interaction module is designed with a dual purpose: This block displays the suitable level of challenge and game type, as indicated by the Interaction manager, along with the recording of user performance and raw gaze data. Figure 3.7 displays the inner workings and sub-modules that compound this part of the system.

Fig. 3.7 Low level design - Interaction module - Architecture



Whenever a new player is presented into the system, this module is in charge of login him/her into the system. If this new user has a previous playing history registered, it is retrieved for tailoring the gaming experience to his/her gaming profile. If not, a new one is created. After the initialization process, the player is presented with the appropriate game and level that suits best his or her skills and learning requirements. As soon as the game starts every interaction with the system is recorded and stored in the database together with player's relevant information.

The eyetracker recorder runs in parallel with the game interaction, recording all the raw gaze data information. Listing 3.1 shows the stored gaze data for each participant and exercise. These data consists of the (x, y) coordinates recorded by the eye tracking sensor, the timestamp in which they were received, the pupil size for each eye and the exercise details.

Listing 3.1 Raw gaze data example

```
<user date="2014-05-22" id="12" sessionid="899" time="09:20:15" >
  <exercise id="puzzle" level="Level5" mode="performance">
    <eyedata>
      <timestamp time="1167610729217997">
        <left_eye pupil_diam="3.0435333252" validity="0" x="134.831732304" y="64.8299084174"/>
        <right_eye pupil_diam="2.89215087891" validity="0" x="69.9537996816" y="59.426052033"/>
      </timestamp>
      ...
    </eyedata>
  </exercise>
</user>
```

These raw data is used for analysis and processing so as to obtain meaningful information about eye fixation locations, fixation durations, saccades and saccadic durations.

The interaction factory receives the recommendation from the interaction manager, as it is displayed in Figure 3.8. Based on this information and the available resources, this module determines which is the most suitable game and level to play next. All the games available in this system have been developed following a modular fashion that allow future games to be included easily into the architecture.

Fig. 3.8 Low level design - Interaction module - Interaction factory

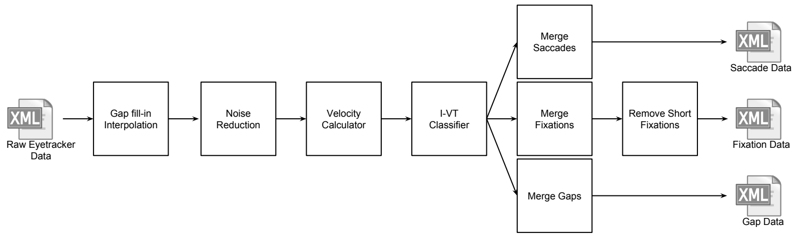
Following the interaction manager recommendation, this sub-module determines on each iteration which is the most suitable game and level from the available curricula, as it is displayed in Figure 3.8.

3.2.2 B. Data processing

These raw data is used for analysis and processing so as to obtain meaningful information about eye fixation locations, fixation durations, saccades, saccadic durations, gaps and gaps durations. Fixations are the period of time when the eyes remain fairly still and new information is acquired from visual array [Rayner 09], while saccades are the eye movements themselves. During saccades, no information is retrieved by the brain, since vision is suppressed under most normal circumstances [Matin 74]. Gaps are defined as the period of time where no eye position was registered, excluding blinks. Blinks are filtered out for each user. Gaps may be produced because the user changes the visual focus from the screen to other external actuator, or because she or he gets too close to the monitor, failing to practice good postures.

In order to detect saccades, fixations and gaps, processing techniques need to be applied to the raw data file. These steps are based on the Tobii I-VT Fixation Filter algorithm [Olsen 12a]. They have been all implemented in the Python programming language and are outlined in Figure 3.9.

Fig. 3.9 Low level design - Data processing module - Architecture



As Figure 3.9 shows, the first step in the processing algorithm is to apply the gap fill-in interpolation function. This step consists of filling in data where data is missing due to tracking problems which are not related to participants' behaviour (such as blinks, or when the user looks away from the screen). In order to distinguish between tracking problems and users' behaviour, a max gap length is set which limits the maximum length of the gap to be filled in.

After the gaps are filled in, the noise reduction function is applied. This function is based on a low-pass filter which aims to smooth out the noise.

The third step is the velocity calculator, which relates each sample with its velocity, in terms of visual angle (degrees per second). In order to reduce the impact of noise, velocity for each sample is calculated as the average velocity of a period of time, taking as central data input the current sample. This is done using a window length of 20 ms, which, according to the literature, has been found to handle a reasonable level of noise without distorting the signal [Olsen 12a].

The I-VT classifier applied to the signal is based on the one described by Komogortsev *et al.* [Komogortsev 10] and outlined in the Tobii White Paper [Olsen 12a]. The classifier determines which samples belong to a saccade, fixation or gap, based on a velocity threshold and the angle velocities calculated in the previous step. It also groups together consecutive samples using the same classification. The velocity threshold is set to 30 °/s [Olsen 12a, Olsen 12b].

The merge fixations function aims to merge adjacent fixations that have been split up. This is done taking into account two different thresholds, the max-time between fixations which is set to 75 ms [Olsen 12a],which is lower than the normal blink duration [Komogortsev 10] [Ingre 06, Volkmann 80], and the max-angle between fixations which is set at 0.5° [Olsen 12a, Komogortsev 10, Najemnik 05, Nakatani 08, Kliegl 04]

Shorter fixations are filtered and removed. For the purposes of this analysis, 100 ms was set as the lower limit for fixation duration. This value was chosen based on the work of McConkie et al., who concluded that 60 ms must pass before current visual information becomes available to the visual cortex for processing [McConkie 85]. R. Tai *et al* arrived at the lower limit of 100 ms by adding 30 ms, which is the time that elapses, at the end of a fixation, between when a command to move the eyes is sent and the onset of that saccade is reported. They allowed also 10 ms for the processing of any currently observed stimuli, reaching the 100 ms threshold [Tai 06].

After all the processing functions have been applied to the current data, three different gaze data files are created. One with all the fixations, other with all the saccadic movements each participant performed and a third one with relevant information about data loss. As shown in the three listings below, the three different processed data files have a similar structure to raw data.

– **Fixation data file**

The stored fixation data saves all the fixations recorded during the exercise, along with the current activity information, user data and the duration, start time, end time and position of each fixation. Its information outline is displayed in Listing 3.4.

Listing 3.2 Fixation data example

```
<user date="2014-05-22" id="12" sessionid="897" time="09
    :15:27">
  <exercise id="puzzle" level="Level1" mode="performance">
    <fixationData>
      <fixation>
        <time duration="212.356933594" end_time="
            1.1676104137e+12" start_time="1.16761041349e+12
            ">
          <position x="54.2251062717" y="130.508713537" />
        </time>
      </fixation>
      ...
    </fixationData>
  </exercise>
</user>
```

– Saccade data file

The stored saccadic data saves all the saccadic movements recorded during the exercise, along with the current activity information, user data and the duration, start time, end time, initial position and final position. Its information outline is displayed in Listing 3.3.

Listing 3.3 Saccade data example

```
<user date="2014-05-22" id="12" sessionid="897" time="09
    :15:27">
  <exercise id="puzzle" level="Level1" mode="performance">
    <saccadeData>
      <saccade>
         <time duration="94.0268554688" end_time="
             1.16761118002e+12" start_time="1.16761117993e
             +12">
            <init_position start_x="33.3166545786" start_y=
                "284.608986725" />
            <end_position end_x="34.0874523538" end_y="
                223.394850807" />
         </time>
      </saccade>
       ...
    </saccadeData>
  </exercise>
</user>
```

– Gap data file

The stored gap data saves all the different gaps recorded during the exercise, along with the current activity information, user data and the duration, start time, end time, and screen position registered before each gap. The latest valid position of the gaze is useful to identify the area in which the eyetracker lost tracking, which could potentially help to determine the cause of the gap (bad posture, loss of interest...). Its information outline is displayed in Listing 3.3.

Listing 3.4 Gap data example

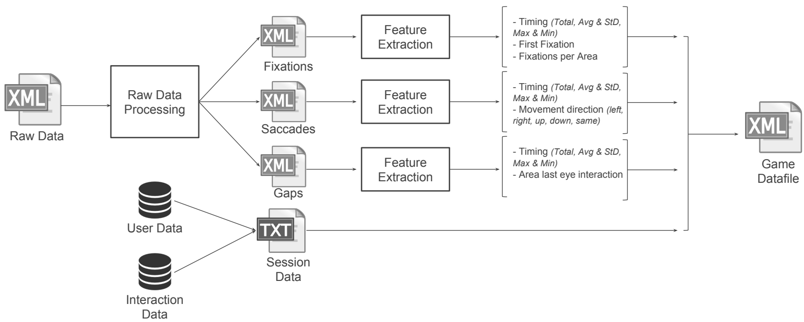
```
<user date="2014-05-22" id="12" sessionid="897" time="09:15:27">
  <exercise id="puzzle" level="Level1" mode="performance">
    <gapData>
      <gap>
        <time duration="226.244140625" end_time="1.16761273771e+12" start_time="1.16761273748e+12">
          <prev_position prev_x="-3.82225275329" prev_y="401.516704147" />
        </time>
      </gap>
      ...
    </gapData>
  </exercise>
</user>
```

3.2.3 C. Feature extraction

The data collected during interaction with the puzzle game is analysed in this block. Features of different nature are then computed and extracted. This section details each of the extracted features. Features are divided into three different categories:

eye movements, performance metrics and users' details, as it is displayed in Figure 3.10.

Fig. 3.10 Low level design - Feature extraction module - Architecture



– **Eye movements**

In order to further analyse eye movements, the display was divided in 9 different areas, as it is shown in Figure 3.11. This division may help to identify the *'hot zones'* in the display, so as to determine where the core process of gaze activity took place.

Fixations Fixations are computed and processed for each user, game and level, and several parameters were drawn from this data: Number of fixations per area in the screen, fixation average duration and standard deviation, fixation maximum and minimum duration.

Saccades Saccadic movements are computed and processed for each user, and various parameters were drawn from this data:

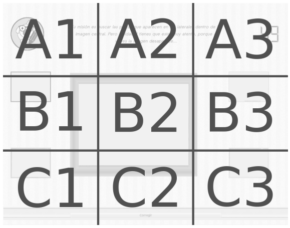


Fig. 3.11 Low level design - Feature extraction module - Image regions

Number of saccades, saccade average duration and standard deviation, saccade maximum and minimum duration, and saccade direction (up, left, right, down, up-right, up-left, down-left, down-right or same quadrant).

Gaps Gaps are the period of time where no eye position was registered, excluding blinks. Various parameters were drawn from this data: Gaps per area in the screen, gap mean duration and standard deviation and the area where the last fixation before gap was located. The last eye movement before the gap was analysed. If it was a saccade near screen's borders, it is more likely than the user have changed the focus of attention. If the last movement is registered in a central position of the monitor, the user may be getting too close to the computer.

– **Performance metrics**

Performance metrics describes the interaction data of each user with the system. These data comprises traditional in-game parameters, that are later computed along with eye gaze data.

Score Score data refers to the number of good and bad choices that users performed per level.

Time Time data refers to the final amount of time that participants employed per level (with a maximum of 50 seconds, which is the pre-set time per exercise).

– **User details**

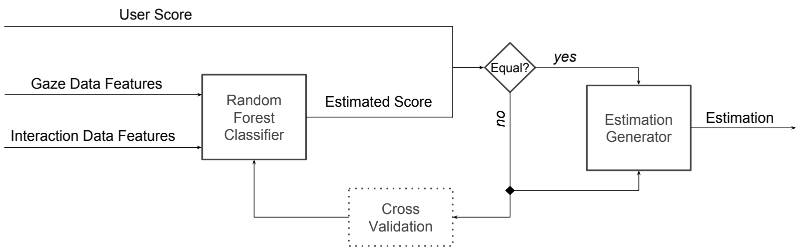
User relevant details are also considered for building the Game Global Profile and for the later classification process: Age, gender and whether the user is diagnosed with an attention related disorder or learning disability.

Feature selection is a determining factor when classifying patterns. Features need to be insensitive to noise and separated from each other. Their main purpose is to objectively describe certain aspects, in this case of the attention and performance process in intelligent therapies aimed at children.

3.2.4 D. Classification

This section outlines the steps taken in the classification process. The aim of this block is to assess the feasibility of using a set of combined features to evaluate user performance. These features are related to user interaction, timing and visual attention, as it has been described in subsection 3.2.3. The architecture of this modules is described in Figure 3.12.

Fig. 3.12 Low level design - Classification module - Architecture



As it is displayed in Figure 3.12, interaction and gaze features extracted in the previous model are input into a random forest classifier. This classifier belongs to the family of ensemble learning algorithms that work by running a base learning algorithm multiple times, voting out the resulting hypotheses [Dietterich 02]. Ensemble learning has received an increasing interest, since it is more accurate and robust to noise than single classifiers [Breiman 96, Rodriguez-Galiano 12]. This classifier is defined as a combination of tree predictors. Each tree depends on the values of a random vector sampled independently and with the same distribution for all trees [Breiman 01]. Using the random selection of features yields error rates that compare favorably to AdaBoost [Schapire 98], but are more robust with noise handling [Breiman 01].

In this system architecture, this classifier receives both interaction and gaze features as inputs. Based on prior knowledge acquired by the system through the different game iterations, it gives an estimation of the answer scored by the player. This estimation is then compared with the actual score obtained by the player for that interaction data.

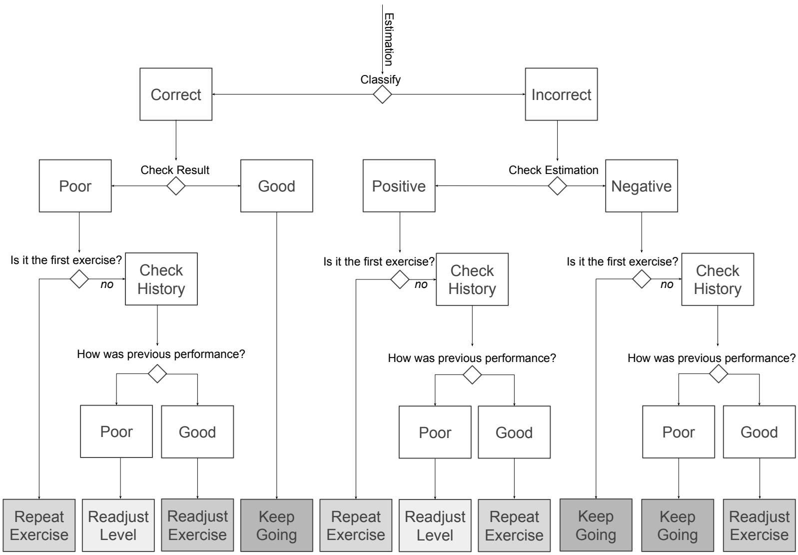
If the estimation is correct, this data is passed to the interaction manager, for this module to determine the next suitable level. If the estimation is incorrect, this data is still being send to the interaction manager together with a qualifier: it can be a positive (the result was over estimated) or negative (the result was under estimated) qualification. In case of an incorrect estimation, this data is added to the prior knowledge of the system and by applying each time a cross-validation classification process of 100 iterations to all of the available data.

3.2.5 E. Interaction manager

This module is in charge of determining which is the most suitable level and game option for the next iteration of the game. It gives a customized option for each player every time he or she completes, successfully or not, a level. Figure 3.13 displays the hierarchical tree that compounds this module's architecture.

This module is based on the estimation given by the classifier block. This estimation can be labelled as correct: if the automatic classification of the interaction and gaze data matches the real score obtained by the player, or incorrect: if the classifier misinterprets the given features because they do not match with its previous knowledge.

Fig. 3.13 Low level design - Interaction manager module - Architecture



Incorrect estimations can be due to a over estimation of the given features, which are labelled as positive incorrect estimations, or due to an under estimation of the given features, which are named as negative incorrect estimations.

Based on the estimation given by the classifier block and the previous history of the current player stored in the system, in case there is any, the hierarchical tree architecture of this module selects the most suitable level.

“Who in the world am I? Ah, that’s the great puzzle."
Alice in Wonderland and through the Looking Glass

Lewis Carroll

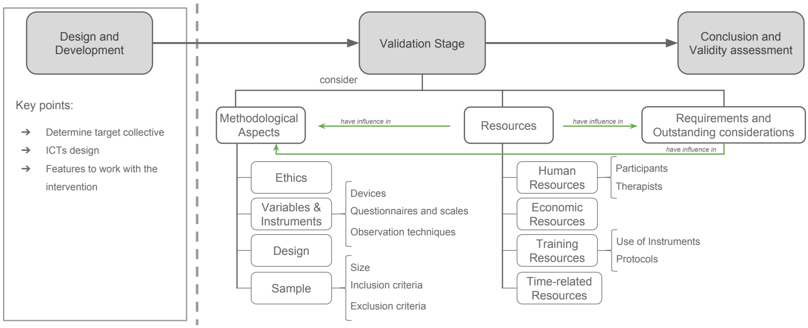
4
Experiment Design

This chapter explains the methodology behind each of the different user studies that compound this dissertation. This work can be divided into three main user studies, two of whom have been labelled as pilot studies. The aim of this chapter is to give an overview about the different protocols used, along with the overall research methodology followed for carrying out each of these experiments.

4.1 Overall methodology

Prior to the performance of the different studies presented in this dissertation, a research methodology is defined. Figure 4.1 summarizes the whole process followed during the design of the two pilot studies and final research study.

Fig. 4.1 Methodology stages



Key factors of the design and development stage are to determine the target collective, ICT design and features to work with the intervention. In this case, the target collective of the study are children and teenagers with and without learning disabilities.

The performance of some usability tests during the design and development of the intervention may provide developers with data which may help them to adjust the system to focus on their target audience. Once the design and development stages are over, the pre-defined validation stage starts. This is where the evaluation criteria detailed on each article takes place. At this stage, it is important to take into account several aspects.

Special attention need to be paid to methodological considerations, and perform the study in the same conditions with every user. In order to ensure this, environment ethics and sample related features must be considered along with the selection of appropriate variables and instruments.

Resource management refers not only the human resources but also economic, training and time related resources. It is important to assess the impact of the study in all these different scales. Researchers need to be sure that the outcomes expected from the study are worth the total cost of conducting the investigation.

Bearing all this criteria in mind and establishing the collaboration channels for being able to work with the selected sample of users, the different methodologies were implemented for working in all the studies. The following sections will give an overview of the methods followed in each of them.

4.2 Pilot Study I: Adaptive Tele-Therapies based on serious games

This study is the first step in determining the suitability of adaptive intelligent interventions based on serious games. It presents an intelligent tele-therapy tool based on serious games for health.

It is aimed at the improvement of time management skills and the prioritisation of tasks in children and teenagers. It was designed as a proof of concept prior to the development of the presented framework for serious games using biofeedback. Its main objective was to assess and evaluate the use of adaptive tele-therapies within a group of typically developing children and adolescents aged between 12 and 19 years old. This proof of concept tool was designed for objectively evaluating user predisposition to adaptive systems based on serious games.

The following paragraphs detail the materials and methods used in the analysis of this user study. The tool consisted on a virtual interactive balance which, by making use of decision trees and user-entered parameters, is capable of prioritising between two proposed activities.

4.2.1 Participants

Preliminary evaluation of the tool was made with a group of typically developing children and adolescents aged between 12 and 19 years old, with an average age of 16.23 years old.

Table 4.1 Pilot Study I - Participants

Centre	**Gender**		**Total**
Volunteers	Females	7	17
	Males	10	

Seventeen randomly selected participants (7 women and 10 men), selected from a group of volunteers, took part in these trials. These users are resident in the Basque Country, Spain, have not been diagnosed with any learning disability and have Spanish as their mother tongue. For individuals under 18 years old the approval of parents or guardians was requested prior to conducting the surveys.

Users responded to questionnaires and tests independently, they used their own online devices (mobiles, tablets and/or smartphones) to do so. Participants were allowed to choose the time of day when they wanted to perform the evaluation of the tool. This evaluation took place during the months of July and August 2013.

4.2.2 Methods

The online tool was produced in Django, the high-level Web framework for Python. The results obtained and the necessary parameters were stored in a MySQL database. User interface and user interaction were developed using JQuery, Javascript and CSS, with Touch Punch being used to adapt the tool to touch screens. The priority activity decision algorithm was calculated by using a decision tree implemented in Python. The technologies and techniques employed here were selected in order to boost user-involvement and the adaptation of the system to users' needs.

For the evaluation of users' time management skills prior to the test, the Time Management Behavior Questionnaire (hereinafter TMBQ) by Macan et al. [Macan 90, Macan 94, Macan 96] was used, adapted to the Spanish language. This scale was chosen for its prestige and validity in the measurement of time management skills, especially for the age range selected in the test. The Spanish version was validated and accepted [García-Ros 12, PuJoL 13]. The aim of this questionnaire was to learn about and analyse the management skills of the participating users. It is composed of 34 items divided into four sections. Items were evaluated using a Likert scale ranging from 1 (“never") to 5 (“always"). This questionnaire was available online, by means of the Google Forms tools during the months of July and August 2013.

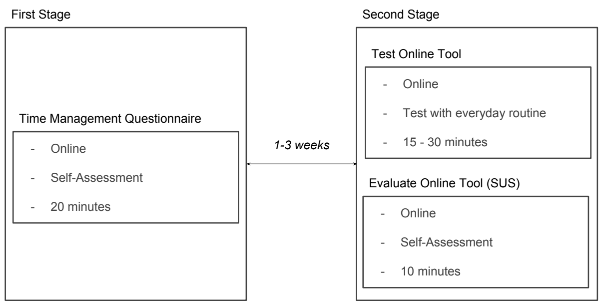
The discussed tool was evaluated by a user satisfaction test based on the *System Usability Scale (SUS)* [Brooke 96]. Items were evaluated by using a Likert scale ranging from 1 (“strongly agree") to 5 (“strongly disagree"). The completion of the questionnaire aims to continue to adapt the system to the users' final needs. This questionnaire was available online, by means of the Google Forms tools during the months of July and August 2013.

4.2.3 Experimental procedure

The tool's complete testing procedure was carried out online, outside of the laboratory environment, and in the course of users' everyday life. These conditions were considered appropriate due to the nature of the system.

Each user answered the questions and tested the tool on their own device, in order to verify the accessibility of the system and whether it fulfilled the criterion of being multi-platform. Figure 4.2 shows the procedure followed during tests, with the time interval left between the two testing stages.

Fig. 4.2 Pilot study I - Research procedure



As shown in Figure 4.2, the system's testing procedure was carried out in two mutually independent stages, between one to three weeks apart, on the basis of the results obtained in the first stage and adapting it to users' availability.

The first stage involved conducting a questionnaire about time management habits. This questionnaire was prepared online, making use of Google Forms. It consisted of 34 items divided into 4 different areas, each one focused on a specific skill (effective time management, preference for disorganisation, establishing concrete objectives and perception of time). Items were evaluated by using a Likert scale ranging from 1 (“never") to 5 (“always"). The purpose of this analysis was to establish the global situation of the user in terms of time management skills. This test took around 15 to 20 minutes to complete.

The second stage was comprised of two sections: firstly, the use of the application, and secondly, the performance of a systematic evaluation of the questionnaire.

The application test was conducted in the course of a full day, integrating it in a natural way into the users’ routine. This testing required various easy questions in the application, on the basis of the activities carried out. It took between 15 to 30 minutes, depending on the number of attempts users made.

Once this battery test was finished, the users took the online usability questionnaire SUS, translated into Spanish. This usability questionnaire was composed of 10 items evaluated by a Likert scale ranging from 1 (“totally disagree") to 5 (“totally agree") and aimed at measuring system usability. This survey was conducted in order to prove if the design and purpose of the system was easy to learn and user-friendly. This test took around 5 to 10 minutes to complete.

All the questionnaires and tests of the application were carried out in an autonomous way by users, outside of the laboratory environment. The two testing stages were carried out on their own devices, adapting the tests to the aim of the application; they took between 35 minutes to 1 hour to be completed successfully.

The results obtained in these two stages were uploaded and analysed in the IBM – SPSS Statistics predictive analysis software tool.

4.3 Pilot Study II: Assessing Visual Attention Using Eye Tracking Sensors in Intelligent Cognitive Therapies Based on Serious Games

Once the usability and viability of the use of adaptive therapies based on serious games has been analysed, a gaze analysis study was designed. This study examines the use of eye tracking sensors as a mean to identify children's behaviour in attention-enhancement therapies. This study sets a step further in the development of this dissertation, trying to give an answer to the suitability of using gaze behaviour for determining different children's visual attention profiles while playing. This study took place after the confirmation of the suitability of adaptive therapies based on serious games given by the first pilot study.

For the purpose of gaze behavioural data analysis, a set of data collected from 32 children with different attention skills was analysed during their interaction with a set of puzzle games. We hypothesize that participants with better performance may have quantifiable different eye-movement patterns from users with poorer results.

The following paragraphs detail the materials and methods used in for the analysis of this user study.

4.3.1 Participants

The process for assessing attention was performed with a group of typically developing children. This process relies on data recorded with an eye tracking sensor. Participants were aged between 8 and 12 years, with an average age of 10.0 (SD = 1.34).

Table 4.2 Pilot Study II - Participants

Centre	Gender		Total
Colegio Vizcaya	Females	13	32
	Males	19	

Thirty-two randomly-selected participants (13 girls and 19 boys) were selected from a group of 83 volunteers by their teachers. This sample size was considered adequate for the purpose of the outlined pilot study [Hertzog 08].

These children live in the Basque Country, Spain, have not been diagnosed with any attention-related disorder and speak Spanish as their mother tongue. All of the participants were recruited from the Colegio Vizcaya School.

Since they were mature minors, the approval of parents or guardians was requested prior to conducting the study. This approval consisted of an informed consent following receipt of a detailed description of the study, distributed via the school's regular newsletter.

4.3.2 Materials

All participants in the study completed the same assessment, which consisted of a puzzle exercise with four different levels of difficulty. Users have to connect each of the four slices presented in the exercise with its corresponding part in the main image. As Figure 4.3 displays, all of the participants were presented with the same image for each level, and all of the elements in the user interface appeared in the same part of the screen at each level.

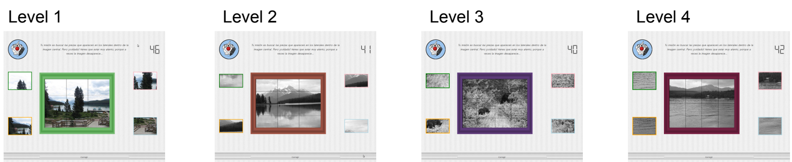


Fig. 4.3 Pilot study II - Task's different levels

The main image and the slices appeared in the middle of the screen, occupying the whole display from left to right. The question stem appeared in the upper middle part of the screen. The button to advance to the next level appeared at the lower middle part of each screen. The consistent layout of the screen was intended to minimize wide eye movements.

Different levels' settings are outlined in Table 4.3. All of the users had a maximum pre-set time of 50 seconds to complete each of the levels. However, if they finished the level before the time ended, they could go on to the next exercise. Depending on the level, the displayed image was labeled as easy, medium or hard. Only Level 1 is displayed in color. Hard images have very similar slices and are more complicated to complete. Table 4.3 displays the different levels' settings.

Table 4.3 Different levels' settings.

	Time (s)	Grid Size	No. of Slices	Display		Image Level		
				Color	Greyscale	Easy	Medium	Hard
Level 1	50	3×3	9	x		x		
Level 2	50	3×4	12		x	x		
Level 3	50	4×4	16		x		x	
Level 4	50	5×4	20		x			x

4.3.3 Devices and technologies

All of the data for this study were collected on the same device, which was located at the children's school outside the laboratory environment. These conditions were considered appropriate due to the nature of the system.

The set of puzzles was developed in Python. The results obtained and the necessary parameters were stored in a SQLite database. The user interface and user interaction were developed using PyQT4. The classification process was implemented using the Scikit-learn library for machine learning in Python. Fixation heat maps were produced based on the implementation developed by jjguy.

The puzzles were displayed on a 19-inch Lenovo monitor interface with an Acer Aspire Timeline X laptop running on Ubuntu 12.04. All of the text in the different exercises was displayed as black text against a light-grey background following normal grammatical conventions in Spanish. Images were inserted as JPEG digital pictures scaled from their original versions. Response selection and any changes were stored by monitoring the user interaction and recording eye movements with a Tobii X1 Light eye tracker sensor.

The eye tracker is a non-invasive sensor with remote function. Participants were not required to remove their glasses or contact lenses during the tests. Accuracy under ideal conditions is 0.5deg of the visual angle, while the sampling rate in this study was typically 28–32 Hz. The Tobii X1 light sensor was located beneath the computer monitor with the headrest fastened to the front edge of the desk, monitoring the participant's head. The laptop was located behind the monitor, without interfering with the participants' field of vision.

A typical experimental trial including calibration lasted less than 20 min for each participant.

4.3.4 Experimental procedure

Prior to this study, participants' teachers responded autonomously to the EDAH scale for the evaluation of ADHD in the questionnaire on children between 6 and 12 years old [Farré 01]. Farré and Narbona designed this scale based on their experience with the adapted Conners questionnaire [Conners 98]. The EDAH measures the main characteristics of ADHD and the behavioral problems that may coexist with attentional deficit. This questionnaire was used to ensure that participants did not exhibit any ADHD-related behavior.

After completing the exercises, participants themselves were asked to fill in a usability questionnaire. The usability of the system was evaluated by a user satisfaction test based on the System Usability Scale [Brooke 96]. This questionnaire consists of 10 items, which were evaluated by using a Likert scale ranging from 1, strongly agree, to 5, strongly disagree. Through feedback from this questionnaire, researchers will be able to continue to adapt the system to users' final needs.

Before completing the usability questionnaire, participants were seated in front of the eye tracking sensor to permit data collection. Users were seated opposite the center of the monitor, after adjusting the seating position to their height. Once they were aligned with the screen, the calibration process started, which took between 2 and 5 min per child. This calibration entails a visual target that moves around the screen. Participants were asked to follow this target with their gaze for a period of time. The target consists of a calibration grid with 5 positions, one on each corner of the screen and the last one right in the screen's center. The target consists of different calibration bullet points that appeared one after the other in the same order for all participants, starting from the top left corner.

Prior to the start of the exercises, participants were told in which kind of tasks they were taking part. They were also introduced to the eye tracking technology, and the sensor functionality was explained.

Participants used the system and filled in the questionnaire in a controlled environment, with a researcher observing and keeping track of all of the behavioural aspects of the study, but not interfering in the experimental setting.

4.4 Study: Gaze Behaviour Analysis in Cognitive Therapies based on Serious Games

The purpose of this study is to keep exploring the use of eye tracking sensors to evaluate, classify and assess the behaviour of children in attention-related cognitive therapies based on serious games. The final aim of this research is to determine the utility of eye-related data as an input biofeedback signal for attention improvement therapies. Authors will try to give an answer to the following research question: Is visual interaction an efficient way of determining the attention degree and/or the performance interaction of different users? This article is the continuation of the previous Pilot Study II detailed in section 4.3.

For the purpose of gaze behavioural data analysis, a set of data collected from 82 children with different attention skills was analysed during their interaction with a set of games. Interaction data and gaze behaviour patterns were recorded for further analysis.

The following paragraphs detail the materials and methods used in for the analysis of this user study.

4.4.1 Participants

The process for assessing attention was performed with a group of children with different levels of attention capacities. This process relies on data recorded with an eye tracking sensor. Participants were aged between 8 and 12 years, with an average age of 10 ($SD = 1.14$).

Eighty-two randomly selected participants (38 girls and 44 boys), were selected from a group of 173 volunteers by their teachers and psychologists. This sample size was considered adequate for the purpose of the outlined study [Hertzog 08].

71 children (40 boys and 31 girls) out of the 82 total were recruited from the Colegio Vizcaya School. Five of them, all

Table 4.4 Study - Participants

Centre	Gender		Total
Colegio Vizcaya	Females	40	71
	Males	31	
Albor Cohs	Females	4	11
	Males	7	

boys, were diagnosed with some kind of learning difficulty. The remaining 11 children (4 boys and 7 girls) were recruited from the Albor-Cohs psychology cabinet, all of them were diagnosed with some kind of learning difficulty.

These children live in the Basque Country, Spain. 16 were diagnosed with an attention-related disorder or learning difficulty, all of them were taking medication during the test. All children have Spanish as their mother tongue, and were enrolled either in language model A (only Spanish) or B (Spanish and Basque) at school.

Since they were mature minors, the approval of parents or guardians was requested prior to conducting the study. This approval consisted of an informed consent following receipt of a detailed description of study, distributed via the school's or cabinet's regular newsletter.

4.4.2 Materials

All of the data for this study was collected on the same device, which was located at the children's school or at the psychology cabinet, outside the laboratory environment. These conditions were considered appropriate due to the nature of the system.

Materials used in the performance of this study were the same of the ones described in previous section 4.3.2.

4.4.3 Devices and technologies

Devices and technologies used in the performance of this study were the same of the ones described in previous section 4.3.3.

4.4.4 Experimental procedure

Being this study a continuation and extension of Pilot Study II, the experimental procedure carried out in this study was the same of the one described in previous section 4.3.4.

"But, I nearly forgot, you must close your eyes otherwise you won't see any-thing."
Alice in Wonderland and through the Looking Glass

Lewis Carroll

5 Results

This chapter includes the three user studies that belong to the main development of this dissertation. These studies were conceived as a mean to validate the hypothesis of this work, the suitability of visual attention in player profiling in serious games. Each of the studies included in this section will try to go one step further in the objective determination of the validity of this hypothesis. They have been included with the format they have been published in three different journals with impact factor. One of the articles, the final study, is currently under revision and un-published.

For ease of reference, each article is preceded by a summary of its main contributions and objectives, its purpose inside the whole dissertation work and the details of where and when was it published.

5.1 ARTICLE II: Adaptive Tele-Therapies based on serious games

This article presents an intelligent tele-therapy tool based on Serious Games for Health, aimed at the improvement of time management skills and the prioritisation of tasks. The use of techniques employed in the design of videogames, combined with the establishment of certain goals, guidelines and rules which not only encourage the use of these technologies, but also serve to reinforce the training of such skills as working memory, stimulation of attention, concentration and the aforementioned management skills.

The developed tele-system is based on the use of decision trees within Django, a high-level Python Web framework. The technologies and techniques used were selected so as to boost user involvement and to enable the system to be easily customised.

This article was designed as a first proof of concept for determining the suitability of adaptive intelligent intervention based on serious games in children and teenagers. The main contribution of this article is the assessment of the usability of this kind of systems with the focus group of youngsters between 12 and 19 years old.

To do so, I relied on the collection of parameters and the conduct of surveys for assessing time management skills, as well as measuring system usability and availability, while players were interacting with a key-turn adaptive serious games development for fostering time management.

The final calculations based on the usability questionnaire resulted in an average score of 78.75 out of 100. The main conclusion of this study was that the creation of a customisable tool capable of working with different skills, in conjunction with the replication of the current study, may help to understand users' needs, as well as boosting time management skills among teenagers.

This study laid the foundations of the presented framework for intelligent adaptive interventions based on serious games. This article was published on the *International Journal of Environmental Research and Public Health (IF: 2.063 [Q2])*, Volume 11, Issue 1 in January 2014 [Frutos-Pascual 14]. This article has been included on its original full version that is available through the journal web page.

Int. J. Environ. Res. Public Health **2014**, *11*, 749-772; doi:10.3390/ijerph110100749

OPEN ACCESS

International Journal of
Environmental Research and Public Health
ISSN 1660-4601
www.mdpi.com/journal/ijerph

Article

Adaptive Tele-Therapies Based on Serious Games for Health for People with Time-Management and Organisational Problems: Preliminary Results

Maite Frutos-Pascual *, Begoña García Zapirain and Amaia Méndez Zorrilla

DeustoTech Life (eVIDA), Faculty of Engineering, University of Deusto, Bilbao 48007, Spain;
E-Mails: mbgarciazapi@deusto.es (B.G.Z.); amaia.mendez@deusto.es (A.M.Z.)

***** Author to whom correspondence should be addressed; E-Mail: maitefrutos@deusto.es;
Tel.: +34-944-139-000 (ext. 2051).

Received: 7 October 2013; in revised form: 16 December 2013 / Accepted: 17 December 2013 / Published: 7 January 2014

Abstract: Attention Deficit with Hyperactivity Disorder (ADHD) is one of the most prevalent disorders within the child population today. Inattention problems can lead to greater difficulties in completing assignments, as well as problems with time management and prioritisation of tasks. This article presents an intelligent tele-therapy tool based on Serious Games for Health, aimed at the improvement of time management skills and the prioritisation of tasks. This tele-system is based on the use of decision trees within Django, a high-level Python Web framework. The technologies and techniques used were selected so as to boost user involvement and to enable the system to be easily customised. This article shows the preliminary results of the pilot-phase in an experiment performed to evaluate the use of adaptive tele-therapies within a group of typically developing children and adolescents aged between 12 and 19 years old without ADHD. To do so, we relied on the collection of parameters and the conduct of surveys for assessing time management skills, as well as measuring system usability and availability. The results of a time management survey highlighted that the users involved in the trial did not use any specific or effective time management techniques, scoring 1.98 and 2.30 out of 5 points in this area for ages under 15 and over 16 years old, respectively. The final calculations based on the usability questionnaire resulted in an average score of 78.75 out of 100. The creation of a customisable tool capable of working with different skills, in conjunction with the replication of the current study, may help to understand these users' needs, as well as boosting time management skills among teenagers with and without ADHD.

Keywords: ADHD; tele-therapy; time management; Serious Games; intelligent systems

1. Introduction

Attention deficit with hyperactivity disorder (ADHD) is one of the most prevalent disorders within the child population today, affecting an estimated 5.29% of children worldwide [1]. At present the diagnosis and treatment of ADHD should be performed in a multimodal environment in which information is provided to parents, teachers and patients with ADHD. In addition, the latter must have psycho-educational support at school, as well as psychological support in cases where it is strictly necessary (for the individual, the family and the community) and pharmacological treatment, if required by the specific symptoms [2]. Over-diagnosis must be minimised and a strict control of the administration of medication must be ensured [3].

ADHD usually manifests before the child is 7 years of age and is characterised by a certain degree of impulsiveness, inattention and excess of activity which does not match the child's developmental age and is not appropriate in all situations that may arise [4]. In addition to these hyperactivity and inattention problems, organisational difficulties such as time management may also occur, especially within the school setting [5], as well as in adulthood, in the individual's professional life.

Time management is meant as the ability to use the resource of time in the most efficient way. This efficiency depends upon task management and prioritisation, routine planning and the ability to remember what is still to be done. These management, organisation and perception-related problems are detailed in the current definitions of ADHD [6–8]. Toplak *et al*., in their review of the literature, reported the most relevant features associated with these management problems. These include delays in the performance of specific tasks, problems related to turn-taking and premature answers [9], as well as difficulty in prioritising tasks.

Time management and the estimation of the likely duration of tasks differ among the users diagnosed with ADHD and those belonging to the control group, both in adulthood and in childhood. This difference has been reflected in many studies, especially by assessing the estimation of the length of short periods of time while performing specific psychological tests [10–15].

Parents, doctors and teachers all report these types of behaviour, especially when they are linked to a lack of organisation and the acquisition of the appropriate time-planning strategies [16]. Organisational problems such as the ones derived from poor planning of an academic course or a subject, lead to bad academic results that can end in school failure in the most dramatic cases.

It is important to work on these management skills from childhood, providing the tools and the knowledge necessary to empower users. The improvement and promotion of these skills is directly connected, in most cases, to an improvement in school performance [17].

Succeeding in motivating a child affected by ADHD to undertake academic activities is a key element in their development. There is a need to encourage their learning process and memory, concentration and time management skills using structured activities, clear rules and striking materials. The possible solutions or improvements include developing the interest of the group towards these activities, which, together with an early and continuous intervention [18], can achieve a behavioural reduction of these problems and the negative consequences that arise in the mid- to long-term.

The use of techniques employed in the design of videogames, combined with the establishment of certain goals, guidelines and rules which not only encourage the use of these type of technologies, but also serve to reinforce the training of such skills as working memory, stimulation of attention, concentration and the aforementioned management skills, represent a new type of effective therapy in their application to ADHD [19,20].

The online performance of these therapies provides the users involved with greater autonomy, fostering communication between participating groups (doctors, psychologists, parents, teachers, and children, among others) [21]. Moreover, tele-therapies can be used and implemented on a simultaneous basis, and they are available in different places and on different devices [22].

The aim of this article is to assess the time management abilities of a group of typically developing children and adolescents, and to make a preliminary evaluation of the utility, usability and availability of the tool with respect to that group. These tests are intended to serve as guidance for adapting the tool presented in this article to final users, helping them to evaluate whether it could be used on a daily basis with users with and without ADHD. An important additional aim of the present study is to discuss whether the use of online game-based interventions fosters motivation and engagement in therapies. A supplementary objective is the evaluation of the use of online therapies to promote availability and convenience while using the tool. A final and future objective is to further assess the impact on the mid-term usage of this tool in teenagers with ADHD.

To accomplish the stated aims, this article presents an online interactive tool aimed to assist in the prioritisation of tasks by children and adolescents with ADHD, based on the techniques used in Serious Games for Health. This tool has been designed to strengthen time management skills related to day planning, and to help parents and instructors to determine which areas cause more confusion when establishing organisation strategies. The ultimate goal of this application is to serve as guidance to final users, providing them with advice and recommendations about management, as well as giving them sufficient autonomy to plan their day, by prioritising certain tasks above others.

In Section 2, the use of games for health in the diagnosis and treatment of ADHD will be placed in context, and the use of time management and organisation strategies in people with ADHD will also be studied. Subsequently, materials and method sections will be introduced along with the results, and followed by a final discussion and conclusion.

2. Background

This section reviews the previous work available in the literature on the use of Serious Games for Health in the field of ADHD, as well as recent studies on the evaluation of time management skills in people with ADHD.

2.1. Serious Games for Health and ADHD

This section analyses the use of Serious Games for Health in the diagnosis and treatment of ADHD. These games are specially designed to evaluate, diagnose, and be employed in therapies with children affected by ADHD, and have captured the attention of schools in recent years. The reason for this interest lies in the fact that many children who do not inhibit their hyperactive and/or impulsive behaviour, are capable of regulating it while playing videogames that motivate and foster their concentration [23,24].

It also should not escape notice, especially in relation to commercial videogames, that there seems to be evidence concerning the vulnerability of certain subgroups diagnosed with ADHD to videogame addiction [25,26]. Nonetheless, recent studies show that there is no direct link between the exposure to the use of videogames and attention problems, but there are other risk factors such as an inadequate family environment or anxiety problems [27].

2.1.1. Diagnosis and Evaluation

This section details the most relevant publications related to the diagnosis and evaluation of ADHD using Serious Games for Health. A game designed and tested to help in the diagnosis of ADHD called *The Supermarket Game* has been developed in Brazil. It has been tested with 80 children diagnosed with ADHD and it has proven to be an efficient way of distinguishing among children who have, and who have not been diagnosed with ADHD. Decision-making algorithms are currently being optimised in order to discriminate among different types of ADHD [28,29]. Several years earlier, Rizzo, Bewerly *et al*. designed a 3D virtual classroom to help in the diagnosis of ADHD based on the interaction with the system. This system was tested with eight children with ADHD and 10 non-diagnosed children with positive and accurate results [30]. The concept of the virtual classroom was also used in the AULA Nesplora project, which implemented a version of Conners' Continuous Performance Test using a 3D virtual classroom. It was initially tested with 57 children and it is currently being used in some Spanish assessment environments [31].

Cyber Cruiser is a car rally-style game oriented towards the assessment of the executive function and prospective memory in children with ADHD, and has been used to establish differences in these skills among diagnosed and control groups in a sample of 80 children [32], with positive results as assessed by the Conners' Parent Rating Scale [33]. The improvement of executive memory has undergone several developments based on games over the past few years. The most current one is the design and implementation of two sets of Serious Games, *Cognitive Carnival* and *Caribbean Quest* in 2012 [34,35].

2.1.2. Treatment

At present, there exist some evidence indicating that videogames can contribute to the improvement, regulation and standardisation of some symptoms related to ADHD. The first therapeutic videogame available to do so was developed by Pope and Bogart in 1996, and is a modification of software developed by NASA for the training of pilots, based on the progressive adaptation of a computer programme to user attention levels [36]. In 2001, Pope and Palsson went further and carried out a broader development of this patent as an intervention for the improvement of

ADHD by modifying commercial videogames in combination with the use of the measurements obtained by an EEG [37].

Another alternative is *The Journey to Wild Divine*, a 3D virtual world created to be controlled through biofeedback, and available for purchase. This game is controlled by relying on relaxation techniques, which are very useful in the regulation or monitoring of hyperactivity and impulsivity in children and adolescents with ADHD [38]. At the end of 2012 a study started in London to evaluate self-management skills in 25 children with ADHD, in which a virtual helicopter was controlled by using Images by Magnetic Resonance (IMR) [39] to measure the activity on certain areas of the brain. In 2008 a game called *Self City* was released. It is oriented towards the improvement of social skills in adolescents diagnosed with ADHD and/or Pervasive Developmental Disorder. This game was developed in the shape of a 3D virtual world in which the adolescent is forced to cope with several situations [40].

The games not specifically designed for ADHD, but oriented towards a wide spectrum of disorders in children include *Play Mancer* and *Personal Investigator* (*PI*). *Play Mancer* emerged from a European Cooperation Project in 2007, aimed at the creation of a common framework for the development of therapy-oriented Serious Games [41,42]. *PI* is a 3D game specially designed to help adolescents with disorders through the use of Focused Solution Therapies—SFT and therapy games. It was initially tested with 4 adolescents with positive results [43].

In addition to the games specifically designed for the treatment and/or diagnosis of ADHD, among another disorders, commercial videogames adapted to ADHD therapy programs have also been used over the past few years, including *Robomemo* (CogMed, Stockholm, Sweden) [44]. Table 1 shows a summary of the results obtained.

Building on the background developed in this section along with the information available in Table 1, the authors would like to highlight several projects classified into two different groups: Number of users involved; and systems used outside of a laboratory setting.

Due to the number of users involved in the testing of the project, which strengthens the systems described, *The Supermarket Game* and *CyberCruiser* can be featured. These two projects have been tested with over 50 users.

The next group of projects have been selected due to their use outside of a laboratory setting. In this section *Aula Nesplora*, *The Journey to Wild Divine* and *Play Attention* need to be considered. These three systems are currently being used outside of the laboratory by professionals in psychotherapeutic or scholastic settings (*Aula Nesplora* and *Play Attention*). *The Journey to Wild Divine* is a commercial title available for online purchase.

Table 1. Review of Serious Games and ADHD.

Authors	Game Name	Year	Country	Type	Goal	Technologies
Pope and Bogart [36]	-	1994	US	Development	Fostering attention	EEG, PC
Peter Freer [45]	*Play Attention*	2000	US	Development/In Use	Fostering attention	EEG, PC
Kerns [32]	*Cyber-Cruiser*	2000	CA	Development	Evaluating executive functioning and prospective memory	PC
Pope and Palsson [37]		2001	US	Commercial videogame adaptation	Fostering attention	Play Station, EEG
Bell, Smith *et al.* [38]	*The Journey to Wild Divine*	2003	US	Commercial	Relaxation, Mindfulness	Biofeedback, PC
Rizzo, Bewerly *et al.* [30]	-	2004	US	Development	ADHD diagnosis	3D virtual classroom, PC
Coyle, Sharry *et al.* [43]	*Personal Investigator*	2005	IE	Development	Focused Solution Therapies	Virtual world, PC
Andrade *et al.* [28,29]	*Supermarket game*	2006	BR	Development	ADHD diagnosis	PC
Conconi, Jiménez *et al.* [41,42]	*Play Mancer*	2007	EU	Development	Creating a common framework for Serious Games-based therapies	Virtual world, PC
Van Dijk, Hunneman *et al.* [40]	*Self-City*	2008	NL	Development	Fostering social skills	PC
Bartle [34,35]	*Cognitive Carnival*	2012	CA	Development	Executive memory	PC
Bartle [34,35]	*Caribbean Quest*	2012	CA	Development	Executive memory	PC
Rubia *et al.* [39]	-	2012	GB	Development	Self-control	MRI, Virtual helicopter
Díaz-Orueta, García-López *et al.* [31]	*AULA Nesplora*	2013	ES	Development/In Use	Testing attention: Conners' CPT	3D virtual classroom PC

Table 2. Review of Time Management and ADHD.

Authors	Year	Country	Participants	Method	Assessment	Result
Abikoff, Gallagher *et al.* [46]	2013	US	158 Children: 64 Organisational Skills Training 61 Precluded Skills training 33 Wait-List Control	20 individual clinic-based sessions 10–12 weeks	1 month post-treatment 2 in the following school year Measures: Organizational Skills Training (OST) Children's Organisational Skills Scale for Parents (COSS-Parent) Children's Organisational Skills Scale for Teachers (COSS-Teacher)	Promises clinical utility in children with ADHD and organisational deficits
Langberg, Becker *et al.* [47]	2013	US	23 Children	Homework, Organisation and Planning Skills Intervention (HOPS) Included demographic and child characteristics	Parent-rated materials organisation and planning skills	Importance of teaching students with ADHD to use structured organisation systems
Pfiffner, Villodas *et al.* [48]	2013	US	57 Children (mean age 8.1 years) 17 girls 40 boys	Collaborative Life Skills Programme (CLS Programme) 10 schools 3 integrated components over 12 weeks	-	Support the focus of CLS on both ADHD symptom reduction and organisational skills improvement
Parker, Hoffman *et al.* [49]	2013	US	19 students	10 different US campuses One-on-one interviews	Learning and Strategies Inventory	ADHD helped participants to enhance their self-control
Field, Parker *et al.* [50]	2013	US	160 college students: 70 Female 90 Male	Weekly phone-based coaching interviews	Learning and Study Strategies Inventory (LASSI) College Well-Being Scale	Statistically significant higher total scores in both scales

Table 2. *Cont.*

Authors	Year	Country	Participants	Method	Assessment	Result
Hart, Radua *et al.* [51]	2012	UK	150 patients 145 healthy controls	Peak coordinates extracted from: Case-control activation differences Demographic, clinical and methodological variables	Meta regression analyses	Suggests potential normalisation effects on the function of the pre-frontal region with long-term psycho-stimulant treatment
Bioulac, Lallemand *et al.* [52]	2012	US	36 boys 20 ADHD 16 controls	Virtual classroom task	Continuous Performance Test (CPT)	Children with ADHD are vulnerable to time-on-task effect on performance
Langberg, Epstein *et al.* [53]	2012	US	37 middle school students with ADHD 24 HOPS intervention 13 wait-list control	HOPS Intervention 3-month follow-up	Parent and teacher ratings of organisational skills and homework problems were collected. School grades were also collected	Intervention participants did not make significant improvements relative to the comparison group according to teacher ratings
Gureasko-Moore, DuPaul *et al.* [54]	2008	US	3 male students enrolled in a public secondary school	Training in self-management procedures.	-	Results were consistent across the 3 participants.

By examining Table 1 it can be seen that very little work on time managment skills has been done, the analysed references being mostly focused on attention problems. As was shown in Section 1, the literature has established that developing these management skills constitutes a key factor for children and teenagers diagnosed with ADHD, due to their issues with estimating and prioritising tasks. Upon review of the literature, it followed that there is clearly a need to work on these problems effectively, which resulted in the solution presented in this article.

This solution is focused on techniques used in the field of Serious Games, combined with work on time management skills, particularly centred on prioritising tasks. Furthermore, the fact that these exercises can be carried out online makes this tool accessible anywhere, and could serve as a consultation tool in everyday life.

2.2. Time Management and Organisational Skills in ADHD

This section reviews the latest available interventions that are focused on the enhancement of time management skills and organisational behaviour in children and teenagers with ADHD. People diagnosed with ADHD may express organisational and time estimation problems [55,56]. Upon review of the literature, several authors have centred their attention on working and improving these skills with children and teenagers with ADHD. The following paragraphs review the latest studies about time management and organisational skills intervention.

Abikoff and his colleagues have been working for the last decade on the analysis and impact of various approaches to organisational capacities and time management skills on children with ADHD. This paragraph outlines some examples of their research. In 2003, Abikoff, Gallagher *et al*. performed an assessment, analysis and treatment of time management skills and planning deficits in children with ADHD [57]. Later, in 2009, Abikoff, Nissley *et al*. evaluated the effect of medication (methylphenidate-osmotic-release oral system [MPH-OROS]) on Organisation, Time Management and Planning (OTMP) interventions. Results showed that some children remained resistant to treatment. In 2013 they concluded that the importance of performing OTMP interventions with children with ADHD was due to the promising clinical results [46].

In 2013, Langberg, Becker *et al*. demonstrated the importance of teaching organisational and academic planning and management guidelines to children with ADHD. They administered a Homework, Organization and Planning Skills (HOPS) intervention in a school setting [47]. In the same year, Pfiffner, Villodas *et al*. developed a new school-home collaborative intervention (Collaborative Life Skills Program–CLS) for youngsters with attention and/or behavioural problems. This study concluded that there was a reduction of symptomatology related to ADHD and there was an enhancement of organisational abilities [48]. Parker, Field and other colleagues published several articles identifying undergraduate students with ADHD perceptions and needs that were related to academic coaching. Participants provided positive feedback about academic coaching, as they perceived an improvement in their performance and self-organisation abilities [49,50].

In 2012, a considerable number of studies were published on the implications of time management and estimation deficits in people with ADHD. Hart, Radua *et al*. assessed time estimation abilities in people with ADHD by performing the meta-analysis of fMRI images. They found potential normalisation effects in pre-frontal region linked with the use of long-term psycho-stimulant treatments [51]. In this same year, Bioulac, Lallemand, Rizzo *et al*. evaluated the use of a 3D virtual classroom and concluded that children are vulnerable to a time-on-task effect on their performance. They stated that virtual reality is a reliable method to test ADHD children's ability to sustain performance over time [52]. Prior to the research published in 2013, Langberg, Epstein *et al*. studied the use of HOPS intervention with a 3-month follow-up parent-rated organised action, on planning and homework completion behaviour. They stated that participants made significant improvements relative to the wait-list comparison, taking into account parent-rated assessment, while participants did not make significant improvements relative to the comparison group according to teacher ratings [53].

In previous years other references can be found in the literature to the use of specific interventions for people with ADHD focused on time management skills. In 2008, Langberg, Epstein *et al*., underlined the need to perform controlled studies on the use of specific interventions aimed at fostering executive functions [17]. In 2006, Gureasko-Moore, DuPaul *et al*., engaged in the controlled monitoring of a small group of secondary students with ADHD, which showed an improvement in the use of organisational techniques in a school setting [54].

Table 2 outlines the main characteristics of each of the studies included in this review, with specific information about experiment setting, results and participants' description. Please refer to this for a further analysis. As described in this section and also stated in Table 2, previous studies show the efficacy of performing specific interventions with children and teenagers with ADHD. Nevertheless, the authors have found that there is currently a lack of support tools to help these users in their day-to-day lives, beyond specific interventions taking place at a specific time.

The authors have identified a need to develop tools that are fully available online at any time of the day. These tools should be ready to help and assess users outside of the specific interventions. That is why it is interesting to develop tools intended to complement the interventions focused on time management. These developments should be aimed at answering specific inquiries or doubts that may eventually arise. The system developed fits this description, as it provides final users with a consultation tool that fosters the autonomous self-management of time.

3. Materials and Methods

This section shows the materials and methods used in the development and testing of the online application discussed in this article. The tool consists of a virtual interactive balance which, by making use of decision trees and user-entered parameters, is capable of prioritising between two proposed activities.

3.1. Participants

Preliminary evaluation of the tool was made with a group of typically developing children and adolescents aged between 12 and 19 years old, with an average age of 16.23 years old.

Seventeen randomly selected participants (seven women and 10 men), selected from a group of volunteers, took part in these trials. These users are resident in the Basque Country, Spain, have not been diagnosed with ADHD and have Spanish as their mother tongue. For individuals under 18 years old the approval of parents or guardians was requested prior to conducting the surveys.

Users responded to questionnaires and tests independently, and used their own online devices (mobile, tablet and/or smartphone) to do so. Participants were allowed to choose the time of day when they wanted to perform the evaluation of the tool. This evaluation took place during the months of July and August 2013.

3.2. Methods

This section details the technical features of the system and describes the questionnaires delivered to the users before and after using the application. The online tool was produced in Django [58], the high-level Web framework for Python [59]. The results obtained and the necessary parameters were stored in a MySQL database [60]. User interface and user interaction were developed using JQuery [61], Javascript and CSS, with Touch Punch being used to adapt the tool to touch screens [62]. The priority activity decision algorithm was calculated by using a decision tree implemented in Python. The technologies and techniques employed here were selected in order to boost user-involvement and the adaptation of the system to users' needs.

For the evaluation of users' time management skills prior to the test, the *Time Management Behavior Questionnaire* (hereinafter TMBQ) by Macan *et al.* [63–65] was used, adapted to the Spanish language. This scale was chosen for its prestige and validity in the measurement of time management skills, especially for the age range selected in the test. The Spanish version was validated and accepted [66,67]. The aim of this questionnaire was to learn about and analyse the management skills of the participating users. It is composed of 34 items divided into four sections, as shown in Table 3. Items were evaluated using a Likert scale ranging from 1 ("never") to 5 ("always"). This questionnaire was available online, by means of the Google Forms tools [68] during the months of July and August 2013.

Table 3. TMBQ—Sections [66].

Section	Interpretation
F1	*This section describes students' willingness to prioritise and select tasks, in order to obtain specific goals. High scores in this section show an effective task prioritisation.*
F2	*Area 2 contains the items related to the use of specific techniques associated with effective time management. High scores on this section indicate effective time management skills using specific techniques.*
F3	*This section evaluates the way subjects organise their time, and how their study environment is structured. High scores in this section indicate a preference for a disorganised environment.*
F4	*This area shows the degree of control users perceive to have over their own time. High scores in this area indicate a high control over their time.*

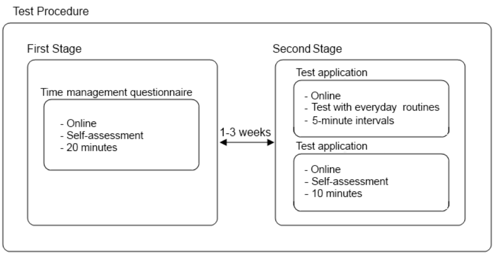
The discussed tool was evaluated by a user satisfaction test based on the *System Usability Scale* [69] (henceforth *SUS*). This questionnaire consists on 10 items. These items were evaluated by using a Likert scale ranging from 1 ("strongly agree") to 5 ("strongly disagree"). The completion of the questionnaire aims to continue to adapt the system to the users' final needs. This questionnaire was available online, by means of the Google Forms tools during the months of July and August 2013.

Scoring for the SUS was as follows: each item is ranged from 0 to 4. For items 1, 3, 5, 7 and 9, the score contribution was scale position minus 1. For items 2, 4, 6, 8 and 10, the score contribution was 5 minus the scale position. The sum scores were multiplied by 2.5 to obtain the overall value for the SUS [69].

3.3. Experiment Description

The tool's complete testing procedure was carried out online, outside of the laboratory environment, and in the course of users' everyday life. These conditions were considered appropriate due to the nature of the system. Each user answered the questions and tested the tool on their own device, in order to verify the accessibility of the system and whether it fulfilled the criterion of being multi-platform. Figure 1 shows the procedure followed during tests, with the time interval left between the two testing stages.

Figure 1. Test procedure.



As shown in Figure 1, the system's testing procedure was carried out in two mutually independent stages, between one to three weeks apart, on the basis of the results obtained in the first stage and adapting it to users' availability.

The first stage involved conducting a questionnaire about time management habits. This questionnaire was prepared online, making use of Google Forms. It consisted of 34 items divided into four different areas, each one focused on a specific skill (effective time management, preference for disorganisation, establishing concrete objectives and perception of time). Items were evaluated by using a Likert scale ranging from 1 ("never") to 5 ("always"). The purpose of this analysis was to establish the global situation of the user in terms of time management skills. This test took around 15 to 20 minutes to complete.

The second stage was comprised of two sections: firstly, the use of the application, and secondly, the performance of a systematic evaluation of the questionnaire. The application test was conducted in the course of a full day, integrating it in a natural way into the users' routine. This testing required various easy questions in the application, on the basis of the activities carried out. It took between 15 to 30 min, depending on the number of attempts users made.

Once this battery test was finished, the users took the online usability questionnaire *SUS* [69], translated into Spanish. This usability questionnaire was composed of 10 items evaluated by a Likert scale ranging from 1 ("totally disagree") to 5 ("totally agree") and aimed at measuring system usability. This survey was conducted in order to prove if the design and purpose of the system was easy to learn and user-friendly. This test took around 5 to 10 min to complete.

All the questionnaires and tests of the application were carried out in an autonomous way by users, outside of the laboratory environment. The two testing stages were carried out on their own devices, adapting the tests to the aim of the application; they took between 35 min to 1 h to be completed successfully. The results obtained in these two stages were uploaded and analysed in the *IBM—SPSS Statistics* [70] predictive analysis software tool.

4. Results

This section shows the technical results of the system, and the ones obtained after the user- experience evaluation process.

4.1. Technological Results: Tele-Therapy System

In this section the application design and development is explained. The application is an online platform developed in Django, which aids in the improvement of time management skills, and more particularly, in task prioritising.

A tool was designed which consists of three parts: training, consultation and monitoring. All of them are subject to the same format, but their purposes differ; each one of them will now be explained in detail.

a. Tool for queries

The consultation tool is a virtual balance in which users drag two activities that cause confusion by means of drag & drop techniques. Figure 2 shows the flow diagram of this tool for queries.

In addition to dragging the activities, as it is shown in Figure 2, they have to enter relevant information about the activities, such as the deadline for performing them, and, in the case of specific activities, they have to report whether someone (father, mother or someone else) asked them to do the activities.

As it is shown in the decision tree in Figure 3, a decision can be made on the basis of whether an activity needs to be a priority or not. Each branch of the tree is associated with a specific and unique weight. If two activities are labelled as a priority, the system decides, according to which one of them has a greater weight, which is the high-preference activity.

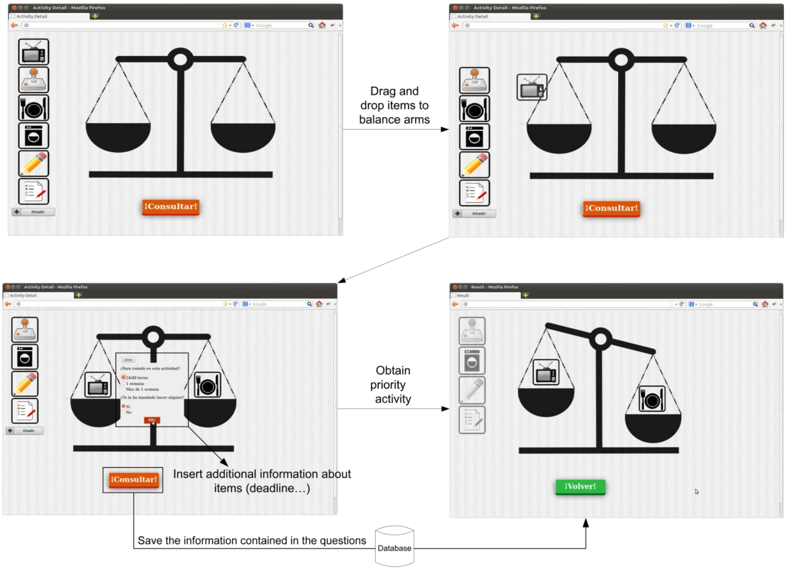
The decision tree was made following the ID3 algorithm [71], which determines how to divide the information and when to stop dividing it [72]. In order to decide which feature is the first one to divide the tree, Shannon entropy [73] was used, defined by Equation (1):

$$\mathrm{H} = -\sum_{i=1}^{n} p(x_i)\, log_2 p(x_i) \tag{1}$$

The tool is fully configurable, and allows the user to add new activities to the system by entering a series of parameters. This way it is possible to scale the size of the system and cover new application activities and areas. The final purpose is to serve as a consultation system when working on task prioritising skills.

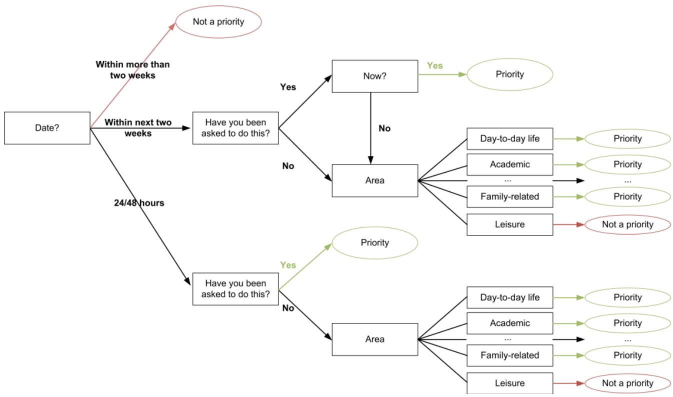
Int. J. Environ. Res. Public Health **2014**, *11* **762**

Figure 2. Tool for queries flow.



On the basis of these parameters and of the area in which the tasks are classified, the system is responsible for evaluating them, and determining which one needs to be a priority by making use of the decision tree shown in Figure 3.

Figure 3. Implemented Decision Tree.



b. Tool for training

In the training tool the user is presented with two activities already placed in the balance, with a series of features associated with each one of them (execution date, whether or not it has been requested by someone, among others). The user must discern, using the data provided, which activity needs to be a priority.

This game mode was developed in the way of a "quiz", where users continue working on the task prioritising on the basis of the area (academic, personal, family, leisure, among others) and the given parameters. Figure 4 outlines the flow diagram of the tool.

Figure 4. Tool for training diagram.

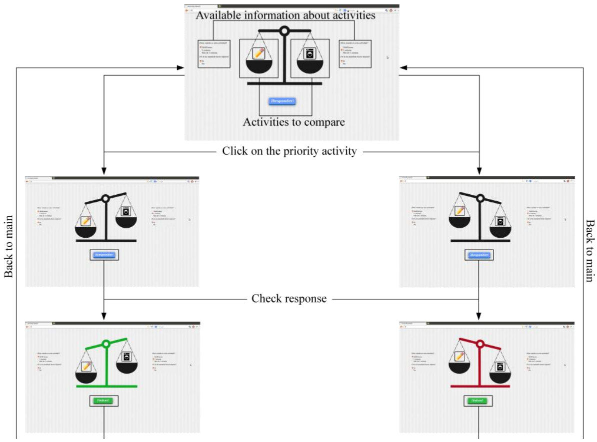


Figure 4 shows the flow diagram of the "Quiz" section. The activities are customised by means of the queries the user may make, adapting automatically to those areas that cause the most difficulty.

c. Tool for supervisors

The supervisor tool keeps a history of all the queries made and the results obtained by each user from the training tool, showing the questions with their answers, together with the date of completion, in order to determine which areas are most difficult for each user. Furthermore, this tool allows users' exercises to be customised, enabling questions to be added to the tool concerning the specific working skills of each user. This way the system is not only adapted to the exercises contained in the tool, but it is also fully customisable from the supervision area.

4.2. User Experience Evaluation

This section describes the preliminary results of the tool shown in this article. These results were taken during the months of July and August 2013. The tests consisted of three parts: completing a questionnaire in Spanish about time management skills, performing some exercises by using the online tool, and finally, answering an online questionnaire about the usability of this tool.

Participants were 17 users randomly selected from a group of volunteers. These users performed the referred tests on their online devices in the summer of 2013. Seven women and 10 men participated in these tests, as shown in Figure 5a.

Figure 5. (**a**) Gender segmentation; (**b**) Device segmentation.

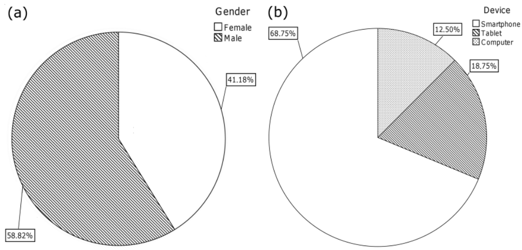


Figure 5b shows the segmentation of devices among users during the tests. Participants had the flexibility to choose the device and platform on which to perform the tests, and the only requirement was an Internet connection. The majority of the group chose to use their smartphones and only five participants decided to use other devices, such as a computer (two users) or a tablet (three users).

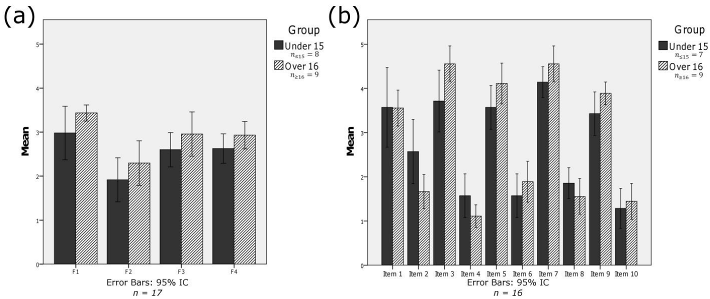
In order to thoroughly analyse the results, the sample was divided into two age-groups. Due to the number of users that took part in this study, a Mann-Whitney non-parametric test was applied. This test was significant ($p < 0.05$) for the 12–15 year-old and the 16–19 year-old groups. The following sections describe each of the tests, along with the results obtained.

a. Test 1: TMBQ

The evaluation of participants' time management skills was done by using a Spanish validated version of the *TMB Questionnaire* [66]. This questionnaire consists of 34 items divided into four different areas; please refer to Table 3 for details. Items were evaluated by using a Likert scale ranging from 1 ("never") to 5 ("always"). Average results obtained in each of the areas described are those shown in Figure 6a. These mean values were obtained by calculating the average of each of the items within each area and the responses of the 17 participants divided into two groups.

As shown in Figure 6a, the average rating per block varies between 2.98 and 1.92 points, out of a total of 5, for the group under 15 years old and between 3.44 and 2.30 for the group over 16 years old. The areas that make these extremes are those corresponding to the willingness of students to set specific goals (area 1; corresponding to F1) and the use of specific effective techniques in time management (area 2; corresponding to F2).

Figure 6. (**a**) TMBQ—responses; (**b**) SUS questionnaire—responses.



b. Online tool evaluation

The section of the tool used for evaluating user experience is available online [74]. This is a multi-platform tool which allows participants to select when and how they want to perform the tests.

Figure 7. Online tool using a tablet device.

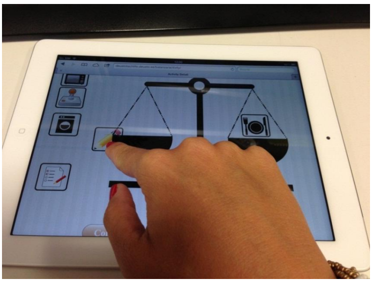


Figure 7 shows one of the seventeen participants performing the usability evaluation of the described tool from a tablet device.

c. Test 2: SUS Questionnaire

In order to evaluate system usability, the SUS questionnaire [69] was used, which consists of 10 items. The link for the completion of this questionnaire was provided to users alongside with the link to the online tool. Items were evaluated by using a Likert scale ranging from 1 ("strongly disagree") to 5 ("strongly agree").

As far as the answers are concerned, 16 participants, seven women and 9 men, out of the 17 initial participants in the trial, answered the system usability questionnaire. Due to schedule issues, one of the participants left the testing after the completion of the first questionnaire. This user belonged to the group under 15-years old. The inclusion of this user in the first test was not considered decisive in order to obtain the results. Please refer to Figure 6b for average answers per item obtained for each age-group. The SUS mean overall score was 78.75 out of 100, with a standard deviation of 6.39.

Users under 15-years-old scored an average result of 73.93 out of 100, with a standard deviation of 4.97 while participants over 16-years-old rated the system with an average result of 82.5, with a standard deviation of 4.68.

5. Discussion and Conclusions

This article addresses the issue of time management and the prioritisation of tasks in teenagers with and without ADHD diagnosis. In this section, authors attempted to answer the following questions, drawn from the section of this article which set out the objectives: Is it possible that the use of interactive and personalised content, such as serious games, encourage the use of tools to improve time management skills? Does the development of online tools foster their availability? Do users feel comfortable using these tools?

Mobile devices are increasingly used in daily life, relegating computers to the background when it comes to short, specific activities [75]. Results obtained in this study show that 15 out of 17 participants used mobile devices to perform the activities proposed in this trial. The rise in the use of these devices, along with access to a permanent Internet connection, increases the demand for interactive online content, enabling the emergence of tools like the one presented in this article.

According to the literature mentioned in Sections 1 and 2, serious games within these interactive tools can also help to improve specific skills through the use of gaming techniques and appealing and accessible content.

As explained in Section 1, time management skills and task prioritisation capabilities need to be developed during childhood and adolescence, in order to prepare users for adult life and the management of their own time. Moreover, they could also be key skills in obtaining good academic results.

Based on the review conducted by the authors in Section 2.2 about the strengthening of time management and organisational skills within the ADHD collective, it is clear than specific approaches and measurements should be taken. After analysing the literature, it can be determined that there is an agreement about the efficiency of targeted face-to-face therapies focused on the enhancement of time management skills [46–54].

Bioulac, Lallemand *et al.* proposed a virtual development that implied the first steps in the evolution of face-to-face therapies to more innovative interventions. These new therapies included the use of new technologies such as virtual environments and computer-based therapies, thus providing users with a greater autonomy [52].

The described system presented in this paper goes one step beyond, by including the use of online therapies which promote availability and convenience while using the tool. These therapies serve as a supporting tool for children and adolescents with time management and organizational problems. The development of new online solutions could help to complement traditional therapies fostering their effects and helping to create efficient management and organisational skills.

Results obtained from the TMBQ show scores below 3.5 points out of 5 in every area. These results are especially low when it comes to the use of specific techniques for efficient time management skills.

These results suggest the need for the creation of new, efficient, and attractive content that provides guidance and effective techniques to work on time management skills. It is therefore necessary to create an accessible tool for time management habits and the prioritising of tasks for adolescents in general, and not only for those diagnosed with ADHD.

The described tool is focused on working on these activities, particularly on prioritising tasks. When analysing its usability, results were in 78.75 out of 100. Although this result means a good acceptance of the system, the authors have analysed its weaknesses in order to implement possible improvements. Segmenting participants by age shows that there are differences related to system usability and users' comfort and confidence. Even though results were acceptable in both groups (please refer to Section 4 for details), the groups of 12–15 year-old users showed a greater divergence of views when they were asked to evaluate whether they felt comfortable with the system or not. These results suggest that some users may not feel comfortable with the use of tools that can be used to control their daily routines, leaving a record of their activities.

As a conclusion to this study, it was confirmed that there is a need for new interactive content in order to work on time management skills in teenagers with and without ADHD. Key skills to be worked on lie not only in the field of the prioritisation of tasks, but also in the effective utilisation of specific techniques to that effect. Nevertheless, authors consider that this kind of adaptive tele-therapies should be adopted as a support tool for traditional therapies, not as a substitute for conventional interventions.

Future directions for this study in the field of the design and development of the system are expanding the tool to include new skills, such as effective time organisation, that continues along the lines of the current tool. Moreover, the system should be adapted to be suitable for all age-ranges, trying to minimise divergences between them. Additionally, providing a greater degree of customisation to the tool may help to improve the acceptance by both age groups.

Finally, the main direction for the future is to replicate this study:

- With a higher number of users.
- With users with and without an ADHD diagnosis.
- With the segmentation of participants into two age groups.
- Developing a bilingual or trilingual tool that allows the study to be replicated in other national and international territories in where reported diagnosis of ADHD is significantly different from the Basque Country, Spain.
- Performing a longitudinal study on the impact of online game-based tools on time management skills in teenagers with and without ADHD.

The creation of a completely new tool capable of working on different skills, in conjunction with the replication of the current study, may help to understand these users' needs and boost time management skills among teenagers with and without ADHD.

Acknowledgments

This work was partially supported by the Basque Country Department of Education, Universities and Research. The authors would like to acknowledge the DEIKER Agency for Research Management and Promotion at the University of Deusto, and also, Kattalin Camara, Celia Alvarez, Xabier Larrakoetxea, Jessica Frutos and Jose Maeso for their support and collaboration.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Polanczyk, G.; de Lima, M.S.; Horta, B.L.; Biederman, J.; Rohde, L.A. The worldwide prevalence of ADHD: A systematic review and metaregression analysis. *Amer. J. Psychiat.* **2007**, *164*, 942–948.
2. *Guía de Práctica Clínica sobre el Trastorno por Déficit de Atención con Hiperactividad (TDAH) en Niños y Adolescentes*; Ministry of Health, Social Policy and Equality: Madrid, Spain, 2010.
3. Pardey, M.C.; Kumar, N.N.; Goodchild, A.K.; Clemens, K.J.; Homewood, J.; Cornish, J.L. Long-term effects of chronic oral Ritalin administration on cognitive and neural development in adolescent Wistar Kyoto rats. *Brain Sci.* **2012**, *2*, 375–404.
4. *Diagnostic and Statistical Manual of Mental Disorders*; American Psychiatric Association: Arlington, VA, USA, 2000; Volume 4, p. 943.
5. Houghton, S.; Durkin, K.; Ang, R.P.; Taylor, M.F.; Brandtman, M. Measuring temporal self-regulation in children with and without attention deficit hyperactivity disorder: Sense of time in everyday contexts. *Eur. J. Psychol. Assess.* **2011**, *27*, 88–94.
6. Barkley, R.A. Behavioral inhibition, sustained attention, and executive functions: Constructing a unifying theory of ADHD. *Psychol. Bull.* **1997**, *121*, 65–94.
7. Barkley, R.A.; Koplowitz, S.; Anderson, T.; McMurray, M.B. Sense of time in children with ADHD: Effects of duration, distraction, and stimulant medication. *J. Int. Neuropsychol. Soc.* **1997**, *3*, 359–369.
8. Barkley, R.A.; Murphy, K.R.; Bush, T. Time perception and reproduction in young adults with attention deficit hyperactivity disorder. *Neuropsychology* **2001**, *15*, 351–360.
9. Toplak, M.E.; Rucklidge, J.J.; Hetherington, R.; John, S.C.F.; Tannock, R. Time perception deficits in attention-deficit/hyperactivity disorder and comorbid reading difficulties in child and adolescent samples. *J. Child Psychol. Psychiat.* **2003**, *44*, 888–903.
10. Carelli, M.G.; Forman, H.; Mäntylä, T. Sense of time and executive functioning in children and adults. *Child Neuropsychol.* **2007**, *14*, 1–15.
11. Toplak, M.E.; Dockstader, C.; Tannock, R. Temporal information processing in ADHD: Findings to date and new methods. *J. Neurosci. Meth.* **2006**, *151*, 15–29.
12. González-Garrido, A.A.; Gómez-Velázquez, F.R.; Zarabozo, D.; López-Elizalde, R.; Ontiveros, A.; Madera-Carrillo, H.; Vega, O.L.; de Alba, J.L.; Tuya, J.M. Time reproduction disturbances in ADHD children: An ERP study. *Int. J. Neurosci.* **2008**, *118*, 119–135.
13. Meaux, J.B.; Chelonis, J.J. Time perception differences in children with and without ADHD. *J. Pediatr. Health Care* **2003**, *17*, 64–71.
14. Meaux, J.B.; Chelonis, J.J. The relationship between behavioral inhibition and time perception in children. *J. Child Adolesc. Psychiatr. Nurs.* **2005**, *18*, 148–160.
15. Smith, A.; Taylor, E.; Rogers, J.W.; Newman, S.; Rubia, K. Evidence for a pure time perception deficit in children with ADHD. *J. Child Psychol. Psychiat.* **2002**, *43*, 529–542.

16. Radonovich, K.J.; Mostofsky, S.H. Duration judgments in children with ADHD suggest deficient utilization of temporal information rather than general impairment in timing. *Child Neuropsychol.* **2004**, *10*, 162–172.
17. Langberg, J.M.; Epstein, J.N.; Graham, A.J. Organizational-skills interventions in the treatment of ADHD. *Expert Rev. Neurother.* **2008**, *8,* 1549–1561.
18. Turgay, A.; Goodman, D.W.; Asherson, P.; Lasser, R.A.; Babcock, T.F.; Pucci, M.L.; Barkley, R. Lifespan persistence of ADHD: The life transition model and its application. *J. Clin. Psychiat.* **2012**, *73*, 192–201.
19. Dovis, S.; van der Oord, S.; Wiers, R.W.; Prins, P.J. Can motivation normalize working memory and task persistence in children with attention-deficit/hyperactivity disorder? The effects of money and computer-gaming. *J. Abnorm. Child Psychol.* **2011**, *40*, 669–681.
20. Prins, P.J.; Dovis, S.; Ponsioen, A.; ten Brink, E.; van der Oord, S. Does computerized working memory training with game elements enhance motivation and training efficacy in children with ADHD? *Cyberpsychol. Behav. Soc. Netw.* **2011**, *14*, 115–122.
21. Nelson, E.L.; Duncan, A.B.; Peacock, G.; Bui, T. Telemedicine and adherence to national guidelines for ADHD evaluation: A case study. *Psychol. Serv.* **2012**, *9*, 293–297.
22. Palmer, N.B.; Myers, K.M.; van der Stoep, A.; McCarty, C.A.; Geyer, J.R.; Desalvo, A. Attention-deficit/hyperactivity disorder and telemental health. *Curr. Psychiatry Rep.* **2010**, *12*, 409–417.
23. Wilkinson, N.; Ang, R.P.; Goh, D.H. Online video game therapy for mental health concerns: A review. *Int. J. Soc. Psychiatr*. **2008**, *54*, 370–382.
24. Geurts, H.M.; Luman, M.; van Meel, C.S. What's in a game: The effect of social motivation on interference control in boys with ADHD and autism spectrum disorders. *J. Child Psychol. Psychiat.* **2008**, *49*, 848–857.
25. Bioulac, S.; Arfi, L.; Bouvard, M.P. Attention deficit/hyperactivity disorder and video games: A comparative study of hyperactive and control children. *Eur. Psychiat.* **2008**, *23*, 134–141.
26. Frölich, J.; Lehmkuhl, G.; Döpfner, M. Computer games in childhood and adolescence: Relations to addictive behavior, ADHD, and aggression. *Z. Kinder Jugendpsychiatr. Psychother.* **2009**, *37*, doi:10.1024/1422-4917.37.5.393.
27. Ferguson, C.J. The influence of television and video game use on attention and school problems: A multivariate analysis with other risk factors controlled. *J. Psychiatr. Res.* **2011**, *45*, 808–813.
28. Santos, F.E.G.; Bastos, A.P.Z.; Andrade, L.C.V.; Revoredo, K.; Mattos, P. *Assessment of ADHD through a Computer Game: An Experiment with a Sample of Students*. In Proceedings of Third International Conference on Games and Virtual Worlds for Serious Applications (IEEE), Athens, Greece, 4–6 May 2011; pp. 104–111.
29. De Andrade, L.C.V.; Carvalho, L.A.V.; Lima, C.; Cruz, A.; Mattos, P.; Franco, C.; Grieco, B. *Supermarket Game: An Adaptive Intelligent Computer Game for Attention Deficit/Hyperactivity Disorder Diagnosis*. In Proceedings of Fifth Mexican International Conference on Artificial Intelligence (IEEE), Apizaco, Mexico, November 2006; pp. 359–365.
30. Rizzo, A.A.; Bowerly, T.; Shahabi, C.; Buckwalter, J.G.; Klimchuk, D.; Mitura, R. Diagnosing attention disorders in a virtual classroom. *Computer* **2004**, *6*, 87–89.

31. Díaz-Orueta, U.; Garcia-López, C.; Crespo-Eguílaz, N.; Sánchez-Carpintero, R.; Climent, G.; Narbona, J. AULA virtual reality test as an attention measure: Convergent validity with Conners' Continuous Performance Test. *Child Neuropsychoogy* **2013**, 1–15, doi:10.1080/09297049.2013. 792332.
32. Kerns, K.A. The CyberCruiser: An investigation of development of prospective memory in children. *J. Int. Neuropsychol. Soc.* **2000**, *6*, 62–70.
33. Kerns, K.A.; Price, K.J. An investigation of prospective memory in children with ADHD. *Child Neuropsychol.* **2001**, *7*, 162–171.
34. Bartle, D.W. Development of cognitive video games for children with attention and memory impairment, PhD Thesis, University of Victoria, Victoria, BC, Canada, 2010.
35. Pei, J.; Kerns, K. Using games to improve functioning in children with fetal alcohol spectrum disorders. *Game. Health. J.* **2012**, *1*, 308–311.
36. Pope, A.T.; Bogart, E.H. Method of Encouraging Attention by Correlating Video Game Difficulty with Attention Level. U.S. Patent 5377100; 27 December 1994.
37. Palsson, O.S.; Harris, R.L., Sr.; Pope, A.T. Method and Apparatus for Encouraging Physiological Self-Regulation through Modulation of an Operator's Control Input to a Video Game or Training Simulator. U.S. Patent 6450820, 17 September 2002.
38. Bell, C. *The Journey to Wild Divine*; The Wild Divine Project: Las Vegas, NV, USA, 2003.
39. Statements, F.; Lives, T.; Bear, P.; McCall, D.; Bruce, F.; Harper, D.; McCabe, K.; Murnaghan, D.; Hadley, T.; Fry, M.; *et al. Helping children affected by disability and infections Action Medical Research for children*; Action Medical Research: Horsham, UK, 2012
40. Van Dijk, D.; Hunneman, R.; Wildlevuur, S. *Self City*: *Training Social Skills in a Game*. In Proceedings of Second European Conferences on Game-based Learning, Barcelona, Spain, October 2008; pp. 481–488.
41. Conconi, A.; Ganchev, T.; Kocsis, O.; Papadopoulos, G.; Fernández-Aranda, F.; Jiménez-Murcia, S. PlayMancer: A Serious Gaming 3D Environment. In Proceedings of the International Conference on Automated Solutions for Cross Media Content and Multi-Channel Distribution, Florence, Italy, 17–19 November 2008; pp. 111–117.
42. Jiménez-Murcia, S.; Fernández-Aranda, F.; Kalapanidas, E.; Konstantas, D.; Ganchev, T.; Kocsis, O.; Lam, T.; Santamaría, J.J.; Raguin, T.; Breiteneder, C.; *et al*. PlayMancer project: A serious videogame as an additional therapy tool for eating and impulse control disorders. *Stud. Health Technol. Inform.* **2009**, *144*, 163–166.
43. Coyle, D.; Matthews, M.; Sharry, J.; Nisbet, A.; Doherty, G. Personal investigator: A therapeutic 3D game for adolecscent psychotherapy. *Interact. Technol. Smart Educ*. **2007**, *2*, 73–88.
44. Klingberg, T.; Fernell, E.; Olesen, P.J.; Johnson, M.; Gustafsson, P.; Dahlström, K.; Gillberg, C.G.; Forssberg, H.; Westerberg, H. Computerized training of working memory in children with ADHD—A randomized, controlled trial. *J. Am. Acad. Child Adolesc. Psychiatry* **2005**, *44*, 177–186.
45. Siglin, J.A.; Logic, U. Play Attention-focusing on success. *Intervention School Clinic* **2000**, *36*, 122–124.

46. Abikoff, H.; Gallagher, R.; Wells, K.C.; Murray, D.W.; Huang, L.; Lu, F.; Petkova, E. Remediating organizational functioning in children with ADHD: Immediate and long-term effects from a randomized controlled trial. *J. Consult. Clin. Psychol*. **2013**, *81*, 113–128.
47. Langberg, J.M.; Becker, S.P.; Epstein, J.N.; Vaughn, A.J.; Girio-Herrera, E. Predictors of response and mechanisms of change in an organizational skills intervention for students with ADHD. *J. Child Fam. Stud.* **2013**, *22*, doi:10.1007/s10826-012-9662-5.
48. Pfiffner, L.J.; Villodas, M.; Kaiser, N.; Rooney, M.; McBurnett, K. Educational outcomes of a collaborative school-home behavioral intervention for ADHD. *Sch. Psychol. Q.* **2013**, *28*, 25–36.
49. Rolands, L.; Sawilowsky, S.; Hoffman, S.F.; Parker, D.R. Self-control in postsecondary settings: Students' perceptions of ADHD college coaching. *J. Atten. Disord.* **2011**, *17*, 215–232.
50. Field, S.; Parker, D.R.; Sawilowsky, S.; Rolands, L. Assessing the impact of ADHD coaching services on university students learning skills, self-regulation, and well-being. *J. Postsecond. Educ. Disabil.* **2013**, *26*, 67–81.
51. Hart, H.; Radua, J.; Mataix-Cols, D.; Rubia, K. Meta-analysis of fMRI studies of timing in attention-deficit hyperactivity disorder (ADHD). *Neurosci. Biobehav. Rev.* **2012**, *36*, 2248–2256.
52. Bioulac, S.; Philip, P.; Lallemand, S.; Fabrigoule, C.; Rizzo, A.; Bouvard, M.P. Impact of time on task on ADHD patient's performances in a virtual classroom. *Eur. J. Paediatr. Neurol.* **2012**, *16*, 514–521.
53. Langberg, J.M.; Epstein, J.N.; Becker, S.P.; Girio-Herrera, E.; Vaughn, A.J. Evaluation of the homework, organization, and planning skills (HOPS) intervention for middle school students with attention deficit hyperactivity disorder as implemented by school mental health providers. *Sch. Psych. Rev.* **2012**, *41*, 342–364.
54. Gureasko-Moore, S.; Dupaul, G.J.; White, G.P. The effects of self-management in general education classrooms on the organizational skills of adolescents with ADHD. *Behav. Modif.* **2006**, *30*, 159–183.
55. Prior, H.; Walg, M.; Oepen, J. Adjustment of time perception in the range of seconds and milliseconds: The nature of time-processing alterations in children with ADHD. *J. Atten. Disord*. **2012**, doi:10.1177/1087054712454570.
56. Carelli, M.G.; Wiberg, B. Time out of mind: Temporal perspective in adults with ADHD. *J. Atten. Disord.* **2012**, *16*, 460–466.
57. Abikoff, H.; Gallagher, R.; Ma, J. Measuring and treating organizational, time management, and planning deficits in children with ADHD. *J. Child Adolesc. Psychopharmacol.* **2003**, *13*, 422–423.
58. Django. Available online: https://www.djangoproject.com/ (accessed on 19 July 2013).
59. Python. Available online: http://www.python.org/ (accessed on 19 July 2013).
60. MySQL. Available online: http://www.mysql.com/ (accessed on 19 July 2013).
61. JQuery. Available online: http://jquery.com/ (accessed on 27 August 2013).
62. Touch Punch. Available online: http://touchpunch.furf.com/ (accessed on 27 August 2013).
63. Macan, T.H.; Shahani, C.; Dipboye, R.L.; Phillips, A.P. College students' time management: Correlations with academic performance and stress. *J. Educ. Psychol.* **1990**, *82*,760–768.
64. Macan, T.H. Time management: Test of a process model. *J. Appl. Psychol.* **1994**, *79*, 381–391.

65. Macan, T.H. Time-management training: Effects on time behaviors, attitudes, and job performance. *J. Psychol.* **1996**, *130*, 229–236.
66. García-Ros, R.; Pérez-González, F. Spanish version of the time management behavior questionnaire for university students. *Span. J. Psychol.* **2012**, *15*, 1485–1494.
67. Aponte, E.D.; Pujol, L. Manejo del tiempo académico en jóvenes que inician estudios en la Universidad Simón Bolívar. *Rev. Latinoamer. De Cienc. Sociales Niñez y Juventud.* **2013**, *11*, 93–108.
68. Google Forms. Available online: http://www.google.com/drive (accessed on 19 July 2013).
69. Brooke, J. SUS—A quick and dirty usability scale. *Usability Eval. Ind.* **1996**, *189*, 189–194.
70. IBM SPSS Software. Available online: http://www-01.ibm.com/software/analytics/spss/ (accessed on 19 July 2013).
71. Quinlan, J.R. Induction of decision trees. *Mach. Learn.* 1986, *1*, 81–106.
72. Harrington, P. *Machine Learning in Action*; Manning Publications Co.: Shelter Island, NY, USA, 2012.
73. Shannon, C.E. The mathematical theory of communication. 1963. *M. D. Comput.* **1948**, *14*, 306–317.
74. Balance Online Tool. Available online: http://deustotechlife.deusto.es/balanza/activity/ (accessed on 28 August 2013).
75. Wu, W.-H.; Wu, Y.-C.J.; Chen, C.-Y.; Kao, H.-Y.; Lin, C.-H.; Huang, S.-H. Review of trends from mobile learning studies: A meta-analysis. *Comput. Educ.* **2012**, *59*, 817–827.

5.2 ARTICLE III: Assessing Visual Attention Using Eye Tracking Sensors in Intelligent Cognitive Therapies Based on Serious Games

This second pilot study is the first step taken in this dissertation into the analysis of gaze pattern behaviours in children while using intervention exercises based on serious games.

This article records and analyses visual interaction patterns through the use of eye tracking sensors as a means to identify children's behaviour in attention-enhancement therapies.

In this study, 32 children aged between 8 and 12 years with different attention skills were asked to solve a set of puzzles while their gaze patterns and interaction were recorded using an eye-tracking sensor. The recorded eye information includes the location of gaze fixation on the computer screen, the duration of fixations and saccades (the path of the eye movements), along with interaction information regarding performance during the exercise.

We hypothesized that participants with better performance in the proposed exercises would demonstrate patterns of eye-movements that are quantifiably different from individuals with a weaker performance.

Identification of these differences would be especially advantageous for teachers and psychologists, as this study may provide new insight into the strategies for the improvement of attention skills. Moreover, a study of the relation between gaze patterns and the degree of expertise was also analysed to determine if there is any difference, in terms of gaze behaviours, between the first approach to an exercise and the subsequent ones.

The use of gaze data constitutes a new information source in intelligent therapies that may help to build new approaches that are fully-customized to final users' needs, which is the main objective of this dissertation.

This study was conceived as the first step for assessing the viability of visual attention as a mean for determining, in junction with interaction patterns, user gaming profiles in serious games for attention enhancement. This article was published on the *Sensors International Journal (IF: 2.245 [Q1])* [Frutos-Pascual 15a], Volume 15, Issue 5 in May 2015. This article has been included on its original full version that is available through the journal web page.

Sensors **2015**, *15*, 11092-11117; doi:10.3390/s150511092

OPEN ACCESS
sensors
ISSN 1424-8220
www.mdpi.com/journal/sensors

Article

Assessing Visual Attention Using Eye Tracking Sensors in Intelligent Cognitive Therapies Based on Serious Games

Maite Frutos-Pascual * and Begonya Garcia-Zapirain

DeustoTech Life [eVIDA] Faculty of Engineering University of Deusto, Avda de las Universidades 24, Bilbao 48015, Spain; E-Mail: mbgarciazapi@deusto.es

* Author to whom correspondence should be addressed; E-Mail: maitefrutos@deusto.es; Tel.: +34-94-4139000 (ext. 2051).

Academic Editor: Gianluca Paravati

Received: 24 February 2015 / Accepted: 27 April 2015 / Published: 12 May 2015

Abstract: This study examines the use of eye tracking sensors as a means to identify children's behavior in attention-enhancement therapies. For this purpose, a set of data collected from 32 children with different attention skills is analyzed during their interaction with a set of puzzle games. The authors of this study hypothesize that participants with better performance may have quantifiably different eye-movement patterns from users with poorer results. The use of eye trackers outside the research community may help to extend their potential with available intelligent therapies, bringing state-of-the-art technologies to users. The use of gaze data constitutes a new information source in intelligent therapies that may help to build new approaches that are fully-customized to final users' needs. This may be achieved by implementing machine learning algorithms for classification. The initial study of the dataset has proven a 0.88 ($\pm$0.11) classification accuracy with a random forest classifier, using cross-validation and hierarchical tree-based feature selection. Further approaches need to be examined in order to establish more detailed attention behaviors and patterns among children with and without attention problems.

Keywords: eye tracker; attention; intelligent therapies; serious games; children

1. Introduction

In recent years, the usage of video game-related content in areas, such as education, therapies and training, has risen sharply. Several studies suggest that the future of pedagogy will inevitably be linked to the proposal of combined play and learning in order to promote creativity in future generations [1]. The boom in serious games brings together the potential available invideo games, devoting it fully to the enhancement of specific abilities, skills and aptitudes in children and adults.

Moreover, the design and development of new adaptive serious games whose content changes based on user interaction make therapies, training and education more customized. These techniques provide systems with an efficient way of learning based on the users themselves, providing them with customized and personal experiences, which may increase their potential effects [2].

One of the most widely-used forms of adaptive intervention consists of helping students to complete some educational activities when they have specific difficulties proceeding on their own [3].

The purpose of this study is to explore the use of eye tracking sensors to evaluate the behavior of children in attention-related cognitive therapies based on serious games to determine the utility of eye-related data as an input biofeedback signal for attention improvement therapies.

Eye movements are a natural information source for proactive systems that analyze user behavior, where the goal is to infer implicit relevance feedback from gaze [4]. Moreover, following the eye-mind hypothesis put forth by Carpenter in 1980, there is a close link between the direction of the human gaze and the focus of attention [5], provided that the visual environment in front of the eyes is pertinent to the task that we want to study [6]. Eye tracking sensors collect information about the location and duration of an eye fixation within a specific area on a computer monitor.

In this study, normal developing children aged between eight and 12 years and with different attention skills are asked to solve a set of puzzles while their gaze patterns and interaction are recorded using an eye-tracking sensor. The recorded eye information includes the location of gaze fixation on the computer screen, the duration of fixations and saccades (the path of the eye movements), along with interaction information regarding performance during the exercise. We hypothesize that participants with better performance in the proposed exercises would demonstrate patterns of eye-movements that are quantifiably different from individuals with a weaker performance. Identification of these differences would be especially advantageous for teachers and psychologists, as this study may provide new insight into the strategies for the improvement of attention skills. Moreover, the authors would like to study the relation between gaze patterns and the degree of expertise. This will be done by determining if there are any differences between the first approach to an exercise and the subsequent ones.

This article is outlined as follows: First, the use of eye tracking sensors in the field of serious games will be studied and placed in context. Subsequently, the Materials and Methods Section will be introduced, in which the authors discuss the form and function of the data collected from the eye tracking sensor. Next, a discussion of the collected data and the approaches to data analysis are examined. Finally, the manuscript concludes with a discussion of possibilities for further research into the uses of eye tracking sensor and data as a biofeedback input to intelligent therapies.

2. Literature Review

The observation of eye-movements is not a new area of research within psychology-related fields, having been studied in depth over the last few decades [7–9].

Research using eye tracking sensors affords a unique opportunity to test aspects of theories about multimedia learning concerning processing during learning [10]. Moreover, the use of this approach may help in understanding where players focus their attention during game play [11], as well as how they confront unfamiliar games and software [12].

However, it was not until recently that researchers began to analyze and introduce eye tracking sensors and techniques in serious games and computer games [13–15]. Games that can be controlled solely through eye movement would be accessible to persons with decreased mobility or control. Moreover, the use of eye tracking data can change the interaction with games, producing new input experiences based on visual attention [15].

Eye tracking devices have been used in the design of educational games, in terms of assessing usability based on user gaze behaviors when interacting with the game [16,17]. El-Nasr and Yan used eye tracker sensors to analyze attention patterns within an interactive 3D game environment, so as to improve game level design and graphics [18].

Kickmeier-Rust *et al.* focused on assessing the effectiveness and efficiency of serious games. For this purpose, they assessed these variables with gaze data and gaze paths, in order to obtain interaction strategies in specific game situations [19]. Sennersten and Lindley also evaluated the effectiveness of virtual environments in games through the analysis of visual attention using eye tracking data [20]. Johansen *et al.* discussed the efficiency of eye tracker sensors in assessing users' behavior during game play [21].

Józsa and Hamornik used recorded eye tracking data to evaluate learning curves in university students while using a seven hidden differences puzzle game. They used this data to assess similarities and differences in information acquisition strategies considering gender- and education-dependent characteristics [22]. Dorr *et al.* conducted a similar study concluding that expert and novice players use different eye movement strategies [23]. Muir *et al.* used eye-tracking data to capture user attention patterns and to present results on how those patterns were affected by existing user knowledge, attitude towards getting help and performance while using the educational game, Prime Club [3].

Radoslaw *et al.* used eye tracker sensors for assessing render quality in games. They argued that gaze-dependent rendering was especially important when immersed in serious games, where players in virtual environments played a primary role [24]. Smith and Graham and Hillaire *et al.* concluded that use of an eye tracker increases video game immersion, altering the game play experience [25,26].

Chang *et al.* developed the game WAYLAas a means to evaluate the potential to offer new interaction experiences based on eye tracking and visual attention. These authors took advantage of the popularity and arrival of more affordable eye tracker sensors [27].

Li and Zhang used eye-movement analysis to assess patients' mental engagement in a rehabilitation game. Therapists use this feedback to adjust rehab exercises to users' needs [28]. Continuing with the health-related field, Lin *et al.* developed an eye-tracking system for eye motion disability rehabilitation as a joystick-controlled game [29]. Vickers *et al.* developed a framework that integrated automatic modification of game tasks, interaction techniques and input devices according to a user ability profile [30].

Walber *et al.* presented EyeGrab, a game for image classification controlled by the players' gaze. The main purpose of this game was to collect eye tracking data to enrich image context information [31].

Other studies, such as those conducted by Nacke *et al.*, evaluated the use of eye tracker sensors as an alternative way of controlling interaction with games, obtaining favorable outcomes where this challenge results in positive affection and feelings of flow and immersion [32]. Ekman *et al.* goes one step further, discussing the limitations of using pupil-based interaction and providing suggestions for using pupil size as an input modality [33].

Table 1 shows the experimental conditions for the most relevant articles included in this section.

Table 1. Literature review: Experimental conditions.

Authors	Citation	Year	Country	Game	Device	Display	Participants			
							Total	Female	Male	Age (SD)
Kiili *et al.*	[17]	2014	FI	Animal Class	Tobii T60	17"	8	2	6	7–13 years
Chang *et al.*	[27]	2013	PT	Wayla	Tobii REX	–	–	–	–	–
Muir *et al.*	[3]	2012	US	Prime Club	Tobii T120	17"	12	6	6	10–12 years
Walber *et al.*	[31]	2012	DE	EyeGrab	–	–	24	7	17	15–32 years
Kickmeier *et al.*	[19]	2011	AT	80 Days	Tobii 1750	–	9	4	5	13 (1.61)
Józsa and Hammornik	[22]	2011	HU	7 Hidden Differences	Tobii T120	17"	43	14	29	19–26 years
Pretorious *et al.*	[12]	2010	ZA	Timez Attack	Tobii 1750	17" (1280 × 1024)	8	4	4	9–12 years over 40
Sennersten and Lindley	[20]	2010	SE	FPS computer game	Tobii 1750	–	–	–	–	–
Nacke *et al.*	[32]	2010	CA	Half-Life 2	Tobii T120	–	30	2	28	18.67 (4.26)
Hillaire *et al.*	[26]	2008	FR	Quake III	ASL6000	Cylindrical Screen (1280 × 1025)	8	0	8	25.8 (4.3)
Dorr *et al.*	[23]	2007	DE	Breakout Game	SensoMotoric IViewX Hi-Speed	20"	9	–	–	–
El Nasr *et al.*	[18]	2006	US	Game Soul Caliber	ISCAN ETL-500 (head-mounted)	–	6	–	–	20–30 years
Smith and Graham	[25]	2006	CA	Custom build scene	RED250	22" (1680 × 1050)	21	1	20	21–24 years

3. Materials and Methods

This section presents the methodology used in this study along with participants' characteristics and the selection procedure.

3.1. Participants

The process for assessing attention was performed with a group of typically developing children. This process relies on data recorded with an eye tracking sensor. Participants were aged between 8 and 12 years, with an average age of 10.0 (SD = 1.34). Thirty-two randomly-selected participants (13 girls and 19 boys) were selected from a group of 83 volunteers by their teachers. This sample size was considered adequate for the purpose of the outlined pilot study [34].

These children live in the Basque Country, Spain, have not been diagnosed with any attention-related disorder and speak Spanish as their mother tongue. All of the participants were recruited from the Colegio Vizcaya School.

Since they were mature minors, the approval of parents or guardians was requested prior to conducting the study. This approval consisted of an informed consent following receipt of a detailed description of the study, distributed via the school's regular newsletter.

3.2. Materials

All participants in the study completed the same assessment, which consisted of a puzzle exercise with four different levels of difficulty. Users have to connect each of the four slices presented in the exercise with its corresponding part in the main image. As Figure 1 displays, all of the participants were presented with the same image for each level, and all of the elements in the user interface appeared in the same part of the screen at each level.

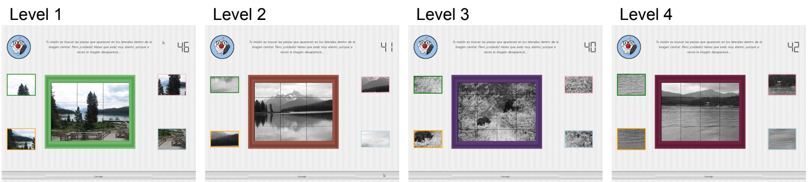


Figure 1. Different levels of the task.

The main image and the slices appeared in the middle of the screen, occupying the whole display from left to right. The question stem appeared in the upper middle part of the screen. The button to advance to the next level appeared at the lower middle part of each screen. The consistent layout of the screen was intended to minimize wide eye movements.

Different levels' settings are outlined in Table 2. All of the users had a maximum pre-set time of 50 seconds to complete each of the levels. However, if they finished the level before the time ended, they could go on to the next exercise. Depending on the level, the displayed image was labeled as easy, medium or hard. Only Level 1 is displayed in color. Hard images have very similar slices and are more complicated to complete. Table 2 displays the different levels' settings.

Table 2. Different levels' settings.

	Time (s)	Grid Size	No. of Slices	Display		Image Level		
				Color	Greyscale	Easy	Medium	Hard
Level 1	50	3×3	9	x		x		
Level 2	50	3×4	12		x	x		
Level 3	50	4×4	16		x		x	
Level 4	50	5×4	20		x			x

3.3. Devices and Technologies

All of the data for this study were collected on the same device, which was located at the children's school outside the laboratory environment. These conditions were considered appropriate due to the nature of the system.

The set of puzzles was developed in Python [35]. The results obtained and the necessary parameters were stored in a SQLitedatabase [36]. The user interface and user interaction were developed using

PyQT4 [37]. Fixation heat maps were produced based on the implementation developed by jjguy [38]. The classification process was implemented using the Scikit-learn library for machine learning in Python [39].

The puzzles were displayed on a 19-inch Lenovo monitor interface with an Acer Aspire Timeline X laptop running on Ubuntu 12.04. All of the text in the different exercises was displayed as black text against a light-grey background following normal grammatical conventions in Spanish. Images were inserted as JPEG digital pictures scaled from their original versions. Response selection and any changes were stored by monitoring the user interaction and recording eye movements with a Tobii X1 Light eye tracker sensor. Figure 2 shows the study setting while one of the participants was interacting with the system.

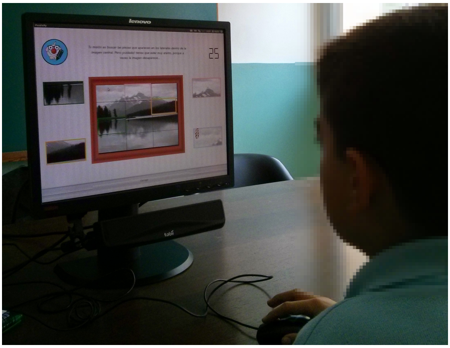


Figure 2. Participant using the system while his gaze is being recorded.

The eye tracker is a non-invasive sensor with remote function. Participants were not required to remove their glasses or contact lenses during the tests. Accuracy under ideal conditions is 0.5 deg of the visual angle, while the sampling rate in this study was typically 28–32 Hz. As Figure 2 displays, the Tobii X1 light sensor was located beneath the computer monitor with the headrest fastened to the front edge of the desk, monitoring the participant's head. The laptop was located behind the monitor, without interfering with the participants' field of vision.

A typical experimental trial including calibration lasted less than 20 min for each participant.

3.4. Experimental Procedure

Prior to this study, participants' teachers responded autonomously to the EDAH scale for the evaluation of ADHD in the questionnaire on children between 6 and 12 years old [40]. Farré and Narbona designed this scale based on their experience with the adapted Conners questionnaire [41]. The EDAH measures the main characteristics of ADHD and the behavioral problems that may coexist with attentional deficit. This questionnaire was used to ensure that participants did not exhibit any ADHD-related behavior.

After completing the exercises, participants themselves were asked to fill in a usability questionnaire. The usability of the system was evaluated by a user satisfaction test based on the System Usability Scale [42]. This questionnaire consists of 10 items, which were evaluated by using a Likert scale ranging from 1, strongly agree, to 5, strongly disagree. Through feedback from this questionnaire, researchers will be able to continue to adapt the system to users' final needs.

Before completing the usability questionnaire, participants were seated in front of the eye tracking sensor to permit data collection. Users were seated opposite the center of the monitor, after adjusting the seating position to their height. Once they were aligned with the screen, the calibration process started, which took between 2 and 5 min per child. This calibration entails a visual target that moves around the screen. Participants were asked to follow this target with their gaze for a period of time. The target consists of a calibration grid with 5 positions, one on each corner of the screen and the last one right in the screen's center. The target consists of different calibration bullet points that appeared one after the other in the same order for all participants, starting from the top left corner.

Prior to the start of the exercises, participants were told in which kind of tasks they were taking part. They were also introduced to the eye tracking technology, and the sensor functionality was explained.

Participants used the system and filled in the questionnaire in a controlled environment, with a researcher observing and keeping track of all of the behavioral aspects of the study, but not interfering in the experimental setting.

3.5. Data Analysis, Processing and Classification

Recorded gaze data during the exercises has been processed, analyzed and used in order to identify the set of features that may help to build a classifier, as shown in Figure 3.

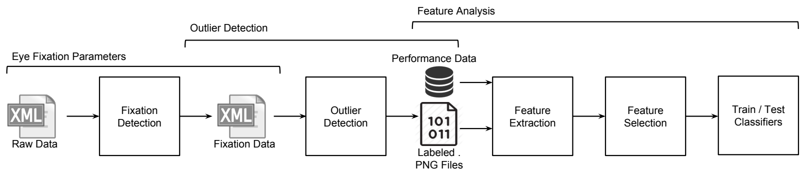


Figure 3. Raw data processing.

This section will explain in detail the different steps involved in the data analysis and feature identification process, so as to contribute to the core of intelligent therapies based on visual attention and user interaction.

3.5.1. Eye Fixation Parameters

The analysis of fixations and saccadic movements during the performance of certain tasks is related to attention in various ways. Several studies support this hypothesis [43–45], concluding that oculomotor mechanisms rely on attention for some aspects of eye movement control [46].

During the performance of the study, raw gaze data were recorded with the eye tracking sensor. These raw gaze data were stored as .xml files in the system, with information related to the level of the exercise that was currently running.

Listing 1 shows the stored gaze data for each participant and exercise. These data consist of the (x, y) coordinates recorded by the eye tracking sensor, the timestamp in which they were perceived, the pupil size for each eye and the exercise; the level and the mode the coordinates belong to were also stored for matching the raw gaze data with other interaction recordings.

These raw data were used for analysis and processing so as to obtain meaningful information about eye fixation locations, fixation durations, saccades and saccadic durations. Fixations are the period of time when the eyes remain fairly still and new information is acquired from the visual array [9], while saccades are the eye movements themselves. During saccades, no information is retrieved by the brain, since vision is suppressed under most normal circumstances [47].

Listing 1: Raw gaze data example

```
<user date="2014-05-22" id="12" sessionid="899" time="09:20:15" >
    <exercise id="puzzle" level="Level5" mode="performance">
        <eyedata>
            <timestamp time="1167610729217997">
                <left_eye pupil_diam="3.0435333252" validity="0" x="134.831732304" y="64.8299084174"/>
                <right_eye pupil_diam="2.89215087891" validity="0" x="69.9537996816" y="59.426052033"/>
            </timestamp>
            ...
        </eyedata>
    </exercise>
</user>
```

In order to detect the saccades and fixations, some processing techniques need to be applied to the raw data file. These steps are based on the Tobii I-VTfixation filter algorithm [48], have all been implemented in the Python programming language and are outlined in Figure 4.

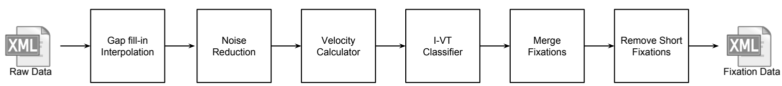


Figure 4. Raw data processing [48].

As Figure 4 shows, the first step in the processing algorithm is to apply the gap fill-in interpolation function. This step consists of filling in data where data are missing due to tracking problems that are not related to participants' behavior (such as blinks or when the user looks away from the screen). In order to distinguish between tracking problems and users' behavior, a max gap length is set, which limits the maximum length of the gap to be filled in. Following Tobii's white paper for the I-VT fixation filter and the value used by Komogortsev, this value was set at 75 ms [48,49].

After the gaps are filled in, the noise reduction function is applied. This function is based on a low-pass filter, which aims to smooth out the noise. The third step is the velocity calculator, which relates each sample with its velocity, in terms of visual angle (degrees per second). In order to reduce the impact of noise, the velocity for each sample is calculated as the average velocity of a period of time, taking as the central data input the current sample. This is done using a window length of 20 ms, which, according to the literature, has been found to handle a reasonable level of noise without distorting the signal [48].

The I-VT classifier applied to the signal is based on the one described by Komogortsev *et al.* [49] and outlined in the Tobii white paper [48]. The classifier determines which samples belong to a saccade, fixation or gap, based on a velocity threshold and the angle velocities calculated in the previous step. It also groups together consecutive samples using the same classification. The velocity threshold is set to 30 deg/s [48,50].

The merge fixations function aims to merge adjacent fixations that have been split up. This is done taking into account two different thresholds, the max-time between fixations, which is set to 75 ms [48], that is lower than the normal blink duration [49,51,52], and the max-angle between fixations, which is set at 0.5 deg [48,49,53–55]

Once all of the fixations have been identified, the shorter ones are removed. For the purposes of this analysis, 100 ms was set as the lower limit for fixation duration. This value was chosen based on the work of McConkie *et al.*, who concluded that 60 ms must pass before current visual information becomes available to the visual cortex for processing [56]. R. Tai *et al.* arrived at the lower limit of 100 ms by adding 30 ms, which is the time that elapses, at the end of a fixation, between when a command to move the eyes is sent and the onset of that saccade is reported. They allowed also 10 ms for the processing of any currently-observed stimuli, arriving at the 100-ms threshold [57].

After all of the processing functions have been applied to the current data, a new gaze data file is created with all of the fixations for the current exercise and participant. As shown in Listing 2, fixation data have a similar structure to raw data.

Listing 2: Fixation data example

```
<user date="2014-05-22" id="12" sessionid="897" time="09:15:27">
  <exercise id="puzzle" level="Level1" mode="performance">
    <fixationData>
      <fixation>
        <time duration="212.356933594" end_time="1.1676104137e+12" start_time="1.16761041349e+12">
          <position x="54.2251062717" y="130.508713537" />
        </time>
      </fixation>
      ...
    </fixationData>
  </exercise>
</user>
```

The stored fixation data save all of the fixations recorded during the exercise, along with the current activity information, user data and the duration, start time, end time and position of each fixation.

3.5.2. Outlier Detection Process

Once the processing stage is over, the fixation data are used to determine the outliers among the recorded data. This process is outlined in Figure 5.

As Figure 5 shows, a fixation heat map is created for each file. The fixation count heat map shows the accumulated number of fixations for each puzzle level and for each participant. Each fixation made adds a value to the color map at the location of the fixation [58].

The alpha layers of the images stored are then analyzed as a measure to identify the location and amount of fixations and saccades. All of the images are the same size and dimensions.

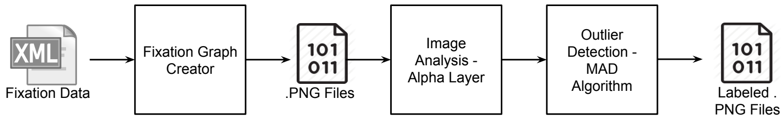


Figure 5. Outlier detection process.

The alpha information per image is stored in order to be processed by the median absolute deviation (MAD) algorithm for outlier detection implemented in Python.

The median deviation is a measure of scale based on the median of the absolute deviations from the median of the distribution [59]. The formula is shown in Equation (1).

$$MAD = median_i(|X_i - median_j(X_j)) \tag{1}$$

Moreover, the heat maps were analyzed taking into account users' overall performance during the entire study, so as to have another feature to determine outlier detection.

3.5.3. Classification

This section outlines the first steps taken in the classification process. The aim of this part is to assess the feasibility of using a set of combined features to evaluate user performance. These features are related to user interaction, timing and visual attention, as well as image-related data obtained directly from the heat maps.

This part explains the theoretical insights taken in this process. Please refer to the same section in the Results part for the mathematical outcomes of this process.

3.5.3.1. Feature Identification

Feature selection is a determining factor when classifying patterns. Features need to be insensitive to noise and separated from each other. Their main purpose is to objectively describe certain aspects, in this case of the attention and performance process in intelligent therapies aimed at children.

A collection of 34 features was selected based on image characteristics and user performance related to the current exercise. Features were selected based on the recorded data. The authors, in conjunction with the multidisciplinary team taking part in this project, took into consideration performance variables, as well as gaze pattern recordings. The subset of selected features for analysis from the pilot phase is outlined in Table 3.

Table 3. Selected features for analysis.

Feature	Data Type		Feature	Data Type		Feature	Data Type	Feature	Data Type
Classification	outlier/normal		Fixation avgduration	ms		Fixation No. in A1–C3 (9 features)	Integer	Time per level (4 features)	seconds
s		Fixation No. in A1–C3 (9 features)	Integer		Time per level (4 features)	seconds			
Global alpha	percentage		Fixation max duration	ms		Fixation avg duration in A1–C3 (9 features)	ms	Total correct answers	integer
s		Total correct answers	integer						
s		Fixation avg duration in A1–C3 (9 features)	ms		Total correct answers	integer			
s		Total correct answers	integer						
Fixation total No.	Integer		Fixation min duration	ms		Total time in exercise	seconds	Correct answers per level (4 features)	integer

Heat maps were divided into 9 quadrants in order to obtain detailed data about the location and density of fixations per participant and level.

The selected features were chosen for further analysis and consideration, so as to determine if they are suitable for use in an automatic classifier, capable of discerning the users' performance based on their interaction and gaze patterns.

3.5.3.2. Feature Selection

Feature selection creates a subset of features, improving their predictive performance and constructing patterns more efficiently. This helps to avoid multidimensionality, which may otherwise have an adverse effect on the decision making process [60].

Several techniques were used in this process. In order to assess the success rate of the classifier while obtaining the most accurate set of features, a set of different ensemble classifiers was used and compared with a traditional decision tree classifier.

- Sequential search: This process works by selecting the best features based on univariate statistical tests [39]. Inside this topic, the select k-best feature selection algorithm was applied. This process removes all but the k highest scoring features.
- L1-based feature selection: This was applied to assess the feasibility of discarding the zero coefficients. This is a means of reducing the dimensionality of data [39].
- Hierarchical feature selection: In these feature selection processes, the set of features is divided into smaller subsets until only one remains in each node [61]. Tree-based estimators were applied to compute feature importance, so as to discard the irrelevant ones [39].

3.5.3.3. Classifier Performance Analysis

Ensemble learning algorithms works by running a base learning algorithm multiple times, voting out the resulting hypotheses [62]. Ensemble learning has received an increasing interest recently, since it is more accurate and robust to noise than single classifiers [63,64].

This article compares the performance capabilities of 3 different ensemble algorithms when they are applied to the real dataset recorded in this study. The aim of this experiment is to assess the feasibility of building a classifier able to determine user performance using an adequate set of features of a different nature recorded during the therapy.

All of the classifiers were evaluated using cross-validation. The studied classifiers were:

- Random forest: This classifier is defined as a combination of tree predictors. Each tree depends on the values of a random vector sampled independently and with the same distribution for all trees [65]. Using the random selection of features yields error rates that compare favorably to AdaBoost [66], but are more robust with noise handling [65].
- Extremely randomized trees: A tree-based ensemble method for supervised classification and regression. It is a strongly randomized attribute selection method. This algorithm is accurate and computationally efficient [67].
- AdaBoost: This algorithm is an iterative procedure that tries to approximate the Bayes classifier by combining several weaker classifiers. A score is assigned to each classifier, and the final classifier is defined as the linear combination of the classifiers from each stage [68].

Moreover, a regular decision tree classifier was applied in order to assess the potential and improvement in accuracy, if any, of the previously mentioned tree-based ensemble methods.

4. Results

The recordings for the results explained in this section were taken during the month of May, 2014, at the Colegio Vizcaya school in Biscay, Spain.

4.1. Analysis of User Performance: Outcome Scores and Response Times

Although the present study is focused on the use of gaze data to analyze performance in attention-related cognitive therapies, we feel that it is also important to address commonly-used measurements to categorize user performance in this type of exercise: outcome scores and response times. These measures might be quite general in some cases where they show only a vague impression of the user's performance.

Participants' responses were recorded through the system implemented in Python. Their overall number of correct responses, as well as their number of correct responses per level are shown in Table 4.

The overall mean of correct responses is 11.937 (SD = 2.20) out of a possible score of 16. When the results are examined by levels, there are some differences in performance between the first two levels, which participants considered much easier, and the last two, which they found more difficult.

Users had a maximum of 50 seconds to complete each exercise. However, they were able to finish the level before time ran out. Considering the response times, *i.e.*, total time spent on test questions, the data show that the majority of participants took most of the entire time available at all of the levels.

Levels can be segmented into two groups, according to difficulty. The first two are considered the easiest ones, while the last two are trickier. There is a tendency between the two groups; users tend to perform slower with Levels 1 and 3 than with Levels 2 and 4. This may be because they tend to be more careful with novelty exercises or when the difficulty suddenly changes.

Figure 6 shows the overall performance of participants, regarding total time *versus* correct answers. As is displayed in Figure 6, users tend to respond correctly to more than half of the possible answers, while using 75% or more than the available time. When analyzing the group with the weakest performance, with a number of total correct answers below 10, it is clear that 75% of the participants in this group have a higher performance time.

Sensors **2015**, *15* **11104**

Table 4. Participants outcome scores and response times for each level.

Participant	Level 1		Level 2		Level 3		Level 4		Total (16 max.)	
ID (*n* = 32)	Time (s)	Items Correct	Time (s)	Items Correct	Time (s)	Items Correct	Time (s)	Items Correct	Time (s)	Items Correct
12	40.58	4	34.62	4	47.94	2	48.13	3	171.28	13
15	36.79	4	36.83	2	48.66	0	49.03	3	171.33	9
13	34.76	4	48.28	4	48.29	3	48.33	3	179.67	14
14	40.87	4	30.72	4	48.06	2	47.29	3	166.95	13
16	44.99	4	48.77	1	48.35	0	47.53	1	189.66	6
17	47.48	4	48.08	3	48.05	4	48.26	2	191.89	13
18	48.81	3	47.71	3	47.82	2	38.03	2	182.39	10
19	42.01	4	33.93	4	47.95	2	24.86	4	148.77	14
20	43.62	4	40.49	4	47.99	3	42.85	4	174.96	15
21	29.76	3	24.35	4	48.12	2	21.47	4	123.72	13
22	31.21	4	33.36	4	43.60	2	40.81	2	148.98	12
23	48.71	3	33.31	4	48.09	4	46.19	3	176.32	14
24	39.25	4	41.33	4	48.24	1	39.53	2	168.37	11
25	25.63	4	29.38	4	48.03	4	29.73	3	132.79	15
26	30.11	4	48.50	3	48.28	1	39.94	3	166.84	11
27	48.90	3	44.77	4	48.07	2	47.92	3	189.67	12
28	43.09	4	41.57	4	47.94	3	35.00	4	167.61	15
29	48.53	4	48.12	3	47.82	0	48.07	2	192.56	9
30	48.65	4	37.18	0	41.56	4	39.36	4	166.77	12
31	39.92	4	33.26	3	47.76	1	44.94	2	165.91	10
32	48.72	4	46.81	3	48.09	2	48.25	2	191.89	11
33	42.77	4	46.85	4	47.90	2	37.31	3	174.85	13
34	46.62	4	31.48	4	47.93	3	43.86	2	169.92	13
35	48.60	3	48.95	4	48.42	2	48.79	2	194.78	11
36	48.69	3	46.78	3	48.29	1	48.57	1	192.34	8
37	49.08	4	48.84	3	48.52	1	47.88	2	194.34	10
38	35.96	4	46.81	4	46.56	3	48.62	3	177.97	14
39	46.30	4	39.33	3	47.39	1	46.84	2	179.88	10
40	40.34	4	36.38	3	48.07	3	48.20	1	172.99	11
41	38.17	4	45.31	3	48.08	2	39.87	4	171.45	13
42	48.50	4	27.59	4	47.95	2	48.02	2	172.07	12
43	48.03	3	23.81	4	38.89	4	38.89	4	149.63	15
Average (*SD*)	42.36 (6.70)	3.75 (0.42)	39.80 (7.97)	3.37 (0.94)	47.40 (2.09)	2.12 (1.18)	42.89 (7.16)	2.65 (0.94)	172.45 (17.18)	11.93 (2.20)

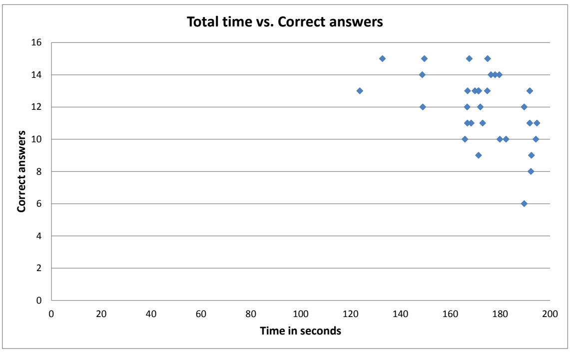


Figure 6. Time *vs.* correct answers: user performance.

Further analysis of user performance will be outlined in the following sections. The correct items' mean (11.93 out of 16) and standard deviation (SD = 2.20) values were used for obtaining the threshold for the weakest performers. This results in the value 9.73; since the study needs an entire threshold, this value was rounded up to 10. Participants with scores lower than this threshold were classified as the weakest performers. A total of four participants matched this criteria, so they were paired with the four best performers to obtain two balanced groups for further analysis. In order to address the research question stated in the Introduction, the four best performers (users with IDs 20, 25, 28 and 43) and the weakest four (users with Is 15, 16, 29 and 36) will be analyzed.

4.2. Fixation Heat Maps

Fixations were analyzed for each of the participants. Fixations were defined as a gaze longer than 100 ms. In order to address the research question stated in the Introduction, the most accurate and the weakest performers were selected for further analysis.

Fixations were displayed as heat maps, which were created based on the entire time participants took for each level. Red spots indicate higher levels of fixation, with yellow and green indicating decreasing amounts of fixations. Areas without color were not fixated upon. The most accurate performers are displayed in Figure 7, while the four with the weakest performance are displayed in Figure 8.

Figure 7. Participants with the best performance results.

When comparing the heat maps of both groups, there are some differences between the number, density and clustering of fixations. In Figure 7, where the total score results of the participants are 15 correct answers out of 16 possible ones for every case, the number of fixations is lower than for the participants with a weaker performance. Not only is it lower among participants, it also seems to decrease when analyzing the intra-level gaze behavior for each of them.

It is important to bear in mind that an overall lower number of total fixations suggests less time spent viewing specific areas of the assessment item.

Regarding Figure 8, where the total score for these participants ranges between six and nine correct answers out of 16 possible ones, the fixation density is higher for all the cases, except for the participant with ID 29.

Figure 8. Participants with the worst performance results.

This hypothesis agrees with R. Tai *et al.*, who found an inverse relation between the fixation and saccade amount and the degree of expertise of the participants [57].

Sensors **2015**, *15* **11107**

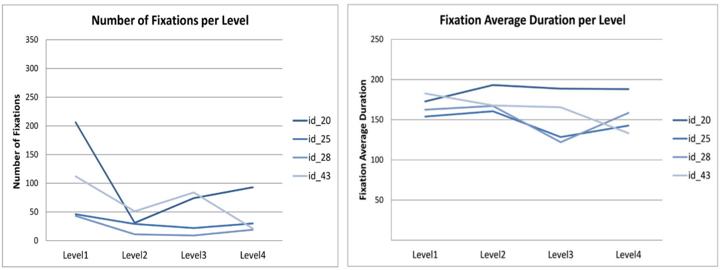
4.3. Quantitative Analysis

This section includes a quantitative analysis of the data regarding various features, such as the number of fixations per level, their average duration and the gender and age of the selected subgroups of participants, in order to analyze the feasibility of establishing some behavioral patterns.

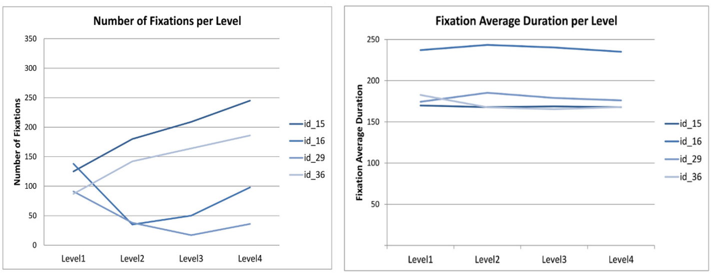
Table 5 displays the four participants with the best results. When processing the number of fixations, we observe that they decrease in number with the progression of the levels for all of the users, as displayed in Figure 9a. Since the exercise has the same visual layout for every level, this may be related to their having achieved a certain degree of expertise with each new level.

Table 5. Participants gender, age, number of fixations and average duration of these per level: best performers.

Participant ID	Age	Gender	Level 1			Level 2			Level 3			Level 4			Total	
			No. Fix.	Fix. Avg Duration (ms)	SD (ms)	No. Fix.	Fix. Avg Duration (ms)	SD (ms)	No. Fix.	Fix. Avg Duration (ms)	SD (ms)	No. Fix.	Fix. Avg Duration (ms)	SD (ms)	Time (s)	Correct Ans
20	9	Female	206	172.72	67.03	31	193.10	74.89	74	188.59	68.65	93	188.08	67.93	174.96	15
25	12	Male	46	153.85	58.53	29	160.49	70.62	22	128.40	37.64	30	142.66	72.09	132.79	15
28	11	Male	43	162.42	59.04	11	167.07	57.24	9	121.98	9.82	19	158.42	55.27	167.61	15
43	11	Female	112	182.61	50.90	51	167.70	90.66	84	165.43	78.35	21	133.15	39.96	149.63	15



(a)



(b)

Figure 9. Fixation data: best and weakest performers. (**a**) No. of fixations *vs.* fixation avg. duration, best performers; (**b**) No. of fixations *vs.* fixation avg. duration, weaker performers.

Table 6 displays the four participants with the weakest performance results. When processing the number of fixations, we find no specific relationship among them either, as displayed in Figure 9b, which may be related to the lack of appropriate techniques for solving the puzzle task.

Table 6. Participants gender, age, number of fixations and average duration of them per level: weakest performers.

Participant ID	Age	Gender	Level 1			Level 2			Level 3			Level 4			Total	
			No. Fix.	Fix. Avg Duration (ms)	SD (ms)	No. Fix.	Fix. Avg Duration (ms)	SD (ms)	No. Fix.	Fix. Avg Duration (ms)	SD (ms)	No. Fix.	Fix. Avg Duration (ms)	SD (ms)	Time (s)	Correct Ans
15	11	Female	125	169.91	80.86	180	167.95	74.16	209	168.85	71.15	245	167.82	69.23	171.33	9
16	9	Female	138	237.16	128.35	35	243.44	94.41	50	240.26	90.68	98	235.15	84.44	189.66	6
29	12	Male	91	174.38	76.58	38	185.29	91.79	17	179.01	70.67	36	176.08	70.79	192.56	9
36	8	Male	87	163.61	69.03	142	159.41	60.23	164	165.25	71.36	186	167.99	76.02	192.34	8

Further analysis of the results was made on the best *vs.* weakest performers' data. Due to the number of users that were used for the further analysis of the results, a Mann–Whitney non-parametric test was applied. The results of the test are outlined in Table 7.

Table 7. Best *vs.* weakest performers by level: Mann–Whitney analysis of the results.

	Level 1		Level 2		Level 3		Level 4		Global	
	p	Mann–Whitney U-Value	*p*	Mann–Whitney U-Value	*p*	Mann–Whitney U-Value	*p*	Mann–Whitney U-Value	*p*	Mann–Whitney U-Value
Fix Number	0.33	6.0	0.05	2.0	0.23	5.0	0.03	1.0	0.03	1.0
Fix Avg. Time	0.23	5.0	0.33	3.0	0.15	4.0	0.09	3.0	0.01	0.0
Time	0.15	4.0	0.09	3.0	0.09	3.0	0.01	0.0	0.09	3.0
Correct Answers	0.42	8.0	0.43	8.0	0.01	0.0	0.01	5.0	0.09	3.0

Table 7 shows that there are some significant differences ($p \leq 0.05$) in performance between groups. These differences appeared in the number of fixations in Levels 2 and 4 and globally. Moreover, there are other significant differences for fixation average time (global), time (Level 4) and the number of correct answers (Levels 3 and 4). However, these results may not be enough to conclude that there are consistent differences regarding the level of expertise of the participants.

R. Tai *et al.* [57] and Chi *et al.* [69] hypothesized that fixation duration data did not produce clear and consistent differences regarding the level of expertise of the participants, which agrees with the results obtained in this section.

4.4. Classification

This section outlines the first steps taken in the classification process. The aim of this part is to assess the feasibility of using a set of combined features to evaluate user performance. These features are related to user interaction, timing and visual attention, as well as image-related data obtained directly from the heat maps.

This part explains the mathematical outcomes of this process. Please refer to Section 3.5.3 for the theoretical insights.

In order to further assess the number of optimal features for the classification part, a recursive feature elimination process with cross-validation was applied. Table 8 displays the existing relation between the number of features and the classifier's accuracy. The number of features depends on the feature selection algorithm applied. These algorithms were outlined in Section 3.5.3.2.

Table 8. Performance comparison of feature selection algorithms using selected classifiers.

	Feature Selection Algorithms							
		Select *k*-best					L1-Based	Hierarchical
No. of Features	All	22	16	12	6	4	7	Btw10–14
Decision Tree	0.76 (±0.21)	0.80 (±0.18)	0.80 (±0.14)	0.81 (±0.15)	0.80 (±0.14)	0.80 (±0.15)	0.79 (±0.15)	0.82 (±0.17)
Random Forest	0.84 (±0.17)	0.87 (±0.11)	0.86 (±0.12)	0.87 (±0.11)	0.84 (±0.11)	0.83 (±0.14)	0.86 (±0.11)	0.88 (±0.11)
Extra Tree	0.80 (±0.18)	0.85 (±0.12)	0.82 (±0.14)	0.82 (±0.14)	0.81 (±0.14)	0.80 (±0.14)	0.83 (±0.14)	0.84 (±0.14)
AdaBoost	0.78 (±0.21)	0.85 (±0.14)	0.85 (±0.15)	0.84 (±0.15)	0.82 (±0.14)	0.81 (±0.15)	0.85 (±0.14)	0.86 (±0.13)

The accuracy results displayed in Table 8 were obtained by applying a cross-validation process of 100 iterations to all of the available data. These user data were divided as follows: 60% of the data for training and 40% for testing the classifier inside the cross-validation process.

With this setting, feature selection seems to be beneficial for building any type of analyzed classifier. However, when employing all of the available features, the accuracy rate falls below 0.80 for decision trees and AdaBoost classifiers. The select K-best features algorithm improves the classifiers' accuracy, especially when using 22 features. The L1-based algorithm displays good accuracy results for all of the ensemble methods and falls below 0.80 for the decision tree classifier. The tree-based hierarchical algorithm employed gives good results in accuracy with a limited number of features that range between 10 and 14, depending on the classifier employed.

The authors compared the accuracy performance of the selected ensemble classifiers with the overall performance of the a decision tree classifier. Since the data did not follow a normal distribution, a Mann–Whitney analysis was used. The results of comparing the performance of every ensemble classifier (with all features) with the decision tree classifier (with all features) is displayed in Table 9.

Table 9. Ensemble methods *vs.* decision trees: Mann–Whitney analysis of their accuracy.

Ensemble Method	(accuracy)	Decision Tree (accuracy)	*p*	Mann–Whitney U-Value
Random Forest	0.84		<0.001	2843.5
Extra Trees	0.8	0.76	0.003	3919.5
AdaBoost	0.78		0.02	4206

As is displayed in Table 9, the use of ensemble classifier methods significantly improves the overall performance of the classifier, regardless of the number of features employed. In the case of using all of the available features, the best classifier for the recorded data is the random forest.

Analyzing the difference in intra-classifier performance, Table 10 displays the Mann–Whitney analysis of the different feature-selection algorithms, comparing their performance with the accuracy obtained with the all features approach.

Table 10 illustrates that for almost all of the analyzed settings in this article, the use of a smaller set of features significantly improves the overall accuracy of all of the ensemble classifiers and the decision tree.

Table 10. All features *vs.* feature selection algorithms: Mann–Whitney analysis of their accuracy.

	All Features *vs.* Selected K best (22 features)		**All Features *vs.* L1-based (7 features)**		**All Features *vs.* Hierarchical (btw. 10–14 features)**	
	p	**Mann–Whitney U-Value**	***p***	**Mann–Whitney U-Value**	***p***	**Mann–Whitney U-Value**
Decision Trees	0.001	3550.5	0.002	3979	<0.001	3134.5
Random Forest	0.0007	3798.5	0.162	4201.5	0.003	3901.5
Extra Trees	<0.001	3228.5	0.0003	3638.5	<0.001	3308
AdaBoost	<0.001	3096.5	<0.001	3075	<0.001	2740.5

After carrying out all of the detailed experimental tests based on the recorded data, it can be concluded that accurate classification of different user performance according to their interaction and visual attention is possible.

5. Discussion and Conclusions

In the Conclusion, we intend to give an answer to the research questions outlined in the Introduction, as well as put forth new thoughts and trends about the present and future of assessing visual attention using eye tracker sensors in serious games.

According to the literature, there are several theories that link eye-movements with attentional processes [5,6], linking eye movements with cognitive processes, such as reading, visual search and scene perception. However, regarding intelligent therapies, eye movements do not always tell the whole story about the attentional process [70]. These resources should be complemented with other interaction records, as well as with relevant data about the participant. The higher the system information, the more accurate its customization to users' final needs.

In the Introduction, we hypothesized that participants with better performance may demonstrate patterns of eye-movements quantifiably different from individuals with weaker performance. Although some differences were found during the exercises, it is necessary to extend the study or to replicate it, in order to make stronger assumptions.

A comparison of the fixation duration data did not produce clear and consistent differences corresponding to the level of performance. These results corresponded with those related to the expertise level found by R. Tai *et al.* [57] and Chi *et al.* [69].

Regarding Figure 8, which shows the fixation heat maps for the weaker performers, fixation density is higher for all of the cases, except for the participant with ID 29. Moreover, the fixation density in Figure 7 decreases with the performance of new levels. These findings agree with R. Tai *et al.*, who found an inverse relationship between the fixation and saccade amount and the participants' degree of expertise [57].

When analyzing performance data, there are some differences between the two groups for which the puzzle levels are classified into according to difficulty. Table 4 shows the performance results. When changing the exercise type or level of challenge, users tend to spend more time and perform the exercise with taking more time to think. When the tasks are repeated, the ability level increases and the time to complete them drops. This may be related to the acquisition of specific problem-solving skills, which become more accurate with repetition. Further studies need to be carried out about the users' ability and performance capabilities in repetitive tasks.

Intelligent therapies that dynamically adapt themselves to users' needs and performance based on their interaction with the system have been proven to be efficient in terms of improvement comparisons [71]. A good set of collected data may provide improved means for obtaining adapted and efficient intelligent cognitive data. Researchers should be very careful with the selected and recorded features. Several different approaches need to be followed in order to obtain the most accurate set of performance data.

Moreover, a deeper analysis of timing per exercise may also prove to be interesting for study. As a future approach, the reading instructions stage will be separated from the performance of the exercise, so that we can obtain explicit performance timing, with and without the reading stage. This could give further information about whether there are any differences between the first performance of an exercise and the subsequent ones. This new approach may also help in further assessment of attention in the performance and instruction reading stages.

Reviewing the literature, there are several studies published linking the size of the pupils with cognitive processes [72–74]. Although, this response in the pupils is slow [75]. Current eye trackers measure pupil size and give it as another parameter, so it is easy to analyze this feature during the performance of tasks. This parameter was not analyzed in this study, and it may be an interesting additional feature in future research about this topic.

In recent years, the popularity of eye trackers has increased, and there are some open-source projects offering tools for gaze data analysis [76–80], while some manufacturers offer low-cost devices, such as the EyeTribe [81]. There are also several DIY approaches for building custom eye trackers [82–84]. The accuracy of these systems may sometimes be slightly inferior to high-end eye trackers, but they may be a viable solution for use outside the laboratory setting [85]. The use of eye trackers outside the research community may help to extend its potential with available intelligent therapies, bringing state-of-the-art technologies to users.

This study may expand in future directions, such as the design and development of the system, so that the tool includes new skills that continue along the lines of the current tool, for work on new capabilities, such as working-memory or processing speed.

Moreover, future lines should include the design and development of a robust classifier, with the selected features outlined in Section 3.5.3.1. The initial study of the classifier capabilities of ensemble methods with the available user data has produced positive results, especially when implementing a feature selection algorithm beforehand (see Section 4.4 for further information about the ensemble classifiers performance). Other classifiers need to be studied and tested, in order to consider others that may be more accurate, alone or in combination with others. This approach will help to create an autonomous system able to discern user implication based on visual attention and performance records.

Finally, some directions for the future are to replicate this study:

- with a greater number of users;
- with users with and without attention-related problems;
- developing a bilingual or trilingual tool that allows the study to be replicated in other areas in Spain and abroad where reported diagnosis of attention-related problems are significantly different from the Basque Country, Spain.

The use of gaze data constitutes a new information source in intelligent therapies that may help to build new approaches that are completely customized to final users' needs. Further studies need to be carried out in order to establish more detailed attention behaviors and patterns among children with and without attention problems. The replication of this study, along with the extension of the current system with new exercises, may help to build personalized performance profiles per user. These profiles may help in creating new customized therapies, while providing a new degree of information to the children themselves, therapists, psychologists, teachers and family.

Acknowledgments

The authors would like to express their appreciation to DEIKER, the Agency for Research Management and Promotion at the University of Deusto. They would also like to thank Patricia Clemente and Mari Jose Pecharroman and their team at Colegio Vizcaya School for their support and collaboration during the study and the AHIDA association for their advice in defining the given puzzle. The authors would like to thank professors Quasim Mehdi and Georgios Paltoglou at the University of Wolverhampton for their support and advice and researcher Philip Heaton at Deustotech-Life for his help with revising this article.

Author Contributions

Maite Frutos-Pascual is a PhD candidate whose research area is intelligent cognitive therapies based on biofeedback and serious games. She designed and developed the system and carried out the data analysis. Begonya Garcia-Zapirain is her PhD supervisor, she helped with the design of the experiment and she were on charge of participants recruitment. All the authors were responsible of the final version of this manuscript and all of them contributed extensively to this study.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Samuelsson, I.P.; Carlsson, M.A. The playing learning child: Towards a pedagogy of early childhood. *Scand. J. Educ. Res.* **2008**, *52*, 623–641.
2. Tobail, A.; Crowe, J.; Arisha, A. Learning by gaming: Supply chain application. In Proceedings of the IEEE Proceedings of the 2011 Winter Simulation Conference (WSC), Phoenix, AZ, USA, 11–14 December 2011; pp. 3935–3946.
3. Muir, M.; Conati, C. An analysis of attention to student—Adaptive hints in an educational game. In *Intelligent Tutoring Systems*; Springer: Crete, Greece, 2012; pp. 112–122.
4. Klami, A. Inferring task-relevant image regions from gaze data. In Proceedings of the 2010 IEEE International Workshop on Machine Learning for Signal Processing (MLSP), Kittilä, Finland, 29 August–1 September 2010; pp. 101–106.

5. Just, M.A.; Carpenter, P.A. A theory of reading: From eye fixations to comprehension. *Psychol. Rev.* **1980**, *87*, 329–354.
6. Hyönä, J. The use of eye movements in the study of multimedia learning. *Learn. Instr.* **2010**, *20*, 172–176.
7. Rayner, K. Eye movements in reading and information processing. *Psychol. Bull.* **1978**, *85*, 618–660.
8. Rayner, K. Eye movements in reading and information processing: 20 years of research. *Psychol. Bull.* **1998**, *124*, 372–422.
9. Rayner, K. Eye movements and attention in reading, scene perception, and visual search. *Q. J. Exp. Psychol.* **2009**, *62*, 1457–1506.
10. Mayer, R.E. Unique contributions of eye-tracking research to the study of learning with graphics. *Learn. Instr.* **2010**, *20*, 167–171.
11. Sundstedt, V.; Stavrakis, E.; Wimmer, M.; Reinhard, E. A psychophysical study of fixation behavior in a computer game. In Proceedings of the 5th symposium on Applied Perception in Graphics and Visualization, Los Angeles, CA, USA, 9–10 August 2008; pp. 43–50.
12. Pretorius, M.; Gelderblom, H.; Chimbo, B. Using eye tracking to compare how adults and children learn to use an unfamiliar computer game. In Proceedings of the 2010 Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists, Bela Bela, South Africa, 11–13 October 2010; pp. 275–283.
13. Deng, S.; Kirkby, J.A.; Chang, J.; Zhang, J.J. Multimodality with Eye tracking and Haptics: A New Horizon for Serious Games? *Int. J. Serious Games* **2014**, *1*, 17–33.
14. Almeida, S.; Veloso, A.; Roque, L.; Mealha, O. The eyes and games: A survey of visual attention and eye tracking input in video games. In Proceedings of the SBGames 2011—X Brazilian Symposium on Computer Games and Digital Entertainment—Arts & Design Track, Salvador, Brazil, 7–9 November 2011.
15. Isokoski, P.; Joos, M.; Spakov, O.; Martin, B. Gaze controlled games. *Univ. Access Inf. Soc.* **2009**, *8*, 323–337.
16. Zain, N.H.M.; Razak, F.H.A.; Jaafar, A.; Zulkipli, M.F. Eye tracking in educational games environment: evaluating user interface design through eye tracking patterns. In *Visual Informatics: Sustaining Research and Innovations*; Springer: Selangor, Malaysia, 2011; pp. 64–73.
17. Kiili, K.; Ketamo, H.; Kickmeier-Rust, M.D. Eye Tracking in Game-based Learning Research and Game Design. *Int. J. Serious Games* **2014**, *1*, 51–65.
18. El-Nasr, M.S.; Yan, S. Visual attention in 3D video games. In Proceedings of the 2006 ACM SIGCHI International Conference On Advances in Computer Entertainment Technology, Hollywood, CA, USA, 14–16 June 2006; p. 22.
19. Kickmeier-Rust, M.D.; Hillemann, E.; Albert, D. Tracking the UFO's Paths: Using Eye-Tracking for the Evaluation of Serious Games. In *Virtual and Mixed Reality-New Trends*; Springer: Orlando, FL, USA, 2011; pp. 315–324.
20. Sennersten, C.C.; Lindley, C.A. Real time eye gaze logging in a 3D game/simulation world. *Key Eng. Mat.* **2010**, *437*, 555–559.

21. Johansen, S.A.; Noergaard, M.; Rau, J. Can eye tracking boost usability evaluation of computer games. In Proceedings of the CHI, Florence, Italy, 5 April 2008.
22. Józsa, E.; Hámornik, B. Find the difference: Eye tracking study on information seeking behavior using an online game. *J. Eye Track. Vis. Cognit. Emot.* **2011**, 2, 27–35.
23. Dorr, M.; Böhme, M.; Martinetz, T.; Barth, E. Gaze beats mouse: A case study. In Proceedings of the 2nd Conference on Communication by Gaze Interaction, Turin, Italy, 4 September 2006; pp. 16–19.
24. Turner, J.; Velloso, E.; Gellersen, H.; Sundstedt, V. EyePlay: applications for gaze in games. In Proceedings of the First ACM SIGCHI Annual Symposium On Computer-Human Interaction in Play, London, UK, 19 October 2014; pp. 465–468.
25. Smith, J.D.; Graham, T. Use of eye movements for video game control. In Proceedings of the 2006 ACM SIGCHI International Conference On Advances in Computer Entertainment Technology, Montreal, QC, Canada, 22 April 2006; p. 20.
26. Hillaire, S.; Lécuyer, A.; Cozot, R.; Casiez, G. Using an eye-tracking system to improve camera motions and depth-of-field blur effects in virtual environments. In Proceedings of the IEEE Virtual Reality Conference (VR'08), Reno, NE, USA, 8 March 2008; pp. 47–50.
27. Chang, W.; Shen, P.A.; Ponnam, K.; Barbosa, H.; Chen, M.; Bermudez, S. WAYLA: Novel gaming experience through unique gaze interaction. In Proceedings of the ACM SIGGRAPH 2013 Emerging Technologies, San Francisco, CA, USA, 23 July 2013; p. 16.
28. Li, S.; Zhang, X. Eye-movement-based objective real-time quantification of patient's mental engagement in rehabilitation: A preliminary study. In Proceedings of the 2014 IEEE International Conference on Mechatronics and Automation (ICMA), Tianjin, China, 3–6 August 2014; pp. 180–185.
29. Lin, C.S.; Huan, C.C.; Chan, C.N.; Yeh, M.S.; Chiu, C.C. Design of a computer game using an eye-tracking device for eye's activity rehabilitation. *Opt. Lasers Eng.* **2004**, *42*, 91–108.
30. Vickers, S.; Istance, H.; Heron, M.J. Accessible gaming for people with physical and cognitive disabilities: a framework for dynamic adaptation. In Proceedings of the CHI'13 Extended Abstracts on Human Factors in Computing Systems, Paris, France, 27 April 2013; pp. 19–24.
31. Walber, T.; Neuhaus, C.; Scherp, A. EyeGrab: A Gaze-based Game with a Purpose to Enrich Image Context Information. In Proceedings of the EuroHCIR, Nijmegen, Netherlands, 24 August 2012; pp. 63–66.
32. Nacke, L.E.; Stellmach, S.; Sasse, D.; Lindley, C.A. Gameplay experience in a gaze interaction game. In Proceedings of the 5th Conference on Communication by Gaze Interaction—COGAIN 2009, Lyngby, Denmark, 26 May 2009; pp. 49–54.
33. Ekman, I.M.; Poikola, A.W.; Mäkäräinen, M.K. Invisible eni: Using gaze and pupil size to control a game. In Proceedings of the CHI'08 Extended Abstracts on Human Factors in Computing Systems, Florence, Italy, 5 April 2008; pp. 3135–3140.
34. Hertzog, M.A. Considerations in determining sample size for pilot studies. *Res. Nurs. Health* **2008**, *31*, 180–191.
35. Python Programming Language. Available online: https://www.python.org/ (accessed on 9 January 2015).

36. SQLite Database Engine. Available online: http://www.sqlite.org/ (accessed on 9 January 2015).
37. PyQT4. Available online: http://www.riverbankcomputing.com/software/pyqt/ (accessed on 9 January 2015).
38. Heatmaps - jjguy. Available online: http://jjguy.com/heatmap/ (accessed on 2 May 2015).
39. Scikit-Learn. Available online: http://scikit-learn.org/stable/index.html (accessed on 3 February 2015).
40. Farré, A.; Narbona, J. EDAH. Escalas para la evaluación del trastorno por déficit de atención con hiperactividad. In *Madrid: TEA Ediciones*; TEA Ediciones: Madrid, Spain, 2001.
41. Conners, C.K.; Sitarenios, G.; Parker, J.D.; Epstein, J.N. The revised Conners' Parent Rating Scale (CPRS-R): Factor structure, reliability, and criterion validity. *J. Abnorm. Child Psychol.* **1998**, *26*, 257–268.
42. Brooke, J. SUS-A quick and dirty usability scale. *Usabil. Eval. Ind.* **1996**, 189–194.
43. Kowler, E.; Zingale, C. Smooth eye movements as indicators of selective attention. *Atten. Perform.* **1985**, *11*, 285–300.
44. Schneider, W.X.; Deubel, H. Visual attention and saccadic eye movements: Evidence for obligatory and selective spatial coupling. *Stud. Vis. Inf. Process.* **1995**, *6*, 317–324.
45. Klein, R.; Kingstone, A.; Pontefract, A. Orienting of visual attention. In *Eye Movements and Visual Cognition*; Springer: New York, NY, USA, 1992; pp. 46–65.
46. Hoffman, J.E.; Subramaniam, B. The role of visual attention in saccadic eye movements. *Percept. Psychophys.* **1995**, *57*, 787–795.
47. Matin, E. Saccadic suppression: A review and an analysis. *Psychol. Bull.* **1974**, *81*, 899–917.
48. Olsen, A. *The Tobii I-VT Fixation Filter*; Tobii AB: Danderyd, Sweden, 20 March 2012.
49. Komogortsev, O.V.; Gobert, D.V.; Jayarathna, S.; Koh, D.H.; Gowda, S.M. Standardization of automated analyzes of oculomotor fixation and saccadic behaviors. *IEEE Trans. Biomed. Eng.* **2010**, *57*, 2635–2645.
50. Olsen, A.; Matos, R. Identifying parameter values for an I-VT fixation filter suitable for handling data sampled with various sampling frequencies. In Proceedings of the Symposium on Eye Tracking Research and Applications, Santa Barbara, CA, USA, 28–30 March 2012; pp. 317–320.
51. Ingre, M.; ÅKerstedt, T.; Peters, B.; Anund, A.; Kecklund, G. Subjective sleepiness, simulated driving performance and blink duration: Examining individual differences. *J. Sleep Res.* **2006**, *15*, 47–53.
52. Volkmann, F.C.; Riggs, L.A.; Moore, R.K. Eyeblinks and visual suppression. *Science* **1980**, *207*, 900–902.
53. Najemnik, J.; Geisler, W.S. Optimal eye movement strategies in visual search. *Nature* **2005**, *434*, 387–391.
54. Nakatani, C.; van Leeuwen, C. A pragmatic approach to multi-modality and non-normality in fixation duration studies of cognitive processes. *J. Eye Mov. Res.* **2008**, *1*, 1–12.
55. Kliegl, R.; Grabner, E.; Rolfs, M.; Engbert, R. Length, frequency, and predictability effects of words on eye movements in reading. *Eur. J. Cogn. Psychol.* **2004**, *16*, 262–284.
56. McConkie, G.W.; Underwood, N.R.; Zola, D.; Wolverton, G. Some temporal characteristics of processing during reading. *J. Exp. Psychol. Hum. Percept. Perform.* **1985**, *11*, 168–186.

57. Tai, R.H.; Loehr, J.F.; Brigham, F.J. An exploration of the use of eye-gaze tracking to study problem-solving on standardized science assessments. *Int. J. Res. Method Educ.* **2006**, *29*, 185–208.
58. Bojko, A.A. Informative or misleading? Heat maps deconstructed. In *Human-Computer Interaction. New Trends*; Springer: Berlin, Germany, 2009; pp. 30–39.
59. Howell, D.C. Median Absolute Deviation. In *Encyclopedia of Statistics in Behavioral Science*; Taylor & Francisr: Hoboken, NJ, USA, 2005.
60. Guyon, I.; Elisseeff, A. An introduction to variable and feature selection. *J. Mach. Learn. Res.* **2003**, *3*, 1157–1182.
61. Freeman, C.; Kulic, D.; Basir, O. Feature-Selected Tree-Based Classification. *IEEE Trans. Cybern.* **2013**, *43*, 1990–2004.
62. Dietterich, T.G. Ensemble learning. In *The Handbook of Brain Theory and Neural Networks*; MIT Press: Cambridge, MA, USA, 2002; pp. 110–125.
63. Breiman, L. Bagging predictors. *Mach. Learn.* **1996**, *24*, 123–140.
64. Rodriguez-Galiano, V.; Ghimire, B.; Rogan, J.; Chica-Olmo, M.; Rigol-Sanchez, J. An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS J. Photogramm. Remote Sens.* **2012**, *67*, 93–104.
65. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32.
66. Schapire, R.E.; Freund, Y.; Bartlett, P.; Lee, W.S. Boosting the margin: A new explanation for the effectiveness of voting methods. *Ann. Stat.* **1998**, 26, 1651–1686.
67. Geurts, P.; Ernst, D.; Wehenkel, L. Extremely randomized trees. *Mach. Learn.* **2006**, *63*, 3–42.
68. Zhu, J.; Zou, H.; Rosset, S.; Hastie, T. Multi-class adaboost. *Stat. Its Interface* **2009**, *2*, 349–360.
69. Chi, M.T.H.; Glaser, R.; Rees, E. Expertise in problem solving. In *Advances in the Psychology of Human Intelligence*; Sternberg, R.J., Ed.; Lawrence Erlbaum Associates: Hillsdale, NJ, USA, 1982; pp. 1–75.
70. Borji, A.; Itti, L. State-of-the-art in visual attention modeling. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 185–207.
71. Solana, J.; Cáceres, C.; García-Molina, A.; Chausa, P.; Opisso, E.; Roig-Rovira, T.; Menasalvas, E.; Tormos-Muñoz, J.M.; Gómez, E.J. Intelligent Therapy Assistant (ITA) for cognitive rehabilitation in patients with acquired brain injury. *BMC Med. Inform. Decis. Mak.* **2014**, *14*, doi:10.1186/1472-6947-14-58.
72. Binda, P.; Pereverzeva, M.; Murray, S.O. Pupil size reflects the focus of feature-based attention. *J. Neurophysiol.* **2014**, *112*, 3046–3052.
73. Sebastián-Gallés, N. Eyes wide shut: Linking brain and pupil in bilingual and monolingual toddlers. *Trends Cogn. Sci.* **2013**, *17*, 197–198.
74. Gabay, S.; Pertzov, Y.; Henik, A. Orienting of attention, pupil size, and the norepinephrine system. *Atten. Percept. Psychophys.* **2011**, *73*, 123–129.
75. Wierda, S.M.; van Rijn, H.; Taatgen, N.A.; Martens, S. Pupil dilation deconvolution reveals the dynamics of attention at high temporal resolution. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 8456–8460.

76. OGAMA—Open Gaze and Mouse Analyzer. Available online: http://www.ogama.net/ (accessed on 23 January 2015).
77. PyGaze—The Open-Source Toolbox for Eye Tracking. Available online: http://www.pygaze.org/ (accessed on 23 January 2015).
78. OpenEyes—Open-Source Open-Hardware Toolkit for Low-Cost Real-Time Eye Tracking. Available online: http://thirtysixthspan.com/openEyes/ (accessed on 23 January 2015).
79. OpenGazer—Open-Source Gaze Tracker for Ordinary Webcams. Available online: http://www.inference.phy.cam.ac.uk/opengazer/ (accessed on 23 January 2015).
80. Pupil Labs—Open-Source Mobile Eye Tracking Platform. Available online: http://pupil-labs.com/pupil/ (accessed on 23 January 2015).
81. EyeTribe—Ther World's First 99 Dollar Eye Tracker. Available online: http://www.theeyetribe.com/ (accessed on 23 January 2015).
82. Mantiuk, R.; Kowalik, M.; Nowosielski, A.; Bazyluk, B. *Do-It-Yourself Eye Tracker: Low-Cost Pupil-Based Eye Tracker For Computer Graphics Applications*; Springer: Klagenfurt, Austria, 2012.
83. Huang, C.W.; Jiang, Z.S.; Kao, W.F.; Huang, Y.L. Low-Cost and High-Speed Eye Tracker. In *Intelligent Technologies and Engineering Systems*; Springer: New York, NY, USA, 2013; pp. 421–427.
84. Ho, H.F. Low cost and better accuracy eye tracker. In Proceedings of the IEEE 2014 International Symposium on Next-Generation Electronics (ISNE), Kwei-Shan Tao-Yuan, Taiwan, 7 May 2014; pp. 1–2.
85. Johansen, S.A.; San Agustin, J.; Skovsgaard, H.; Hansen, J.P.; Tall, M. Low cost *vs.* high-end eye tracking for usability testing. In Proceedings of the CHI'11 Extended Abstracts on Human Factors in Computing Systems, Vancouver, BC, Canada, 7 May 2011; pp. 1177–1182.

5.3 ARTICLE IV: Gaze Behaviour Analysis in Cognitive Therapies based on Serious Games

The purpose of this study is to keep exploring the use of eye tracking sensors to evaluate, classify and assess the behaviour of children in attention-related cognitive therapies based on serious games to determine the utility of eye-related data as an input biofeedback signal for attention improvement therapies.

In this study, 82 randomly selected participants (38 girls and 44 boys), aged between 8 and 12 years with different attention skills were asked to solve a set of puzzles while their gaze patterns and interaction metrics were recorded using an eye-tracking sensor.

We tried to give an answer to the following research questions: Is the gaze interaction with the system consistent in children between 8-12 years old? Is visual interaction an efficient way of determining the attention degree and/or the performance interaction of different users? This article is the continuation of the published pilot study introduced in section 5.2.

Gaze behaviour was recorded and statistically analysed for the different main gaze-related parameters. Performance metrics and gaze behaviours were then classified for determining its suitability for intelligent adaptive therapies. Statistical differences were found in gaze data between different performance levels, this relation however was not strong enough in isolation, and performance metrics are also needed for creating the whole picture. Random forest classification process with cross-validation obtained an accuracy of 0.82 (=/- 0.09). This accuracy result is promising, suggesting that gaze patterns, in combination with performance, could be a good starting point in creating intelligent therapies based on serious games.

This study explored the suitability of using gaze behaviour and performance metrics as input data for customizing serious games. This article was submitted to the *Computers and Education Journal (IF: 2.556 [Q1])* as it is currently under review.

Gaze Behavior Analysis in Cognitive Therapies based on Serious Games

Maite Frutos-Pascual*, Begonya Garcia-Zapirain

DeustoTech Life [eVIDA] Faculty of Engineering University of Deusto, Avda de las Universidades 24, Bilbao, Spain

1,∗

Abstract

The purpose of this study is to explore the use of eye tracking sensors to evaluate and classify children's behaviour in attention-related cognitive therapies based on serious games to determine the utility of eye data as an input signal for adaptive learning experiences. This process was performed with a group of 82 children aged between 8 and 12 years old while they were interacting with a set of puzzle games. Gaze behaviour was recorded and statistically analysed for the analysis of different features such as fixations and saccadic movements. Performance metrics and gaze behaviours were then classified for determining its suitability for intelligent adaptive cognitive learning experiences. Statistical differences were found in gaze data between different performance levels, this relation however was not strong enough in isolation, and performance metrics are also needed for creating the whole picture. None of the gaze features analysed between levels displayed significant differences, suggesting that gaze interaction is consistent when only the level of challenge is altered and not the visual stimulus. Random forest classification process with cross-validation obtained an accuracy of 0.82 (=/- 0.09). This accuracy result is promising, suggesting that gaze patterns, in combination with performance, could be a good starting point in creating intelligent therapies based on serious games.

Keywords: Gaze Behaviour, Serious Games, Cognitive Intelligent Therapies

*Corresponding author
Email address: maitefrutos@deusto.es ()

1. Introduction

Popularity of video games is by no means a new trend in European households. The health of the video and computer games industry, together with the variety of genres and technologies available, mean that videogame concepts and programmes are being applied in numerous different disciplines. This growth in games as a mainstream entertainment has raised the question of how to take advantage of this digital games trend for educational purposes [1].

Several studies suggest that the future of pedagogy will inevitably be linked to the proposal of combined play and learning, in order to promote creativity in future generations [2]. The main assumption under this idea is that digital games enable learners to learn in an engaging, motivating and pleasant way [3].

One of the promising ways that games can be used for fostering education is by adapting themselves to each child individually [4]. Players have different learning abilities and training needs, however, serious games do not usually take player individuality into account. This may lead to the generation of stereotyped training conditions, affecting engagement and replay value of games [5].

Generating games that are flexible and customizable enough to adapt themselves autonomously to different users' needs is a key advance in the development of game-based learning experiences [6]. These adaptive digital educational games able to customize user interaction can enhance gameplay experience [3]. New game developments inside this trend provide systems with an efficient way of learning based on the users themselves, customizing and personalizing their experience with the system, which may increase their potential effects [7].

The first step for obtaining this customization is by analysing and understanding players' interaction behaviours. Research using eye tracking provides an opportunity to test theories about multimedia learning concerning processing during learning [8]. Moreover, the use of this approach may help in understanding where players focus their attention during game play [9].

Inferring relevance feedback from gaze it is done by capturing eye movements with proactive systems that may help to analyze user behavior [10]. Moreover, following the eye-mind hypothesis put forth by Carpenter in 1980, there is a close link between the direction of the human gaze and the focus of attention [11], provided that the visual environment inf front of the eyes is pertinent to the task that we want to study [12]. Eye tracking sensors collect information about the location and duration of an eye fixation within a specific area on a computer monitor.

Recently, researchers began to introduce eye-trackers devices and gaze data processing techniques in serious games and computer games [13, 14, 15]. Studies throughout the literature, such as those conducted by Nacke *et al.*, evaluated the use of gaze data as an alternative way of controlling interaction with games. They obtained favorable outcomes where this challenge results in positive affection and feelings of flow and immersion [16].

The purpose of this study is to keep exploring the use of eye tracking sensors to evaluate, classify and assess the behaviour of children in attention-related cognitive therapies based on serious games to determine the utility of eye-related data as an input biofeedback signal for attention improvement therapies. Authors will try to give an answer to the following research questions: Is the gaze interaction with the system consistent in children between 8-12 years? Is visual interaction an efficient way of determining the attention degree and/or the performance interaction of different users? This article is the continuation of the published work by the same authors *Assessing Visual Attention Using Eye Tracking Sensors in Intelligent Cognitive Therapies based on Serious Games* [17].

This article is outlined as follows: Materials and Methods Section is introduced, in which the authors discuss the form and function of the data collected from the eye tracker, the experiment description and the algorithms behind the proposed system. Next, a discussion of the collected data and the approaches to data analysis are examined. Finally the manuscript concludes with a discussion of possibilities for further research into the uses of eye tracking sensor and data as a biofeedback input to intelligent therapies.

2. Related Work

The observation of eye-movements is not a new area of research within psychology-related fields, having been studied in depth over the last few decades [18, 19, 20].

Research using eye tracking sensors affords a unique opportunity to test aspects of theories about multimedia learning concerning processing during learning [8]. Moreover, the use of this approach may help in understanding where players focus their attention during game play [9], as well as how they confront unfamiliar games and software [21].

However, it was not until recently that researchers began to analyse and introduce eye tracking sensors and techniques in serious games and computer games [13, 14, 15]. Games that can be controlled solely through eye movement would be accessible to persons with decreased mobility or control. Moreover, the use of eye tracking data can change the interaction with games, producing new input experiences based on visual attention [15].

Eye tracking devices have been used in the design of educational games, in terms of assessing usability based on user gaze behaviours when interacting with the game [22, 23]. El-Nasr and Yan used eye tracker sensors to analyse attention patterns within an interactive 3D game environment, so as to improve game level design and graphics [24].

Kickmeier-Rust *et al.* focused on assessing the effectiveness and efficiency of serious games. For this purpose, they assessed these variables with gaze data and gaze paths, in order to obtain interaction strategies in specific game situations [25]. Sennersten and Lindley also evaluated the effectiveness of virtual environments in games through the analysis of visual attention using eye tracking data [26]. Johansen *et al.* discussed the efficiency of eye tracker sensors in assessing users' behaviour during game play [27].

Jzsa and Hamornik used recorded eye tracking data to evaluate learning curves in university students while using a seven hidden differences puzzle game. They used this data to assess similarities and differences in information acquisition strategies considering gender- and education-dependent characteristics [28]. Dorr *et al.* conducted a similar study concluding that expert and novice players use different eye movement strategies [29]. Muir *et al.* used eye-tracking data to capture user attention patterns and to present results on how those patterns were affected by existing user knowledge, attitude towards getting help and performance while using the educational game, Prime Club [10].

Radoslaw *et al.* used eye tracker sensors for assessing render quality in games. They argued that gaze-dependent rendering was especially important when immersed in serious games, where players in virtual environments played a primary role [30]. Smith and Graham and Hillaire *et al.* concluded that use of an eye tracker increases video game immersion, altering the game play experience [31, 32].

Chang *et al.* developed the game WAYLA as a means to evaluate the potential to offer new interaction experiences based on eye tracking and visual attention. These authors took advantage of the popularity and arrival of more affordable eye tracker sensors [33].

Li and Zhang used eye-movement analysis to assess patients' mental engagement in a rehabilitation game. Therapists use this feedback to adjust rehab exercises to users' needs [34]. Continuing with the health-related field, Lin *et al.* developed an eye-tracking system for eye motion disability rehabilitation as a joystick-controlled game [35]. Vickers *et al.* developed a framework that integrated automatic modification of game tasks, interaction techniques and input devices according to a user ability profile [36].

Walber *et al.* presented EyeGrab, a game for image classification controlled by the players' gaze. The main purpose of this game was to collect eye tracking data to enrich image context information [37].

Other studies, such as those conducted by Nacke *et al.*, evaluated the use of eye tracker sensors as an alternative way of controlling interaction with games, obtaining favourable outcomes where this challenge results in positive affection and feelings of flow and immersion [38]. Ekman *et al.* goes one step further, discussing the limitations of using pupil-based interaction and providing suggestions for using pupil size as an input modality [16].

3. Materials and Methods

This section presents the methodology used in this study along with participants' characteristics and selection procedure.

3.1. Participants

The process for assessing attention was performed with a group of children with different levels of attention capacities. This process relies on data recorded with an eye tracking sensor. Participants were aged between 8 and 12 years, with an average age of 10 ($SD = 1.14$).

Eighty-two randomly selected participants (38 girls and 44 boys), were selected from a group of 173 volunteers by their teachers and psychologists. This sample size was considered adequate for the purpose of the outlined study [39].

71 children (40 boys and 31 girls) out of the 82 total were recruited from the Colegio Vizcaya School. Five of them, all boys, were diagnosed with some kind of learning difficulty. The remaining 11 children (4 boys and 7 girls) were recruited from the Albor-Cohs psychology cabinet, all of them were diagnosed with some kind of learning difficulty.

These children live in the Basque Country, Spain. 16 were diagnosed with an attention-related disorder or learning difficulty, all of them were taking medication during the test. All children have Spanish as their mother tongue, and were enrolled either in language model A (only Spanish) or B (Spanish and Basque) at school.

Since they were mature minors, the approval of parents or guardians was requested prior to conducting the study. This approval consisted of an informed consent following receipt of a detailed description of study, distributed via the school's or cabinet's regular newsletter.

3.2. Materials

All participants in the study completed the same assessment, which consisted of different parts. First, participants were asked to complete the first level of a puzzle exercise with four different levels of difficulty. In the first level users were able to familiarize themselves with the system, learning all the puzzle mechanics. Users have to connect each of the four slices presented in the exercise with its corresponding part in the main image. All participants were presented with the same image for each level, and all of the elements in the user interface appeared in the same part of the screen in each level.

The main image and the slices appeared in the middle of the screen, occupying the whole display from left to right. The question stem appeared in the upper middle part of the screen. The button to advance to the next level appeared at the lower middle part of each screen. The consistent layout of the screen was intended to minimize wide eye-movements.

Different levels' settings are outlined in table 1. All the participants had a maximum pre-set time of 50 seconds to complete each of the levels. However, if they finish the level before the time ends, they can go on to the next exercise. Depending on the level, the displayed image was labelled as easy, medium or hard. Only level one is displayed in colour, and it is only used for training purposes. Hard images have very similar slices, and are more complicated to complete.

Table 1: Different levels' settings [17]

	Time (secs)	Grid Size	Nº Slices	Display		Image Level		
				Color	Greyscale	Easy	Medium	Hard
Level 1	50	3 x 3	9	x		x		
Level 2	50	3 x 4	12		x	x		
Level 3	50	4 x 4	16		x		x	
Level 4	50	5 x 4	20		x			x

3.3. Devices and technologies

All of the data for this study was collected on the same device, which was located at the children's school or at the psychology cabinet, outside the laboratory environment. These conditions were considered appropriate due to the nature of the system.

The set of puzzles was developed in Python [40]. The results obtained and the necessary parameters were stored in a SQLite database [41]. User interface and user interaction were developed using PyQT4 [42]. The classification process was implemented using *Scikit-learn* library for machine learning in python [43]. Statistical analysis was performed using SciPy [44].

The puzzles were displayed in a 19 inch Lenovo monitor interface with a Acer Aspire Timeline X laptop running on Ubuntu 12.04. All the text in the different exercises was displayed as black text against a light-grey background following normal grammatical conventions in Spanish. Images were inserted as JPEG digital pictures scaled from their original versions. Response selection and any changes were stored by monitoring the user interaction and recording eye movements with a Tobii X1 Light eye tracker sensor. Figure **??** shows the study setting while one of the participants was interacting with the system.

The eye tracker is a non-invasive sensor with remote function. Participants were not required to remove their glasses or contact lenses during the tests. Accuracy under ideal conditions is 0.5 of the visual angle, while the sampling rate in this study was typically 28-32 Hz. The Tobii X1 light sensor was located beneath the computer monitor with the headrest fastened to the front edge of the desk, monitoring the participants head. The laptop was located behind the monitor, without interfering in the participants' field of vision.

A typical experimental trial including calibration lasted less than 25 minutes for each participant.

3.4. Experimental procedure

After completing the exercises, participants themselves were asked to fill in a usability questionnaire. Usability of the system was evaluated by a user satisfaction test based on the *System Usability Scale* [45]. This questionnaire consists on 10 items which were evaluated by using a Likert scale ranging from 1, strongly agree to 5 strongly disagree. Through feedback from this questionnaire researchers will be able to continue to adapt the system to users' final needs.

Before completing the usability questionnaire, participants were seated in front of the eye tracking sensor to permit data collection. Users were seated opposite the centre of the monitor, after adjusting the seating position to their height. Once they were aligned with the screen, the calibrating process started, which took between 2 and 5 minutes per child. This calibration entails a visual target that moves around the screen. Participants were asked to follow this target with their gaze for a period of time. The target consists of a calibration grid with 5 positions, one on each corner of the screen and a last one right in the screen's centre. The target consists of different calibration bullet points that appeared one after the other in the same order for all the participants, starting from the top left corner.

Prior to the start of the exercises, participants were told which kind of tasks were they taking part in. They were also introduced to the eye tracking technology and the sensor functionality was explained.

Participants used the system and filled in the questionnaire in a controlled environment, with a researcher observing and keeping track of all the behavioural aspects of the study but not interfering in the experimental setting.

3.5. Data recording and processing system

A modular system was designed for recording gaze behaviour and interaction parameters while players were taking part in the experiment.

The high level block diagram of the data recording and processing system is displayed in Figure 1.

The main functions of the different blocks displayed in Figure 1 are outlined below:

- **A. Interaction -** This module is in charge of displaying the appropriate level and serious games type inside from the available options. As dis-

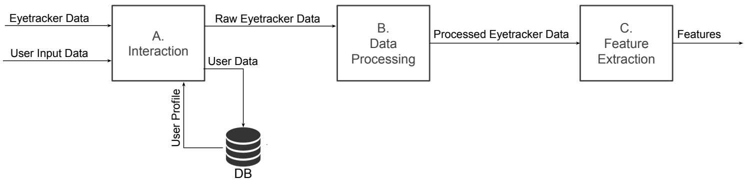


Figure 1: System block diagram

played in Figure 1. This module is also in charge of recording both gaze and interaction data streams, storing them in the appropriate format for enabling the analysis module to extract all the meaningful information.

– **B. Data processing -** These raw data was used for analysis and processing so as to obtain meaningful information about eye fixation locations, fixation durations, saccades, saccadic durations, gaps and gaps durations. In order to detect the saccades, fixations and gaps, some processing techniques need to be applied to the raw data file. These steps are based on the Tobii I-VT Fixation Filter algorithm [46], have been all implemented in the Python programming language and are outlined in detail in our previous work [17].

– **C. Feature extraction -** This is a feature extractor module of the processed gaze data and the performance recording generated by the player on each interaction with the game.Data collected during users interaction with the puzzle game was analysed. Features of different nature were then computed and extracted as detailed in section 3.6.

Once the data has been recorded and the different features detailed in subsection 3.6 have been extracted, they are statistically analysed in order to provide an answer to the research questions exposed in the introduction section.

3.6. Features

Data collected during users interaction with the puzzle game was analysed. Features of different nature were then computed and extracted.

Subsequent paragraphs details each of the extracted features. Features are divided into three different categories, eye movements, performance metrics and users' details.

3.6.1. Eye movements

Raw gaze data was recorded with the eye tracking sensor during the testing procedure. These recordings were stored as '.xml' files in the system. In order to obtain meaningful information, raw data was processed for eye-related movements.

In order to further analyse eye movements, the display was divided in 9 different regions, as it is shown in Figure 2. This division may help to determine which are the *'hot zones'* in the display, so as to determine where the gaze activity took place.

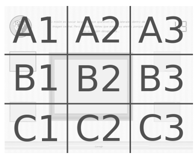


Figure 2: Image regions

- **Fixations** Fixations are the period of time when the eyes remain fairly still and new information is acquired from visual array [20]. Fixations were computed and processed for each user, and several parameters were drawn from this data: Number of fixations per area in the screen, fixation average duration and standard deviation, fixation maximum and minimum duration.

- **Saccades** Saccadic movements are rapid changes in position of the eyeballs typically found between fixational pauses during reading [47]. Saccades were computed and processed for each user, and various parameters were drawn from this data: Number of saccades, saccade average duration and standard deviation, saccade maximum and minimum duration, and saccade direction (up, left, right, down, up-right, up-left, down-left, down-right or same quadrant).

- **Gaps** Gaps were the period of time where no eye position was registered, excluding blinks. Blinks were filtered out for each user. Gaps may be produced because the user changes the visual focus from the screen to other external actuator, or because she or he gets too close to the monitor. When they get too close to the monitor, they are failing to practice good postures while using the computer. Various parameters were drawn from this data: Gaps per area in the screen, gap mean duration and standard deviation and the area where the last fixation os saccade before gap was located. The last eye movement before the gap was analysed. If it was a saccade near screen's borders, it is more likely than the user have changed the focus of attention. If the last movement is registered in a central position of the monitor, the user may be getting too close to the computer.

3.6.2. Performance metrics

Performance metrics describes the interaction data of each of the users with the system. These data comprises traditional in-game parameters, that are later computed along with eye gaze data.

- **Score** Score data refers to the number of good and bad choices that users performed per level.

- **Time** Time data refers to the final amount of time that participants employed per level (with a maximum of 50 seconds, which is the pre-set time per exercise).

3.6.3. User details

User relevant details (age, gender and whether the user was diagnosed or not with an attention-related disorder or learning difficulty) were also considered for the later classification process.

3.7. Gaze behaviour and performance statistical analysis

Gaze behaviour was statistically analysed for the different main gaze-related parameters introduced in subsection 3.6.1. Kruskal Wallis H test [48] was used for analysis over the ANOVA test as our data did not follow a normal distribution. A posthoc test for multiple comparisons using Mann-Whitney U post hoc test [49] was used for statistically significant results of the Kruskal Wallis H test, to check for statistical differences using pairwise comparisons between groups of independent variables. Pearson correlation coefficient was calculated for addressing potential relationships between gaze parameters and performance levels. Most relevant comparisons are outline in the results section.

3.8. Classification

Ensemble learning algorithms works by running a base learning algorithm multiple times, voting out the resulting hypotheses [50]. Ensemble learning have received an increasing interest, since it is more accurate and robust to noise than single classifiers [51, 52].

A random forest classifier using cross-validation was used in this article. This classifier is defined as a combination of tree predictors. Each tree depends on the values of a random vector sampled independently and with the same distribution for all trees [53]. Using random selection of features yields error rates that compare favorably to Adaboost [54], but are more robust with noise handling [53].

4. Results

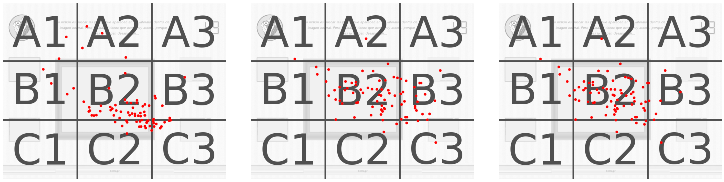
This section details the most relevant results obtained after the statistical analysis and classification process.

4.1. Fixation analysis

Fixations were analysed to determine if they could be an indicator of visual attention and user performance. The statistical analysis and main key findings are detailed in the following sections.

4.1.1. First fixation

The location of first fixation was reported per user and level, as it is displayed in Figure 3.



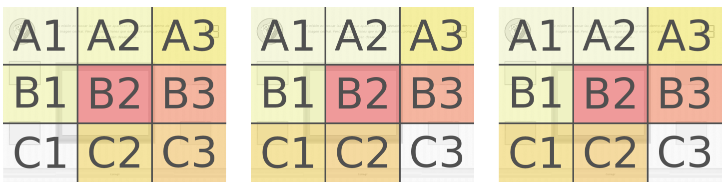
(a) Level 1 (b) Level 2 (c) Level 3

Figure 3: Fixation concentration heatmap by area and level

The location of the first fixation was widespread between users. However, Figure 3 displays a similar first fixation patterns between levels 2 (3b) and 3 (3c). However in Level 3 (3c) fixations were more clustered in the central quadrant B2.

4.1.2. Fixations per quadrant

Fixation per quadrant in the screen were extracted and analysed. Figure 4 displayed the amount quadrants in the screen with the higher amount of fixations in red, with stronger colours indicating a higher concentration of fixations in that quadrant.



(a) Level 1 (b) Level 2 (c) Level 3

Figure 4: Fixation concentration heatmap by area and level

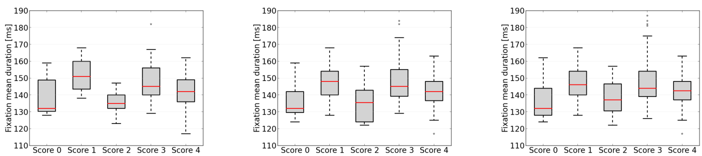
This values were obtained by the total amount of fixations registered per user in each of the quadrants. This calculation was applied in every level. The most-fixated quadrant in the screen does not vary in-between levels, however areas with no registered fixations on them varies from the first level (4a) to the rest, being consistent between Level 2 (4b) and Level 3 (4c).

4.1.3. Fixation mean duration time

Fixation average duration time per level was analysed. The results were as follows: Level 1 - Mean duration of fixation 143.18 ms *(SD 11.21 ms)*; Level 2 - Mean duration of fixation 143.25 ms *(SD 10.80 ms)*; Level 3 - Mean duration of fixation 143.59 ms *(SD 11.46 ms)*. Mann-Whitney U analysis displayed no significant differences in mean duration of the fixation between users in the different levels.

4.1.4. Fixation mean duration vs. score

The differences in fixation behaviours per performance levels were questioned. Participants were clustered based on the score they obtained per level and their average fixation duration was calculated. This result is displayed in Figure 5. Weak Pearson correlation was found between mean duration of fixations and score ($\rho = 0.38$).



(a) Level 1 (b) Level 2 (c) Level 3

Figure 5: Mean duration of fixations per level vs. participants' score

The statistical analysis of the results displayed in Figure 5 were as follows:

– **Level 1 (Figure 5a) -** Significant differences between the five groups ($p < 0.05$ KW) were obtained when comparing the mean of the fixation average time values for the five different score groups in level 1. In comparison according to pairs, there were significant differences between scores 0 and

1 ($p < 0.05$ MW, U value = 7.5), between scores 1 and 2 ($p < 0.05$ MW, U value = 3.0), between scores 1 and 4 ($p < 0.05$ MW, U value = 1.95) and between scores 2 and 3 ($p < 0.05$ MW, U value = 13.5).

– **Level 2 (Figure 5b) -** Significant differences between the five groups ($p < 0.001$ KW) were obtained when comparing the mean of the fixation average time values for the five different score groups in level 2. In comparison according to pairs, there were significant differences between scores 0 and 1 ($p < 0.05$ MW, U value = 35), between scores 0 and 3 ($p < 0.05$ MW, U value = -2.0), between scores 0 and 4 ($p < 0.05$ MW, U value = -2.2), between scores 1 and 2 ($p < 0.05$ MW, U value = 4.4), between scores 1 and 4 ($p < 0.05$ MW, U value = 1.9), between scores 2 and 3 ($p < 0.05$ MW, U value = 10.2), between scores 2 and 4 ($p < 0.05$ MW, U value = 46) and between scores 3 and 4 ($p < 0.05$ MW, U value = 90).

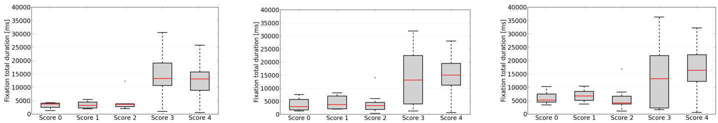
– **Level 3 (Figure 5c) -** Significant differences between the five groups ($p < 0.001$ KW) were obtained when comparing the mean of the fixation average time values for the five different score groups in level 2. In comparison according to pairs, there were significant differences between scores 0 and 1 ($p < 0.05$ MW, U value = 75), between scores 0 and 3 ($p < 0.05$ MW, U value = -3.0), between scores 0 and 4 ($p < 0.05$ MW, U value = -2.8), between scores 1 and 2 ($p < 0.05$ MW, U value = 13.2), between scores 2 and 3 ($p < 0.05$ MW, U value = 37.6) and between scores 2 and 4 ($p < 0.05$ MW, U value = 13.50).

4.1.5. Fixation total duration

Fixation total duration time per level was analysed. The results were as follows: Level 1 - Total duration of fixation 10,78 s *(SD 7.2 s)*; Level 2 - Mean duration of fixation 11.3 s *(SD 8.1 s)*; Level 3 - Mean duration of fixation 13.1 s *(SD 7.2 s)*. Even though the average time spent in fixation seems to increase with every level Mann-Whitney U analysis displayed no significant differences in mean duration of the fixation between users in the different levels.

4.1.6. Fixation total duration vs. score

The differences in fixation total duration per performance levels were questioned. Participants were clustered based on the score they obtained per level and their fixation total duration was calculated. This result is displayed in Figure 6. Moderate Pearson correlation was found between total duration of fixations and score ($\rho = 0.63$).



(a) Level 1 (b) Level 2 (c) Level 3

Figure 6: Fixation total time vs. participants' score per level

The statistical analysis of the results displayed in Figure 6 were as follows:

- **Level 1 (Figure 6a) -** Significant differences between the five groups ($p < 0.05$ KW) were obtained when comparing the total fixation duration time values for the five different score groups in level 1. In comparison according to pairs, there were significant differences between scores 0 and 3 ($p < 0.05$ MW, U value = -2.0), between scores 0 and 4 ($p < 0.05$ MW, U value = -2.7), between scores 1 and 3 ($p < 0.05$ MW, U value = 15.0), between scores 1 and 4 ($p < 0.05$ MW, U value = -2.6), between scores 2 and 3 ($p < 0.05$ MW, U value = 15.0), and between scores 2 and 4 ($p < 0.05$ MW, U value = 55.0).

- **Level 2 (Figure 6b) -** Significant differences between the five groups ($p < 0.05$ KW) were obtained when comparing the total fixation duration time values for the five different score groups in level 2. In comparison according to pairs, there were significant differences between scores 0 and 3 ($p < 0.05$ MW, U value = -2.0), between scores 0 and 4 ($p < 0.05$ MW, U value = -3.1), between scores 1 and 3 ($p < 0.05$ MW, U value = 23.0), between scores 1 and 4 ($p < 0.05$ MW, U value = -3.1), between scores 2 and 3 ($p < 0.05$ MW, U value = 32.0), and between scores 2 and 4 ($p < 0.05$ MW, U value = 58.0).

- **Level 3 (Figure 6c) -** Significant differences between the five groups ($p < 0.05$ KW) were obtained when comparing the total fixation duration time values for the five different score groups in level 3. In comparison according to pairs, there were significant differences between scores 0 and 4 ($p < 0.05$ MW, U value = -2.8), between scores 1 and 4 ($p < 0.05$ MW, U value = -2.3) and between scores 2 and 4 ($p < 0.05$ MW, U value = 65.0).

4.1.7. Key findings in fixations

All fixation measurements were extracted and analysed to give an insight of gaze behaviour and interaction. The main findings are detailed below:

- Location of the first fixation provide a preference for the central area of the screen, especially when the level progresses and the users become more confident with the outline of the interface.

- The analysis of fixations per quadrant displays a huge preference of users for the central areas of the screen, which are the ones that contain the main information and task of the presented system.

- Mean duration of fixations per level did not provide any significant difference. However, when users were segmented per level and score obtained, significant differences were found between certain scores, suggesting a partial relationship between the score and the average time spent per fixation, however clear and consistent data can not be extracted from this duration.

- Fixation total duration per level did not provide any significant differences. However, when participants were segmented per level and score, significant differences were found. Although participants with higher scores were more widespread in values, the ones with a lower amount of points were more clustered, suggesting a fixation pattern relation between performance and fixations.

Key findings in fixations suggest that it could exist a pattern relation between performance and eye fixations. This relation has been explored before in the literature [55, 56].

4.2. Saccade analysis

Eye saccadic movements were analysed following a similar fashion as the one presented in the fixation analysis section 4.1. The statistical analysis and main key features are detailed in the following sections.

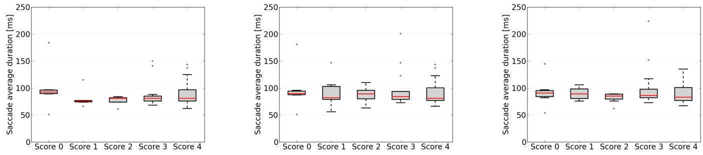
4.2.1. Saccade mean duration

Saccade average duration time per level was analysed. The results were as follows: Level 1 - Mean duration of saccade 94.70 ms *(SD 45.33 ms)*; Level 2 - Mean duration of fixation 101.49 ms *(SD 48.80 ms)*; Level 3 - Mean duration of fixation 102.12 ms *(SD 42.86 ms)*. Mann-Whitney U analysis displayed no significant differences in mean duration of the saccades between users in the different levels.

4.2.2. Saccade mean duration vs. score

The differences in saccade behaviours per performance level were also questioned. Participants were clustered based on the score they obtained per level and their average saccade duration was calculated. This result is displayed in Figure 7. No Pearson correlation was found for mean duration of saccadic movements versus performance score.

No significant differences between the five groups were found when comparing the average duration of saccadic eye movements for the five different score groups in Levels 1 (Figure 7a), 2 (Figure 7b) or 3 (Figure 7c) ($p > 0.05$ KW). In comparison according to pairs, there were no significant differences between score pairs ($p > 0.05$ MW).



(a) Level 1 (b) Level 2 (c) Level 3

Figure 7: Mean duration of fixations per level vs. participants' score

4.2.3. Key findings in saccades

The analysis of saccadic movements did not provide any significant results when segmented per level and/or per score. However, as it has been previously studied in the literature, saccadic eye movements are related to the final fixation of the attentional process, driving gaze to the different locations, while they are not able to attend to stimulus outside this path [57, 58].

4.3. Gap analysis

Although our definition of gap is not directly related with gaze movements but with the lack of them, we considered its analysis as an interesting part of this study. Its analysis could give an insight into the different behaviour that participants maintain while interacting with the system. The statistical analysis and main key features are detailed during the following sections.

4.3.1. Gaps per quadrant

Gap direction was analysed in the form of the analysis of the last gaze position registered before a gap. In this way of analysis we can determine the direction of the gap (i.e. Loosing gaze data when the last fixation position was registered in the middle of the screen is likely to represent a user getting too close to the monitor, failing to perform a good posture). These heatmaps per level are displayed in Figure 8.

Figure 8 shows that the majority of registered gaps were made when the users were moving their head to the right, or when they were getting too close

(a) Level 1 (b) Level 2 (c) Level 3

Figure 8: Quadrant-based heatmap with the last fixation position before a gap was registered per level

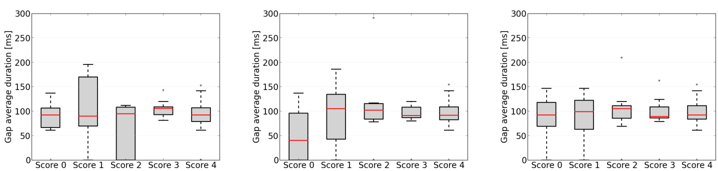
to the display monitor.

4.3.2. Gap mean duration

Gap average duration time per level was analysed. Mean duration of gaps in all levels was under 100 ms with a SD below 50 ms. Mann-Whitney U analysis displayed no significant differences in mean duration of the gaps between users in the different levels.

4.3.3. Gap mean duration vs. score

The differences in gap recorded data per performance level were also questioned. Participants were clustered based on the score they obtained per level, then their average gap duration was calculated. This result is displayed in Figure 9. No Pearson correlation was found for mean duration of gaps in recorded gaze data versus performance score.



(a) Level 1 (b) Level 2 (c) Level 3

Figure 9: Mean duration of gaps per level vs. participants' score

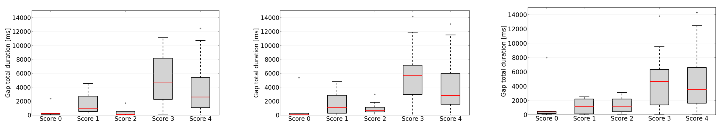
No significant differences between the five groups were found when comparing the average duration of gap recorded data for the five different score groups in Levels 1 (Figure 9a), 2 (Figure 9b) or 3 (Figure 9c) ($p > 0.05$ KW). In comparison according to pairs, there were no significant differences between score pairs ($p > 0.05$ MW).

4.3.4. Gap total duration

Gap total duration time per level was analysed. Total duration of gaps in all levels was under 4 s with a SD below 3 s. Mann-Whitney U analysis displayed no significant differences in mean duration of the fixation between users in the different levels.

4.3.5. Gap total duration vs. score

The differences in gap total duration per performance levels were questioned. Participants were clustered based on the score they obtained per level and their gap total duration was calculated. This result is displayed in Figure 10. Weak Pearson correlation was found between total duration of gaps and score ($\rho = 0.41$).



(a) Level 1 (b) Level 2 (c) Level 3

Figure 10: Total duration of gaps per level vs. participants' score

The statistical analysis of the results displayed in Figure 10 were as follows:

– **Level 1 (Figure 10a) -** Significant differences between the five groups ($p < 0.05$ KW) were obtained when comparing the total gap duration time values for the five different score groups in level 1. In comparison according to pairs, there were significant differences between scores 0 and 3 ($p < 0.05$ MW, U value = -2.8), between scores 0 and 4 ($p < 0.05$ MW, U value = -2.9), between scores 1 and 3 ($p < 0.05$ MW, U value = 16.0), between scores 2 and 4 ($p < 0.05$ MW, U value = 4.0) and between scores 2 and 4 ($p < 0.05$ MW, U value = 33.5).

- **Level 2 (Figure 10b) -** Significant differences between the five groups ($p < 0.05$ KW) were obtained when comparing the total gap duration time values for the five different score groups in level 2. In comparison according to pairs, there were significant differences between scores 0 and 3 ($p < 0.05$ MW, U value = -2.2), between scores 0 and 4 ($p < 0.05$ MW, U value = -2.8), between scores 1 and 3 ($p < 0.05$ MW, U value = 20.0), between scores 1 and 4 ($p < 0.05$ MW, U value = -2.2), between scores 2 and 3 ($p < 0.05$ MW, U value = 22.0), and between scores 2 and 4 ($p < 0.05$ MW, U value = 64.0).

- **Level 3 (Figure 10c) -** Significant differences between the five groups ($p < 0.05$ KW) were obtained when comparing the total gap duration time values for the five different score groups in level 3. In comparison according to pairs, there were significant differences between scores 0 and 4 ($p < 0.05$ MW, U value = -2.5), between scores 1 and 4 ($p < 0.05$ MW, U value = -1.8) and between scores 2 and 4 ($p < 0.05$ MW, U value = 89.0).

4.3.6. Key findings in gaps

Gap measurements were extracted and analysed to give an insight of gaze behaviour and interaction. The main findings are detailed below:

- The analysis of movements before a gap per quadrant displays a pattern that is maintained throughout the three levels. Users lost their gaze mainly when they were looking at the right hand side of the monitor and when they were too close to the display. The researcher performing the test was located at the right hand side of the display, so this gaps were probably due to participants asking questions during the performance of the test.

- Mean duration of gaps per level and per score did not provide any significant difference.
- Gap total duration per level did not provide any significant differences. However, when participants were segmented per level and score, significant differences were found. This results followed the same pattern as in the analysis of fixation total duration.

4.4. Classification

This subsection outlines the first steps taken in the classification process. Gaze analysed during previous subsections suggest that there is a potential in using gaze behaviour for assessing in performance scores in this cognitive intelligent therapy based on serious games. The features used in the classification process were the ones outlined in section 3.6.

Accuracy result of the classification process was obtained applying a cross-validation process of 100 iterations to all the available feature data. These user data was divided as follows 60% of the data for training and 40% for testing the classifier inside the cross-validation process. With this settings, an accuracy classification level of 0.82 (+/- 0.09) was obtained.

5. Conclusion

In the Conclusion, we intend to give an answer to the research questions outlined in the Introduction, as well as put forth new thoughts and trends about the present and future of assessing visual attention using eye tracker sensors in serious games.

According to the literature, there are several theories that link eye-movements with attentional processes [11, 12], linking eye movements with cognitive processes, such as reading, visual search and scene perception. However, regarding intelligent therapies, eye movements do not always tell the whole story about the attentional process [59]. In this study, fixations and gaps were found to have a link with the score performance level. However this relation was not considered to be strong enough when using it in isolation, we need a set of performance metrics as well as gaze data for creating the whole picture of user profile.

In the Introduction, we hypothesized that participants gaze interaction with the system may be consistent across different levels. None of the gaze features analysed between levels displayed significant differences between them, suggesting that gaze interaction is consistent when the same visual stimulus is presented, only altering the challenge level.

We also hypothesized that visual interaction may be an efficient way of determining the attention or performance degree of interaction with the system. The analysed gaze patterns displayed some differences between different performance levels. In our previous study, a comparison of the fixation duration data did not produce clear and consistent differences corresponding to the level of performance. These results corresponded with those related to the expertise level found by R. Tai *et al.* [55] and Chi *et al.* [56]. In this new extended study, fixation mean duration results were not strong enough to determine a trend between the performance of the participants and their average fixation duration per level and score obtained, this agrees with the findings in the literature.

R. Tai *et al.*, who found an inverse relationship between the fixation and saccade amount and the participants' degree of expertise [55]. They performed their study with a group of teachers at a secondary school. In contrast to this earlier findings, however, no evidence of this was detected in this case, with 82 participants between 8 and 12 years. We did find, however, that the amount of time spent in fixations in users with lower scores is less spread and more consistent that the ones with higher scores, suggesting a differentiation pattern between them.

Intelligent therapies that dynamically adapt themselves to users' needs and performance based on their interaction with the system have been proven to be efficient in terms of improvement comparisons [60].

The classification process applied obtained an accuracy of 0.82 when trying to determine, based on the set of performance and gaze features, the final score of the participant. This accuracy result is promising, suggesting that gaze patterns, in combination with performance, could be a good starting point in creating intelligent therapies based on serious games. A good set of collected data may provide improved means for obtaining adapted and efficient intelligent cognitive data. Researchers should be very careful with the selected and recorded features. Several different approaches need to be followed in order to obtain the most accurate set of performance data. New extended studies and replications need to be carried out to ascertain this result.

In recent years, the popularity of eye trackers has increased, and there are some open-source projects offering tools for gaze data analysis [61, 62, 63, 64, 65], while some manufacturers offer low-cost devices, such as the EyeTribe [66]. There are also several DIY approaches for building custom eye trackers [67, 68, 69]. The accuracy of these systems may sometimes be slightly inferior to high-end eye trackers, but they may be a viable solution for use outside the laboratory setting [70]. The use of eye trackers outside the research community may help to extend its potential with available intelligent therapies, bringing state-of-the-art technologies to users.

The use of gaze data constitutes a new information source in intelligent therapies that may help to build new approaches that are completely customized to final users' needs. The replication of this study, along with the extension of the current system with new exercises, may help to build personalized performance profiles per user. These profiles may help in creating new customized therapies, while providing a new degree of information to the children themselves, therapists, psychologists, teachers and family.

Acknowledgments

The authors would like to express their appreciation to DEIKER, the Agency for Research Management and Promotion at the University of Deusto. They would also like to thank Patricia Clemente and Mari Jose Pecharroman and their team at Colegio Vizcaya School and Angela Magaz and her team at Albor-Cohs for their support and collaboration during the study.

References

[1] J. L. Plass, B. D. Homer, C. K. Kinzer, Foundations of game-based learning, Educational Psychologist 50 (4) (2015) 258–283.

[2] I. P. Samuelsson, M. A. Carlsson, The playing learning child: Towards a pedagogy of early childhood, Scandinavian Journal of Educational Research 52 (6) (2008) 623–641.

[3] E. L.-C. Law, X. Sun, Evaluating user experience of adaptive digital educational games with activity theory, International Journal of Human-Computer Studies 70 (7) (2012) 478–497.

[4] E. Andersen, Optimizing adaptivity in educational games, in: Proceedings of the International Conference on the Foundations of Digital Games, ACM, 2012, pp. 279–281.

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

[6] S. Göbel, V. Wendel, C. Ritter, R. Steinmetz, Personalized, adaptive digital educational games using narrative game-based learning objects, in: Entertainment for Education. Digital Techniques and Systems, Springer, 2010, pp. 438–445.

[7] A. Tobail, J. Crowe, A. Arisha, Learning by gaming: supply chain application, in: Simulation Conference (WSC), Proceedings of the 2011 Winter, IEEE, 2011, pp. 3935–3946.

[8] R. E. Mayer, Unique contributions of eye-tracking research to the study of learning with graphics, Learning and instruction 20 (2) (2010) 167–171.

[9] V. Sundstedt, E. Stavrakis, M. Wimmer, E. Reinhard, A psychophysical study of fixation behavior in a computer game, in: Proceedings of the

5th symposium on Applied perception in graphics and visualization, ACM, 2008, pp. 43–50.

[10] M. Muir, C. Conati, An analysis of attention to student–adaptive hints in an educational game, in: Intelligent Tutoring Systems, Springer, 2012, pp. 112–122.

[11] M. A. Just, P. A. Carpenter, A theory of reading: from eye fixations to comprehension., Psychological review 87 (4) (1980) 329.

[12] J. Hyönä, The use of eye movements in the study of multimedia learning, Learning and Instruction 20 (2) (2010) 172–176.

[13] S. Deng, J. A. Kirkby, J. Chang, J. J. Zhang, Multimodality with eye tracking and haptics: a new horizon for serious games?, International Journal of Serious Games 1 (4).

[14] S. Almeida, A. Veloso, L. Roque, O. Mealha, The eyes and games: A survey of visual attention and eye tracking input in video games, Proceedings of the SBGames, Salvador BA, Brazil (2011) 7–9.

[15] P. Isokoski, M. Joos, O. Spakov, B. Martin, Gaze controlled games, Universal Access in the Information Society 8 (4) (2009) 323–337.

[16] I. M. Ekman, A. W. Poikola, M. K. Mäkäräinen, Invisible eni: using gaze and pupil size to control a game, in: CHI'08 Extended Abstracts on Human Factors in Computing Systems, ACM, 2008, pp. 3135–3140.

[17] M. Frutos-Pascual, B. Garcia-Zapirain, Assessing visual attention using eye tracking sensors in intelligent cognitive therapies based on serious games, Sensors 15 (5) (2015) 11092–11117.

[18] K. Rayner, Eye movements in reading and information processing., Psychological bulletin 85 (3) (1978) 618.

[19] K. Rayner, Eye movements in reading and information processing: 20 years of research., Psychological bulletin 124 (3) (1998) 372.

[20] K. Rayner, Eye movements and attention in reading, scene perception, and visual search, The quarterly journal of experimental psychology 62 (8) (2009) 1457–1506.

[21] M. Pretorius, H. Gelderblom, B. Chimbo, Using eye tracking to compare how adults and children learn to use an unfamiliar computer game, in: Proceedings of the 2010 Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists, ACM, 2010, pp. 275–283.

[22] N. H. M. Zain, F. H. A. Razak, A. Jaafar, M. F. Zulkipli, Eye tracking in educational games environment: evaluating user interface design through eye tracking patterns, in: Visual Informatics: Sustaining Research and Innovations, Springer, 2011, pp. 64–73.

[23] K. Kiili, H. Ketamo, M. D. Kickmeier-Rust, Eye tracking in game-based learning research and game design, International Journal of Serious Games 1 (2) (2014) 51–65.

[24] M. S. El-Nasr, S. Yan, Visual attention in 3d video games, in: Proceedings of the 2006 ACM SIGCHI international conference on Advances in computer entertainment technology, ACM, 2006, p. 22.

[25] M. D. Kickmeier-Rust, E. Hillemann, D. Albert, Tracking the ufos paths: Using eye-tracking for the evaluation of serious games, in: Virtual and Mixed Reality-New Trends, Springer, 2011, pp. 315–324.

[26] C. C. Sennersten, C. A. Lindley, Real time eye gaze logging in a 3d game/simulation world, Key Engineering Materials 437 (2010) 555–559.

[27] S. A. Johansen, M. Noergaard, J. Rau, Can eye tracking boost usability evaluation of computer games, in: Proceedings of CHI, Vol. 8, 2008.

[28] E. Józsa, B. Hámornik, Find the difference!: eye tracking study on information seeking behavior using an online game.

[29] M. Dorr, M. Böhme, T. Martinetz, E. Barth, Gaze beats mouse: a case study, in: Proceedings of the 2nd Conference on Communication by Gaze Interaction, 2007, pp. 16–19.

[30] J. Turner, E. Velloso, H. Gellersen, V. Sundstedt, Eyeplay: applications for gaze in games, in: Proceedings of the first ACM SIGCHI annual symposium on Computer-human interaction in play, ACM, 2014, pp. 465–468.

[31] J. D. Smith, T. Graham, Use of eye movements for video game control, in: Proceedings of the 2006 ACM SIGCHI international conference on Advances in computer entertainment technology, ACM, 2006, p. 20.

[32] S. Hillaire, A. Lécuyer, R. Cozot, G. Casiez, Using an eye-tracking system to improve camera motions and depth-of-field blur effects in virtual environments, in: Virtual Reality Conference, 2008. VR'08. IEEE, IEEE, 2008, pp. 47–50.

[33] W. Chang, P.-A. Shen, K. Ponnam, H. Barbosa, M. Chen, S. Bermudez, Wayla: novel gaming experience through unique gaze interaction, in: ACM SIGGRAPH 2013 Emerging Technologies, ACM, 2013, p. 16.

[34] S. Li, X. Zhang, Eye-movement-based objective real-time quantification of patient's mental engagement in rehabilitation: A preliminary study, in: Mechatronics and Automation (ICMA), 2014 IEEE International Conference on, IEEE, 2014, pp. 180–185.

[35] C.-S. Lin, C.-C. Huan, C.-N. Chan, M.-S. Yeh, C.-C. Chiu, Design of a computer game using an eye-tracking device for eye's activity rehabilitation, Optics and Lasers in Engineering 42 (1) (2004) 91–108.

[36] S. Vickers, H. Istance, M. J. Heron, Accessible gaming for people with physical and cognitive disabilities: a framework for dynamic adaptation, in: CHI'13 Extended Abstracts on Human Factors in Computing Systems, ACM, 2013, pp. 19–24.

[37] T. Walber, C. Neuhaus, A. Scherp, Eyegrab: A gaze-based game with a purpose to enrich image context information., in: EuroHCIR, Citeseer, 2012, pp. 63–66.

[38] L. E. Nacke, S. Stellmach, D. Sasse, C. A. Lindley, Gameplay experience in a gaze interaction game, arXiv preprint arXiv:1004.0259.

[39] M. A. Hertzog, Considerations in determining sample size for pilot studies, Research in nursing & health 31 (2) (2008) 180–191.

[40] Python programming language, `https://www.python.org/`, accessed: 15-may-2016.

[41] Sqlite database engine, `http://www.sqlite.org/`, accessed: 15-may-2016.

[42] Pyqt4, `http://www.riverbankcomputing.com/software/pyqt/`, accessed: 15-may-2016.

[43] Scikit-learn, `http://scikit-learn.org/stable/index.html`, accessed: 15-may-2016.

[44] Scipy, `https://www.scipy.org/`, accessed: 14-may-2016.

[45] J. Brooke, Sus-a quick and dirty usability scale, Usability evaluation in industry 189 (1996) 194.

[46] A. Olsen, The tobii i-vt fixation filter, Tobii Technology.

[47] G. Westheimer, Mechanism of saccadic eye movements, AMA Archives of Ophthalmology 52 (5) (1954) 710–724.

[48] W. H. Kruskal, W. A. Wallis, Use of ranks in one-criterion variance analysis, Journal of the American statistical Association 47 (260) (1952) 583–621.

[49] R. L. Iman, W. Conover, Multiple comparisons procedures based on the rank transformation, in: Proceedings of the Joint Statistical Meetings, Houston Texas, August, 1983.

[50] T. G. Dietterich, Ensemble learning, The handbook of brain theory and neural networks 2 (2002) 110–125.

[51] L. Breiman, Bagging predictors, Machine learning 24 (2) (1996) 123–140.

[52] V. Rodriguez-Galiano, B. Ghimire, J. Rogan, M. Chica-Olmo, J. Rigol-Sanchez, An assessment of the effectiveness of a random forest classifier for land-cover classification, ISPRS Journal of Photogrammetry and Remote Sensing 67 (2012) 93–104.

[53] L. Breiman, Random forests, Machine learning 45 (1) (2001) 5–32.

[54] R. E. Schapire, Y. Freund, P. Bartlett, W. S. Lee, Boosting the margin: A new explanation for the effectiveness of voting methods, Annals of statistics (1998) 1651–1686.

[55] R. H. Tai, J. F. Loehr, F. J. Brigham, An exploration of the use of eye-gaze tracking to study problem-solving on standardized science assessments, International journal of research & method in education 29 (2) (2006) 185–208.

[56] M. T. Chi, et al., Expertise in problem solving.

[57] J. E. Hoffman, B. Subramaniam, The role of visual attention in saccadic eye movements, Perception & psychophysics 57 (6) (1995) 787–795.

[58] E. Kowler, E. Anderson, B. Dosher, E. Blaser, The role of attention in the programming of saccades, Vision research 35 (13) (1995) 1897–1916.

[59] A. Borji, L. Itti, State-of-the-art in visual attention modeling, Pattern Analysis and Machine Intelligence, IEEE Transactions on 35 (1) (2013) 185–207.

[60] J. Solana, C. Cáceres, A. García-Molina, P. Chausa, E. Opisso, T. Roig-Rovira, E. Menasalvas, J. M. Tormos-Muñoz, E. J. Gómez, Intelligent therapy assistant for cognitive rehabilitation in patients with acquired brain injury, BMC medical informatics and decision making 14 (1).

[61] Ogama - open gaze and mouse analyzer, `http://www.ogama.net/`, accessed: 15-may-2016.

[62] Pygaze - the open-source toolbox for eye tracking, `http://www.pygaze.org/`, accessed: 15-may-2016.

[63] Openeyes - open-source open-hardware toolkit for low-cost real-time eye tracking, `http://thirtysixthspan.com/openEyes/`, accessed: 15-may-2016.

[64] Opengazer - open-source gaze tracker for ordinary webcams, `http://www.inference.phy.cam.ac.uk/opengazer/`, accessed: 15-may-2016.

[65] Pupil labs - open-source mobile eye tracking platform, `http://pupil-labs.com/pupil/`, accessed: 23-jan-2015.

[66] Eyetribe - ther world's first 99 dollar eye tracker, `http://www.theeyetribe.com/`, accessed: 15-may-2016.

[67] R. Mantiuk, M. Kowalik, A. Nowosielski, B. Bazyluk, Do-it-yourself eye tracker: Low-cost pupil-based eye tracker for computer graphics applications, Springer, 2012.

[68] C.-W. Huang, Z.-S. Jiang, W.-F. Kao, Y.-L. Huang, Low-cost and high-speed eye tracker, in: Intelligent Technologies and Engineering Systems, Springer, 2013, pp. 421–427.

[69] H.-F. Ho, Low cost and better accuracy eye tracker, in: Next-Generation Electronics (ISNE), 2014 International Symposium on, IEEE, 2014.

[70] S. A. Johansen, J. San Agustin, H. Skovsgaard, J. P. Hansen, M. Tall, Low cost vs. high-end eye tracking for usability testing, in: CHI'11 Extended Abstracts on Human Factors in Computing Systems, ACM, 2011, pp. 1177–1182.

"It's ridiculous to leave all the conversation to the pudding!"
Alice in Wonderland and through the Looking Glass

Lewis Carroll

6 Conclusion

This chapter introduces the most relevant conclusions drawn after the development of this work. These findings will be introduced following the completion of the different objectives outlined in the Introduction chapter.

This dissertation studied the relation between gaze pattern behaviours and player interaction while using a serious games approach. This project was undertaken to design a profiling framework for serious games development. This dissertation has also evaluated the implications of using gaze pattern behaviours together with interaction data to adapt serious games, customizing game developments. The conclusions obtained after the completion of this work will be detailed throughout this section.

The hypothesis stated in section 1.1, inside the Introduction chapter was:

> *Adaptive multi-variable algorithms based on gaze behaviour and user interaction for tailoring intelligent systems inside the serious games field can be used to profile users′ activity patterns within these systems, enabling the development of intelligent frameworks for the real time customization of the game-play.*

Returning to the hypothesis at the beginning of this study, it is now possible to state that the use of gaze data may constitute a new information source in intelligent therapies, helping to build new approaches that are fully-customized to final users' needs.

This dissertation work has provided a modular framework for processing, analysing and creating tailored user interventions based on serious games. This framework takes into account players' current state and previous history with the system by using artificial intelligence techniques and machine learning algorithms.

The research work behind the design and development of this framework has involved the process and analysis of gaze pattern behaviours recorded from a total of 82 participants aged between 8 and 12 years old.

The completion of the different stages and contributions of this dissertation has been made possible by the fulfilment of the objectives presented in section 1.1.All the five main specific objectives of this dissertation were tackled and fulfilled during the research process.

- *SO1: Define the current state of intelligent serious games* - This objective was successfully completed by the publication of an article that consisted of an analysis of serious games, offering a literary review of their use combined with certain artificial intelligence techniques in the area of *decision making* and *machine learning*. Please refer to 2.3 for the complete article.

- *SO2: Select the different questions that are going to be tackled during the research process* - Through the different contributions to the literature and the research work performed throughout the whole development process, the author of this dissertation tried to address different questions. These questions emerged during the process of determining the suitability of visual attention in the development of customizable serious games based on player profiles. These research questions are summarized and answered below.

– *SO3: Design and implementation of the technique* - The complete design and implementation of the final adaptation framework has been detailed in Chapter 3 - System Design. This framework was supported by the the analysis of the literature and the current state-of-the-art and trends in gaze analysis. It has been designed in a modular fashion that allows its re-implementation and its use in other developments by the research community.

– *SO4: Configuration of the test-bed environment for the exposed technique* - Testing protocol and methodology was carefully selected, designed and carried out in each of the different pilot studies performed during the development of this dissertation. Chapter 4 - Methodology detailed the complete process carried out in each of the performed studies.

– *SO5: Analysis and evaluation of the results* - All the data collected during the pilot studies was analysed and evaluated. The analysis of this data lead the author of this work to the definition of different conclusion and discussion that were introduced in Chapter 5 - Results.

6.1 Research questions

Different questions were addressed during this dissertation process. The different contributions to the literature and the research work performed while the creation of this work led to the proposal of the following questions.

Which is the current state of the art of intelligent serious games?

This question was addressed by the development of a categorization framework for classifying all the relevant articles published during the last decade. This article created a trend analysis about the use of certain artificial intelligence algorithms related to decision making and learning in the field of serious games.

These AI techniques offer significant potential in the development of serious games, enhancing the player experience in all the stages of its gaming experience. As it has been reviewed in one of the article that compound this dissertation, please refer to section 2.3, AI algorithms may help to improve several stages inside the serious games development:

- **Inner working of the game:** Serious games are constantly moving closer to modern games development, following the same pattern with regard to AI techniques. Most modern games addresses three basic game-related needs when implementing intelligent algorithms: move characters, make decisions inside the gameflow, and think tactically [Millington 09].

– **Personalized gaming experience:** The design and development of adaptive intelligent serious games with content changes based on user interaction makes player experience, training and education more customized. AI techniques provide systems with an efficient way of learning based on the users themselves, providing them with customized personal experiences, which may increase their potential effects [Muir 12].

In summary, AI techniques are not only not limited in the context of serious games but also they have a promising potential in the future of this area. The revised literature highlights the potential of intelligent serious games, and the wide range of possibilities they provide to researchers, professionals, and final users. The future of serious games will probably be closer to modern games development involving more AI algorithms, arts and animations, ending with serious games that resemble more to modern video-games, engaging users and game professionals into their path.

What do players think about customized serious game developments?

This question was addressed through the development of a customized adaptive tool based on serious games for evaluating time management skills in teenagers between 12 and 19 years old. A research study was designed and performed for giving a first insight into the use of customizable systems. This was analysed by a user centred design approach, in terms of user experience and usability. Please refer to section 5.1 inside the results chapter for the complete published article.

As a conclusion to this study, it was confirmed that there was a need for new interactive and adaptive content in order to work on different skills in teenagers with and without learning difficulties. One of the final conclusion of the pilot study was to have a development that automatically adapt itself to be suitable for all age-ranges, trying to minimise divergences between them.

Usability results of the evaluated tool gave a positive 78.5 out of 100 possible in the System Usability Scale. This results means a good acceptance of the system, which laid the foundations for the development of customizable serious games in this dissertation.

Which are the potential benefits of mixing biofeedback with serious games?

As it has been stated in the introduction section, the use of video game-related content in areas such as educational therapies and training has risen sharply. Several studies suggested that the future of pedagogy will inevitably be linked to the proposal of combined play and learning, for promoting creativity in future generations [Samuelsson 08]. The boom of serious games bring together the potential available in video games, devoting it fully to the enhancement of specific abilities, skills and aptitudes in children and adults.

Moreover, the design and development of new adaptive serious games whose content changes based on user interaction make therapies, training and education more customized. These techniques provide systems with an efficient way of learning based on the users themselves, providing them with customized and personal experiences, which may increase their potential effects [Tobail 11].

One potential way of doing this is by the combination of performance metrics with objective data recorded from user interaction. Eye movements are a natural information source for proactive systems that analyse user behaviour, where the goal is to infer implicit relevant feedback from gaze [Schwartz 03]. Moreover, following the eye-mind hypothesis put forth by Carpenter in 1980, there is a close link between the direction of the human gaze and the focus of attention [dOrnellas 14], provided that the visual environment in front of the eyes is pertinent to the task that we want to study [Schönauer 11]. Eye tracking sensors collect information about the location and duration of an eye fixation within a specific area on a computer monitor.

Are gaze patterns linked somehow with the degree of expertise of players? Do player with better performance visually interact quantifiably different from individuals with a weaker performance?

This question was addressed through the performance of a user study examining the use of eye tracking sensors as a means to identify children's behaviour in attention-enhancement therapies. For this purpose, a set of data collected from 32 children with different attention skills was analysed during their interaction with a set of puzzle games. The main research hypothesis of this study was that participants with better performance may have quantifiable different eye-movement patterns from users with poorer results.

Although some differences were found during the performance of this study, it is necessary to extend the study or to replicate it, in order to make stronger assumptions. A comparison of the fixation duration data did not produce clear and consistent differences corresponding to the level of performance.

These results corresponded with those related to the expertise level found by R. Tai *et al.* [Tai 06] and Chi *et al.* [Chi 81]. Fixation density was found to decrease with the performance of new levels belonging to the same exercise and with the same exercise outline. These findings agree with R. Tai *et al.*, who found an inverse relationship between the fixation and saccade amount and the participants' degree of expertise [Kliegl 04].

In recent years, the popularity of eye trackers has increased, and there are some open-source projects offering tools for gaze data analysis [Wierda 12, oga 09, pyg 13, ope 11, ope 07], while some manufacturers offer low-cost devices, such as the EyeTribe [eye 11]. There are also several DIY approaches for building custom eye trackers [Mantiuk 12, Huang 13]. The accuracy of these systems may sometimes be slightly inferior to high-end eye trackers, but they may be a viable solution for use outside the laboratory setting [Ho 14]. The use of eye trackers outside the research community may help to extend its potential with available intelligent therapies, bringing state-of-the-art technologies to users.

The use of gaze data constitutes a new information source in intelligent therapies that may help to build new approaches that are completely customized to final users' needs. Further studies need to be carried out in order to establish more detailed attention behaviours and patterns among children with and without attention problems. The replication of this study, along with the extension of the current system with new exercises, may help to build personalized performance profiles per user. These profiles may help in creating new customized therapies, while providing a new degree of information to the children themselves, therapists, psychologists, teachers and family.

Can the combination of gaze patterns and traditional interaction measurements be used into a profile generator engine for adapting the gameflow?

When analysing performance data in Pilot Study II (please refer to section 5.2), some differences were found between the group of best and weaker performance players for the puzzle games in terms of performance interaction and gaze behaviour. When changing the exercise type or level of challenge, users tend to spend more time and perform the exercise with taking more time to think. When tasks are repeated, the ability level increases and the time to complete them drops. This may be related to the acquisition of specific problem-solving skills, which become more accurate with repetition. Further studies need to be carried out about the users' ability and performance capabilities in repetitive tasks. This conclusion led to the definition of the current research question and the design and performance of the final study in this dissertation.

6.2 Social impact

Learners' motivation has been determined by engagement during gameplay, which turns into the development of problem-solving competences [Eseryel 14]. Adaptive levels of challenge in game based learning has been proved to have a positive effect on learning via the increased level of engagement [Hamari 16].

The creation of a modular framework for the automatic generation of player profiles in real time may help to the creation of personalised game-based learning experiences and cognitive intelligent therapies based on serious games.

Intelligent therapies that dynamically adapt themselves to users needs and performance based on their interaction with the system have been proven to be efficient in terms of improvement comparison [Chi 81].

The research work carried out in this dissertation contributes with a new insight into the role visual attention plays in the performance level. It also contributes with the definition of a real time customization framework that may help in the creation of tailored experiences, boosting learning capabilities in serious games.

6.3 Scientific contribution

What follows is the complete relation of the different publications that are part of this research work. This dissertation has been published as a PhD by publication document compounded by 3 published and one under-review article in international journals with impact factor. However, other different contributions have also been made to the scientific community in the shape of communications to different international conferences and a book chapter. All these contributions are summarized below.

6.3.1 Articles in international journals with impact factor

The articles detailed in this section are the ones that compound this PhD by Publication dissertation. Three of them have already been accepted and published in international journals while the last one is currently under review.

Table 6.1 Publication I - International Journal with Impact Factor

Title	Review of the Use of AI Techniques in Serious Games: Decision making and Machine Learning		
Authors	*Maite Frutos-Pascual, Begoña Garcia-Zapirain*		
Journal	IEEE Transactions on Computational Intelligence and AI in Games		
Impact Factor	1.481 *(2014)*	**Quartile**	Q1
Date	25 Dec 2015		
DOI	10.1109/TCIAIG.2015.2512592		

Table 6.2 Publication II - International Journal with Impact Factor

Title	Assessing Visual Attention Using Eye Tracking Sensors in Intelligent Cognitive Therapies Based on Serious Games		
Authors	*Maite Frutos-Pascual, Begoña Garcia-Zapirain*		
Journal	Sensors		
Impact Factor	2.245 *(2014)*	**Quartile**	Q1
Date	12 May 2015		
DOI	10.3390/s150511092		

Table 6.3 Publication III - International Journal with Impact Factor

Title	Adaptive Tele-Therapies Based on Serious Games for Health for People with Time-Management and Organisational Problems: Preliminary Results		
Authors	*Maite Frutos-Pascual, Begoña Garcia-Zapirain, Amaia Méndez Zorrilla*		
Journal	International Journal of Environmental Research and Publich Health		
Impact Factor	2.063 *(2014)*	**Quartile**	Q2
Date	7 Jan 2014		
DOI	10.3390/ijerph110100749		

Table 6.4 Publication IV - International Journal with Impact Factor

Title	Gaze Behavior Analysis in Cognitive Therapies based on Serious Games		
Authors	*Maite Frutos-Pascual, Begoña Garcia-Zapirain*		
Journal	Computers and Education		
Impact Factor	2.556 *(2014)*	**Quartile**	Q1
Date	Under Review		
DOI	–		

6.3.2 Communications in international conferences

Although the articles in this section have not been specifically included as part of this dissertation, all of them were produced while this work was being carried on. All the articles in this section are related to the topic and research field of this dissertation.

Table 6.5 Publication V - Conference

Title	Adaptive cognitive rehabilitation interventions based on serious games for children with ADHD using biofeedback techniques: assessment and evaluation		
Authors	*Maite Frutos-Pascual, Begoña Garcia-Zapirain, Kattalin Camara Buldian*		
Conference	8th International Conference on Pervasive Computing Technologies for Healthcare		
Year	2014	**Location**	Germany
Publisher	ICST		
DOI	10.4108/icst.pervasivehealth.2014.255249		

Table 6.6 Publication VI - Conference

Title	Guided Crossword-Puzzle Games aimed at Children with Attention Deficit: Preliminary results		
Authors	*Maite Frutos-Pascual, Begoña Garcia-Zapirain*		
Conference	Computer Games: AI, Animation, Mobile, Multimedia, Educational and Serious Games (CGAMES)		
Year	2014	**Location**	USA
Publisher	IEEE		
DOI	10.1109/CGames.2014.6934137		

Table 6.7 Publication VII - Conference

Title	Where do they look at? Analysis of gaze interaction in children while playing a puzzle game		
Authors	*Maite Frutos-Pascual, Begoña Garcia-Zapirain*		
Conference	Computer Games: AI, Animation, Mobile, Multimedia, Educational and Serious Games (CGAMES)		
Year	2015	**Location**	USA
Publisher	IEEE		
DOI	10.1109/CGames.2015.7272954		

6.3.3 Book chapters

The following book chapter was also produced during the development of this dissertation, however, it has not been included inside the PhD by publication research work by itself.

Table 6.8 Publication VIII - Book Chapter

Title	Evolution and Use of Serious Games for Health Review and Practical Case about People with ADHD		
Authors	*Maite Frutos-Pascual, Begoña Garcia-Zapirain*		
Book	Advances in Medicines and Biology Volume 69		
Year	2013	**Chapter**	VIII
Publisher	Nova Science Publishers		
ISSN	2157-5398		

6.4 Recommendations for future research

While this study adds to the existing body of literature on the potential of real time profiling of digital game-based learning and cognitive intelligent therapies based on serious games, there are some limitations to the study that should be considered. This section will identify these limitations and provide suggestions for future research to address these limitations. The section will also include suggestions on ways that future research can build upon the findings of this study.

All the 82 participants in the two gaze-related studies were children aged between 8 and 12 years old, all of them residents in the Basque Country, Spain, all of these are factors that can limit the generalizability of the findings. The study setting, which requires of a one-researcher to one-children approach, also limited the number of participants in the study, reducing the statistical power of it. A study similar to the ones presented in this dissertation but with larger sample sizes and including students from other geographical regions of Spain, Europe or other international locations would provide more generalizable results.

The current design of the study, the system and the availability of the children did not allow the performance of a longitudinal study. A longitudinal study analysing the effect sizes of learning and engagement in the long term, with different experimental groups will help to establish the final suitability of adaptive intelligent game based learning experiences.

A third limitation is in the amount of prior knowledge of digital games that students possessed. Prior experience with technology has been associated with the effective use of that technology as a learning tool [ref]. None of the participants were experienced in the presented puzzle game before, but they did have experience playing similar games. Future studies should address this limitation by providing opportunities for students to become familiar with game mechanics.

A fourth limitation is linked to the novelty of the study per se. Sun and Rueda have noted that integrating new technology into the classroom has a positive effect on student engagement [Sun 12]. The novelty of performing this game experience inside their school routine may have helped into the final engagement. Future studies should consider the possibility of this effect and seek to minimize or eliminate it.

Future studies could build upon the findings on this research by incorporating the gaze behaviour knowledge gathered during this dissertation on different serious games developments and subject matters. Lengthening the duration of the study to allow students more time on task could also provide a number of benefits, including reducing the effects of novelty and prior knowledge, allowing the students to play more sophisticated and complex games.

6.5 Concluding remarks

In summary this dissertation covered the whole research process from the definition of the research opportunity and the current state of the art in intelligent serious games, to the analysis of the implications of adaptive serious games from a usability standpoint, and the analysis of the usage of visual attention as a mean for profiling player behaviours while interacting with serious games.

The motivation for conducting this study arises from more than simple curiosity. The potential of serious games and the good they can bring is one of the main areas of interest of the PhD candidate, who fully support the use of serious games, specially in the creation of game based learning experiences. Customizing these experiences to the real time needs of the player has been considered a novel approach for addressing engagement of users, a key point in the learning process. This dissertation has tried to set up the scientific basis for the use of objective data together with performance metrics for the creation of customizable serious games based interventions for cognitive stimulation and learning.

References

[Abd-Almageed 02] Wael Abd-Almageed, M Sami Fadali and George Bebis. *A non-intrusive Kalman filter-based tracker for pursuit eye movement*. In American Control Conference, 2002. Proceedings of the 2002, volume 2, pp. 1443–1447. IEEE, 2002.

[Akbulut 12] Yavuz Akbulut and Cigdem Suzan Cardak. *Adaptive educational hypermedia accommodating learning styles: A content analysis of publications from 2000 to 2011*. Computers & Education, Vol. 58, No. 2, pp. 835–842, 2012.

[Al-Aidroos 12] Naseem Al-Aidroos, Christopher P Said and Nicholas B Turk-Browne. *Top-down attention switches coupling between low-level and high-level areas of human visual cortex*. Proceedings of the National Academy of Sciences, Vol. 109, No. 36, pp. 14675–14680, 2012.

[Almeida 11] Samuel Almeida, Ana Veloso, Licínio Roque and O Mealha. *The eyes and games: A survey of visual attention and eye tracking input in video games*. In Proc. SBGames: X Brazilian Symp. Comput. Games Dig. Entertainment Arts Design Track, 2011.

[Andersen 12] Erik Andersen. *Optimizing adaptivity in educational games*. In Proceedings of the International Conference on the Foundations of Digital Games, pp. 279–281. ACM, 2012.

[Anderson 15] Nicola C Anderson, Eduard Ort, Wouter Kruijne, Martijn Meeter and Mieke Donk. *It depends on when you look at it: Salience influences eye movements in natural scene viewing and search early in time*. Journal of vision, Vol. 15, No. 5, pp. 9–9, 2015.

[Andrá 15] Chiara Andrá, Paulina Lindström, Ferdinando Arzarello, Kenneth Holmqvist, Ornella Robutti and Cristina Sabena. *Reading mathematics representations: An eye-tracking study*. International Journal of Science and Mathematics Education, Vol. 13, No. 2, pp. 237–259, 2015.

[Bahill 84] A Terry Bahill and Tom LaRitz. *Why can't batters keep their eyes on the ball*. American Scientist, Vol. 72, No. 3, pp. 249–253, 1984.

[Ballard 95] Dana H Ballard, Mary M Hayhoe and Jeff B Pelz. *Memory representations in natural tasks*. Cognitive Neuroscience, Journal of, Vol. 7, No. 1, pp. 66–80, 1995.

[Bax 13] Stephen Bax. *The cognitive processing of candidates during reading tests: Evidence from eye-tracking*. Language Testing, Vol. 30, No. 4, pp. 441–465, 2013.

[Becker 69] W Becker and AF Fuchs. *Further properties of the human saccadic system: eye movements and correction saccades with and with-*

out visual fixation points. Vision research, Vol. 9, No. 10, pp. 1247–1258, 1969.

[Bernhard 10] Matthias Bernhard, Efstathios Stavrakis and Michael Wimmer. *An empirical pipeline to derive gaze prediction heuristics for 3D action games*. ACM Transactions on Applied Perception (TAP), Vol. 8, No. 1, p. 4, 2010.

[Bernhard 11] M Bernhard, L Zhang and M Wimmer. *Manipulating attention in computer games*. In Ivmsp workshop, 2011 ieee 10th, pp. 153–158. IEEE, 2011.

[Bisson 14] Marie-Josée Bisson, Walter JB Van Heuven, Kathy Conklin and Richard J Tunney. *Processing of native and foreign language subtitles in films: An eye tracking study*. Applied Psycholinguistics, Vol. 35, No. 02, pp. 399–418, 2014.

[Breiman 96] Leo Breiman. *Bagging predictors*. Machine learning, Vol. 24, No. 2, pp. 123–140, 1996.

[Breiman 01] Leo Breiman. *Random forests*. Machine learning, Vol. 45, No. 1, pp. 5–32, 2001.

[Brooke 96] John Brooke. *SUS-A quick and dirty usability scale*. Usability evaluation in industry, Vol. 189, p. 194, 1996.

[Brouwer 09] Anne-Marie Brouwer, Volker H Franz and Karl R Gegenfurtner. *Differences in fixations between grasping and viewing objects*. Journal of Vision, Vol. 9, No. 1, pp. 18–18, 2009.

[Brusilovsky 04] Peter Brusilovsky. *KnowledgeTree: A distributed architecture for adaptive e-learning*.

In Proceedings of the 13th international World Wide Web conference on Alternate track papers & posters, pp. 104–113. ACM, 2004.

[Bundesen 05] Claus Bundesen, Thomas Habekost and Søren Kyllingsbæk. *A neural theory of visual attention: bridging cognition and neurophysiology.* Psychological review, Vol. 112, No. 2, p. 291, 2005.

[Burgos 07] Daniel Burgos, Colin Tattersall and Rob Koper. *How to represent adaptation in e-learning with IMS learning design*. Interactive Learning Environments, Vol. 15, No. 2, pp. 161–170, 2007.

[Buschman 07] Timothy J Buschman and Earl K Miller. *Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices*. science, Vol. 315, No. 5820, pp. 1860–1862, 2007.

[Buswell 35] Guy Thomas Buswell. How people look at pictures. University of Chicago Press Chicago, 1935.

[Calcaterra 05] Andrea Calcaterra, Alessandro Antonietti and Jean Underwood. *Cognitive style, hypermedia navigation and learning*. Computers & Education, Vol. 44, No. 4, pp. 441–457, 2005.

[Camerini 88] PM Camerini, G Galbiati and F Maffioli. *Algorithms for finding optimum trees: Description, use and evaluation*. Annals of Operations Research, Vol. 13, No. 1, pp. 263–397, 1988.

[Carpenter 77] Patricia Ann Carpenter and Marcel Adam Just. *Reading comprehension as eyes see it*. Cognitive processes in comprehension, pp. 109–139, 1977.

[Carrasco 11] Marisa Carrasco. *Visual attention: The past 25 years*. Vision research, Vol. 51, No. 13, pp. 1484–1525, 2011.

[Chang 13] Wein Chang, Po-An Shen, Kushal Ponnam, Helena Barbosa, Monchu Chen and Sergi Bermudez. *WAYLA: novel gaming experience through unique gaze interaction*. In ACM SIGGRAPH 2013 Emerging Technologies, p. 16. ACM, 2013.

[Chen 08] Fred Chen, Hei Kam, Dejan Markovic, Tsu-Jae King Liu, Vladimir Stojanovic and Elad Alon. *Integrated circuit design with NEM relays*. In Computer-Aided Design, 2008. ICCAD 2008. IEEE/ACM International Conference on, pp. 750–757. IEEE, 2008.

[Chi 81] Michelene TH Chi*et al. Expertise in problem solving.* Technical report, 1981.

[Conners 98] C Keith Conners, Gill Sitarenios, James DA Parker and Jeffery N Epstein. *The revised Conners' Parent Rating Scale (CPRS-R): factor structure, reliability, and criterion validity*. Journal of abnormal child psychology, Vol. 26, No. 4, pp. 257–268, 1998.

[Connor 04] Charles E Connor, Howard E Egeth and Steven Yantis. *Visual attention: bottom-up versus top-down*. Current Biology, Vol. 14, No. 19, pp. R850–R852, 2004.

[Corbetta 02] Maurizio Corbetta and Gordon L Shulman. *Control of goal-directed and stimulus-driven attention in the brain*. Nature reviews neuroscience, Vol. 3, No. 3, pp. 201–215, 2002.

[Crouzet 10] Sebastien M Crouzet, Holle Kirchner and Simon J Thorpe. *Fast saccades toward faces: face detection in just 100 ms*. Journal of vision, Vol. 10, No. 4, pp. 16–16, 2010.

[Den Buurman 81] Rudy Den Buurman, Theo Roersema and Jack F Gerrissen. *Eye movements and the perceptual span in reading*. Reading Research Quarterly, pp. 227–235, 1981.

[Deng 14] Shujie Deng, Julie A Kirkby, Jian Chang and Jian Jun Zhang. *Multimodality with Eye tracking and Haptics: A New Horizon for Serious Games?* International Journal of Serious Games, Vol. 1, No. 4, 2014.

[Desimone 95] Robert Desimone and John Duncan. *Neural mechanisms of selective visual attention*. Annual review of neuroscience, Vol. 18, No. 1, pp. 193–222, 1995.

[Dietterich 02] Thomas G Dietterich. *Ensemble learning*. The handbook of brain theory and neural networks, Vol. 2, pp. 110–125, 2002.

[dOrnellas 14] Marcos Cordeiro dOrnellas, Diego João Cargnin and Ana Lucia Cervi Prado. *Thoroughly Approach to Upper Limb Rehabilitation Using Serious Games for Intensive Group Physical Therapy or Individual Biofeedback Training*. In Computer Games and Digital Entertainment (SBGAMES), 2014 Brazilian Symposium on, pp. 140–147. IEEE, 2014.

[Dorr 07] Michael Dorr, Martin Böhme, Thomas Martinetz and Erhardt Barth. *Gaze beats mouse: a case study*. In Proceedings of the 2nd Conference on Communication by Gaze Interaction, pp. 16–19, 2007.

[Drewes 11] Jan Drewes, Julia Trommershäuser and Karl R Gegenfurtner. *Parallel visual search and rapid animal detection in natural scenes*. Journal of vision, Vol. 11, No. 2, pp. 20–20, 2011.

[Duchowski 07] Andrew Duchowski. Eye tracking methodology: Theory and practice, volume 373. Springer Science & Business Media, 2007.

[Ekman 08] Inger M Ekman, Antti W Poikola and Meeri K Mäkäräinen. *Invisible eni: using gaze and pupil size to control a game*. In CHI'08 Extended Abstracts on Human Factors in Computing Systems, pp. 3135–3140. ACM, 2008.

[El-Nasr 06] Magy Seif El-Nasr and Su Yan. *Visual attention in 3D video games*. In Proceedings of the 2006 ACM SIGCHI international conference on Advances in computer entertainment technology, p. 22. ACM, 2006.

[Epelboim 98] Julie Epelboim. *Gaze and retinal-image-stability in two kinds of sequential looking tasks*. Vision research, Vol. 38, No. 23, pp. 3773–3784, 1998.

[Erkelens 95] Casper J Erkelens and Ingrid MLC Vogels. *The initial direction and landing position of saccades*. Studies in Visual Information Processing, Vol. 6, pp. 133–144, 1995.

[Eseryel 14] Deniz Eseryel, Victor Law, Dirk Ifenthaler, Xun Ge and Raymond Miller. *An Investigation of the Interrelationships between Motivation, Engagement, and Complex Problem Solving in Game-based Learning*. Educational Technology & Society, Vol. 17, No. 1, pp. 42–53, 2014.

[eye 11] *EyeTribe - Ther world's first 99 dollar Eye Tracker*. http://www.theeyetribe.com/, 2011. Accessed: 23-jan-2015.

[Farré 01] Anna Farré and JUAN Narbona. *EDAH. Escalas para la evaluación del trastorno por déficit de atención con hiperactividad*. Madrid: TEA ediciones, 2001.

[Frutos-Pascual 14] Maite Frutos-Pascual, Begoña García Zapirain and Amaia Méndez Zorrilla. *Adaptive tele-therapies based on serious games for health for people with time-management and organisational problems: Preliminary results*. International journal of environmental research and public health, Vol. 11, No. 1, pp. 749–772, 2014.

[Frutos-Pascual 15a] Maite Frutos-Pascual and Begonya Garcia-Zapirain. *Assessing visual attention using eye tracking sensors in intelligent cognitive therapies based on serious games*. Sensors, Vol. 15, No. 5, pp. 11092–11117, 2015.

[Frutos-Pascual 15b] Maite Frutos-Pascual and Begonya Garcia-Zapirain. *Review of the Use of AI Techniques in Serious Games: Decision making and Machine Learning*. IEEE Transactions on Computational Intelligence and AI in Games, Vol. Early Access, 2015.

[García-Ros 12] Rafael García-Ros and Francisco Pérez-González. *Spanish version of the time management behavior questionnaire for university students*. The Spanish journal of psychology, Vol. 15, No. 03, pp. 1485–1494, 2012.

[Gazzaley 12] Adam Gazzaley and Anna C Nobre. *Top-down modulation: bridging selective attention and working memory*. Trends in cognitive sciences, Vol. 16, No. 2, pp. 129–135, 2012.

[Giovinco 15] Nicholas A Giovinco, Steven M Sutton, John D Miller, Timothy M Rankin, Grant W Gonzalez, Bijan Najafi and David G Armstrong. *A Passing Glance? Differences in Eye Tracking and Gaze Patterns Between Trainees and Experts Reading Plain Film Bunion Radiographs*. The Journal of Foot and Ankle Surgery, Vol. 54, No. 3, pp. 382–391, 2015.

[Göbel 10] Stefan Göbel, Viktor Wendel, Christopher Ritter and Ralf Steinmetz. *Personalized, adaptive digital educational games using narrative game-based learning objects*. In Entertainment for Education. Digital Techniques and Systems, pp. 438–445. Springer, 2010.

[Goldberg 95] Joseph H Goldberg and Jack C Schryver. *Eye-gaze-contingent control of the computer in-*

terface: Methodology and example for zoom detection. Behavior research methods, instruments, & computers, Vol. 27, No. 3, pp. 338–350, 1995.

[Greene 10] Jeffrey Alan Greene, Lara-Jeane Costa, Jane Robertson, Yi Pan and Victor M Deekens. *Exploring relations among college students' prior knowledge, implicit theories of intelligence, and self-regulated learning in a hypermedia environment*. Computers & Education, Vol. 55, No. 3, pp. 1027–1043, 2010.

[Grindinger 06] Thomas Grindinger. *Eye movement analysis & prediction with the kalman filter*. PhD thesis, Clemson University, 2006.

[Hamari 16] Juho Hamari, David J Shernoff, Elizabeth Rowe, Brianno Coller, Jodi Asbell-Clarke and Teon Edwards. *Challenging games help students learn: An empirical study on engagement, flow and immersion in game-based learning*. Computers in Human Behavior, Vol. 54, pp. 170–179, 2016.

[Hayhoe 00] Mary Hayhoe. *Vision using routines: A functional account of vision*. Visual Cognition, Vol. 7, No. 1-3, pp. 43–64, 2000.

[Hayhoe 05] Mary Hayhoe and Dana Ballard. *Eye movements in natural behavior*. Trends in cognitive sciences, Vol. 9, No. 4, pp. 188–194, 2005.

[Herst 01] AN Herst, J Epelboim and RM Steinman. *Temporal coordination of the human head and eye during a natural sequential tapping* *task*. Vision research, Vol. 41, No. 25, pp. 3307–3319, 2001.

[Hertzog 08] Melody A Hertzog. *Considerations in determining sample size for pilot studies*. Research in nursing & health, Vol. 31, No. 2, pp. 180–191, 2008.

[Hillaire 08] Sébastien Hillaire, Anatole Lécuyer, Rémi Cozot and Géry Casiez. *Using an eye-tracking system to improve camera motions and depth-of-field blur effects in virtual environments*. In Virtual Reality Conference, 2008. VR'08. IEEE, pp. 47–50. IEEE, 2008.

[Ho 14] Hong-Fa Ho. *Low cost and better accuracy eye tracker*. In Next-Generation Electronics (ISNE), 2014 International Symposium on, pp. 1–2. IEEE, 2014.

[Holmqvist 11] Kenneth Holmqvist, Marcus Nyström, Richard Andersson, Richard Dewhurst, Halszka Jarodzka and Joost Van de Weijer. Eye tracking: A comprehensive guide to methods and measures. OUP Oxford, 2011.

[Huang 13] Chi-Wu Huang, Zong-Sian Jiang, Wei-Fan Kao and Yen-Lin Huang. *Low-Cost and High-Speed Eye Tracker*. In Intelligent Technologies and Engineering Systems, pp. 421–427. Springer, 2013.

[Huey 98] Edmund B Huey. *Preliminary experiments in the physiology and psychology of reading*. The American Journal of Psychology, Vol. 9, No. 4, pp. 575–586, 1898.

[Hyönä 10] Jukka Hyönä. *The use of eye movements in the study of multimedia learning*. Learning and Instruction, Vol. 20, No. 2, pp. 172–176, 2010.

[Ingre 06] Michael Ingre, TorbjÖRn ÅKerstedt, BjÖRn Peters, Anna Anund and GÖRan Kecklund. *Subjective sleepiness, simulated driving performance and blink duration: examining individual differences*. Journal of sleep research, Vol. 15, No. 1, pp. 47–53, 2006.

[isf 15] *GameTrack Digest: Quarter 4 2015*. Interactive Software Industry, Vol. 4, 2015.

[Isokoski 09] Poika Isokoski, Markus Joos, Oleg Spakov and Benoît Martin. *Gaze controlled games*. Universal Access in the Information Society, Vol. 8, No. 4, pp. 323–337, 2009.

[Itti 00] Laurent Itti and Christof Koch. *A saliency-based search mechanism for overt and covert shifts of visual attention*. Vision research, Vol. 40, No. 10, pp. 1489–1506, 2000.

[Itti 01] Laurent Itti and Christof Koch. *Computational modelling of visual attention*. Nature reviews neuroscience, Vol. 2, No. 3, pp. 194–203, 2001.

[James 13] William James. The principles of psychology. Read Books Ltd, 2013.

[Johansen 08] Sune Alstrup Johansen, Mie Noergaard and Janus Rau. *Can eye tracking boost usability evaluation of computer games*. In Proceedings of CHI, volume 8, 2008.

[Johansson 01] Roland S Johansson, Göran Westling, Anders Bäckström and J Randall Flanagan. *Eye–hand coordination in object manipulation*. the Journal of Neuroscience, Vol. 21, No. 17, pp. 6917–6932, 2001.

[Józsa 11] E Józsa and BP Hámornik. *Find the difference!: eye tracking study on information seeking behavior using an online game*. 2011.

[Just 80] Marcel A Just and Patricia A Carpenter. *A theory of reading: from eye fixations to comprehension.* Psychological review, Vol. 87, No. 4, p. 329, 1980.

[Kickmeier Rust 11] Michael D Kickmeier Rust, Eva Hillemann and Dietrich Albert. *Tracking the UFO's Paths: Using Eye-Tracking for the Evaluation of Serious Games*. In Virtual and Mixed Reality-New Trends, pp. 315–324. Springer, 2011.

[Kienzle 09] Wolf Kienzle, Matthias O Franz, Bernhard Schölkopf and Felix A Wichmann. *Center-surround patterns emerge as optimal predictors for human saccade targets*. Journal of vision, Vol. 9, No. 5, pp. 7–7, 2009.

[Kiili 14] Kristian Kiili, Harri Ketamo and Michael D Kickmeier-Rust. *Eye Tracking in Game-based Learning Research and Game Design*. International Journal of Serious Games, Vol. 1, No. 2, pp. 51–65, apr 2014.

[Kliegl 04] Reinhold Kliegl, Ellen Grabner, Martin Rolfs and Ralf Engbert. *Length, frequency, and predictability effects of words on eye movements in reading*. European Journal of Cognitive

Psychology, Vol. 16, No. 1-2, pp. 262–284, 2004.

[Koh 10] Do Hyong Koh, Sandeep Munikrishne Gowda and Oleg V Komogortsev. *Real time eye movement identification protocol*. In CHI'10 Extended Abstracts on Human Factors in Computing Systems, pp. 3499–3504. ACM, 2010.

[Komogortsev 07] Oleg V Komogortsev and Javed I Khan. *Kalman filtering in the design of eye-gaze-guided computer interfaces*. In Human-Computer Interaction. HCI Intelligent Multimodal Interaction Environments, pp. 679–689. Springer, 2007.

[Komogortsev 10] Oleg V Komogortsev, Denise V Gobert, Sampath Jayarathna, Do Hyong Koh and Sandeep M Gowda. *Standardization of automated analyses of oculomotor fixation and saccadic behaviors*. Biomedical Engineering, IEEE Transactions on, Vol. 57, No. 11, pp. 2635–2645, 2010.

[Kruger 14] Jan-Louis Kruger and Faans Steyn. *Subtitles and eye tracking: Reading and performance*. Reading Research Quarterly, Vol. 49, No. 1, pp. 105–120, 2014.

[Kunze 13a] Kai Kunze, Hitoshi Kawaichi, Kazuyo Yoshimura and Koichi Kise. *Towards inferring language expertise using eye tracking*. In CHI'13 Extended Abstracts on Human Factors in Computing Systems, pp. 217–222. ACM, 2013.

[Kunze 13b] Kai Kunze, Yuzuko Utsumi, Yuki Shiga, Koichi Kise and Andreas Bulling. *I know what you are reading: recognition of document types using mobile eye tracking*. In Proceedings of the 2013 International Symposium on Wearable Computers, pp. 113–116. ACM, 2013.

[Lai 13] Meng-Lung Lai, Meng-Jung Tsai, Fang-Ying Yang, Chung-Yuan Hsu, Tzu-Chien Liu, Silvia Wen-Yu Lee, Min-Hsien Lee, Guo-Li Chiou, Jyh-Chong Liang and Chin-Chung Tsai. *A review of using eye-tracking technology in exploring learning from 2000 to 2012*. Educational Research Review, Vol. 10, pp. 90–115, 2013.

[Land 99] Michael Land, Neil Mennie and Jennifer Rusted. *The roles of vision and eye movements in the control of activities of daily living*. Perception, Vol. 28, No. 11, pp. 1311–1328, 1999.

[Land 00] Michael F Land and Peter McLeod. *From eye movements to actions: how batsmen hit the ball*. Nature neuroscience, Vol. 3, No. 12, pp. 1340–1345, 2000.

[Land 09] Michael Land and Benjamin Tatler. Looking and acting: vision and eye movements in natural behaviour. Oxford University Press, 2009.

[Law 12] Effie Lai-Chong Law and Xu Sun. *Evaluating user experience of adaptive digital educational games with Activity Theory*. International Journal of Human-Computer Studies, Vol. 70, No. 7, pp. 478–497, 2012.

[Li 14] Songpo Li and Xiaoli Zhang. *Eye-movement-based objective real-time quantification of patient's mental engagement in rehabilitation: A preliminary study*. In Mechatronics and Automation (ICMA), 2014 IEEE International Conference on, pp. 180–185. IEEE, 2014.

[Liang 08] Yulan Liang and John D Lee. *Driver cognitive distraction detection using eye movements*. In Passive Eye Monitoring, pp. 285–300. Springer, 2008.

[Lin 04] Chern-Sheng Lin, Chia-Chin Huan, Chao-Ning Chan, Mau-Shiun Yeh and Chuang-Chien Chiu. *Design of a computer game using an eye-tracking device for eye's activity rehabilitation*. Optics and Lasers in Engineering, Vol. 42, No. 1, pp. 91–108, 2004.

[Lopes 11] Ricardo Lopes and Rafael Bidarra. *Adaptivity challenges in games and simulations: a survey*. Computational Intelligence and AI in Games, IEEE Transactions on, Vol. 3, No. 2, pp. 85–99, 2011.

[Macan 90] Therese H Macan, Comila Shahani, Robert L Dipboye and Amanda P Phillips. *College students' time management: Correlations with academic performance and stress*. Journal of educational psychology, Vol. 82, No. 4, p. 760, 1990.

[Macan 94] Therese Hoff Macan. *Time management: Test of a process model*. Journal of applied psychology, Vol. 79, No. 3, p. 381, 1994.

[Macan 96] Therese Hoff Macan. *Time-management training: Effects on time behaviors, attitudes,* *and job performance*. The Journal of psychology, Vol. 130, No. 3, pp. 229–236, 1996.

[Mack 98] Arien Mack and Irvin Rock. Inattentional blindness, volume 33. MIT press Cambridge, MA, 1998.

[Mampadi 11] Freddy Mampadi, Sherry Y Chen, Gheorghita Ghinea and Ming-Puu Chen. *Design of adaptive hypermedia learning systems: A cognitive style approach*. Computers & Education, Vol. 56, No. 4, pp. 1003–1011, 2011.

[Mantiuk 12] Radosław Mantiuk, Michał Kowalik, Adam Nowosielski and Bartosz Bazyluk. Do-it-yourself eye tracker: Low-cost pupil-based eye tracker for computer graphics applications. Springer, 2012.

[Matin 74] Ethel Matin. *Saccadic suppression: a review and an analysis*. Psychological bulletin, Vol. 81, No. 12, p. 899, 1974.

[Mayer 10] Richard E Mayer. *Unique contributions of eye-tracking research to the study of learning with graphics*. Learning and instruction, Vol. 20, No. 2, pp. 167–171, 2010.

[McConkie 85] George W McConkie, N Roderick Underwood, David Zola and GS Wolverton. *Some temporal characteristics of processing during reading*. Journal of Experimental Psychology: Human Perception and Performance, Vol. 11, No. 2, p. 168, 1985.

[Millington 09] Ian Millington and John David Funge. Artificial intelligence for games. Taylor & Francis US, 2009.

[Muir 12] Mary Muir and Cristina Conati. *An analysis of attention to student–adaptive hints in an educational game*. In Intelligent Tutoring Systems, pp. 112–122. Springer, 2012.

[Nacke 10] Lennart E Nacke, Sophie Stellmach, Dennis Sasse and Craig A Lindley. *Gameplay experience in a gaze interaction game*. arXiv preprint arXiv:1004.0259, 2010.

[Najemnik 05] Jiri Najemnik and Wilson S Geisler. *Optimal eye movement strategies in visual search*. Nature, Vol. 434, No. 7031, pp. 387–391, 2005.

[Nakatani 08] Chie Nakatani and Cees van Leeuwen. *A pragmatic approach to multi-modality and non-normality in fixation duration studies of cognitive processes*. Journal of Eye Movement Research, 1 (2), Vol. 1, pp. 1–12, 2008.

[Nuthmann 10] Antje Nuthmann and John M Henderson. *Object-based attentional selection in scene viewing*. Journal of vision, Vol. 10, No. 8, pp. 20–20, 2010.

[oga 09] *OGAMA - Open Gaze and Mouse Analyzer*. http://www.ogama.net/, 2009. Accessed: 23-jan-2015.

[Olsen 12a] Anneli Olsen. *The tobii i-vt fixation filter*. Tobii Technology, 2012.

[Olsen 12b] Anneli Olsen and Ricardo Matos. *Identifying parameter values for an I-VT fixation filter suitable for handling data sampled with various sampling frequencies*. In Proceedings of the Symposium on Eye Tracking Research and Applications, pp. 317–320. ACM, 2012.

[ope 07] *OpenGazer - Open-source gaze tracker for ordinary webcams*. http://www.inference.phy.cam.ac.uk/opengazer/, 2007. Accessed: 23-jan-2015.

[ope 11] *OpenEyes - Open-source open-hardware toolkit for low-cost real-time eye tracking*. http://thirtysixthspan.com/openEyes/, 2011. Accessed: 23-jan-2015.

[Pantelis 16] Peter C Pantelis and Daniel P Kennedy. *Prior expectations about where other people are likely to direct their attention systematically influence gaze perception*. Journal of vision, Vol. 16, No. 3, pp. 7–7, 2016.

[Park 03] Ok-choon Park and Jung Lee. *Adaptive instructional systems*. Educational Technology Research and Development, Vol. 25, pp. 651–684, 2003.

[Paterno 99] Fabio Paterno and Cristiano Mancini. *Engineering the design of usable hypermedia*. Empirical Software Engineering, Vol. 4, No. 1, pp. 11–42, 1999.

[Plass 15] Jan L Plass, Bruce D Homer and Charles K Kinzer. *Foundations of game-based learning*. Educational Psychologist, Vol. 50, No. 4, pp. 258–283, 2015.

[Pretorius 10] Marco Pretorius, Helene Gelderblom and Bester Chimbo. *Using eye tracking to compare how adults and children learn to use an unfamiliar computer game*. In Proceedings of the 2010 Annual Research Conference of the South African Institute of Computer

Scientists and Information Technologists, pp. 275–283. ACM, 2010.

[PuJoL 13] LYdIa PuJoL. *Manejo del tiempo académico en jóvenes que inician estudios en la Universidad Simón Bolívar*. Revista Latinoamericana de Ciencias Sociales, Niñez y Juventud, Vol. 11, No. 1, 2013.

[Purcell 12] Braden A Purcell, Jeffrey D Schall, Gordon D Logan and Thomas J Palmeri. *From salience to saccades: multiple-alternative gated stochastic accumulator model of visual search*. The Journal of Neuroscience, Vol. 32, No. 10, pp. 3433–3446, 2012.

[pyg 13] *PyGaze - The open-source toolbox for eye tracking*. http://www.pygaze.org/, 2013. Accessed: 23-jan-2015.

[Ramirez-Moreno 13] David F Ramirez-Moreno, Odelia Schwartz and Juan F Ramirez-Villegas. *A saliency-based bottom-up visual attention model for dynamic scenes analysis*. Biological cybernetics, Vol. 107, No. 2, pp. 141–160, 2013.

[Rayner 78] Keith Rayner. *Eye movements in reading and information processing*. Psychological bulletin, Vol. 85, No. 3, p. 618, 1978.

[Rayner 98] Keith Rayner. *Eye movements in reading and information processing: 20 years of research*. Psychological bulletin, Vol. 124, No. 3, p. 372, 1998.

[Rayner 09] Keith Rayner. *Eye movements and attention in reading, scene perception, and visual search*. The quarterly journal of experimental psychology, Vol. 62, No. 8, pp. 1457–1506, 2009.

[Rodrigue 15] Mathieu Rodrigue, Jungah Son, Barry Giesbrecht, Matthew Turk and Tobias Höllerer. *Spatio-Temporal Detection of Divided Attention in Reading Applications Using EEG and Eye Tracking*. In Proceedings of the 20th International Conference on Intelligent User Interfaces, pp. 121–125. ACM, 2015.

[Rodriguez-Galiano 12] VF Rodriguez-Galiano, B Ghimire, J Rogan, M Chica-Olmo and JP Rigol-Sanchez. *An assessment of the effectiveness of a random forest classifier for land-cover classification*. ISPRS Journal of Photogrammetry and Remote Sensing, Vol. 67, pp. 93–104, 2012.

[Salvucci 98] Dario D Salvucci and John R Anderson. *Tracing eye movement protocols with cognitive process models*. 1998.

[Salvucci 00] Dario D Salvucci and Joseph H Goldberg. *Identifying fixations and saccades in eye-tracking protocols*. In Proceedings of the 2000 symposium on Eye tracking research & applications, pp. 71–78. ACM, 2000.

[Samuelsson 08] Ingrid Pramling Samuelsson and Maj Asplund Carlsson. *The playing learning child: Towards a pedagogy of early childhood*. Scandinavian Journal of Educational Research, Vol. 52, No. 6, pp. 623–641, 2008.

[San Agustin 10] Javier San Agustin. *Off-the-shelf gaze interaction*. 2010.

[Schapire 98] Robert E Schapire, Yoav Freund, Peter Bartlett and Wee Sun Lee. *Boosting the margin: A new explanation for the effectiveness of voting methods*. Annals of statistics, pp. 1651–1686, 1998.

[Schönauer 11] Christian Schönauer, Thomas Pintaric and Hannes Kaufmann. *Full body interaction for serious games in motor rehabilitation*. In Proceedings of the 2nd Augmented Human International Conference, p. 4. ACM, 2011.

[Schütz 11] Alexander C Schütz, Doris I Braun and Karl R Gegenfurtner. *Eye movements and perception: A selective review*. Journal of vision, Vol. 11, No. 5, pp. 9–9, 2011.

[Schwartz 03] Mark S Schwartz and Frank Ed Andrasik. Biofeedback: A practitioner's guide . Guilford Press, 2003.

[Sen 84] Tayyar Sen and Ted Megaw. *The effects of task variables and prolonged performance on saccadic eye movement parameters*. Advances in Psychology, Vol. 22, pp. 103–111, 1984.

[Sennersten 10] Charlotte C Sennersten and Craig A Lindley. *Real time eye gaze logging in a 3D game/simulation world*. Key Engineering Materials, Vol. 437, pp. 555–559, 2010.

[Shebilske 83] WL Shebilske and DF Fisher. *Understanding extended discourse through the eyes: How and why*, 1983.

[Siebold 13] Alisha Siebold, Wieske van Zoest, Martijn Meeter and Mieke Donk. *In defense of the* *salience map: Salience rather than visibility determines selection.* Journal of Experimental Psychology: Human Perception and Performance, Vol. 39, No. 6, p. 1516, 2013.

[Smith 06] J David Smith and TC Graham. *Use of eye movements for video game control*. In Proceedings of the 2006 ACM SIGCHI international conference on Advances in computer entertainment technology, p. 20. ACM, 2006.

[Stark 81] Lawrence W Stark and Stephen R Ellis. *Scanpaths revisited: Cognitive models direct active looking*. 1981.

[Sun 12] Jerry Chih-Yuan Sun and Robert Rueda. *Situational interest, computer self-efficacy and self-regulation: Their impact on student engagement in distance education*. British Journal of Educational Technology, Vol. 43, No. 2, pp. 191–204, 2012.

[Sundstedt 08] Veronica Sundstedt, Efstathios Stavrakis, Michael Wimmer and Erik Reinhard. *A psychophysical study of fixation behavior in a computer game*. In Proceedings of the 5th symposium on Applied perception in graphics and visualization, pp. 43–50. ACM, 2008.

[Sundstedt 13] Veronica Sundstedt, Matthias Bernhard, Efstathios Stavrakis, Erik Reinhard and Michael Wimmer. *Visual attention and gaze behavior in games: An object-based approach*. In Game analytics, pp. 543–583. Springer, 2013.

[Tai 06] Robert H Tai, John F Loehr and Frederick J Brigham. *An exploration of the use of eye-gaze tracking to study problem-solving on* *standardized science assessments*. International journal of research & method in education, Vol. 29, No. 2, pp. 185–208, 2006.

[Theeuwes 10] Jan Theeuwes. *Top–down and bottom–up control of visual selection*. Acta psychologica, Vol. 135, No. 2, pp. 77–99, 2010.

[Thorpe 96] Simon Thorpe, Denis Fize, Catherine Marlot*et al. Speed of processing in the human visual system*. nature, Vol. 381, No. 6582, pp. 520–522, 1996.

[Tobail 11] Ayman Tobail, John Crowe and Amr Arisha. *Learning by gaming: supply chain application*. In Simulation Conference (WSC), Proceedings of the 2011 Winter, pp. 3935–3946. IEEE, 2011.

[Tsai 12] Meng-Jung Tsai, Huei-Tse Hou, Meng-Lung Lai, Wan-Yi Liu and Fang-Ying Yang. *Visual attention for solving multiple-choice science problem: An eye-tracking analysis*. Computers & Education, Vol. 58, No. 1, pp. 375–385, 2012.

[Turner 14] Jayson Turner, Eduardo Velloso, Hans Gellersen and Veronica Sundstedt. *EyePlay: applications for gaze in games*. In Proceedings of the first ACM SIGCHI annual symposium on Computer-human interaction in play, pp. 465–468. ACM, 2014.

[Ungerleider 00] Sabine Kastner Ungerleider and Leslie G. *Mechanisms of visual attention in the human cortex*. Annual review of neuroscience, Vol. 23, No. 1, pp. 315–341, 2000.

[Vickers 13] Stephen Vickers, Howell Istance and Michael J Heron. *Accessible gaming for people with physical and cognitive disabilities: a framework for dynamic adaptation*. In CHI'13 Extended Abstracts on Human Factors in Computing Systems, pp. 19–24. ACM, 2013.

[Vidal 12] Mélodie Vidal, Andreas Bulling and Hans Gellersen. *Detection of smooth pursuits using eye movement shape features*. In Proceedings of the symposium on eye tracking research and applications, pp. 177–180. ACM, 2012.

[Volkmann 80] Frances C Volkmann, Lorrin A Riggs and Robert K Moore. *Eyeblinks and visual suppression*. Science, Vol. 207, No. 4433, pp. 900–902, 1980.

[Walber 12] Tina Walber, Chantal Neuhaus and Ansgar Scherp. *EyeGrab: A Gaze-based Game with a Purpose to Enrich Image Context Information.* In EuroHCIR, pp. 63–66. Citeseer, 2012.

[Westheimer 54] Gerald Westheimer. *Mechanism of saccadic eye movements*. AMA Archives of Ophthalmology, Vol. 52, No. 5, pp. 710–724, 1954.

[Widdel 84] Heino Widdel. *Operational problems in analysing eye movements*. Advances in Psychology, Vol. 22, pp. 21–29, 1984.

[Wierda 12] Stefan M Wierda, Hedderik van Rijn, Niels A Taatgen and Sander Martens. *Pupil dilation deconvolution reveals the dynamics of attention at high temporal resolution*. Proceedings of the National Academy of Sciences, Vol. 109, No. 22, pp. 8456–8460, 2012.

[Witzel 12] Naoko Witzel, Jeffrey Witzel and Kenneth Forster. *Comparisons of online reading paradigms: Eye tracking, moving-window, and maze*. Journal of Psycholinguistic Research, Vol. 41, No. 2, pp. 105–128, 2012.

[Zain 11] Nurul Hidayah Mat Zain, Fariza Hanis Abdul Razak, Azizah Jaafar and Mohd Firdaus Zulkipli. *Eye tracking in educational games environment: evaluating user interface design through eye tracking patterns*. In Visual Informatics: Sustaining Research and Innovations, pp. 64–73. Springer, 2011.