Karlsruhe Institute of Technology

Fraunhofer IOSB

Analysis of User Attention for Adaptive Serious Games - Design and Implementation of an Evaluation Framework

MASTER'S THESIS

KIT - KARLSRUHE INSTITUTE OF TECHNOLOGY
FRAUNHOFER IOSB - FRAUNHOFER INSTITUTE OF OPTRONICS, SYSTEM TECHNOLOGIES AND IMAGE EXPLOITATION

Sebastian Leidig

31.10.2016

Main Advisor:	Prof. Dr.-Ing. Jürgen Beyerer
Co-Advisor:	Dipl.-Inf. Alexander Streicher

Interaktive Echtzeitsysteme

Abstract

The "Ideal Path Score" developed in this work is able to improve adaptivity of serious games by more accurately estimating performance and need for help based on players' interactions and eye movements.

The automatic personalization of adaptive e-learning systems supports effective learning for users with varying levels of knowledge and skills. Particularly in games, indicators informing adaptivity, like attention and performance of the player, should be assessed non-invasively to avoid interrupting the player's flow experience. Passive sensors like eye tracking can solve this challenge.

This work examines whether correlations between eye tracking data and attention exist and how eye tracking can improve adaptivity decisions, e.g. when and what should be adapted. To achieve this, an "Ideal Path Score" was developed which assesses performance and attention of players based on eye movements and interactions with the game.

This approach was evaluated through a user study with n=20 participants. The developed approach improves adaptivity in the examined game SaFIRa. Pearson's correlation coefficient between modeled skill level and the players subjective rating of performance increased from 0.19 to 0.30. 46 % of hints needed by the player were displayed automatically by the improved adaptive system (16 % in the original system). A correlation between "Ideal Path Score" and self-assessed attention could not be shown.

Zusammenfassung

Die in der Arbeit entwickelte "Ideal Path Score", basierend auf Spieleraktionen und Eye-Tracking, kann die Adaptivität bei Serious Games und Computersimulationen durch genauere Einschätzung der Leistung und Unterstützungsbedürftigkeit verbessern.

Adaptive E-Learning Systeme und Serious Games ermöglichen Nutzern mit unterschiedlichem Vorwissen und Fähigkeiten durch automatische Personalisierung effektiveres Lernen. Adaptivitätsindikatoren, wie Aufmerksamkeit und Leistung des Nutzers, sollten insbesondere bei Spielen nicht-invasiv vom System gemessen werden, um den "Flow" des Nutzers nicht zu unterbrechen. Passive Sensoren wie etwa Eye-Tracking bieten sich hierfür an.

Diese Arbeit untersucht welche Korrelationen zwischen Eye-Tracking Daten und Aufmerksamkeit bestehen und wie Eye-Tracking Adaptivitätsentscheidungen verbessern kann, etwa wann und was adaptiert werden soll. Dazu wurde eine "Ideal Path Score" entwickelt, die Leistung und Aufmerksamkeit des Spielers basierend auf dessen Augenbewegungen und Interaktionen modelliert.

Im Rahmen eines Experiments mit n=20 Teilnehmern wurde dieser Ansatz evaluiert. Der entwickelte Ansatz verbessert die Adaptivität des untersuchten Spiels SaFIRa. Der Pearson-Korrelationskoeffizient zwischen modellierter Fähigkeit und vom Spieler selbsteingeschätzter Leistung steigt von 0.19 auf 0.30. Die vom Spieler benötigten Hinweise werden in 46 % der Fälle automatisch vom System angezeigt (16 % im ursprünglichen System). Eine Korrelation zwischen "Ideal Path Score" und selbsteingeschätzter Aufmerksamkeit konnte nicht festgestellt werden.

Statement of authorship

I hereby declare that I have produced this work by myself except the utilities known to the supervisor, that I have labeled all used utilities completely and detailed and that I have labeled all material that has been taken with or without modification from the work of others.

Karlsruhe, 28.10.2016 Sebastian Leidig

Contents

1

Introduction

Skills and prior knowledge of learners can vary to a large extent. While traditional teaching was often bound to a “one size fits all” approach due to practical reasons, digital tools provide new opportunities through adaptive, personalized e-learning. By observing and modeling the learner automatically, an adaptive e-learning system can select the appropriate next steps for that individual learner, thereby enabling a more effective and more motivating learning experience [Str16]. This adaptivity is most often based on measuring the abilities of the learner either implicitly or explicitly.

Serious games designed for learning should not only adaptively guide the learner to the next learning goal but in addition also sustain a flow experience for the user [Che07]. This flow, a balance of challenges and skills, can keep the learner self-motivated and is an important aspect of effective serious games. Each individual player’s flow experience is different depending on his abilities, therefore adaptivity of serious games towards the player’s flow and attention is a promising approach.

The vision pursued in this thesis is a system estimating players’ attention and performance based on eye tracking in order to improve their flow and learning through adaquate adaptivity (figure 1.1).

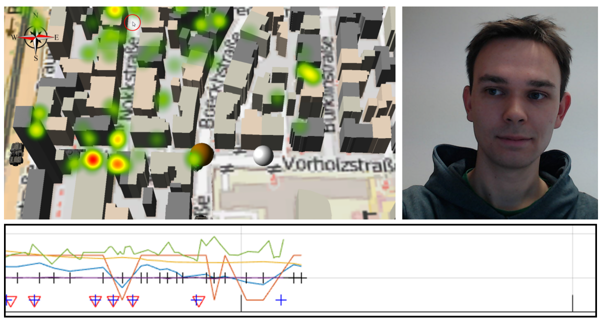


Figure 1.1: In this thesis a system estimating user attention and performance based on eye tracking is developed in order to improve adaptivity of serious games.

Motivation

For effective adaptivity, ideally the adaptive interventions would be guided by a measure of the user’s progress and the purposefulness or goal-orientedness of his actions. A user working efficiently towards the goal does not need further assistance while a user who is lost or moving in a wrong direction should be adaptively assisted.

An approach to measure goal-orientedness is the definition of a metric to measure the distance between an "ideal path" through the game and a user’s observed actions. The ideal path describes all necessary steps to reach the goal without any unnecessary detours. An expert user would follow these steps to solve the game in a straightforward way. Another user may be less experienced or less attentive and divert from this ideal path to some extent.

To accurately judge a player based on an ideal path metric, the user's explicit interactions (e.g. through mouse and keyboard) are not sufficient input. The user may well have looked at the "correct" game object and considered it for a longer time before finally deciding to interact with the wrong object. For questions of evaluation or adaptivity this should be considered differently from a user who did not show any attention towards the relevant next action.

The Attention Adaptivity Framework developed in this thesis integrates data collected through eye tracking and other sensors into the existing e-learning adaptivity system ELAI. This intends to improve existing adaptivity recommendations and to provide additional adaptivity to serious games in order to sustain the learner's game progress and attention.

Using an eye tracker to record eye movements provides a non-invasive way to collect input for adaptivity decisions. This is particularly important in the context of games, as many alternatives like explicitly testing the user through a quiz or questionnaire or offering manual settings for adaptation interrupt gameplay and likely break the user's flow experience. Cognitive science, psychology and usability research have long been studying eye movements as an indicator of cognitive processes. The developed approach builds on these findings, but the Attention Adaptivity Framework does not solely rely on their full accuracy because interpretation of the eye movements is one among several factors, like the user's explicit interactions through mouse and keyboard and the ideal path through the game.

Integrating the Attention Adaptivity Framework into the existing ELAI framework [Str15c] allows the combination of existing didactic factors of ELAI with the factors developed in this work. This should allow for better grounded reasoning towards adaptive actions and avoids reinventing the wheel. Consolidating the adaptivity measures into the ELAI system also eases implementation of attention based adaptivity into other applications as only one framework needs to be understood by the developers to integrate comprehensive adaptivity features.

Background

The department for "Interoperability and Assistance Systems" (IAS) at the Fraunhofer Institute of Optronics, System Technologies and Image Exploitation (Fraunhofer IOSB) researches and develops solutions in the area of computer-supported assistance systems, information management, cooperative work and decision-making processes using interactive technology.

One focus is image interpretation for reconnaissance, i.e. the identification of structures and objects by image interpretation experts on different image types like optical, radar or infra-red images [Rol13]. To improve training of these image interpreters, approaches using simulations and serious games are explored.

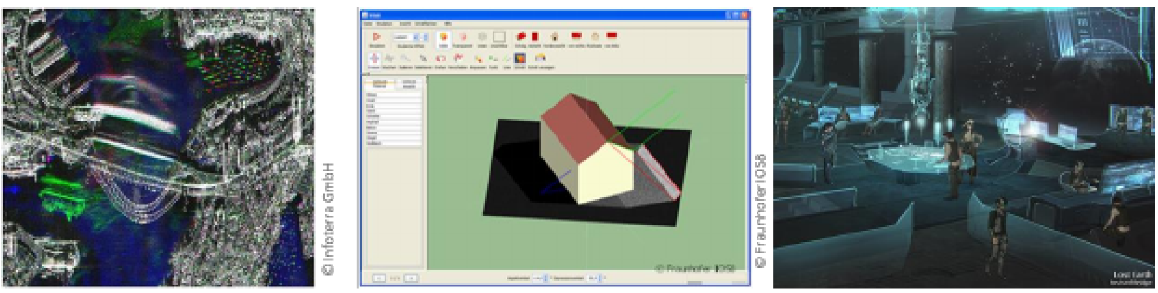


Figure 1.2: Adaptive serious games and simulations for image reconnaissance, image source [Str15a].

Against this background, the "E-Learning A.I." (ELAI) was developed as a framework for intelligent tutoring components with a special focus on interoperability [Str15c]. Its aim is to adapt serious games and simulations "to the experience and knowledge level of the users to keep them in the flow channel" [Str15c]. Interoperability is achieved by separating the components for intelligent adaptivity from the serious game itself and deploying it as a separate server with whom multiple different systems can communicate using standards like xAPI. The ELAI system is described in more detail in chapter 3.5.1. A "proof-of-concept" prototype of the system was implemented for the Seek and Find game "SaFIRa" [Bie16] (see chapter 3.6). My thesis builds on that protype and extends its architecture and functionality towards better extensibility and integration of eye tracking for learner attention estimation.

Objectives and Scope

This thesis deals with the utilization of sensor data from an eye tracking device for non-invasive modeling of user attention and consequently its opportunities for e-learning adaptivity. This focus fits into the overall goal of improving adaptivity for serious games regarding users' flow experience as well as learning outcomes.

The objectives of this work are (a) the design of eye tracking based metrics for learner modeling; (b) the implementation of adaptivity based on these metrics; (c) the development of a framework for easy integration of metrics into adaptive e-learning systems; and (d) the evaluation of these approaches through a user study.

Eye tracking based metrics are to be developed to estimate "goal-orientedness" and attention of users. The focus in this regard is to collect data and investigate correlations between eye movement patterns, user interactions and the learner's (self-reported) attention. The developed metrics are evaluated through a user study.

The overarching question is how eye tracking data can be represented by metrics that provide useful insights into such cognitive states of the learner like attention. This could support decisions on when and what an adaptive e-learning system like ELAI should adapt.

Data from other sensors is collected during the experiments but its analysis is not part of this thesis. Also, basic research, e.g. on eye tracking techniques, is out of its scope.

Adaptivity based on those metrics is to be implemented, improving control of existing adaptivity interventions. However, the implementation serves as a proof of concept and is not the primary focus of the experimental evaluation. Design of new types of adaptive interventions is not subject of this thesis.

An Attention Adaptivity Framework is to be implemented that allows integration of eye tracking into adaptive e-learning applications. The framework should be easily extensible to integrate different sensors into the existing ELAI system. It needs to be platform and game engine independent or easily adaptable using the xAPI standard. The development setup needs to be portable to other machines and the framework has to be configurable through configuration files.

The hypothesis to be studied is that metrics based on eye tracking provide useful information to measure goal-orientedness and attention of users. Furthermore I assume that such measures allow improved adaptivity towards sustained game progress and player attention in serious games.

Solution Approach

After implementation of a framework that allows collection and interpretation of sensor data I adapt the existing serious game "SaFIRa" to use this framework. The designed eye tracking based metrics are evaluated and tuned to this game and improve its existing adaptivity.

The developed Attention Adaptivity Framework enables collection of sensor data, its interpretation and integration into the ELAI system for e-learning adaptivity. The architecture is highly modular. It is used for implementation of the designed metrics and can serve as a foundation for further work to leverage information from eye tracking sensors as well as other sources.

The existing serious game "SaFIRa" (described in chapter 3.6) serves as the basis for a prototypical integration of the framework. This system is then used in a user study to collect various data, particularly from eye tracking, and evaluate the proposed metrics. The data is analysed to evaluate the proposed metrics and derive additional correlations. The primary metric designed, an "Ideal Path Score", combines information from user actions and eye movements to generate deeper insights on the attention and goal-orientedness of the user. I also evaluate indicators to infer user's cognitive states that were proposed in related work. Particularly measures that allow general, interface independent insights on high-level cognitive processes could present good complements to the Ideal Path Score.

The base of this concept is an ideal path as a sequence of ideal actions that serves as a gold standard. A suitable metric then defines distance between this ideal path and an actual user's path. This distance can be used to quantify a user's performance and I term it "Ideal Path Score". Details of the "Ideal Path" model as well as the score definition are described in chapter 4.1.

Outline

Chapter 2 summarizes the state of the art regarding adaptive e-learning, eye tracking for modeling cognitive states and eye tracking for adaptivity in e-learning. The theoretical foundations for the thesis are presented in chapter 3. Chapter 4 describes the developed "Ideal Path Score" to estimate user attention and goal-orientedness in detail and discusses the rationale of that approach. Chapter 5 presents the developed framework that enables integration of sensor data like eye tracking into the adaptive cycle of the ELAI system. The developed approach is evaluated in a user study. The experimental setup as well as detailed results regarding correlations with user attention and improvements of adaptivity decisions are presented in chapter 6. Finally, chapter 7 gives a conclusion and an outlook on future work.

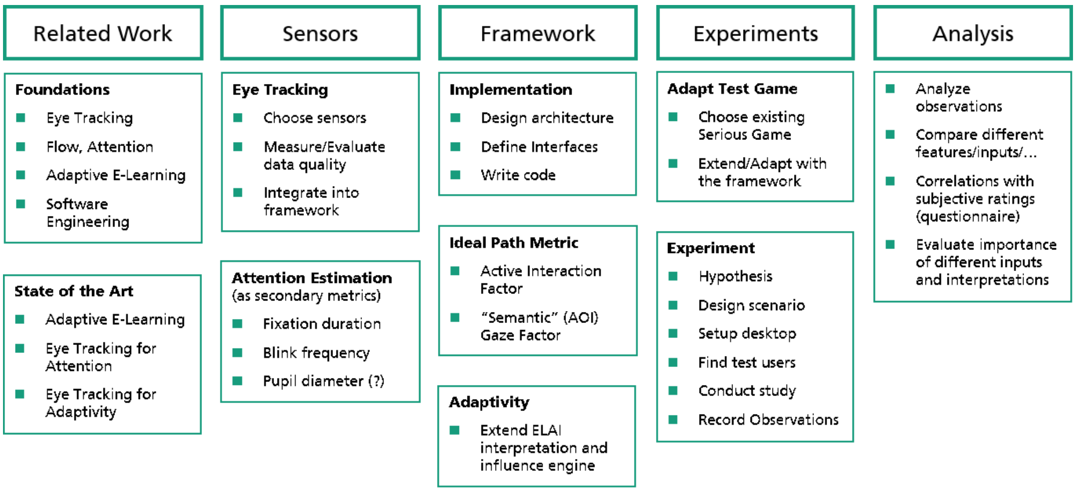


Figure 1.3: Work breakdown structure illustrating the scope of the thesis.

2

State of the Art

This chapter presents the state of the art regarding eye tracking based adaptivity in e-learning, particularly regarding inferences on cognitive states and attention based on eye tracking data as these topics are foundations or closely related approaches to the work in this thesis.

Section 2.1 presents related work on adaptivity of e-learning systems and serious games. Section 2.2 gives an overview of approaches measuring attention and other cognitive processes through data from eye tracking. Approaches using eye tracking based metrics for adaptivity of e-learning systems, closely related to this thesis, are presented in section 2.3.

Adaptive E-Learning and Serious Games

While more basic e-learning facilities like learning management systems (e.g. ILIAS) and Massive Open Online Courses (MOOCs) are nowadays widespread, areas like the application of serious games for learning and intelligent adaptivity towards students' individual needs remain hot topics of research.

Serious Games

Many digital learning games use quiz or card formats but other genres like adventure games have also been shown to be suitable for effective presentation of a complete academic lesson and receive positive evaluation results regarding usability and learner interest [Boe09].

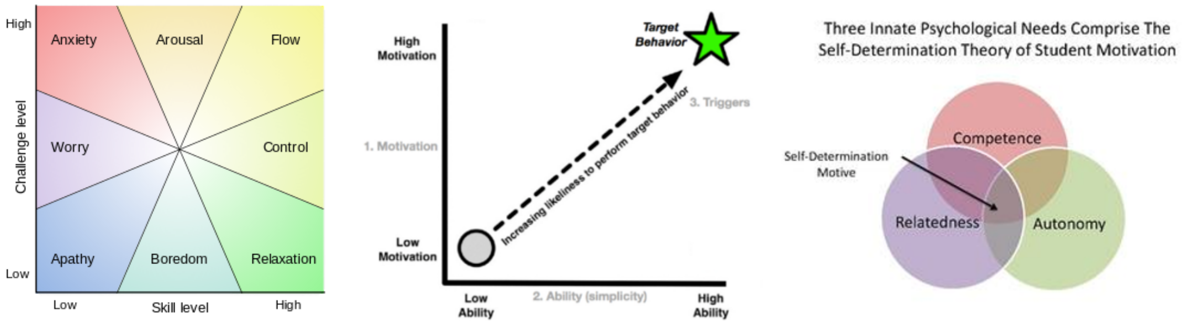
Multiple psychological models are considered to provide a foundation to adaptivity in serious games according to [Str16], including "Flow" [Csí92], Fogg's "Behavior Change" [Fog09] and "Self-Determination Theory" [Dec94] (see figure 2.1).

Adaptation towards an ideal flow experience also needs to consider effects on learning outcomes. "Phases of serious game interaction that are not spent in a state of flow" are also beneficial "in order to allow for self-aware reflection" [Str16].

Adaptive Interventions

A common adaptivity intervention is the display of adaptive hints. However, hints are seldom actively requested by students when playing a learning game but rather presented by the system automatically. Attention to hints does improve game performace but the attention itself is affected by multiple factors like "hint timing and context as well as attitude toward receiving help in general" [Con13]. Similarly, feedback and rewards based on the motivation and knowledge achievement level are used [Ghe10].

Particularly in serious games, systems adaptively select an appropriate game difficulty [Ghe10].



(a) (b) (c)

Figure 2.1: (a) Flow, (b) Behavior Change and (c) Self-Determination Theory are models to consider in the context of adaptivity in serious games, image sources [Bea10], [Fog09], [Wik14].

Figure 2.2: Adaptive hints in Prime Climb [Con13] and SaFIRa [Bie16].

Implementation and Integration of Adaptivity

In addition to questions on what and when to adapt one has to consider the question of how adaptivity can be added to existing education games "in a generic way and with a minimal effort" [Cor13].

Corne [Cor13] uses techniques of aspect-oriented programming to extend existing serious games with adaptivity interventions. This allows the separation of adaptivity concerns from the actual game during development, thereby reducing development effort. [Rol05] develops a user modeling framework following a service-oriented architecture approach.

The ALIGN (Adaptive Learning In Games through Non-invasion) system [Pei08] separates gaming and educational adaptation to allow reusable, abstracted adaptivity based on educational concepts. Such abstract "Adaptive Elements" are then selected for a specific game adaptation using constraints regarding game feasibility and appropriateness (see figure 2.3). This helps to avoid Adaptive Elements that are inconsistent with the game's context and narrative. Rules and other probabilistic methods are used in an "Evidence Interpretation Engine" to map game specific events to facts for educational adaptation. [Pei08]

Semantically Annotated Learning Pathways

To enable didactically interoperable online courses Henning et al. [Hen14] propose individual "Learning Pathways" with large freedom for the learner. These Learning Pathways are determined based on ontological meta-data and logical (OWL) reasoners. Didactic factors are calculated individ-

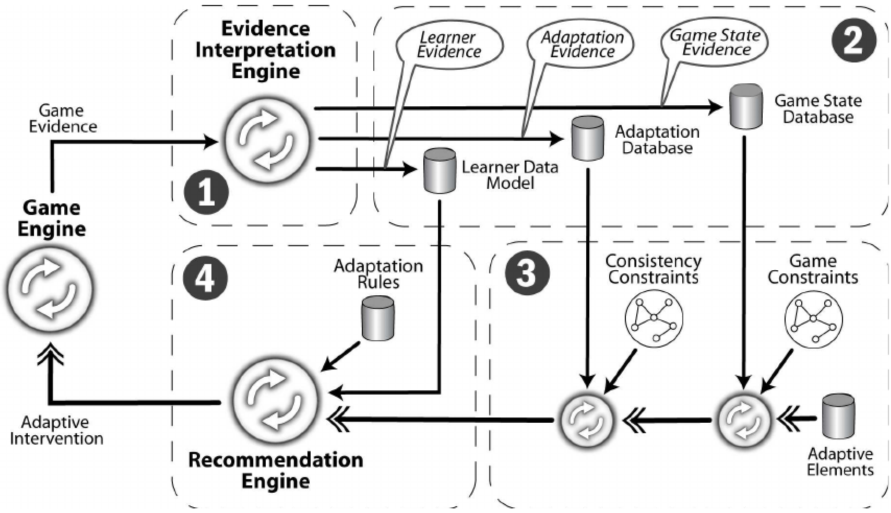


Figure 2.3: The architecture of the adaptive system "ALIGN", image source [Pei08].

ually for each learner and allow incorporation of aspects like motivation or other emotions. These factors can also explain in hindsight why a specific learning pathway was chosen.

Semantic gameplay descriptions in the form of "case-based mappings between content and player experience" (see figure 2.4) have also been used for generation of content individually tuned to a players interests [Lop11]. Such semantic annotations of knowledge objects are also used in more traditional e-learning to dynamically order lessons based on alternative didactic concepts [Swe13].

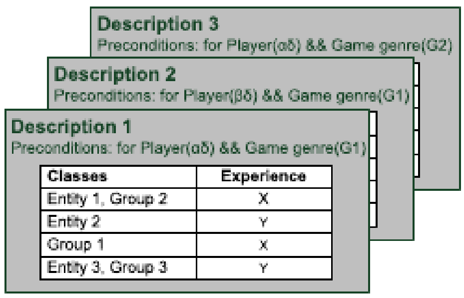


Figure 2.4: Semantic gameplay descriptions by [Lop11].

AdeLE (Adaptive e-Learning with Eye Tracking) is using the SCORM (Sharable Content Object Reference Model) standard to provide interoperability of the learning content. To model aspects required for adaptivity they extend the SCORM specification with additional relations and attributes [Möd04].

Learner Modeling and Classification

Learners can be categorized regarding their representative feature vectors using unsupervised clustering algorithms like k-means. The resulting clusters can then be interpreted using objective information about learning gains of those learners or using expert judgements (i.e. incorporating some supervised classification approach) [Ame07]. The AdeLE system also groups "similar user profiles or user behaviour types" using collaborative filtering [Bar04].

MoGAME (Motivation based Game Level Adaptation Mechanism [Ghe10]) assesses the player's motivation level from interactions logs containing information such as time spent reading a concept, average time interacting with the system, the number of visited pages and the number of help requests. In addition, the knowledge achievement level is estimated from course knowledge and knowledge achievement speed. [Ghe10]

INTUITEL uses a "Learning Progress Model" to integrate information about the user from multiple data sources, including learner data from the LMS and pedagogical and domain knowledge. This Learner Progress Model could also integrate other data items "like emotional or stress measurements involving bodily interfaces to the learner". [Str13]

The ELAI (E-Learning Artificial Intelligence) system also provides a step towards interoperability of adaptivity strategies [Str15c]. This thesis builds upon the ELAI framework and summarizes the system in detail in chapter 3.5.1.

Problem Space as Finite Automata Model

For the ELEKTRA project ([Alb07], [KR10], [KR06]) an approach to microadaptivity in complex learning situations like game-based learning was developed which is based on the information processing theory of human problem solving. This concept has similarities to the Ideal Path model, which is the base of my metrics as described in chapter 4.1.

The problem space, i.e. "the set of all *problem states* a task environment may take", is similar to a finite automata model. It contains an initial state as well as some "*solution states*, in which the problem is considered to be solved". Similar to the approach of calculating the distance to an "ideal path", they propose to score the "correctness" of a problem state by comparing it to a solution state. To avoid an infinite problem space in contexts like games with large freedom for the player to choose and act, they introduce "position categories" to discretize the possible problem states and can thereby work in a finite problem space (see figure 2.5. A correctness value is then manually assigned to each of the position categories to allow quantification. [KR10]

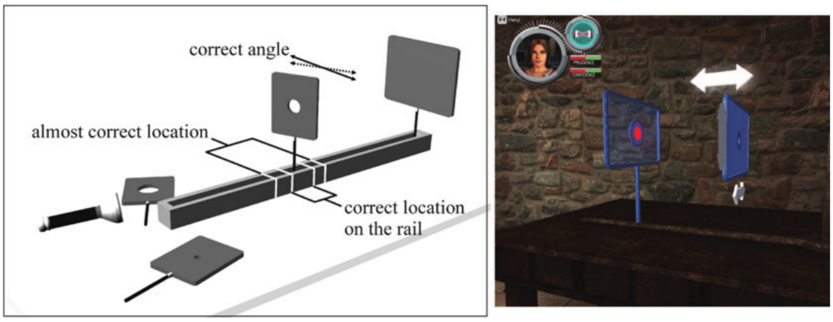


Figure 2.5: Discrete solution states to limit the problem space, image source [KR10].

To model specific skills and a learner's knowledge even when only reaching a partially correct state, an ontological structure of skills and position categories is modelled (see figure 2.6). Specific skills are assigned to a position category as required and allow fine-grained inferences on a learner's knowledge through the skill states of his learner model. This allows a "continuous and implicit assessment" suitable to learning games rather than a traditional assessment through test problems.

The learner model can then inform appropriate adaptive interventions. A range of possible interventions is proposed: "Skill activation" of presumably known skills, "skill acquisition" to teach missing skills, "motivational adaptive interventions" to get the user (re)involved and "assessment clarification" if the system's learner model receives contradicting evidence. [Alb07]

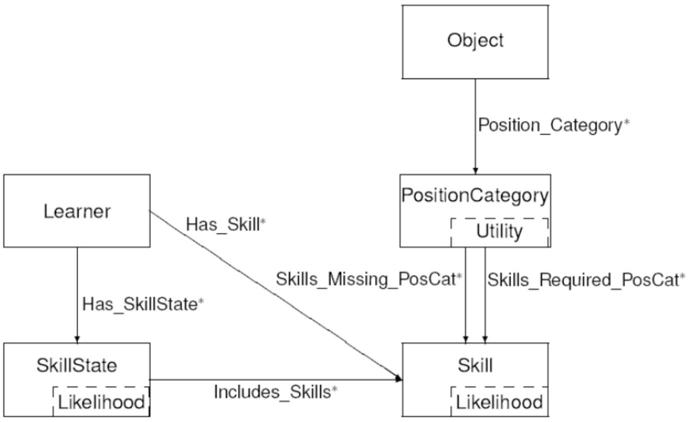


Figure 2.6: Ontology of skills for microadaptivity, image source [Alb07].

Eye Tracking for Modeling of Cognitive States

Eye Tracking for Evaluation of Usability

Traditionally, eye tracking was mainly applied for research in usability. Such studies often use fixations on objects or areas of interest (AOIs) for evaluation of user interfaces. The fixation frequency is used as a measure of the area's importance, the fixation duration as a measure of difficulty of information extraction and interpretation [Jac03].

Good interfaces result in shorter scanpaths (i.e. saccade-fixate-saccade sequences) that cover a smaller area [Gol99].

Eye tracking has also been used to evaluate adaptivity in digital learning games, e.g. regarding students' attention to adaptive hints (figure 2.7), using fixation duration, fixations per word and time to first fixation as parameters for analysis [Con13]. The total fixation time presents a simple and intuitive measure whether (and how much) the user paid attention to a displayed hint.

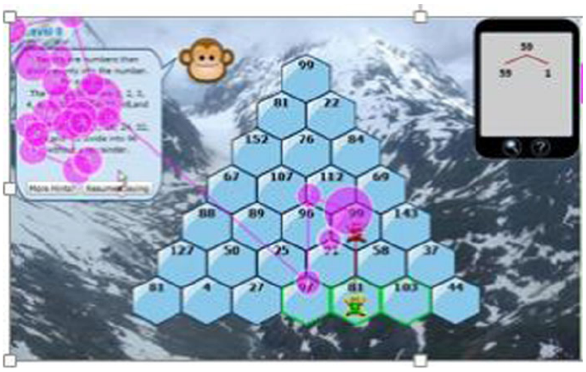


Figure 2.7: Eye movements in a serious game during a study on adaptive hints, image source [Con13].

Eye Movement Patterns

[Gid13] aim to make general inferences that do not rely on what exactly is looked at, thereby dealing with the fact that decision making involves very individual choice of what information is relevant. Instead of semantically interpreting specific AOIs, patterns of fixations and refixations are used for classification. The cognitive process of choosing a product in the supermarket is classified by using eye tracking to identify the transitions between three stages of the decision making process:

orientation, evaluation and verification. For example, a switch from orientation to evaluation is inferred from the first refixation of an object that had already been fixated before. Differences between search and decision tasks were found regarding the number of re-fixations in the second and third stages. [Gid13]

Cognitive Load Estimation

Cognitive or mental load refers to the effort a person puts into a problem solving task.

Pupillometry can be used to estimate this load. The mean pupil diameter change (MPDC), the difference between pupil diameter during a specific period and the overall average pupil diameter for a person, correlates with task performance measures [Pal10]. The pupil is dilating during high cognitive load [Pal10]. Increase in pupil size and blink rate was also reported by [Tak00] to correlate with mental work load, who in addition report that saccade length and rate also corellate with it.

Pupil diameter has peaks when relevant user behaviors in an e-learning environment occur, e.g. when the user discards an incorrect answer option or selects an answer, showing that pupil size is an indicator of mental effort and processing [Can12].

Blink frequency increases over time on a task and [Fuk05] suggest based on their findings that "pupillary measures, blinks, and eye movements assessed together may be used to track aspects of alertness and active information processing."

Velocity of saccades as the ratio of peak velocity to duration (PV/D), relative pupil diameter and a new measure, the bit rate of visual percepton (BRVP) introduced by [Uen04], show high correlations. As decline in PV/D is reported to reflect decline in alertness, saccade dynamics allow assessment of alertness [Uen04].

Saccadic velocity is also said "to increase with increasing task difficulty" [Güt04].

Estimation of Tiredness

Saccadic velocity "is said to decrease with increasing tiredness" and "blink rate, decreasing blink velocity and decreasing degree of openness may be indicators for increasing tiredness" as cited by [Güt04].

Eye Tracking for Measurement of Immersion and Flow

[Jen08] work towards a clearer definition and objective measures for immersion, a term widely used to describe the degree to which players are engaged in a game. They find that immersion can not only be measured subjectively through questionnaires but also objectively through task completion time and eye movements. While eye movements between different games seem hardly comparable, eye movements over time within a game differed for immersive and non-immersive conditions. Eye movements, as the number of fixations, increased over time for the non-immersive condition, while eye movements significantly decreased in the immersive condition or showed no significant change.

However, inferences based on generic, interface independent measures need to consider that eye movements are highly task and context dependent. [Sch11] report from their comprehensive literature review that eye movements are controlled by task demands, resulting in large differences for example between passive viewing and an active task. Studies also showed that eye movements are selected to maximize information gain, thereby correlating with "ideal" eye movement behavior [Sch11].

Eye Tracking for Adaptivity

Eye tracking can be helpful for evaluation of adaptive systems as well as to provide real-time information for adaptivity [Bed05]. Eye tracking is particularly interesting for adaptive systems and user modeling as the data can be recorded non-invasively without interrupting the user. This makes eye tracking suitable for adaptivity in serious games during which players' flow should not be disturbed. Hence eye tracking is also the technique of choice for my thesis. Table 2.2 provides an overview of the different uses of eye tracking data reported in related work which are presented in the following subsections.

For evaluation, eye movements can "quantify whether the decision of the adaptive system were visually attended by the users" [Bed05] and in general analyze user behavior and assess adaptive systems. But going beyond this posterior analysis, real-time gaze data and cognitive processes inferred from it are also suggested as a new source for adaptation, for example providing information on the user's reaction to the adaptivity decision or the user's cognitive workload [Bed05].

Design of suitable metrics could start off with analysis of collected data in a posterior analysis and then integrate promising metrics as data sources for an adaptive system and evaluate them.

For real-time adaptivity, eye movement data can be used in different ways, from directly attentive user interfaces (e.g. displaying translations when detecting difficulties of the reader with a word [Hyr00] or showing contextual information in an e-learning system [Cal08]) to models of abstract meta-cognitive states of the user in order to improve a system's adaptivity decisions.

Visual Patterns Using Areas of Interest

Based on the EVADEG framework, which evaluates serious games regarding aspects of learning performance, gaming experience, game usability and adaptive features, [KR11] use eye tracking to better measure the quality of serious games. They find that high and low performing learners "exhibit different visual patterns". High performers spent more time playing the game, had longer saccades and looked at relevant Areas Of Interest (AOIs) more frequently.

Observing fixations on different AOIs and their relative importance are commonly used to measure learner interest (e.g. [Wei09]).

Students selecting the correct answer from multiple choices exhibit clear gaze patterns (figure 2.8), not revisiting answers they have discarded and looking more frequently at answers they are unsure about [Can12].

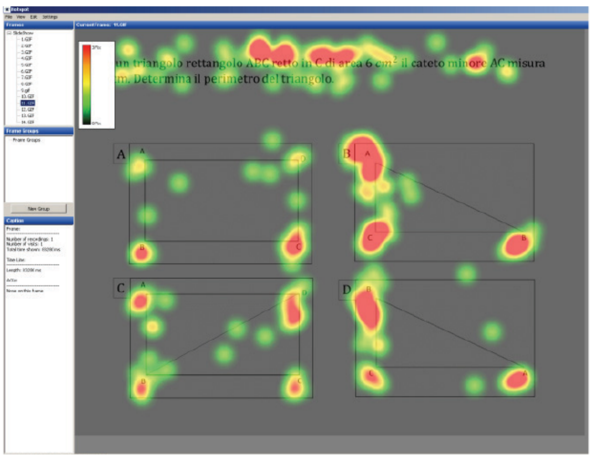


Figure 2.8: Heatmap highlighting the most watched areas during selection of an answer, image source [Can12].

Feature	Inferences
Generic Measures	
Fixation Duration	better learners have longer time played/fixations/saccades [KR11]
Fixation Duration	duration of a fixation correlates with participants' difficultness to process the fixated object [Bed05], [Jac03]
Fixation Frequency	number of fixations over time tend to increase under non-immersive conditions, decrease under immersive conditions [Jen08]
Saccadic Velocity	saccadic velocity decreases with tiredness, increases with difficulty [Bar04], [Güt04]
Saccadic Velocity	decline of velocity of saccades (as ratio of peak velocity to duration) reflects decline in alertness [Uen04]
Blink Frequency	blink velocity/frequency correlates with tiredness [Bar04]
Blink Frequency	increasing blinks over period of time indicates tiredness [Cal08]
Blink Frequency	increase in blink rate and pupil size correlates with mental workload [Tak00]
Pupil Diameter	pupil is dilating during high cognitive load [Pal10], [Tak00]
Pupil Diameter	high increase of pupil size detected when user is disoriented [Cal08]
Pupil Diameter	pupil diameter change peaks occur with important actions (telling answer, identifying/discarding a solution) [Can12]
Pupil Diameter	increasing pupil diameter over period of time indicates tiredness [Cal08]
Pupil Diameter	fatigue indicated by smaller amount of pupil dilation [Wan05]
Pupil Diameter	pupil size could not contribute to predict self-explanation behaviour [Con07]
Mixed	cognitive workload estimated by number of blinks, number of fixations, mean pupil diameter [Cal08]
Mixed	gaze trends (e.g. rate of fixations) only improves classification of learning performance in combination with AOI-based features [Bon13]
Task-Specific Measures	
Fixation Duration	duration looking at region of interest used to classify/highlight sections learned [Cal08]
Fixation Duration	duration looking at textual/graphical objects helps to classify verbal/visual learner style [Meh11]
Fixation Frequency	fixation frequency is a measure of a display's importance [Jac03]
Pupil Diameter	three typical phases of e-learning activities: "searching", "viewing", "preparation"; in searching stage pupil size much larger than in viewing phase. [Can12]
Gaze Shifts	gaze shifts between two related relevant regions indicate "self-explaining" (project ACE: [Mer06], [Con], [Con05])
Gaze Shifts	stages of decision making for choice (orientation, evaluation, verification) can be identified by fixation patterns [Gid13]
Gaze Shifts	frequent refocusing on different cells of the game board interpreted as uncertainty [Wet14]

Table 2.2: Summary of eye tracking features used in related work.

A strong correlation exists between fixation times on AOIs containing textual / graphical learning objects and the student's learner style regarding the Visual / Verbal dimension of the Felder-Silverman Learner Style Model (FSLSM), therefore allowing classification of learning styles based on eye tracking [Meh11].

Estimation of Learning, Attention and Challenge

To predict whether a student falls into the group of high or low learning performance a combination of interface specific AOI-based gaze features and general gaze trend features are most successfull - even without use of interaction behaviors or other not gaze related inputs [Bon13]. While gaze trend features alone (rate and number of fixations, mean and SD of fixation duration, mean and SD of saccade length, mean and SD of relative path angles, mean and SD of absolute path angles) could not predict students to show high or low learning performance, those features did significantly improve predictions when added to AOI-based features (fixation rate in AOI, proportion of fixation time and number in AOI, duration of longest fixation, proportion of transitions from every other AOI). To avoid overfitting [Bon13] performed a cross-validated feature selection. These classifiers are also suitable for real-time, online prediction of student learning that could inform adaptivity.

Virtual agents that react based on information where the user looks (i.e. the number and location of fixations) can help the user to better concentrate on the current topic [Wan05]. Reactions of those virtual agents were also prompted by estimations of the user's declining interest and fatigue, inferred from a smaller amount of pupil dilation and reduced activity of eye movements and mouse/keyboard interactions [Wan05].

Artificial intelligence (AI) for computational opponents in games can also benefit from inputs of the human player's "psychophysical measures" in order to reduce frustration and improve the gameplay experience by choosing an appropriate difficulty level [Wet14]. The AI was artificially limited to fields that received user attention (i.e. fixations). Additionally the AI skill level was parametrized among other factors with the player's confidence. The confidence was inferred from the player's previous moves and progress as well as his eye movements, where frequent refocusing on different cells of the board was interpreted as uncertainty. [Wet14]

The AdELE Framework

The AdELE (Adaptive E-Learning through Eye Tracking) framework combines an Eye Tracking Module, a Content Tracking Module and interactive dialogs to model the learner's state and control appropriate adaptivity [Bar04].

Their eye tracking data provides "hints about concentration, excitement or tiredness of the learner" [Bar04]. Eye tracking helps to distinguish user behaviors (learning, reading, searching in text, observing a picture or reading a text, looking on the navigational elements), but it is "important not to rely exlusively on eye tracking data, but to supplement it also with constant user feedback" and uphold best practice of user experience: To allow the user to take the final decision [Piv05].

To enable teachers to create content suitable for adaptive e-learning, an authoring tool is proposed that provides semi-automatic tagging with meta-data and an adaoption engine that dynamically compiles personalised content [Piv06].

Predicting Learner Self-Explanation: The ACE System

ACE (Adaptive Coach for Exploration) is an intelligent learning environment aimed to "support student-led exploration of mathematical functions", allowing learners to experiment with a formula and its plot.

Eye tracking data improves the modeling of student exploration compared to models solely based on interface features like actions and standard deviation and latency between actions [Ame07]. The eye tracking features employed were gaze shifts between two areas of interest of the e-learning application as well as non-salient regions. "Gaze shifts may be important mostly in discriminating between time spent reflecting on an action's results and idle time" [Ame07].

Gaze shift patterns to observe the relation between a formula at one AOI and a corresponding graph at another AOI are used to classify self-explanation behavior (figure 2.9), improving these models of meta-cognitive behavior [Con05], [Mer06]. The gaze shift feature did not show a clear advantage in classification accuracy compared to a feature based on the duration of interaction but a combination of time and gaze based features is suitable to reach best sensitivity (i.e. detect most cases of true self-explanation) [Con05].

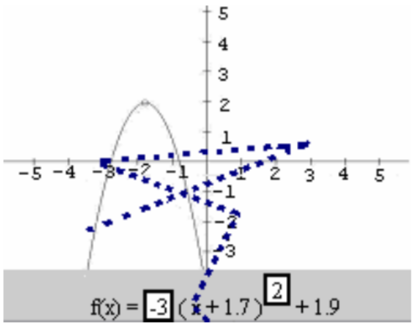


Figure 2.9: The ACE environment uses gaze shifts between a plot and its formula to classify self-exploration, image source [Con07].

Pupil size could not predict self-explanation using any of multiple normalization techniques, which may be related to the specifics of self-explanation or confirm doubts of other studies about pupil size as an indicator of cognitive load in general [Con07].

Good results from a combination of interaction and eye tracking features were confirmed by [Kar11], who built on those previous experiments and demonstrated that their non-task-specific eye tracking features only relying on a few basic AOIs exploited by data mining techniques can lead to an accuracy of 84.51% when used to classify high/low learning performance in combination with interaction features. The eye tracking features consisted of statistics on fixation rate, number and duration, saccade length, relative and absolute saccade angles and transitions between AOIs. Interaction features comprised "usage frequency for each action, as well as mean and standard deviation of time interval between actions".

The e5Learning System

The e5Learning system [Cal08] integrates eye tracking for multiple aspects: A monitor of accessed screen areas, a contextual content generator and an emotion recognizer. The accessed (looked at) screen areas are monitored with regard to an amount (duration) of attention the user should pay to certain portions of content (see figure 2.10). Based on this information the system recommends whether the user should stick with the current lesson or move to the next [Cal08]. Exact evaluation of reading progress is only possible on large text because of the limited precision of the eye trackers, but heuristically progression is checked to lead from left to right, top to bottom while non-textual areas have a number of fixations or duration defined by the content author [Por08].

The contextual content generator of e5Learning dynamically displays context dependent information based on the user's visual focus.

The emotion recognizer is based on changes in the rate of blinks, rate of fixations and mean pupil diameters. A decrease of blinks, increase of fixations or increase of mean pupil diameter are used as indicators for high workload or non-understanding. A "high increase of pupil size between two consecutive samplings" usually signifies that the user is "disoriented". If the mean pupil diameter is monotonically increasing over several intervals, tiredness is suspected and for confirmation it is checked whether the number of blinks is also monotonically increasing. Participant verbally confirmed those cognitive states in 79%-83% of cases during an experiment. [Cal08]

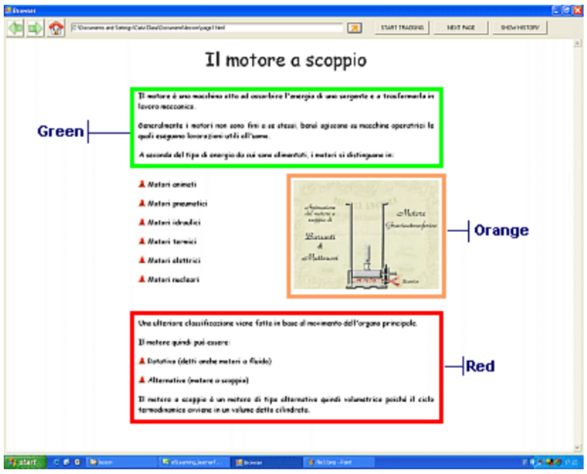


Figure 2.10: The e5Learning system monitors how much specific screen areas are accessed, image source [Cal08].

3

Fundamentals

Attention and Flow

Chapter 4 discusses the overarching goal of this thesis, that is to create a sustained flow experience for which I am are focusing on user attention. In the following the theoretical concepts of flow (section 3.1.1) and attention (section 3.1.2) are summarized. The theoretical model of visual attention (described in section 3.1.2) bring together theories on attention and eye movements.

Flow

The concept of flow [Csí92] describes the state of intense concentration on an activity to an extend that one loses self-consciousness and sense of time. During flow experience one feels in control of one's actions and experiences the activity as intrinsically rewarding. Flow experience results from a system of person, environment and interactions. An important aspect to this is that challenge and skills are subjectively perceived as balanced (see figure 3.2) and above average for the activity [Nak02]. "Flow" represents the high challenge, high skill facet of Csikszentmihalyi's model of mental states (see figure 3.1).

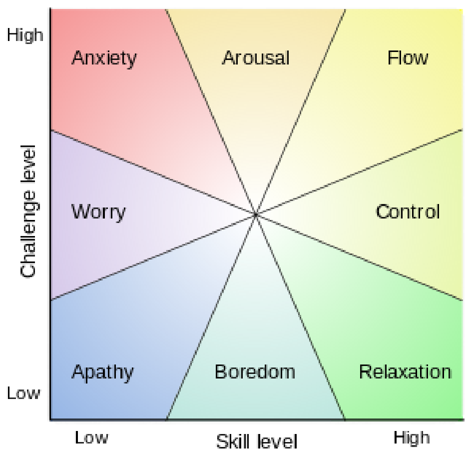


Figure 3.1: Mental states of the Flow Model according to [Csi97], image source [Bea10].

The relation between challenge and abilities is often described as the flow channel or flow zone. Conditions for flow are individual. An expert has a different flow channel than a novice, therefore designers should provide choices that allow to adapt the experience of the individual user (see figure 3.2). However, in games frequent explicit choices could interrupt gameplay. It is best to embed choices into the core activities of the user [Che07].

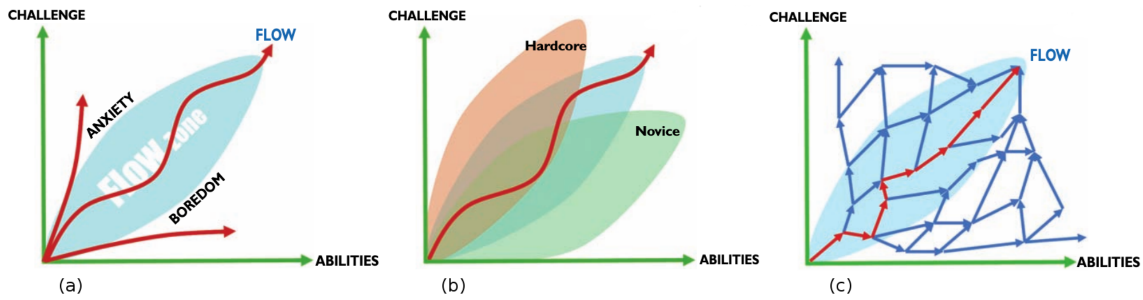


Figure 3.2: Flow zone factors (a) are different for different players (b) therefore designers adapt flow experience through choices (c), image source [Che07].

To measure flow often qualitative interviews are performed or standardised questionnaires are applied. To observe flow experiences over longer periods of time the "Experience Sampling Method" can be used, for which participants are requested to rate their current experience when triggered by a device at random times [Che06].

Studies of flow in software and games mostly use questionnaire techniques to get self-reported levels of concentration, involvement and enjoyment from participants [Nak02], [Chi08], [Nac08].

The "GameFlow" model evaluates games with regard to their enjoyment and flow based on eight elements: concentration, challenge, skills, control, clear goals, feedback, immersion and social interaction [Swe05]. They find that their GameFlow model's focus on aspects of flow is a useful tool to review enjoyment in games. "EGameFlow" adapts this approach specifically to e-learning games [Fu09]. Their resulting dimensions are immersion, social interaction, challenge, goal clarity, feedback, concentration, control, and knowledge improvement.

Attention

Attention as a cognitive process is the selective concentration on one aspect of the environment while ignoring others [Wic00].

People have a limited amount of attention at their disposal [Kah73]. If not enough attention is available (and applied) to a task, mistakes increase and performance in general suffers.

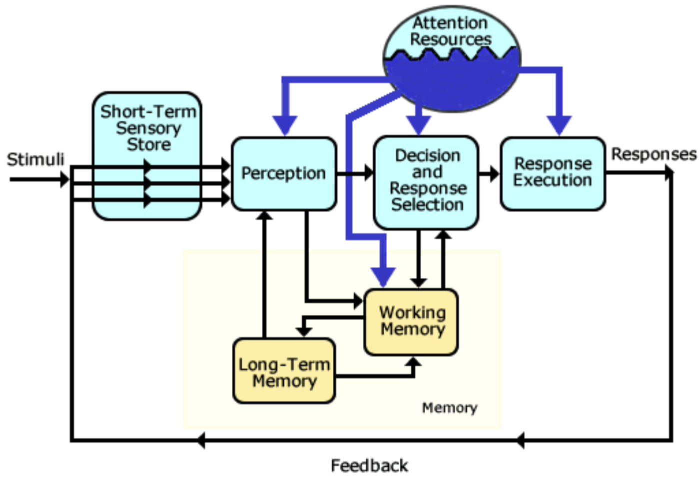


Figure 3.3: Attention seen as a limited resource required for cognitive processes, image source [FAA].

However, human information processing can be classified as "controlled processing" or "automatic processing" [Shi77]. While controlled processing requires consciously directing attention towards the task, automatic processing does not require attention and happens unconsciously. Whether or how much attention is needed for a specific task varies between people. A novice user may need to fully concentrate on some task which an expert hardly thinks about.

The term "visual attention" refers to a person's foveal focus on some object or area. This usually indicates that the person's attention is directed at this point. However, attention can also be directed at areas other than the foveally viewed area [Duc07].

Visual Attention

For this thesis, modeling attention based on eye movements, the visual aspects of attention are particularly relevant. A visual scene is attended in small regions, constructing an overall representation in the mind. Two main components of visual attention are the aspects of "where" and "what" [Duc07].

Where? The gaze continually roams over the visual field. The process of visual selection, determining where one's attention is attracted and one will look next, is often based on peripheral vision. Attention can be controlled consciously to attend peripheral objects without eye movement but eye movements signify overt visual attention as they indicate detailed inspection of an object. [Duc07]

Where visual attention is attracted is not only determined by effects of salience (i.e. how much an object stands out in the given context), task demands and value are also important factors for the eye movements. Also eye movement strategies can differ for experts and novices [Sch11].

What? To perceive details the fovea is focused at the area of interest and allows to inspect features at high resolution [Duc07].

A bottom-up model of visual attention assumes the scene is first screened through peripheral vision to identify interesting features. Attention is then turned to the first area of interest, moving the fovea to this point. Once the movement is complete, the feature is inspected at high resolution before moving to the next area of interest. [Duc07]

While this model is clearly a simplification of the actual processes, it provides a good basis for computational models [Duc07].

Eye Tracking

Eye movements allow inferences about the cognitive state of a person and in particular where (visual) attention is focused. An eye tracker can record this information noninvasively, therefore this is the instrument of choice for my Attention Adaptivity Framework.

In the following, the basic functioning of an eye tracker is summarized (section 3.2.1), relevant terms and measures of eye movement are defined (section 3.2.2) and guidelines on data quality presented (section 3.2.3) The foundations presented in the following sections are based on [Duc07].

Functionality

Common modern eye tracking devices are fixed to the computer display and operate video based. Infrared light emmitted by the device is reflected by the user's eye. This corneal reflection is recorded through the eye tracker's camera and using the position relative to the pupil center x-/y-coordinates are calculated, describing where the eye looks on the screen. [Duc07]

Terms and Metrics

Eye movements consist of *fixations*, when gaze is maintained on one location, and *saccades*, when gaze is shifted to a different location in a quick movement.

Fixations allow the perception of details through the fovea, which is the part of the human eye responsible for high resolution, central vision. About 90% of viewing time is spent in fixations and a fixation usually lasts between 150 and 600 milliseconds. Visual perception is focused on an area of an angle of 1-5°, therefore taking in around 3% of a 21 inch screen at 60 centimeters distance.

However, the eyes are never completely still as microsaccades involuntarily shift the gaze slightly during fixations. This is necessary to overcome adaption of the visual cognitive system because a completely static perception would not generate any neural response. [Duc07]

Saccades are fast eye movements to reposition the fovea, which occur between fixations. During saccades the eyes move at maximum speed and there is no perception during this transition. Saccades usually last around 10-100 milliseconds.

Eye trackers detect saccades by analysing velocity and position variance of the gaze through a sliding window of samples. [Duc07]

Blinks refer to the closing of the eyes for a short time to spread a thin film of tears over them [Gal01]. While blinks need to be filtered out as noise for normal eye tracking [Duc07], blink data itself has been used as an indicator for tiredness [Bar04], [Cal08].

Data Quality

Holmquist et al. [Hol12] argue for standards regarding evaluation and reporting of data quality in eye tracking experiments. With my focus on eye tracking data in this thesis, dependable data quality is important. As part of the conducted experiment I report the quality of my collected eye tracking data in chapter 6.4.1 following Holmquist's guidelines.

The required data quality and accuracy of eye tracker measurements depends on the purpose for which the data is applied. Off-the-shelf eye tracking devices often suffice to observe basic gaze interaction with larger Areas of Interest (AOIs), while research concerning psychology, neurology or reading require better data quality.

Eye tracker quality can be measured through calibrations (i.e. requesting the user to look at known locations of the screen) and should consider varying positions as accuracy usually is best in the middle of the screen. It is measured as *accuracy*, i.e. the distance (in visual degrees) between actual and measure gaze position, and *precision*, i.e. the variance in accuracy, commonly estimated through "standard deviation of the samples" or "root mean square (RMS) of inter-sample angular distances".

Participants or data samples of bad quality need to be excluded from analysis. The criteria used regarding this should be mention in resulting reports. [Hol12]

[Hol12] propose a standard report on data quality for research using eye tracking. The data quality measures publications should report are shown in table 3.1. Also the detection algorithms used should be specified.

Data quality	Average	SD
Calibration accuracy		
Accuracy just before end of recording		
Calibration precision		
Precision just before end of recording		
Accuracy after post-recording processing		
Precision after post-recording processing		
Proportion of dismissed participants		
Proportion of lost data samples in retained data		

Table 3.1: Proposed format by [Hol12] for data quality reports in eye tracking studies.

Serious Games

The term "serious game" refers to games whose primary purpose is not entertainment but rather education or another "serious" aim. Prensky [Pre07] has coined the term "digital game-based learning" for serious games with a focus on utilizing digital games for educational purposes.

Digital Game-Based Learning

Learning is most effective when the activity is active, experiential, problem-based and providing immediate feedback and games often have those characteristics [Con12]. This is particularly promising because games often are much more engaging than traditional teaching.

The fact that games usually provide intrinsic motivation and immersion is central to keep learners engaged. Digital game-based learning can be considered to fall into the category of "Practice Doing" of the learning pyramid (see figure 3.4), which shows a much higher retention rate of students than traditional lectures.

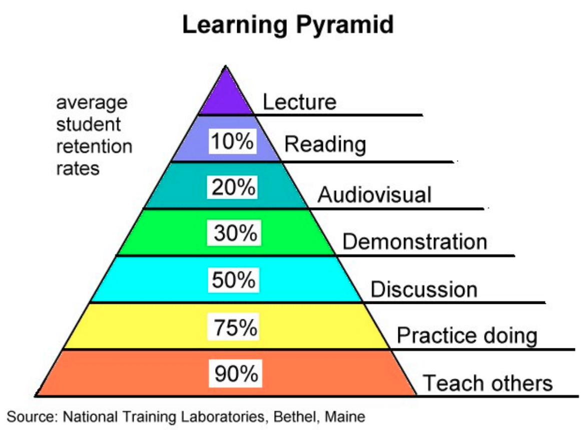


Figure 3.4: Learning games as "Practice Doing" show high retention rates.

A number of studies have shown computer games and serious games to have positive learning outcomes regarding skills, cognitive (e.g. strategic knowledge) and affective (e.g. attitudes) outcomes [Con12].

Immersion and Flow in Games

Multiple terms are used to describe gaming experiences, among the most frequently used are "immersion" and "flow".

Immersion describes the feeling of "being in the game", having almost all attention focused on the game to an extend that one loses track of time and ones surroundings [Jen08]. Immersion is considered to be a gradual experience, with levels from engagement to engrossment to total immersion or "presence".

Different aspects contribute to the immersive experience: Sensory immersion, imaginative immersion (influenced by atmosphere and narrative) and challenge-based immersion (which is comparable to Csikszentmihalyi's concept of flow) [Nac08].

Flow is closely related to immersion. Flow can be seen as an important part of immersion although immersion is considered a more gradual experience and can be observed even without traditional aspects of flow, like clear goals [Jen08].

Adaptive E-Learning

The digital nature of e-learning makes extensive personalization possible. As learners' characteristics like incoming knowledge and skills, cognitive abilities, personality traits, learning styles and interests vary vastly, learning performance can be improved by catering to individual needs and preferences [Shu03].

A commonly adapted parameter, particularly in serious games, is task difficulty (Dynamic Difficulty Adjustment). The task difficulty adjustment strategy in relation to a user's prior video gaming experience does influence learning outcomes. Video game experience leads to higher performance and motivation regarding the learning game, for inexperience gamers careful difficulty adaption influences training performance [Orv08].

Adaptivity in learning games requires special attention as "adapting a game to enhance its educational benefit endangers its intrinsic motivation and flow" [Pei08].

Adaptability and adaptivity describe two slightly different aspects. While "adaptability" indicates that a system can be changed (often manually) for different contexts, "adaptivity" indicates that a system automatically changes [Str16].

Learner Model

A learner model represents the learners' characteristics that are relevant to the adaptive e-learning system. It often is the foundation for personalization of the learning experience. Learner models can be domain-dependent, refering to the present knowledge, or domain-independent, refering to general abilities and personality traits. Microadaptivity, i.e. the custom selection which knowledge element should be presented next, is based on domain-dependent models, while macroadaptivity, i.e. to determine how content should be presented, relies on domain-independent learner profiles. [Shu03]

There are different approaches to model a learner. For a simple "Stereotype Model" learners are simply categorized manually. An "Overlay Model" maps the learner's knowledge "on a concept-by-concept basis", allowing more flexible modeling but also requiring a more complex modularization of the learning content. [Con02]

Content Model

A content model represent a "knowledge map" to be taught. In order to enable automatic inferences and adaptivity, the learning objects often need to be structured and semantically annotated [Shu03].

Adaptive Cycle

Adaptivity in educational systems can be modeled as a cycle of four processes [Shu12]: Capture, Analyze, Select and Present (see figure 3.5). First the system *captures* information about the learner, this can include different aspects and sources like performance data or eye tracking. Then the system *analyzes* the captured data to update a learner model or directly adapt certain aspects of the system. *Selecting* appropriate content with regard to the learner's state and the system goals and *presenting* content in the best way (e.g. the most appropriate media) for the learner conclude the cycle with their adaptions to the system's output.

This cycle can also be implemented partially as needed, for example not selecting content but only using the results of analysis to adapt presentation [Shu12].

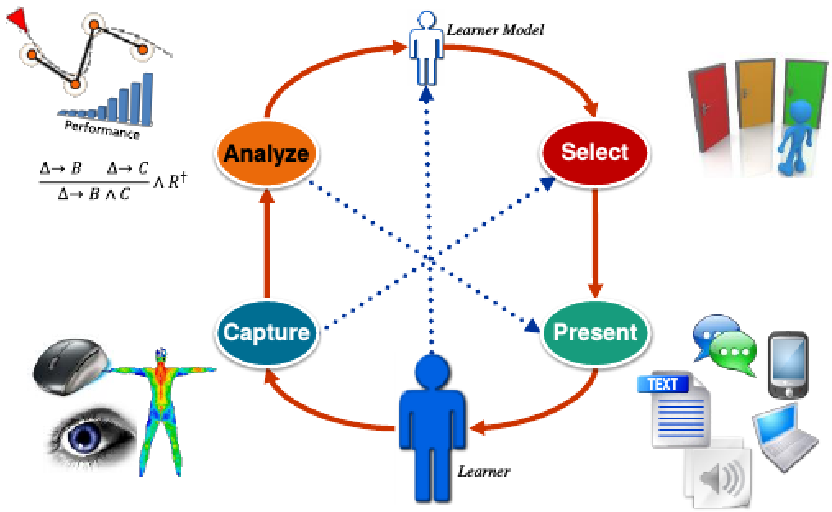


Figure 3.5: Adaptivity in educational systems as a 4 phase adaptive cycle, image source [Str16].

E-Learning Interoperability

With the increasing significance of digital game-based learning systems, the educational ecosystem gets more complex, comprising different learning management systems, simulations and learning games. To allow some integration of those systems and make intelligent adaptivity of learning experiences feasible there is need for an interoperable intelligent tutoring agent framework [Str15c]. This includes extensible frameworks and architectures like ELAI (section 3.5.1) as well as common formats to exchange data, for which xAPI was designed (section 3.5.2).

E-Learning A.I. (ELAI)

This thesis builds on the "E-Learning A.I." (ELAI) system by [Str15c].

The ELAI system provides a way to decouple game and adaptivity concerns. The system uses HLA interfaces, information is mainly transmitted in the xAPI format. "Interpretation Engine" and "Influence Engine" components build learner models and use them to calculate adaptivity parameters respectively, which are then transmitted back to the game implementation to control the game's concrete adaptivity (see figure 3.6).

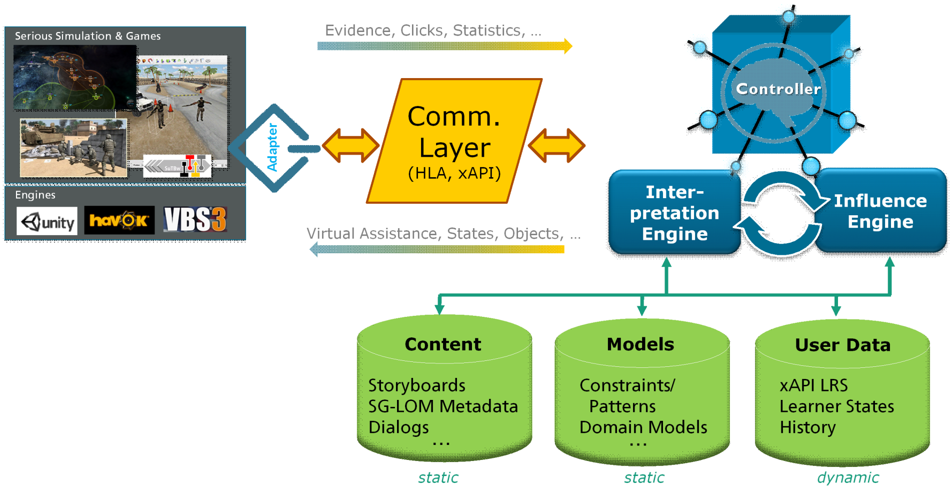


Figure 3.6: The conceptual architecture of the ELAI system, image by Streicher based on [Str15c].

xAPI

The "Experience API" (xAPI, also called "TinCan API") is a standard to record any experience in a common format. Compared to previous specifications for e-learning like SCORM (Sharable Content Object Reference Model), xAPI is simpler and more flexible. It is possible to define relevant Activities and therefore record any kind of learning experience, offline or online, traditional teaching or from learning games [Rus].

xAPI is following the principles of REST (Representational State Transfer), using the standard HTTP methods and data in the JSON format.

The following definitions are taken from the official xAPI specification [xAP].

Every record in xAPI is a *Statement*, which follows the structure "I did this". A Statement comprises at least an Actor, Verb and Object and can include additional data like a Result.

Actors identify an Agent (e.g. a person) or Group who is the subject of the experience. The information of an Actor includes name and a unique identifier like an e-mail address.

Verbs describe the action performed during recorded experiences. A Verb most importantly has an URI to define it uniquely and can include a label for display. The xAPI standard does not define concrete verbs as those need to be specifically relevant to a community of practice. A public directory of verb definitions exists at (https://registry.tincanapi.com/#home/verbs) to allow reuse.

Objects usually are Activities but can also be Sub-Statements, Statement References or Agents. An Activity includes a unique identifier and optionally a definition specifying details like name and description.

Results are optional data objects to describe the measured outcome of Statements. A Result can include information like whether the Statement was successfull and complete, its duration and a Score.

```
{
    "actor":{
        "objectType": "Agent",
        "name":"Example Learner",
        "mbox":"mailto:example.learner@adlnet.gov"
    },
    "verb":{
        "id":"http://adlnet.gov/expapi/verbs/attempted",
        "display":{
            "en-US":"attempted"
        }
    },
    "object":{
        "id":"http://example.adlnet.gov/xapi/example/simpleCBT",
        "definition":{
            "name":{
                "en-US":"simple CBT course"
            },
            "description":{
                "en-US":"A fictitious example CBT course."
            }
        }
    },
    "result":{
        "score":{
            "scaled":0.95
        },
        "success":true,
        "completion":true,
        "extensions": {
            "http://example.com/profiles/meetings/resultextensions/
               minuteslocation": "X:\\meetings\\minutes\\
               examplemeeting.one"
        }
    }
}
```


Listing 3.1: Example xAPI Statement in JSON format including Result.

Extensions allow a flexible way to add additional information to an Activity definition, Statement context or Result. Extensions are defined by a map where the keys must be IRIs. The values of extension fields are specific to an application or standard agreed upon outside of the xAPI specification.

SaFIRa

SaFIRa ("Seek and Find for Image Reconnaissance adaptive") is a serious game developed at Fraunhofer IOSB for demonstration of ELAI in use with an actual learning application. The game is intended for training of experts in image reconnaissance, who have to analyse images from radar, infrared or other sensors. SaFIRa embeds a similar task - map-based orientation to find a specific tank in a city - into a game-like context.

The game was extended by Biegemeier [Bie16] within the scope of his thesis and that version of SaFIRa with its ELAI integration and adaptivity is the foundation for my experiments and the target of my adaptivity improvements using eye tracking. My specific Ideal Path Score (see chapter 4.1) is also customized to the SaFIRa game.

The objective of the game is to find the target tank while navigation through a city with the player avatar. To win the game, the avatar must move directly to this target tank as shown in figure 3.7 (c).

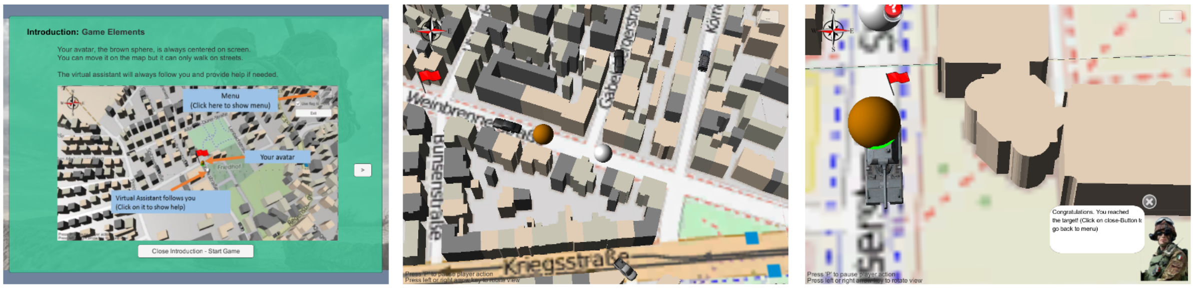
The game contains an introduction comprised of a number of screens with a short explantory text and annotated screenshots to explain the controls and objectives to novice players (see screenshot (a) in figure 3.7). In addition, an in-game virtual assistant can provide the task description and offer concrete hints on how to reach the target, as shown in figure 3.8.

During the game the player has a bird's eye view of a part of the city, which is comprised of an OpenStreetMap map as the ground plane and basic 3D models of the buildings as shown in figure 3.7. The symbolic player avatar, a sphere, is always positioned at the center of the screen and move on the streets as directed by the player. Movements are controlled by double clicking on the map. This positions a flag towards which the avatar then finds its way along the streets (see screenshot (b) in figure 3.7).

By moving the avatar the player gets to see a new part of the city and should get closer to the target tank. The virtual assistant supports the player with hints stating the distance and direction to the target (see figure 3.8). These hints are triggered either by an active request of the player, clicking on the virtual assistant's question mark symbol, or adaptively by the ELAI system whenever the player doesn't make any progress for some time. The adaptivity reasoning is described in more detail in the following section 3.6.

Throughout the city different vehicles can be seen, cars as well as other tanks similar to the target tank, that are of different types however. The player needs to distinguish these game objects from the tank type of the target in order to make efficient progress rather than moving to every visible vehicle.

At the start of a game session, all vehicles as well as the player avatar are randomly placed on the map. Apart from the avatar and the virtual assistant following it, the vehicles and the target tank do not not move during the game.



(a) (b) (c)

Figure 3.7: Screenshots from the SaFIRa game: (a) introduction page, (b) movement to the destination flag, (c) victory after reaching the target tank.

SaFIRa's Adaptivity

The adaptivity of SaFIRa personalizes difficulty and dynamic help of the game based on the user's skill level and current progress. I reuse these adaptivity methods in my version of SaFIRa for my experiment and enhance the adaptivity reasoning by adding eye tracking based features as an additional information source (see chapter 5.4).

On the technical level the game itself is built on the Unity game engine, the adaptivity is controlled by integration with an ELAI server. SaFIRa transmits statements about the player's actions to ELAI in xAPI format. The ELAI server aggregates this information to estimate the player's skill level and determines whether the player currently needs help. SaFIRa periodically polls its ELAI server to receive these influence parameters "needsHelp" and "skillLevel". The calculation of these parameters in the original ELAI system and in my enhanced system are explained in more detail in chapter 5.4.

Adaptive help is displayed by SaFIRa in the form of a hint stating distance and direction to the target as described above and shown in figure 3.8. Such a hint is automatically shown when the ELAI server sets the "needsHelp" influence parameter to "true".

ELAI infers that the player needs help if he has not made any progress during the last 10 seconds, i.e. the player avatar is moving away from the target for more than 10 seconds.

Adaptive difficulty in SaFIRa is based on the skill level of the player modeled by ELAI. ELAI aggregates the information on the player's actions after completion of each game session to estimate the player's current skill level based on duration, difficulty level and number of requested hints during that session. The values are normalized using the respective minimum and maximum values observed during any player's previous session (for details see chapter 5.4). This brings the "skillLevel" parameter to a range of [0,1].

The "skillLevel" influence parameter (or its inverse used as the difficulty level) control the size of the target tank in the game and the appearance of clouds as distractors. Most notably the extensiveness of hints is adapted based on the skill level (see figure 3.8). Players with a medium skill level between 1/3 and 2/3 see hints at help level 2 stating distance and direction to the target as explained above. For players with a skill level above 2/3 (help level 1) only state the distance to the target without a direction, hints for players with a skill level below 1/3 (help level 3) see a line pointing in the direction of the target in addition to the help text.

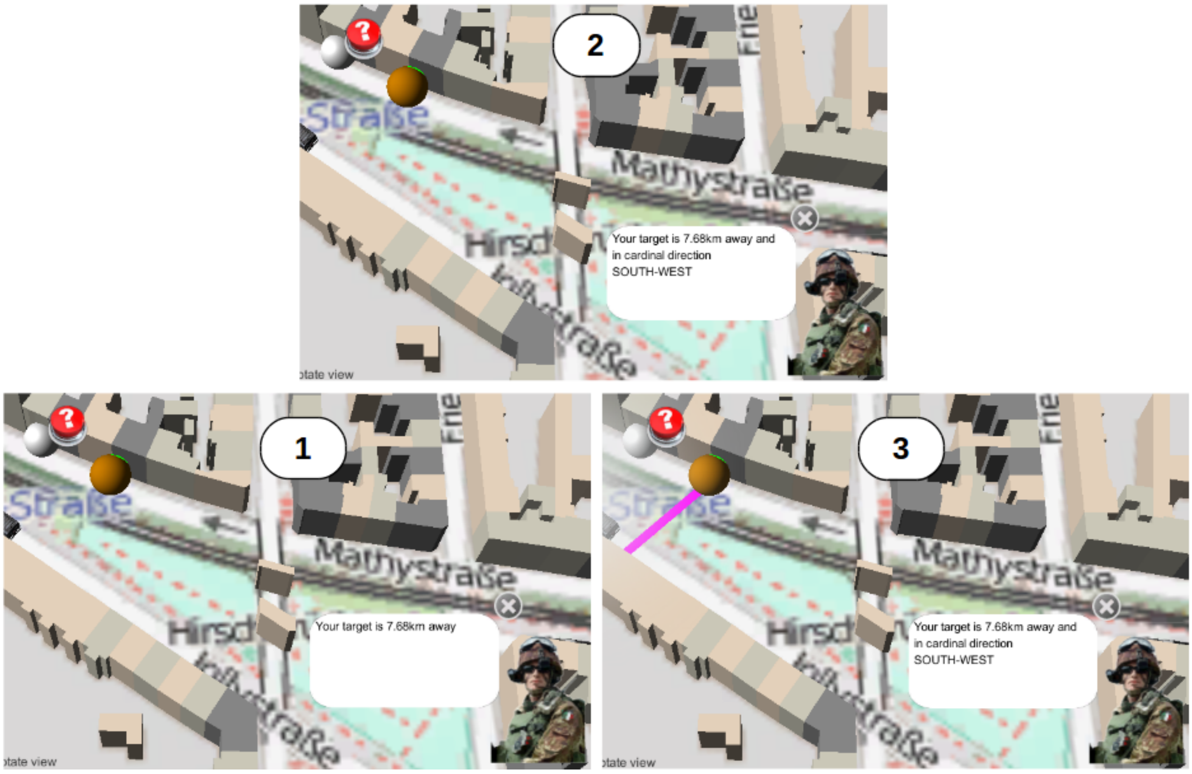


Figure 3.8: The virtual assistant displays hints to help the player. The extensiveness of hints (help level 1/2/3) is adaptively based on the skill level of the player.

Plugin Patterns and Frameworks

The Attention Adaptivity Framework developed as part of this thesis (described in chapter 5 should to be extensible and customizable. Plugins can be an effective approach to meet these objectives and were used on the client as well as the server components of the framework.

Plugins are a common pattern to configure and extend an application at runtime using a central configuration, often through file [Fow03]. This makes extension of an application possible without modifying the application itself and also allows third-parties to adapt an application easily.

The foundation of the plugin pattern are separated interfaces, i.e. interfaces defined in a different package than their implementation [Fow03]. This allows loose coupling of implementation and use of functionality.

The basic plugin pattern is outline in figure 3.9. The main (or host) application defines a service interface which is implemented by a separate plugin component in order to offer functionality of that type of service. The plugin is developed completely independent from the host application apart from implementing the given interface. Plugins are built separatedly and the host application only loads the available plugins at runtime, for example based on a configuration file or through a file discovery process for a given location and file type.

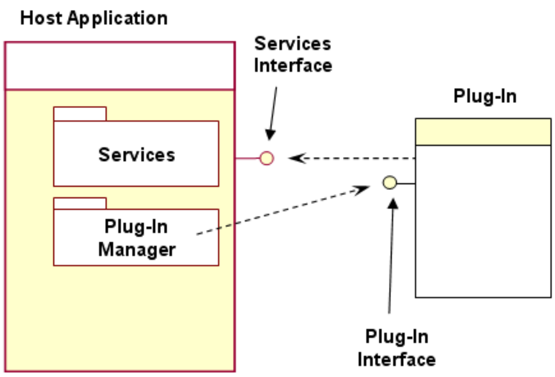


Figure 3.9: The basic Plugin Pattern, image source [Pri08].

The basic plugin pattern is easy to implement [Blo12] and some functionality to support plugins is also already offered by existing frameworks like .NET [Osh03]. OSGi is a widely used standard to implement a dynamic module system for Java [OSG10]. It features a layered approach modules, a life cycle API and more.

The Java ServiceLoader API is a simple plugin loader facility included in the standard Java Platform API. The server-side plugin system for my Attention Adaptivty Framework uses the ServiceLoaded API to dynamically load adaptivity components (described in chapter 5.2.3).

The ServiceLoader API has some limitations, namely that new service providers are only discovered at startup of the application and some functions of service loading are not modifiable. On the other hand the ServiceLoader API offers a straight forward way to implement a plugin system using only standard Java interfaces in addition to the ServiceLoader class from the Java SDK [Ora].

As my framework has no need for more complex plugin features I decided against the complex OSGi standard in favor of a simple implementation of service providers using the ServiceLoader API [Ora].

4

An "Ideal Path Score" based on Interactions and Eye Movements

Measures of user attention, performance and progress are important inputs to an adaptive e-learning system in order to control adaptivity mechanisms appropriately. Features indicating such states have to be defined and calculated. Regarding this, state of the art research shows that a combination of user interactions and eye tracking is a promising and feasible approach, as described in chapter 2.

I therefore designed an "Ideal Path Score" which implements such an approach of combining information on interactions and eye movements. This concept is described in detail in this chapter. Further extensions, like integration of generic eye tracking features may be beneficial and are considered in section 4.2.

Ideal Path Score

The developed "Ideal Path Score" is based on the distance of user actions to an "ideal" path, a model proposed by [Str15b]. To receive a more comprehensive view, eye tracking data as well as user interactions are included in the calculation of the distance of a user's actions to the ideal path.

The underlying "Ideal Path" model is introduced in section 4.1.1 before the general concept of the "Ideal Path Score" is presented in section 4.1.2, including some showcases. The detailed definition of the score (section 4.1.3) and possible variations (section 4.1.4) follow.

Section 4.1.5 discusses the motivation and reasoning behind this "Ideal Path Score" approach on a more general level.

Ideal Path Model

An "Ideal Path" is a sequence of episodes and interactions in a virtual learning environment that most directly leads to the current goal [Str15b]. This ideal path serves as a reference, a gold standard for comparison. Suitable metrics can define a distance between a user's actual path and the ideal path (see figure 4.1). This distance quantifies the user's performance.

A path generally consists of the user's explicit interactions with the virtual learning environment. However, direct interface interactions may not be sufficient to adequately judge the user because they do not offer any insights into the user's cognitive processes and decision making. The assumption for my thesis is that eye movements can contribute valuable additional information to such an "Ideal Path" based score. For example in a case when the user looks at the correct object, considering it, but ultimately decides to interact with a different object eye movements give a more comprehensive picture of the situation. Detecting such situations would allow improved learner models and hence improved adaptivity.

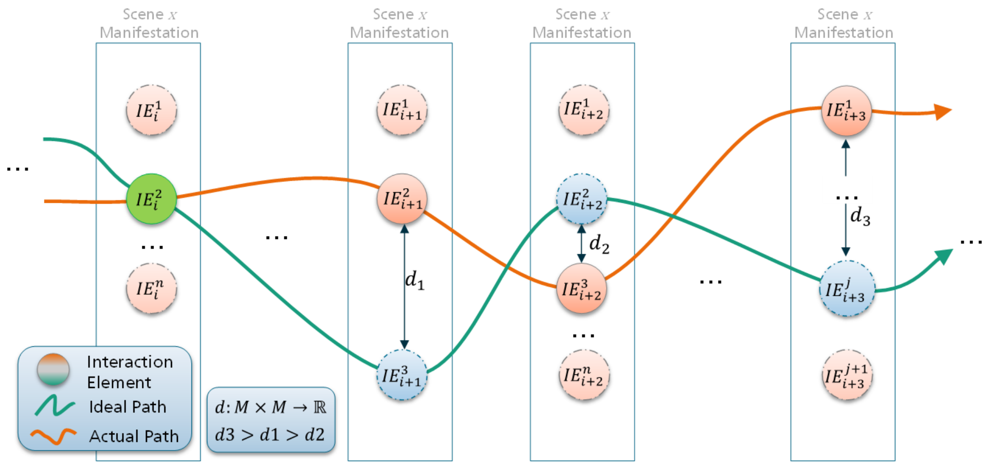


Figure 4.1: Ideal (green) and actual (red) path through a virtual learning environment, the Ideal Path distance can indicate the user's performance, image source [Str15b].

In practice, the calculation of an "Ideal Path" distance requires that an "ideal" action for the current game situation is determined - possibly through heuristics. The actual action of the player can then be compared to this benchmark using a meaningful metric. The definition of ideal actions as well as the definition of a meaningful "distance" measure between such paths needs to be game specific.

The Ideal Path Score Approach

This section describes an "Ideal Path Score" for the serious game SaFIRa, which is also used for the experimental evaluation (see chapter 6). The definitions of "Ideal Path" and "distance" between paths are tuned to specific game context of SaFIRa but should be easy to adapt to similar games.

This section gives a general, exemplary overview, the detailed formulas of the score and its factors are defined in section 4.1.3 and possible variations are presented in section 4.1.4. As argued above, the developed Ideal Path Score combines an interaction based factor and an eye tracking based factor. Eye tracking information is intended to give indications of thought processes, e.g. whether the user is guessing or intensely considering the relevant objects (see scenario in figure 4.4).

A score is calculated for each player action, i.e. for each move the player makes. Thereby each action is directly evaluated and a continuous model of player performance is created. Figures 4.2, 4.3 and 4.4 illustrate Ideal Path Scores in different game situations.

Ideal Path Score Definition

An score is calculated for each of the player's actions in game, rating the quality of it. In the case of the game SaFIRa, an action is a mouse click by the user, setting a new movement destination flag for the player avatar. Input for the computation of the Ideal Path Score is the action itself as well as eye movements since the previous action. Therefore information leading up to the decision of the then executed action is considered in each score.

Normalization makes the score intuitively interpretable and scores from different games comparable. The Ideal Path Score has a range of $[-1.0, 1.0]$ where a score of

1.0 indicates a perfect move

0.0 indicates a move without progress

-1.0 indicates a move after which the player is much worse off than before

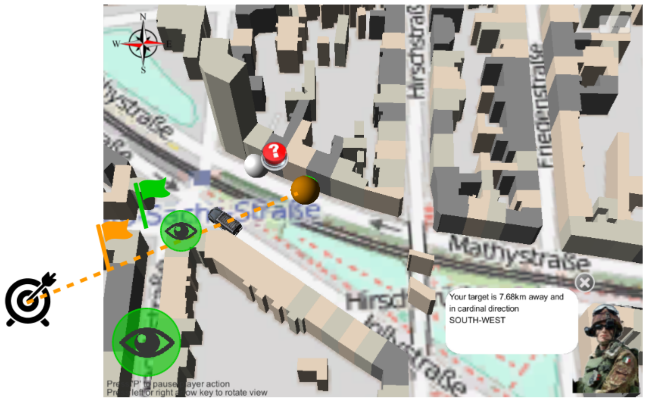


Figure 4.2: The target is outside of the screen area, the ideal move is marked by the golden flag. The attentive, skilled player knows the right direction from previous hints, looks at some points around that direction (green eye symbols) and moves there (green flag). The move results in a high Ideal Path Score as the destination is close to the ideal and fixations were on the relevant screen area close to the ideal movement point.

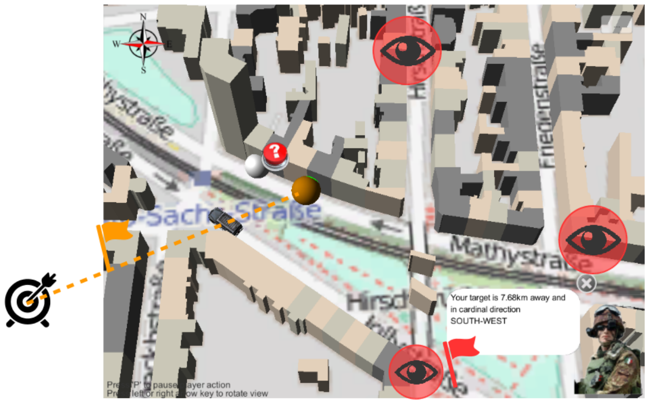


Figure 4.3: The target is outside of the screen area, the ideal move is marked by the golden flag. The confused or unattentive player looks in various directions on the screen (red eye symbols) before moving to a point (red flag) that is actually further from the target than the previous position. The move results in a negative Ideal Path Score as the destination brings the player away from the target and fixations were on screen areas far away from the relevant direction.

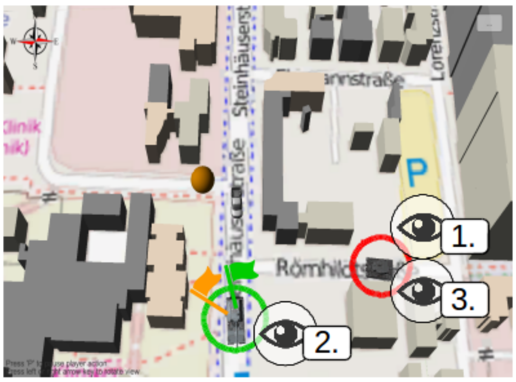


Figure 4.4: The target tank (golden flag) is visible on screen. The inexperienced player is unsure which tank he is the target and looks back and forth between the target and another tank. Due to a lucky guess the player decides to move to the correct tank. The move results in a mediocre Ideal Path Score - although the interaction based factor indicates a near perfect move, the eye tracking based factor shows that the player was unsure about the destination.

Move Factor and Gaze Factor make up the overall Ideal Path Score, representing the interaction based and eye tracking based information respectively. These factors are interpolated linearly:

$$\text{Ideal Path Score} = (1-\alpha)S_{move} + \alpha S_{gaze} \tag{4.1}$$

where $\alpha = 0.25$ is a weighting factor.

S_{move}, the factor based on explicit game actions, reflects how much progress towards the target is made relative to the maximum possible progress. It is calculated as

$$S_{move} = \frac{d_{previous} - d_{actual}}{d_{previous} - d_{ideal}} \tag{4.2}$$

where $d_{previous}$ is the distance to the target before the current action, d_{ideal} is the distance to the target from the best possible movement destination visible on screen and d_{actual} that distance from the actual movement destination of the current action. These distances are illustrated in figure 4.5.

S_{gaze}, the factor based on eye tracking information, consists of the scores for each fixation weighted by its duration relative to the duration of the action. It is calculated as

$$S_{gaze} = \sum_{f \in \text{Fixations}} \text{fixationValue}(f) \frac{\text{duration}(f)}{\text{duration}_{total}} \tag{4.3}$$

where "Fixations" is the set of all fixations recorded between the previous action and the current action, duration_{total} is the time span from the previous until the current action and $\text{fixationValue}(f)$, illustrated in figure 4.6, is defined as

$$\text{fixationValue}(f) = \begin{cases} 1.0 & \text{for relevant game objects} \\ -1.0 & \text{for irrelevant game objects} \\ \text{distance to ideal movement destination} & \text{for fixations on no object} \end{cases} \tag{4.4}$$

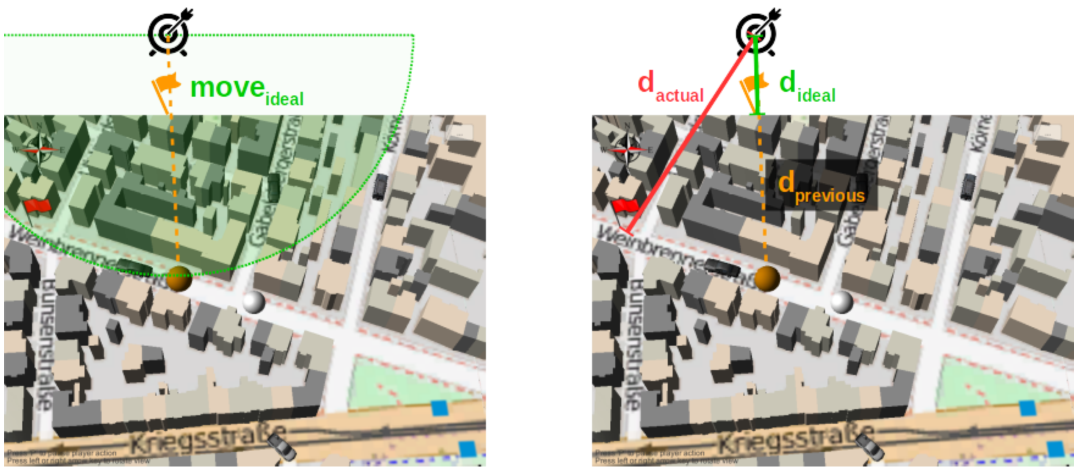


Figure 4.5: The interaction based "Move Factor" of the Ideal Path Score: Moves that bring the player closer to the target receive positive scores (green circle left), moves that lead away receive negative scores. The right image shows the distances used to calculate the score as defined in equation 4.2. The ideal move (golden flag) is the closest point visible on screen on a straight line to the target.

Figure 4.6: The eye tracking based "Gaze Factor" of the Ideal Path Score: Fixations close to the ideal move destination receive positive scores (green eye symbol left), fixations in irrelevant screen areas receive negative scores (red eye symbol left). Fixations on game objects are also considered (right image) - scores reflect whether the player looked at the relevant target object (green circle right) or other irrelevant objects (red circle right) as defined in equation 4.4.

Alternative Versions of an Ideal Path Score

The definitions above are the initial design of the Ideal Path Score and were implemented as such in SaFIRa for the experimental evaluation. Multiple variations and extensions of this design are possible. Some of these were also calculated on the collected data during the experiment and included in the evaluation in chapter 6.

Discretization

Different playing styles can lead to large differences in the Ideal Path Score with the definition above. Making a long move, with its destination close to the edge of the screen, leads to a much higher Ideal Path Score than several shorter moves because the ideal move, making the furthest possible progress, usually is at a screen edge. However, a player using short, quick moves is not necessarily playing worse.

This is particularly problematic when further interpretation relies on the mean Ideal Path Score.

To mitigate this problem, discretized variants of the "Move Factor" and the "Gaze Factor" were developed. Instead of a continuous score - the closer to the ideal move destination, the higher the score - only values of +1, -1 or 0 are assigned to an action or fixation as follows:

$$S_{move,discrete} = \begin{cases} 1.0 & \text{if new destination closer to target} \\ -1.0 & \text{if new destination further away from target} \end{cases} \tag{4.5}$$

S_{gaze} is still defined as it was originally in equation 4.3, with fixation values adapted as

$$\text{fixationValue}_{discrete}(f) = \begin{cases} 0.0 & \text{if f within 250px radius from screen center} \\ 1.0 & \text{if f within 600px radius from ideal movement destination} \\ -1.0 & \text{otherwise} \end{cases} \tag{4.6}$$

for the 1920x1200 pixel resolution of the screen used during the experiment, for which these Ideal Path Score varation was calculated afterwards. This discretized variant $S_{gaze,discrete}$ only considers the position of the fixation on the screen, not whether or what kind or game object is at that location in order to simplify the feature.

Possible Interpretations Based on Fixated Game Objects

The developed framework is able to identify most game objects (in SaFIRa these are primarily vehicles) the player looked at. The current component of the Ideal Path Score is rather naive and only assigns positive values to fixations on the target tank. The knowledge about what game objects were looked at could potentially inform many more specialized adaptivity mechanisms.

Monitoring help usage would be a possible application, for example. [Bie16] had reported in the original evaluation of SaFIRa that some players did not notice updates of the adaptively displayed hints. A simple adaptivity plugin could use the eye tracking based information to check whether a recently displayed hint has been read by the player and trigger a special highlighting of the hint if it is not noticed.

Complex patterns help to infer cognitive processes. A motivation for this is illustrated by the scenario presented in figure 4.4, detecting uncertainty or guessing of a player. Other publications report the use of gaze patterns between areas of interest to infer thought processes like discarding or considering answer choices (see chapter 2.3.1). Modelling such patterns for SaFIRa could lead to interesting insights, although such patterns will probably be highly task and game dependent and therefore difficult to generalize and transfer to other contexts.

To improve the model of "fixation values", i.e. whether the fixation on a specific game object should be rewarded or penalized, a semantic model of the game objects could allow powerful interpretation. Similar to semantically annotated learning content (see chapter 2.1.1), a hierarchical semantic model of SaFIRa's vehicle types could be created. Such a model could provide information whether a fixated vehicle is a tank or car - and based on the similarity to the target tank penalties or rewards for that fixation could be tuned.

Discussion

How to measure attention and goal-oriented progress?

This thesis is part of a research effort whose overarching aim is to improve adaptivity for serious games. A perfect metric would directly measure how successfull and efficient a user acts in relation to the goal of the game. My approach of the "Ideal Path Score" assumes that when a user closely follows the steps of the ideal path without further distractions this is the most direct way to reach the goal of the game and the learning objectives.

This model is in this regard somewhat limited by its focus on "efficient" game progress. It therefore disregards concepts of free play and exploration that can also be valuable modes of entertainment as well as learning.

An important aspect especially for serious games is the user's flow experience that drives intrinsic motivation and concentration. Originally my aim was to measure a user's flow experience in order to sustain it and use it as one factor for controlling adaptive interventions. As flow is a complex phenomenon the decision was taken to focus on estimating the user's attention rather than flow as this should be potentially easier to examine.

My original assumption was that the Ideal Path Score is able to imply the user's attention. However, after further discussions I found it to be unlikely that the Ideal Path Score directly correlates with user attention. For one thing, "attention" in this context first needs to be defined more clearly. A more appropriate term for what I attempt to measure could be "goal-orientedness" or "goal-directed gaze".

Also "attention" or "goal-directedness" likely is not the only factor influencing the Ideal Path Score. Deviations of a user from the ideal path can be explained by different causes: The user's lack of focus or goal-orientedness (i.e. his attention?) but also the user's lack of skills or knowledge to follow the ideal path.

The Ideal Path Score as a metric of overall progress should corellate with a user's skills (modeled through "classic" didactic factors) as well as with a user's "attention" (possibly estimated through data of eye movements or other indicators of distraction, e.g. user's pulse or environmental noise level).

From "goal-orientedness" to attention

Understanding attention as a limited resource needed for controlled cognitive actions, as described in 3.1.2, it can be expected that high goal-orientedness of the user requires sufficient attention. However, low goal-orientedness (i.e. a large distance to the ideal path) can also occur despite high attention of the user if the user is lacking skills or knowledge to complete the task efficiently.

This would suggest that a high Ideal Path Score indicates high attention, while low Ideal Path Scores would not necessarily be connected to low attention:

$$\text{good Ideal Path Score} \Rightarrow \text{high goal-orientedness} \Rightarrow \text{high/sufficient attention} \tag{4.7}$$

$$\text{bad Ideal Path Score} \Rightarrow \text{low goal-orientedness} \not\Rightarrow \text{low attention} \tag{4.8}$$

Generic Eye Movement Patterns

As an alternative or extension to the developed game-specific Ideal Path Score, generic eye movement features can be considered. Features such as mean fixation duration, fixation frequency, blink frequency or features based on pupil diameter change are commonly used in related work as presented in chapter 2.3.

These features are generic in the sense that they only rely on the eye tracking data without specific context information about the content that was viewed. This makes such features potentially easier to transfer between different applications. However, this can also limit their expressiveness. One publication reports that generic eye tracking features were not able to classify learning behaviour by themselves but improved classification in combination with content specific features [Bon13].

While I did not work to improve upon the existing approaches of such features within the scope of my thesis, I evaluated their possible contribution to estimate attention in the context of my work. The features selected for this were chosen based on my summary of common features in table 2.2, chapter 2.3. The analysis based on the collected data during my experiment is presented in chapter 6.4.4.

Integration into Adaptive Systems

This chapter described the details of design and definition of metrics and eye tracking based feature scores. These features serve as indicators about the progress and cognitive state of the user. For them to inform adaptivity decisions, such feature scores may need to be further interpreted by an adaptivity reasoning engine and appropriate adaptivity actions have to be taken based on them.

To achieve this, the features described here are integrated into the existing adaptivity reasoning system ELAI. The details of this integration are described in chapter 5, with details of the actual adaptivity reasoning in section 5.4.

5

Attention Adaptivity Framework

To make eye tracking data usable for adaptivity in e-learning, components to collect eye tracking data, analyze and annotate this raw data, generate interpretations about the learner and influence the actual adaptivity interventions have to be brought together. The developed "Attention Adaptivity Framework" integrates these components in a structured and extensible way.

The ELAI concept, as described in chapter 3.5.1, represents the foundation for the Attention Adaptivity Framework. Through its general ELAI server it allows flexible interoperability between multiple applications and easy integration of new systems. My framework is based on the ELAI implementation developed specifically for SaFIRa [Bie16]. For easy reuse in other applications I modularized this implementation, which was tightly integrated into the SaFIRa Unity game (see figure 5.1). It provides an interface for plugins on the client-side, which can collect and analyze additional data e.g. from eye tracking, as well as a plugin concept on the server-side to adapt and extend ELAI's interpretation and influence engines. An "ELAI Eye Tracking Plugin" (section 5.3) and server-side interpretation and influence plugins (section 5.4) were implemented to realize the Ideal Path Score approach presented in chapter 4.

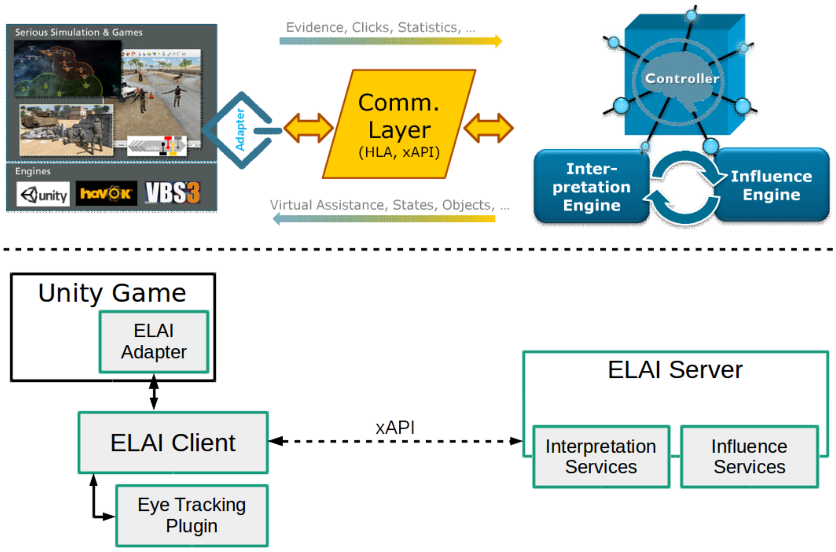


Figure 5.1: The components of the developed Attention Adaptivity Framework (bottom) and the conceptual ELAI system [Str15c] that served as a foundation for its design.

Objectives and Non-Functional Aspects

One of the core concerns of ELAI is the interoperability of e-learning adaptivity between multiple systems. As the developed Attention Adaptivity Framework functions as an extension of the ELAI concept, reusability and generalizability are important aspects for the framework in order for it to be applicable in this broad intended context for different applications.

Reusability in different applications, the ability to implement game-specific interpretations and the extensibility of the general framework to other scenarios not necessarily related to attention or eye tracking are important objectives that guided the design of the Attention Adaptivity Framework.

Reusability is a common concern in software engineering and helps to keep a software ecosystem managable and maintainable. As the ELAI concept specifically aims to integrate multiple e-learning systems, it is to be expected that a lot of basic functionality like communication with an ELAI server is required in a large number of applications.

Therefore a general library should provide these basic functions to ease integration of new applications into the ELAI ecosystem. A library can also keep distribution of potential future changes managable because code is not duplicated and only needs to change in one place. Modularization of the framework isolates dependencies and changes and allows to combine features as needed.

The client-side Attention Adaptivity Framework is available as a collection of plain C# dynamic-link libraries (DLLs) with minimal depencies. They can be used from any .NET compatible game engine or plattform.

The ELAI client library as well as the eye tracking plugin are configurable with a xml file. This allows for example to change the used eye tracking device without the need to recompile the application.

Game-specific interpretation is important despite the focus on building a general and reusable framework. The Ideal Path Score represents a general performance metric that could allow comparison between very different contexts. However, the definition of "ideal paths" and distance measures to such paths necessarily needs to be application-specific for Ideal Path Scores to carry any meaningful insights.

Abstraction through interfaces and the use of approaches like the Strategy Pattern allow an application to provide specific algorithms to the generalized Attention Adaptivity Framework, e.g. for calculation of ideal path scores, while minimizing development effort.

Extensibility of the Attention Adaptivity Framework makes it a feasible foundation for further extensions of the ELAI concept. While I implemented an eye tracking plugin with the aim of modeling user attention and Ideal Path Scores for this thesis, the framework itself is not tied to approaches related to either eye tracking or attention.

The modular and plugin-based design of the framework easily allows further or different extensions. The main ELAI Client and ELAI Server system only provide the basic core functionality and can be extended as needed by independent plugins to customize data collection, interpretation and adaptivity influence.

Architecture

The conceptual architecture of the ELAI system as proposed by [Str15c] and its implementation by [Bie16] consists of an "ELAI Interface" component that is integrated in the specific serious game or e-learning application and the server-side ELAI comprising an "Interpretation Engine" and an "Influence Engine". The architecture of my Attention Adaptivity Framework specifies this structure in more detail and further modularizes it (see figure 5.1).

Overview

The ELAI Client library takes over general client-side functions regarding ELAI, like communication with the ELAI server (see figure 5.1). An "ELAI Client" instance is created and used by an e-learning application's specific "ELAI Adapter". "ELAI Client" plugins (e.g. the "ELAI Eye Tracking Plugin") are also started explicitly by this ELAI Adapter to allow deeper, context-specific integration of these plugins through interactions with the ELAI Adapter.

Communication to the ELAI Server is performed through xAPI statements, thereby following a clearly defined interface to decouple client and server components. The actual networking details are managed by the ELAI Client library while statements are provided or triggered by ELAI plugins or the game's or application's ELAI Adapter. The ELAI server receives xAPI statements, stores them and passes them on to all active Interpretation Services. These can further analyze the information and also generate additional statements containing their interpretations (see sequence diagram, figure 5.2).

The ELAI Client also polls the ELAI server regularly to get updated learner models or influence parameters which are then available to the ELAI Adapter. Actual adaptivity is implemented in the specific application to be appropriate for its context. These implementations are controlled through the ELAI Adapter using the influence parameters provided by the ELAI server. The ELAI server's Influence Engine collects these influence parameters from all active Influence Services wich compute them based on previously stored statements and interpretations. This is illustrated by the sequence diagram in figure 5.3.

Server-side Interpretation and Influence Services are deployed as plugins in independent .jar files thereby providing flexibility to configure the ELAI server in a modular way.

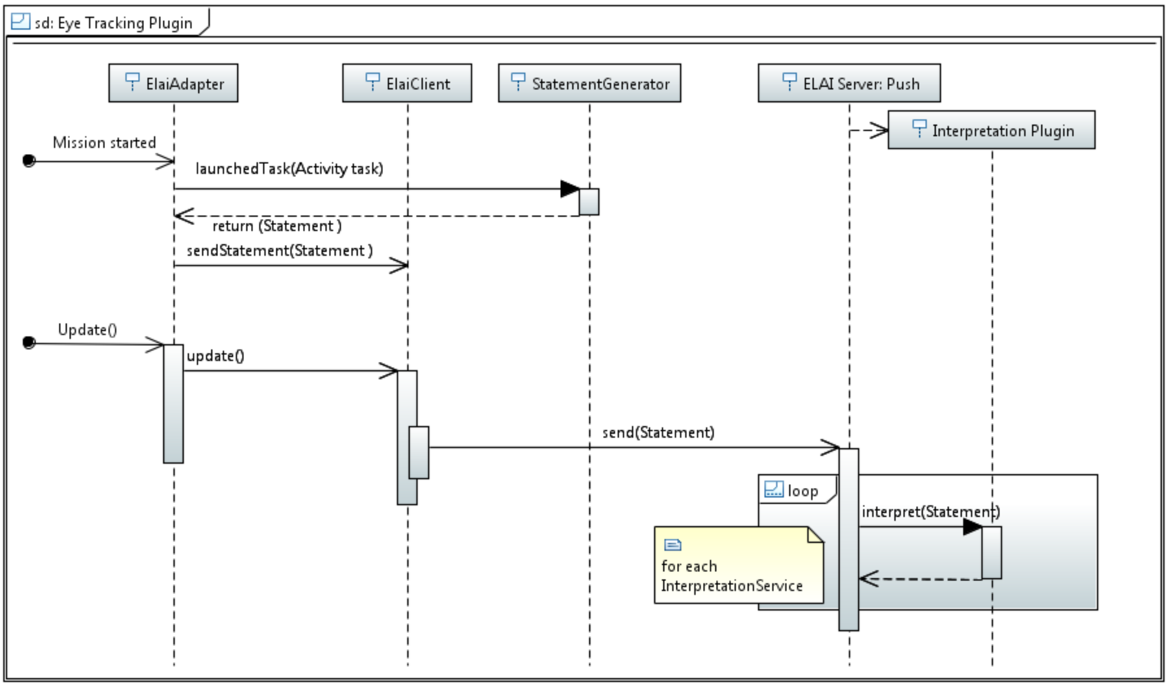


Figure 5.2: The application's ElaiAdapter reports game events ("Mission started") by generating an xAPI statement, supported by the StatementGenerator factory of the ELAI Client library. This statement is passed to the ElaiClient instance, which handles the transmission to the ELAI Server at the next update interval. On receiving the statement the server's Push Servlet passes it on to each available Interpretation Plugin for further classification.

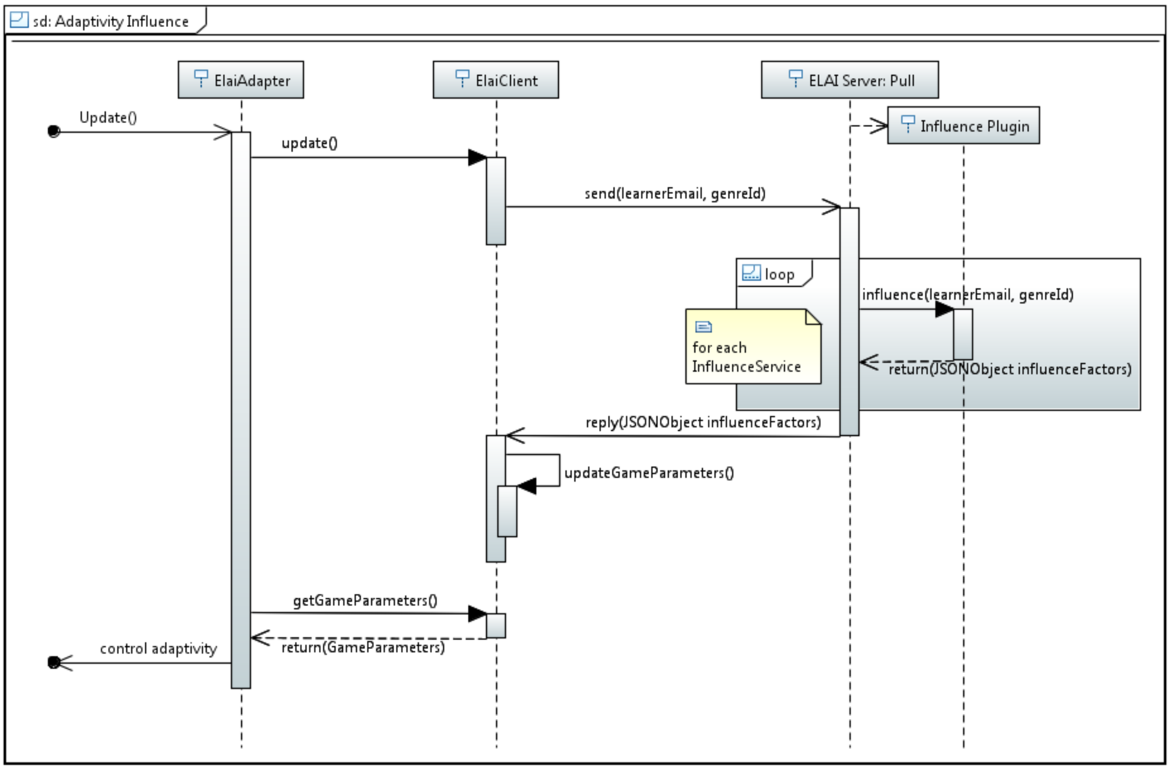


Figure 5.3: The ElaiClient on regular update intervals requests the latest influence parameters from the ELAI Server by sending identification data for the current learner ("learnerEmail") and game genre ("genreId"). The ELAI Server's Pull Servlet then requests all available Influence Plugins to compute their influence factors and returns these to the ElaiClient. The received parameters ("GameParameters") are constantly available to the application's ElaiAdapter to control its adaptivity mechanisms.

ELAI Client and Adapter

The ELAI Client library encapsulates the functionality common to most applications interacting with an ELAI server. As a separate C# dynamic-link library (DLL) it can easily be referenced and used by such applications. It manages the details of network communication with the server, supports composition of proper xAPI statements and offers logging and file configuration. The details and implementations of those features are hidden from the application using the library, allowing simple use and evolution of the ELAI Client library.

The game or e-learning application intending to use it simply needs to create an *ElaiClient* instance. To respect the principle of separation of concerns, this should be done in a class separate from other application code, here called *ElaiAdapter*.

The Adapter pattern suggests wrapping a existing class into an "adapter" to receive a compatible interface [Fre04]. The application-specific *ElaiAdapter*'s role is to adapt the interfaces of the ELAI Client library to the requirements of its application. The *ElaiAdapter* can also instanciate and configure ELAI Client plugins. In this case it more closely resembles a "facade" than an adapter, providing a simplified interface to the whole ELAI Client subsystem for an application [Fre04].

Because the library contains all common functionality the *ElaiAdapter* only requires a small amount of custom code, keeping the effort of integration of an application with ELAI minimal. Figure 5.4 shows this architecture in detail.

An XML configuration file allows changes to settings (e.g. the ELAI server address or which eye tracker model is used) without need to recompile the project. This particularly eases portability between different machines and contexts.

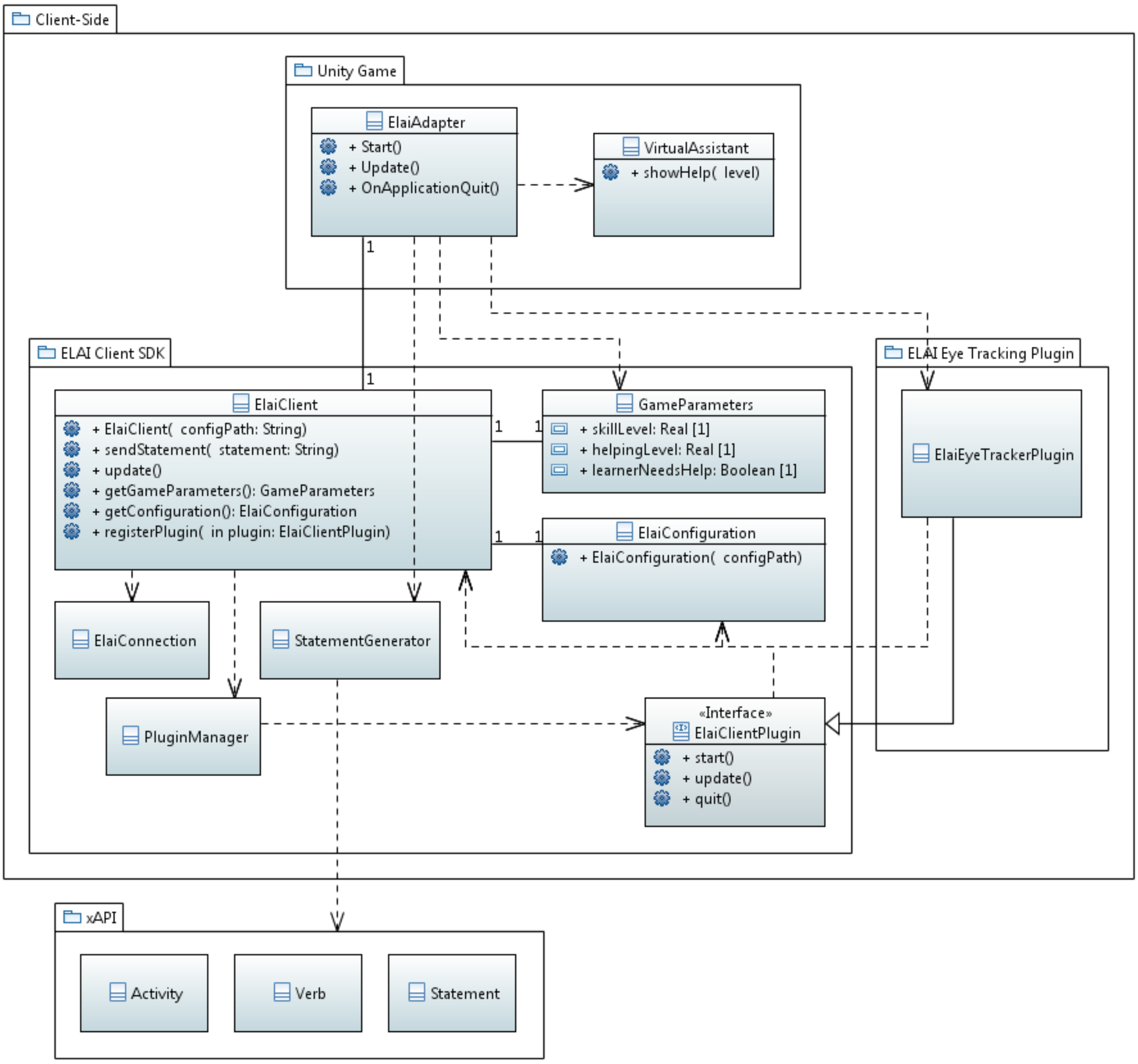


Figure 5.4: Details of the client-side components of the framework: A reusable ELAI Client library and its plugins are simply wrapped and controlled by a game specific ElaiAdapter.

ELAI Client plugins are distributed as separate libraries and can provide additional functionality, particularly to collect other data about the learner's behavior and progress. The ELAI Client library is kept small and provides only core functionality but extensions for use cases like the collection and analysis of eye tracking data can be added and used by any application as needed.

These plugins are not automatically discovered and loaded but instead instanciated by the application specific *ElaiAdapter*. This extra step is intentional to enable each application direct configuration of the plugin. The Strategy pattern is used in these cases, i.e. an implementation of an interface is passed in order to define a specific behavior of the other system [Fre04]. By this approach a plugin can receive application specific strategies to interpret its observations (e.g. the Eye Tracking Plugin can call back to the application to learn which game object is currently at the given screen coordinates the learner looked at).

ELAI Client plugins are built on top of the ELAI Client and use its configuration file and logging features. They are registered with the ELAI Client instance by the ELAI Adapter and from then on their lifecycle (i.e. update intervals and shutdown) is managed automatically by the ELAI Client. The relationship and interaction between ELAI Client and a plugin is illustrated in figure 5.5.

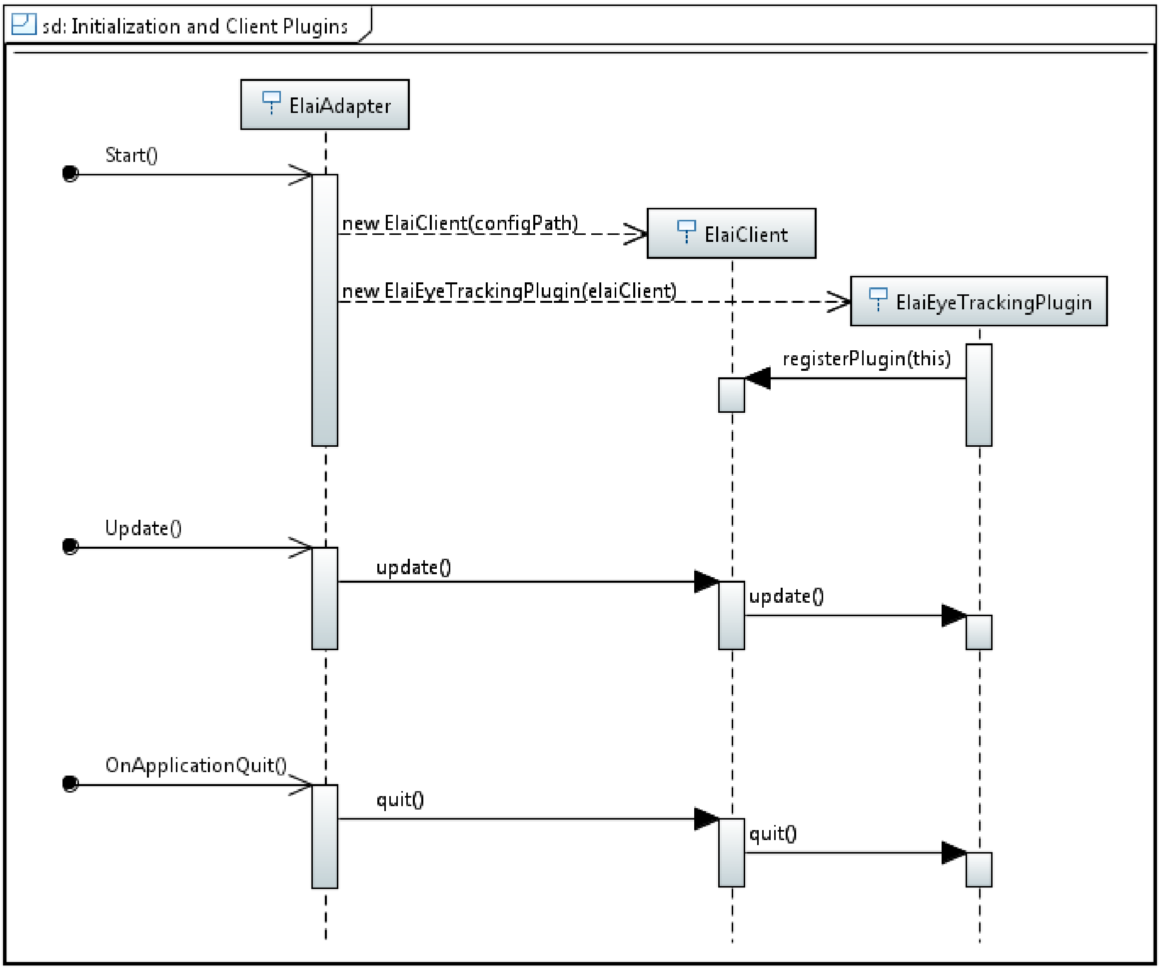


Figure 5.5: ELAI Client plugins are managed by the ElaiClient instance they register with.

ELAI Server

The core ELAI Server provides a "Push" and a "Pull" Servlet for clients to push information about activities of the learner to the server and request computed influence factors to control adaptivity. Updates from e-learning applications are transmitted as xAPI statements. These statements are passed to the "Interpretation Engine" of the ELAI server to be analyzed. All statements are also stored in a Learning Record Store (LRS) to be available for later computations by the "Influence Engine".

The ELAI server is implemented using Java Servlets and runs on an Apache Tomcat web server.

Plugins are taking care of all interpretation and influence computations. The core ELAI server only calls those plugins, stores their interpretation results or transmits their returned influence parameters to the ELAI Client. This ensures separation of concerns for the request handling and interpretation/influence computation in order to keep this complex system maintainable. In addition, this design also allows very flexible configuration and extension of an ELAI server.

The sequence diagrams in figure 5.2 and figure 5.3 illustrate this delegation to Influence and Interpreation Services. Figure 5.6 shows the components of the ELAI server in detail.

The ELAI server plugins located in the defined plugin directory in the filesystem are detected and loaded dynamically at runtime. Plugins are deployed as .jar files containing one or more implementations of the *InterpretationService* or *InfluenceService* interface. A specific configuration file included in the .jar package identifies these classes as plugins to be loaded. These .jar files are then loaded by the server using the Java ServiceLoader API (see chapter 3.7). No changes or configuration in the core ELAI server is necessary when plugins are added or exchanged.

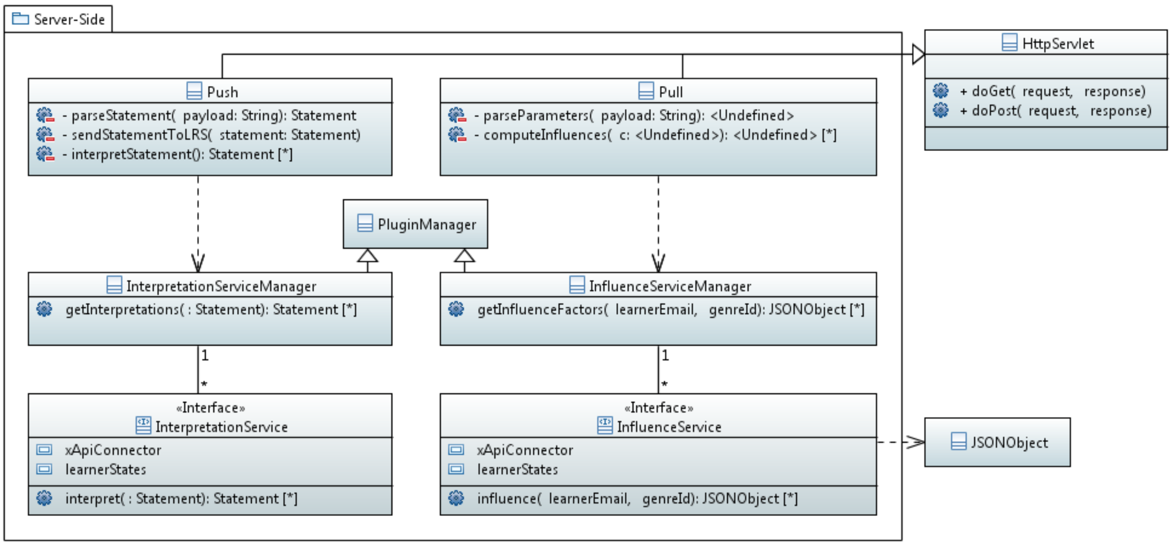


Figure 5.6: The ELAI server's role in the framework is to coordinate Interpretation Service and Influence Service plugins which are responsible for the concrete adaptivity related computations.

Interpretation Services

An Interpretation Service interprets the learner's activities as part of the ELAI Server's Interpretation Engine. It is called every time a statement arrives at the server (also see figure 5.6). The Service receives this statement and also has access to the LRS containing previous statements and interpretations. It then classifies this statement and optionally generates an "interpretation" that is either saved in the database or stored as one or more additional xAPI statements in the LRS.

The interpretation mechanisms implemented for the game "SaFIRa", which were used for the experiment evaluating this thesis, are presented in section 5.4.

Influence Services

An Influence Service as a component of the ELAI server's Influence Engine computes one or more parameters that make up the response to a request for influence factors by an ELAI Client. The Influence Service has access to the database and LRS containing all previous statements and interpretations (see figure 5.6).

All values calculated by Influence Services are sent to the ELAI Client in JSON format and can be used by the e-learning application to control its adaptivity.

The influence parameters implemented for the game "SaFIRa", which were used for the experiment evaluating this thesis, are presented in section 5.4.

Eye Tracking Plugin

The collection and analysis of eye tracking data is encapsulated in an ELAI Client plugin, the "ELAI Eye Tracking Plugin". The plugin records fixations reported by an eye tracker that is connected to the user's computer. The integration of the eye tracker is abstracted by a generic eye tracker interface and the plugin provides implementations for different eye tracking devices. Which eye tracking adapter is used can be switched seamlessly through the local ELAI configuration file, thereby allowing easy portability between different machines. Alternatively, the ELAI Adapter can provide its own eye tracking adapter implementing the required interface, which makes it possible to use any eye tracking device with the framework.

In order to allow more meaningful interpretations of the eye tracking data based on knowledge what objects were looked at, screen coordinates of fixations are passed to the specific application for classification. An implementation of the *RelevantObjectsLocator* has to be provided by the application specific ELAI Adapter during initialization. This implementation checks what object is positioned at the given screen coordinates in the application and returns this semantic information in the form of an xAPI Object to the eye tracking plugin. Such an interaction to identify game objects is illustrated in figure 5.7 and follows the idea of the "Strategy Pattern" [Fre04].

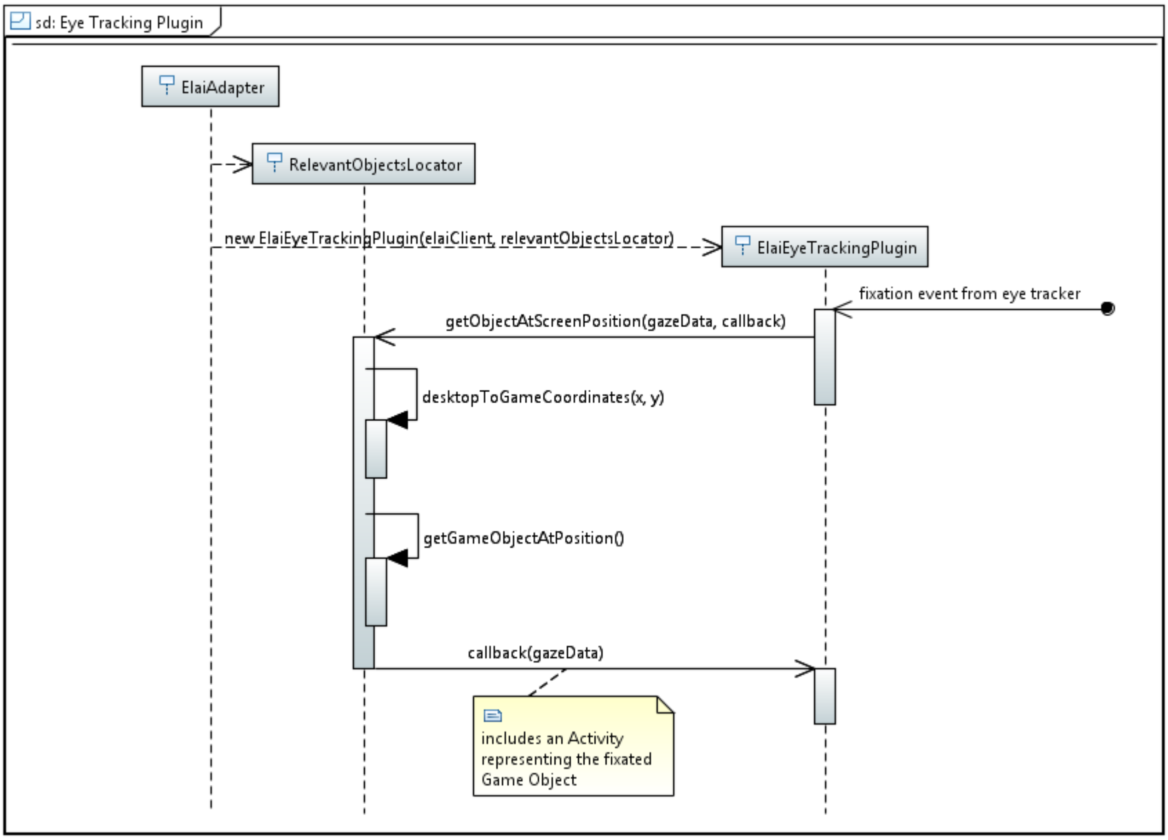


Figure 5.7: An application-specific "RelevantObjectsLocator" strategy is passed to the EyeTrackingPlugin which helps to identify the looked at objects.

Information about the learner's gaze could be transmitted to the ELAI Server in the form of xAPI statements. However, it is easier to interpret eye movements on the client-side through the ELAI Adapter if these interpretations are specific to only one application. The ELAI Server benefits from information that is applicable across multiple contexts. Therefore, the ELAI Eye Tracking Plugin only transmits higher level interpretations of the eye tracking data in the form of Ideal Path Scores to the server.

There may be more detailed gaze information that is useful for the ELAI server's reasoning, e.g. information about the learner looking away from the current application or information about the learner reading the provided help message. In this case future extensions of the Eye Tracking Plugin could sent xAPI statements indicating such specific activities to the ELAI server.

Eye Tracker data is handled using a simple eye tracking framework developed by Fraunhofer IOSB, which was translated from its original version in Java to C# for the ELAI plugin. Fixations are calculated using a dispersion threshold algorithm with a dispersion threshold of 50 pixel.

Adaptivity Reasoning

Reasoning for appropriate adaptivity is application specific, the computations described here are targeted at the serious game SaFIRa (see chapter 3.6). Putting adaptivity measures into practice with the Attention Adaptivity Framework is a four stage process (see figure 5.8) following the adaptive cycle (see chapter 3.4.3).

First the observations on the learner are collected and aggregated on the client-side by the ELAI Adapter (here by computation of an Ideal Path Score, see section 5.4.1) and transmitted to the ELAI server.

Then the ELAI server's Interpretation Services can further interpret this information (here by classifying the player's skill level, see section 5.4.2).

On request the ELAI server's Influence Services compute the influence parameters to be transmitted back to the e-learning application (here by calculating the player's need for immediate help, see section 5.4.3).

The concrete presentation of adaptivity measures like the display of a help message is implemented by the specific e-learning application or serious game client-side and controlled by the influence parameters from the ELAI server.

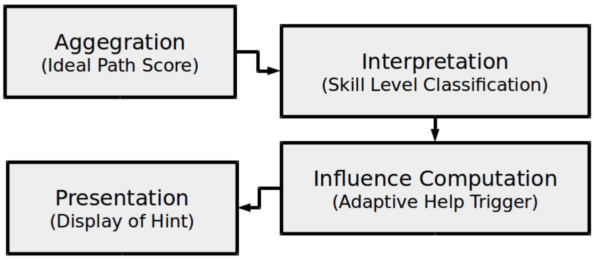


Figure 5.8: Adaptivity is realized in four stages: Aggregation and Presenation on the left are performed client-side, Interpretation and Influence Computation on the right are performed on the ELAI server.

Calculation of Ideal Path Score

The Ideal Path Score aggregates information about eye movements and player actions into a single score that quantifies the player's performance as presented in chapter 4. This score can be interpreted generically by the ELAI and is the primary way how the Eye Tracking Plugin influences adaptivity decisions. The Ideal Path Score replaces the target distance previously used by [Bie16] as a measure of performance for a learner's in-game actions.

For each player action the score is calculated and sent to the ELAI Server in an xAPI statement (see listing 5.1).

```
"actor": { ... },
"verb": {
    "id": "http://iosb.fraunhofer.de/ias/ibis/elai/xapidefinitions/
        verb/played/"
},
"result": {
    "score": {
        "scaled": 0.39908025609066
    },
    "duration": "PT1.33S",
    "extensions": {
        "http://iosb.fraunhofer.de/ias/ibis/elai/xapidefinitions/
            distanceToTarget": 0.8674978017807
    }
}
```

Listing 5.1: Excerpt of an xAPI statement: The Ideal Path Score of a move is transmitted in the statement "result".

Skill Level Classification

A new "skill level" of the learner is estimated after each completed task, i.e. for the game SaFIRa after each game session which is completed by finding the target tank. This estimated skill level is in turned used by the game to set an appropriate difficulty level. The skill level is classified using a linear classifier combining a number of "Didactic Factors". The classification approach remains the same as in the original system of [Bie16]. The following Didactic Factors are used:

- Task Duration (the time the player took to reach to target tank)
- Difficulty Level
- Number of Hints displayed
- *Average Ideal Path Score* (only for the new, improved adaptivity system)

Following Biegemeier's approach [Bie16] each Didactic Factor (x_i) is normalized based on the minimum and maximum value for this Didactic Factor previously observed from any player:

$$normalize(x_i) = \frac{x_i - min(X)}{max(X) - min(X)} \tag{5.1}$$

The skill level is then estimated as a weighted sum of all Didactic Factors (x_i):

$$\text{skillLevel} = \frac{\sum_i x_i w_i}{\sum_i w_i} \tag{5.2}$$

To improve the skill level classification with my Ideal Path Score approach, I introduced an additional Didactic Factor: The mean Ideal Path Score of all actions of the game session.

Adaptive Help

To detect when the player needs help [Bie16] uses the changing distance of the player to the target. If this distance monotonically increases over a time span of 10 seconds, the related Influence Service triggers an automatic help message to be displayed.

To improve this approach I use my developed Ideal Path Score rather than the target distance as an indicator, thereby also including information about eye movements. If the average Ideal Path Score during the last 10 seconds is negative, an automatic help message is triggered.

Use Case

Figures 5.9, 5.10 and 5.11 demonstrate the use of the Attention Adaptivity Framework in practice for a scenario of the SaFIRa game.

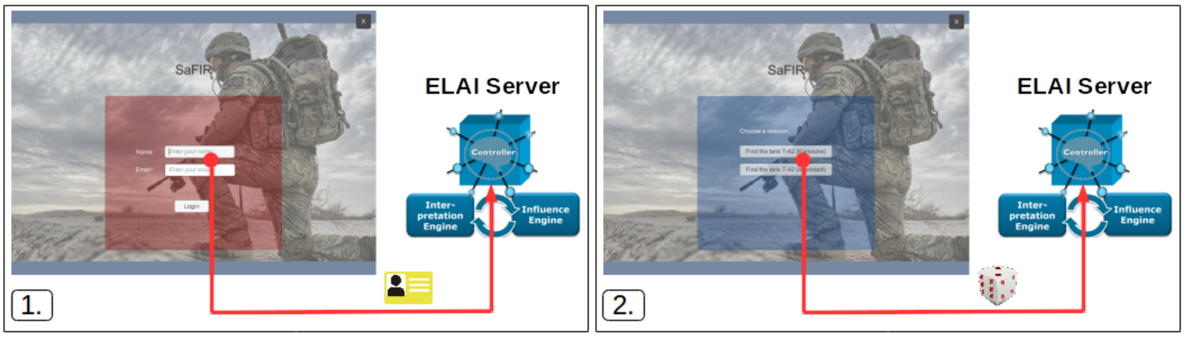


Figure 5.9: (1.) Login information is transmitted to the ELAI Server to uniquely identify the player in order to match his existing learner model. (2.) On starting a mission, the selected game genre is also sent to the ELAI Server to activate the relevant adaptivity mechanisms.

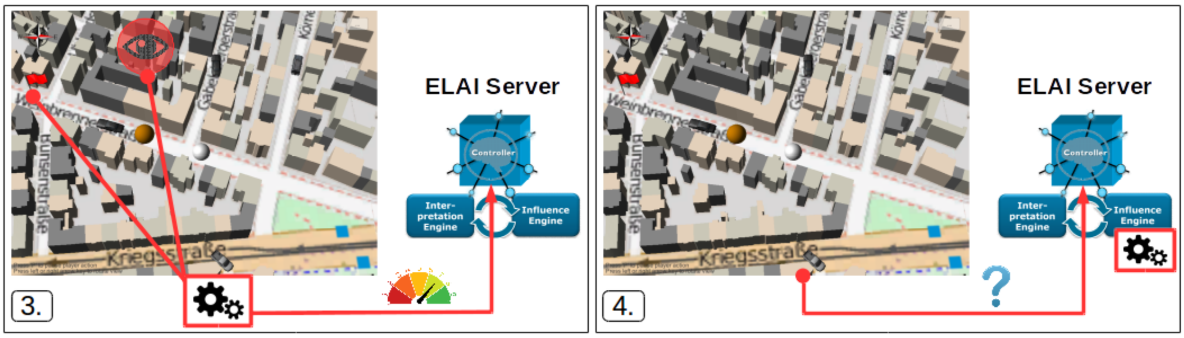


Figure 5.10: (3.) During play the game computes an Ideal Path Score for each move and continuously transmits them to the ELAI Server. (4.) Regularly the game requests updated influence parameters from the server, which triggers their computation by the server's Influence Engine.

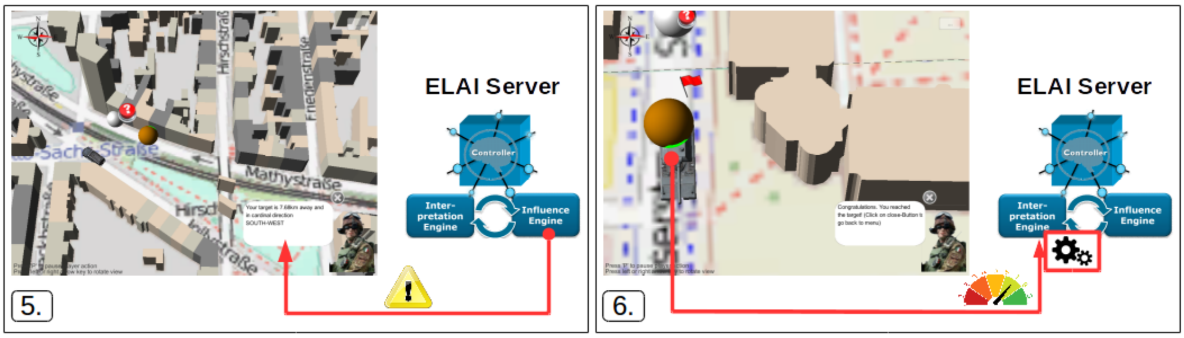


Figure 5.11: (5.) In reply to a request from the game, the server sends updated influence parameters indicating that the player needs help. This causes the game to display an automatic hint. (6.) Supported by the virtual assistant's hints the player reaches the target. Successfull completion of the mission is reported to the ELAI Server and leads to an reclassification of the player's skill level by its Interpretation Engine.

6

Evaluation of the Ideal Path Score Approach

To evaluate the metrics defined in chapter 4, an experiment was conducted. The objectives of the experiment were twofold: (a) To evaluate the Ideal Path Score that was designed in advance and (b) To collect data from a realistic setup for posterior analysis and improvements.

The collected data was used to evaluate and afterwards look for patterns and possible improvements regarding the Ideal Path based approach defined in chapter 4.1 and also to examine the power of other, more generic eye tracking features described in chapter 4.2. This posterior analysis of data uses the experiment as a pilot study and data collection rather than a rigorous evaluation with a system strictly fixed in advance of the evaluation.

After describing the hypotheses (6.1) and the experimental setup (6.2) and procedure (6.3) the results are presented in detail. A general overview of the collected data and questionnaire responses is presented in section 6.4.1 and provides context for the following analysis. The two primary questions of my work are then evaluated: What improvements does the Ideal Path Score deliver for adaptivity decisions (section 6.4.2)? And what correlations exist with user attention (section 6.4.3)? Other generic eye tracking based features are also evaluated (section 6.4.4).

Finally, the experiment and its results are discussed more broadly in section 6.5.

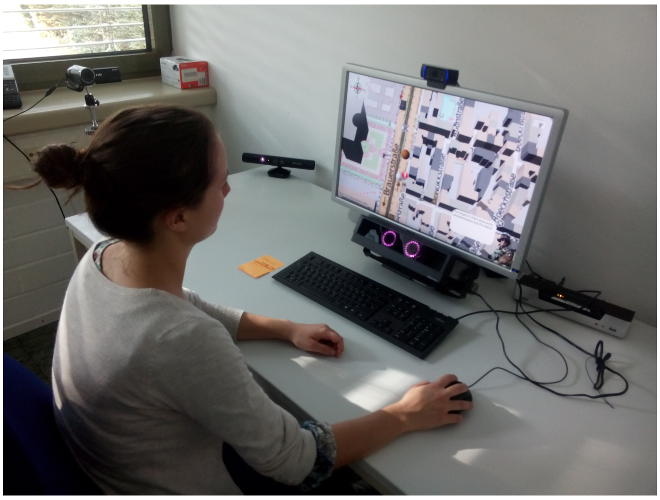


Figure 6.1: A participant of the experiment playing the game SaFIRa.

Hypotheses

The assumption prior to the experiment was that the Ideal Path Metric as designed in chapter 4.1 is able to infer a learner's attention and therefore lead to more appropriate adaptivity decisions of the ELAI system. Apart from observations on the overall approach and its influencing factors, the formally stated hypothesis in table 6.1 were examined.

	Hypothesis
H1.1	Ideal Path Score (IP) correlates with self-reported attention
H1.2	IP correlates with self-reported goal-orientedness
H1.3	The eye movement component of IP leads to higher correlations than a naive, interaction-based IP
H2	Generic eye movement patterns (described in chapter 4.2 and derived from related work in ch 2) correlate with self-reported attention
H3	IP can improve adaptivity decisions
H4.1	The computed "skill level" increases when playing multiple times
H4.2	IP also reflects / contributes to this increase of skill level
H4.3	Learning about different tank types is achieved by playing the game

Table 6.1: Hypothesis to be examined in the course of the experiment

Experimental Setup

The existing serious game SaFIRa, presented in detail in chapter 3.6, was extended with the ELAI adapter and eye tracking plugin, which collect sensor data, semantically annotate it, send it to the ELAI server and implement adaptivity interventions based on the recommendations received from the ELAI server. This functionality as part of the framework was described in detail in chapter 5.

Participants were alternatingly distributed to play a version of SaFIRa with its original adaptivity influence factors from [Bie16] (hereinafter called "legacy" system) and a version using the ideal path based adaptivity influence factor as presented in chapter 4 (hereinafter called "Ideal Path" system). This setup allows the comparison of the new overall system in practice. However, Ideal Path Scores were also calculated by the legacy system and written to log files for posterior analysis. For general analysis of eye movements and Ideal Path Scores the data from all participants is therefore used together, distinctions are only made when discussing adaptive hints and difficulty level which depend on the selection of the "legacy" or "ideal path" variant of the system.

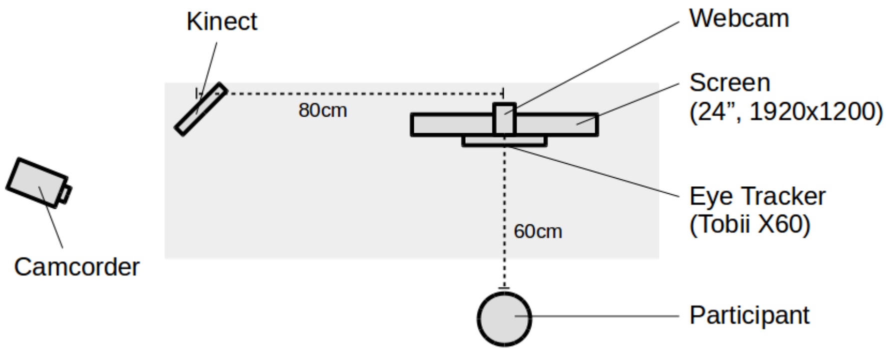


Figure 6.2: Experimental Setup

For the experiment, a Tobii X60 eye tracker was used, which was installed below a 24" monitor with a resolution of 1920x1200. At the beginning of the experiment the participant adjusted his seat to make sure he was in range of the eye tracker. The eye tracker was then calibrated for the participant.

In addition to the eye tracker the experiment was recorded using a Kinect (XBox 360) and a webcam. Possibly, additional non-invasive features could be designed which exploit these sensors, however detailed analysis of the data was not part of this thesis. To support manual analysis, the experiment was also recorded using a camcorder. The complete setup of sensors is shown in figure 6.2, a participant during the experiment can be seen in figure 6.1.

Experimental Procedure

Other publications examining engagement in digital entertainment games most often employed surveys in their research design, physiological responses to gameplay are typically analysed using quasi-experiments, comparing responses to different kinds of games or game events [Boy12]. Studies on games for learning generally preferred quasi-experimental designs [Con12].

For the evaluation of this thesis I chose a survey based approach, asking participants to self-report their subjective levels of attention and goal-orientedness. The responses to this questionnaire were examined for correlations with the data collected, namely Ideal Path Scores and eye tracking data.

In addition to the primarily survey based approach participants were divided into two groups and did the experiment using either a "legacy" version of adaptivity without eye tracking influence or an "ideal path" version of adaptivity which also makes use of eye tracking information. Comparison between those two groups leads to some conclusions about the overall system in practice.

Procedure

The procedure of the experiment for a participant is illustrated in figure 6.3. After calibration of the eye tracker, an initial calibration test (see 6.3.2) was conducted. The participant then had time to read the introduction included in the game, which explains game controls and objectives.

The actual experiment consisted of 5 game sessions after each of which the participant answered a short questionnaire asking for subjective ratings of attention during the previous session (see 6.3.3).

At the end of the experiment, the participant had to answer a concluding questionnaire (see 6.3.3), which checked learning outcomes and recorded demographics. Finally another calibration test was conducted to check eye tracking data quality.

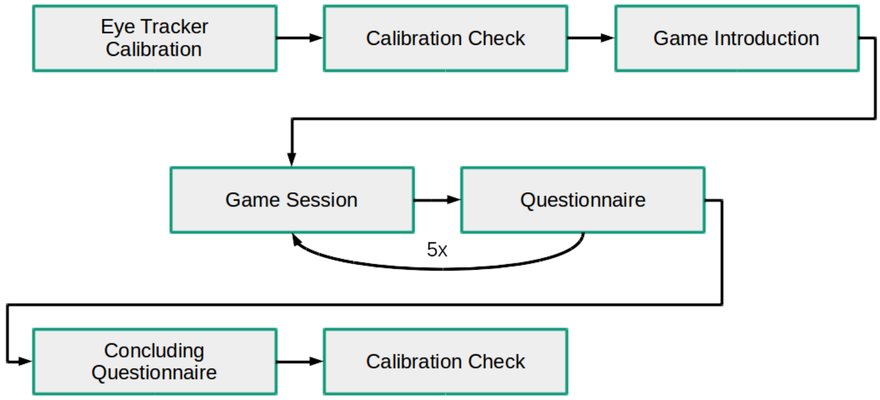


Figure 6.3: The procedure of the experiment for a participant.

Eye Tracking Calibration Tests

The eye tracker was calibrated through the provided third-party tool using 9 test points at the beginning of the experiment for each participant.

As proposed by [Hol12] and summarized in chapter 3.2.3, it is important to check and report the quality of recorded eye tracking data. I therefore test the eye tracker accuracy and precision for each participant right after initial calibration as well as at the end of the experiment. This is done for 6 points at various positions of the screen as shown in figure 6.4.

During this test, a single point is displayed on the otherwise completely black screen. To give the participant enough time to fixate the new point, the eye tracking samples are only recorded starting 1.5 seconds after the point appeared on screen. After a total of 4.0 seconds (1.5 seconds search time + 2.5 seconds recording of samples) the next point is displayed.

The distance (in pixel) of samples from the calibration point is then calculated. Raw gaze position samples rather then computed fixations are used in the calculations.

The results of these data quality measures are presented in the results section in table 6.5.

Figure 6.4: Result of the eye tracking calibration test: Large points represent the fixed calibration points used, clouds of small dots visualize the recorded gaze samples of a participant. During the test only a single point is displayed at a time without any lines.

Questionnaire

During the design of the questionnaire for my experiments I considered best practice guidelines and experiences from studies of related work. One objective also was to keep the questionnaire short and simple as participants needed to answer it multiple times, once for each game session.

Guidelines for Questionnaire Design

In general questionnaires can be problematic because they rely on the participants subjective opinions and understanding of terms and topics, which may distort results [Hol16], [Jen08]. Jennet et al. [Jen08] therefore use a "single question measure of immersion rating (i.e. rate how immersed you felt from 1 to 10)" and a "multiple question measure of immersion (a mixture of questions combining aspects of flow, CA and presence)" together. Pairs of questions are phrased "using negative and positive wording in order to control for wording effects". Questions were rephrased based on experience from previous experiments.

Wording and order of questions is important because participants tend to choose first (or last) option, and tend to agree to questions [Hol16]. It is also recommended to reuse questions from similar studies as this also allows comparability of results [Hol16].

Questionnaire on Attention

Participants of the experiments had to answer a short questionnaire self-assessing their attention after each of the five game sessions they played. Asking participants for an assessment immediately after the experience is similar to the "Experience Sampling" method applied to measure flow experiences [Nak02].

The questions of my questionnaire are in part based on [Jen08], who developed an extensive questionnaire to measure immersion and revised the wording of their questions after user feedback to be simpler and more concise, e.g. by labeling the ends of the scale descriptively rather than using the standard Likert scale approach to label them "strongly disagree" and "strongly agree".

All questions were to be answered on a 8 point scale labeled "0 ... 7" with a descriptive term at either end of the scale. Figure 6.5, a screenshot of part of the original questionnaire, exemplifies the style of these scales.

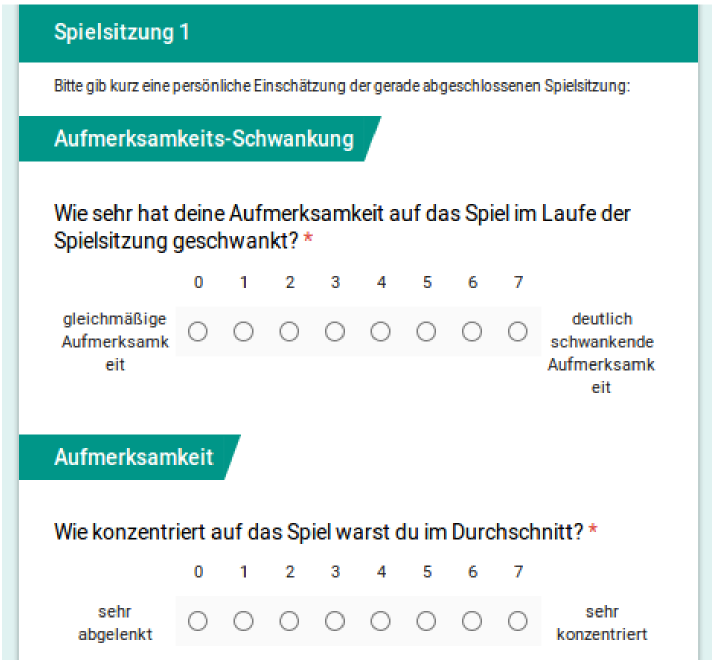


Figure 6.5: Excerpt of the questionnaire illustrating the design of questions with labeled scales.

Table 6.3 shows a translation of the questions and terms used, which can be found in the appendix A1 in its original German version.

The first three questions check for the participants attention. As this is the variable I am most focused on it is tested with two questions (on "attention" and "distraction"), the second question being phrased to ask for non-attention. To keep the questionnaire, which has to be filled repeatedly, short and simple, other aspects are not inquired with additional control questions.

As the Ideal Path Score is conceptually a measure of the goal-orientedness and overall performance of a user, participants are also asked for their self-assessment of these aspects for comparison.

The questions on "Virtual Assistant" and "Difficulty" directly relate to the tuning of the game's adaptivity and are used to evaluate hypothesis H3.

Titel	Question	Label Min (0)	Label Max (7)
Attention Variation	How much did your attention vary during the game session?	even attention	significantly varying attention
Attention	How concentrated on the game have you been on average?	very distracted	very concentrated
Distraction	How much did other thoughts and things in your surroundings engage you during the game session?	not at all	significantly
Virtual Assistant	How did you perceive the number of hints from the virtual assistant?	too few hints	too many hints
Difficulty	As how difficult did you perceive the game?	very easy	very difficult
Goal-Orientedness	How determined did you persue the objective of the game?	not determined	very determined
Performance Overall	How would you rate your performance in the game?	very weak	very good

Table 6.3: Attention Questionnaire: Translated questions and labels of questionnaire ito be answered after each game session.

Concluding Questionnaire

At the end of the experiment, after playing the 5th round of the game the participant has to answer a short additional questionnaire testing his learning and recording demographics.

The questions to check learning effects are taken from the initial evaluation of the SaFIRa game [Bie16] for comparability. The participant is asked which of the three tanks displayed is tank "T42" (the target tank during the game) and which is tank "Tiger 1" (one of the other tanks displayed on the map). Figure 6.6 shows a screenshot of the original test question in German.

Figure 6.6: The questions of the original questionnaire inteded to test learning outcomes.

Demographic information is also asked in this concluding questionnaire. Age, amount of time spent using a computer and extent of playing video games (the former two as a subjective rating on a scale of 1-7) were recorded.

The original questionnaire in German can be found in appendix A2.

Results

The results regarding correlations with user attention and goal-orientedness remain inconclusive because no observed participants showed low levels of attention during the experiment. My Ideal Path Score approach significantly improved the adaptivity mechanisms, however.

Table 6.4 gives a brief summary of findings on each of the stated hypotheses and indicates the sections in which related results are presented in detail. A general overview of the collected data and the distribution of participants self-assessment ratingsis presented in section 6.4.1.

	Hypothesis	Results	Details
H1.1	Correlation Ideal Path Score and Attention	inconclusive	see 6.4.3
H1.2	Correlation Ideal Path Scroe and Goal-Orientedness	inconclusive	see 6.4.3
H1.3	Contribution of eye tracking to correlations	improvements needed	see 6.4.3
H2	Correlation generic eye movements and Attention	inconclusive	see 6.4.4
H3	Contribution of Ideal Path Score to adaptivity	yes	see 6.4.2
H4.1	Increase of computed skill level	not apparent	see 6.4.2
H4.2	Contribution of Ideal Path Score to this increase	-	see 6.4.2
H4.3	Learning about tank types	yes	see 6.5.3

Table 6.4: Overview of results regarding the stated hypotheses.

Overview and Context

Participants

20 students and employees of the academic institution between the age of 22 and 59 participated in the experiment after an initial pre-test. All participants reported to spend very much or much time with computers. Experience with computer games was more varied but the vast majority of participants plays few video games.

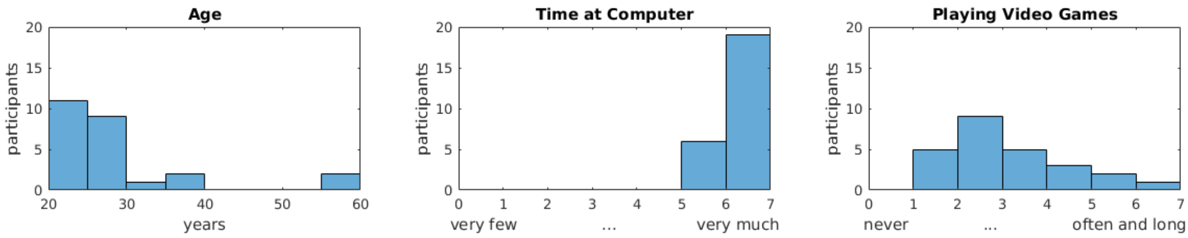


Figure 6.7: Participant Demographics

The participants, all having an academic background and being mostly students, are not necessarily representative for all target groups of the diverse e-learning sector. However, in the scope of this thesis a convenience sampling approach [Mar96], i.e. the selection of easily available participants, was chosen for practical reasons.

Eye Tracking Data Quality

The eye tracking accuracy and precision was tested as describe in section 6.3.2. As shown in table 6.5 these quality measures of the eye tracking data were similar at the beginning and the end of the experiment. This shows that the calibration before start of the experiment was sufficient and recorded eye movements are comparable across the whole duration of the experiment.

Data quality	Average	SD
Accuracy after calibration	58.6 pixel	29.3 pixel
Accuracy just before end of recording	60.9 pixel	41.5 pixel
Precision after calibration	34.7 pixel	25.4 pixel
Precision just before end of recording	35.2 pixel	33.4 pixel
Proportion of dismissed participants	10 %	-
Proportion of lost data samples in retained data (average)	5.6 %	7.5 %

Table 6.5: Eye tracking data quality report for the collected data. Precision values reflect the standard deviation of samples. The proportion of lost data samples only includes the tracking during game sessions and is an average of the proportion of each session and participant.

However, there were significant differences in data quality between participants, which can be explained by varying circumstances, e.g. some participants wearing glasses. My metrics examined in this experiment do not rely on very fine gaze measures (e.g. distinctions which word in a text was fixated), therefore the recorded data quality is sufficient.

Two participants were excluded from the evaluation because of a high proportion of lost data samples (50% and 38%). From observation this was probably due to participants wearing glasses and sitting very bent towards the screen.

Questionnaire Responses

Figure 6.8 shows the distribution of the ratings participants gave themselves after each game session when answering the questions of the questionnaire as described in table 6.3.

After most sessions, participants rated themselves to have had high (a), steady (c) attention and very high goal-orientedness (e), while ratings are still spread out over the scale. This noticable variation will be examined for correlations with the designed metrics in the following sections.

For most sessions the participants saw themselves performing well (d), rating themselves in the upper half of the scale but between medium (rating of 4) to high (rating of 6) performance assessments occur in similar numbers. The difficulty of the game (f) was perceived very differently although participants tended to find the game rather easy than difficult. The distribution of the self-assessment ratings of participants regarding their overall performance (d) and the perceived difficulty of the game (f) are more spread out than the other ratings. These are likely also influenced by confidence and video gaming experience of the participant.

Overall it has to be noted that the very general phrasing of the questionnaire and the inherent nature of self-assessments probably produced some subjective variation that cannot be objectively explained.

To receive a more reliable value of self-assessed attention for each session an additional rating was computed by combining the participants ratings of attention and distraction for that session. The resulting "undistracted attention" rating is defined as follows:

$$\text{"undistracted attention"}_{\text{of session } s} = \frac{attention_s + (7 - distraction_s)}{2} \tag{6.1}$$

The distraction rating is inverted by subtracting it from maximum of the scale. The distribution of the "undistracted attention" rating is shown in figure 6.9. Analysis in the other sections are based on this "undistracted attention" but for simplicty the rating is just called "attention" in the following.

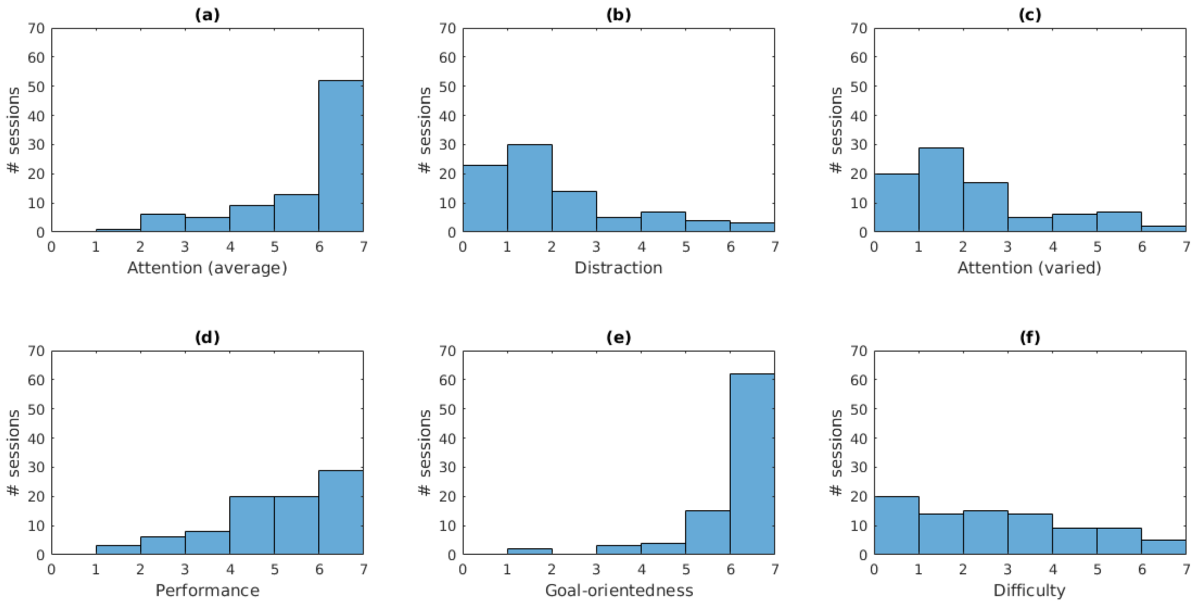


Figure 6.8: Distribution of participants' self-assessment ratings after each session.

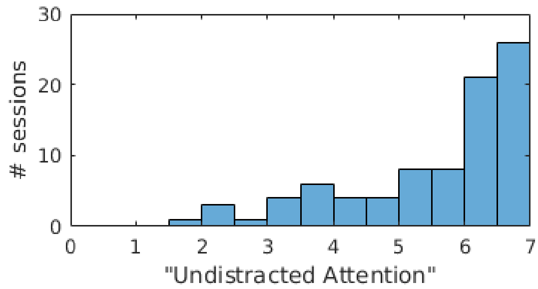


Figure 6.9: Distribution of the computed "undistracted attention" ratings (see 6.1).

The the vast majority of sessions the participants rated their attention as very high: 47 of the 86 sessions have an "undistracted attention" rating of at least 6 on the scale of 0 to 7. Therefore generalizations from this data about really low attention are very limited.

The fact that the variations of this rating, which is based on subjective self-assessments, are so small is one likely reason why any analysis regarding attention turned out to be inconclusive as discussed in section 6.5.4.

Collected Data

A total of 86 sessions from 18 participants comprise the filtered dataset. This excludes the two participants with high eye tracker data loss (see 6.4.1) and four sessions of other participants that had to be aborted due to a bug in the game.

Detailed log files containing user actions, eye movements, adaptivity feature calculations and ELAI communication were written by the framework. These contain timestamps and raw data in addition to aggregated results and are saved in the CSV format to be easily parsable. Figure 6.10 shows an examplary extract from a log file.

The raw data was analyzed and summarized for detailed evaluation. Figure 6.11 visualizes the measured Ideal Path Score and its components in the course of a game session as well as the adaptive help triggered by it.

Timestamp	Category	data 1	data 2	data 3	data 4	data 5
2016-09-23 15:00:41.7457	GAME SCENE	SeekAndFindScene_Karlsruhe				
2016-09-23 15:00:51.4460	EyeTracking AOI	Ended	1596	66	183	{"id":".../xapidefinitions/activity/UNKNOWN"
2016-09-23 15:00:51.6630	EyeTracking AOI	Ended	969	182	167	{"id":".../xapidefinitions/activity/GameObject_BlackCar"
2016-09-23 15:00:52.1278	GAME Move	(-0.7, 0.0, 0.5)				
2016-09-23 15:00:52.1518	IDEAL PATH Fixation Score	0.707922978276535	no object fixated	1596	66	
2016-09-23 15:00:52.1518	IDEAL PATH Gaze Factor	0.707922978276535	0.067100000671	0.0475		
2016-09-23 15:00:52.1518	IDEAL PATH Fixation Score	-0.5	other object			
2016-09-23 15:00:52.1518	IDEAL PATH Gaze Factor	-0.5	0.061233333945667	-0.0306		
2016-09-23 15:00:52.1528	IDEAL PATH moveDistanceFactor	-0.34947937079357	0.75			
2016-09-23 15:00:52.1528	IDEAL PATH eyeMovementsFactor	-0.081977098941462	0.25			
2016-09-23 15:00:52.1528	IDEAL PATH SCORE	-0.282603802830543				
2016-09-23 15:00:52.1538	TargetDistance	1.10678863525391				
2016-09-23 15:00:52.1538	GAME DoubleClick	(848.0, 985.0, 0.0)				
2016-09-23 15:01:28.4594	GAME HELP	VirtualAssistant2D activated	In this mission you ha			
2016-09-23 15:01:30.0925	GAME HELP	User requested help				
2016-09-23 15:01:30.1076	GAME HELP	VirtualAssistant2D activated	Your target is 6.83km			
...						
2016-09-23 15:01:36.1172	GAME HELP	VirtualAssistant2D deactivated				

Figure 6.10: Extract from a log file written by the framework. Logs are stored in CSV format and contain a timestamp and catogory for easy processing.

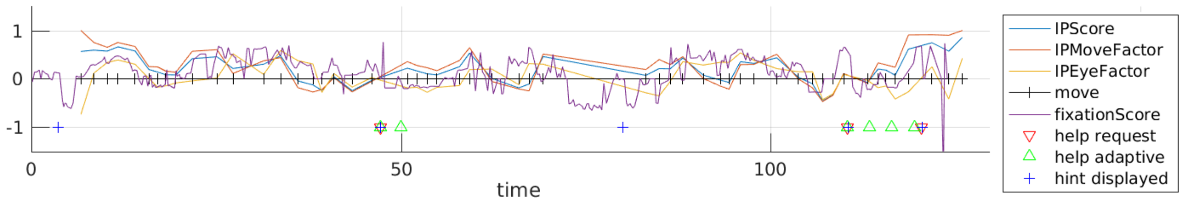


Figure 6.11: Timeline of Ideal Path Scores and its influencing factors over the course of a game session. It can be seen that negative Ideal Path Scores (blue line) over a certain time span trigger the adaptive help mechanism (green triangles).

Session duration (i.e. the time needed to find the target tank) was 109.6 seconds on average. Apart from three exceptional sessions (durations of 436s, 410s and 270s) participants always found the target in less then four minutes. These outliers were caused in one case by the participant explicitly exploring the game rather than searching the target but in another case simply by difficulty to find the target.

The duration not only depends on the player's skill but is also strongly influenced by the random starting positions of each session. Table 6.6 shows the average duration for each session as well as the distance of the player avatar to the target tank at the start of the game session. The experiment was set up in so that each participant played with different starting positions in his or her five sessions but across all participants the same five starting positions were used. Figure 6.12 displays histograms of the session durations of participants for each session.

Session	Duration Mean	Duration Std	Distance to Target
1	125.8 s	87.8 s	9.74 km
2	105.5 s	60.8 s	7.30 km
3	78.6 s	42.3 s	15.80 km
4	133.5 s	90.3 s	14.90 km
5	102.2 s	41.9 s	15.40 km

Table 6.6: Overview of session durations and distance to target tank at beginning of the session.

One would expect reduced durations for later sessions due to learning. However, such a trend is not visible in the data (table 6.6). Because of the different starting and target positions in each session this is not surprising, considering the small number of only 5 sessions. During session 4 the target tank was also partially concealed by buildings in a narrow street, which made it much harder

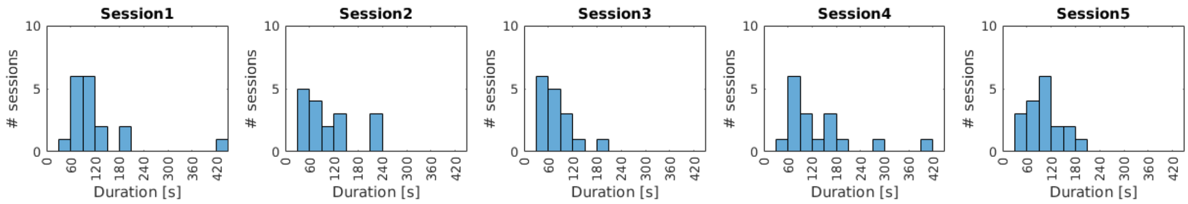


Figure 6.12: Histogram of session durations of all participants for each session. The histogram bins each cover 30 seconds.

to find than during other sessions.

Hints On average, participants were shown 6.0 hints each session. To discount for the varying distance and challenge between the five sessions, table 6.7 also shows means and standard deviations of the number of hints per minute. This value seems to be rather stable across sessions: After participants got used to the game in session 1, the mean of the number of hints per minute stays between 3.3 and 4.0 for sessions 2-5.

However, the standard deviation is extremely high, reflecting observations from the experiment that participants used different strategies playing the game. While some participants challenged themselves to request as little help as possible, others liberally requested hints to avoid running in the wrong direction. To some extend this is caused by the fact that the SaFIRa game is lacking a clear score that could give players an objective indication how they should approach the game, e.g. by penalizing help requests. These issues are further discussed in section 6.5.3.

The median number of hints during a session is 4, which is considerably below the mean value of 6.0 as there were a few sessions with a very large number of hints. In one extreme case 29 hints were displayed during a single session.

Session	Hints/Session Median	Mean	Std	Hints/Minute Mean	Std
overall	4	6.0	5.49	3.3	2.12
session 1	3.5	4.4	3.96	2.1	1.58
session 2	4	5.8	5.08	3.3	1.69
session 3	3	4.9	4.16	3.8	2.10
session 4	6	8.2	8.11	3.6	2.30
session 5	5	6.8	4.91	4.0	2.50

Table 6.7: Statistics on the number of hints displayed during a game session.

The hints at difficulty level 3, which only stated the target distance without a direction, forced participants to change their strategy (refer to figure 3.8 for an illustration of different help levels). Many participants then employed a triangulation approach, requesting many hints in quick succession to check where the target gets closer, which explains the vast difference in the number of hints displayed between sessions at difficulty level 2 and 3, which is shown in table 6.8.

The considerably different playing style at different difficulty levels is not only a challenge for the analysis of this experiment, notably it is also an important factor for adaptive interventions. This suggests that an adaptivity system like ELAI may not only need to be aware of game mechanics in general but also very specifically of the current context and settings in order to make really appropriate adaptivity decisions. On the other hand, this shows that some adaptive settings can have large consequences for the user, like in this case the adaptively estimated skill level which determines the difficulty level of the game.

Difficulty Level	Hints/Session			Hints/Minute	
	Median	Mean	Std	Mean	Std
overall	4	6.0	5.49	3.3	2.12
2	3	4.3	3.15	2.5	1.35
3	11	10.7	7.62	5.3	2.52

Table 6.8: Statistics on the number of hints displayed split by difficulty level. Difficulty level 1 excluded due to low sample size.

Adaptive hints, i.e. hints triggered automatically by ELAI's influence parameter "needsHelp", on average made up 31% of all hints shown (median 20%). Figure 6.13 shows the distribution of this ratio across all sessions, the differences between "Legacy" and "Ideal Path" adaptivity are further examined in section 6.4.2.

Observations during the experiment showed that many participants requested help faster than the adaptive mechanism would trigger them. Therefore adaptivity only played a minor role in those cases. Indeed during 31 sessions, that is more than a third of all sessions, all displayed hints were actively requested by the player.

Also, it is worth considering to infrequently display hints to players although they are making progress. This would explicitly provide confirmation to the player that he is on the right track to assure him. Such a mechanism is currently not provided by the adaptivity and a number of players actively requested hints to get confirmation in such situations.

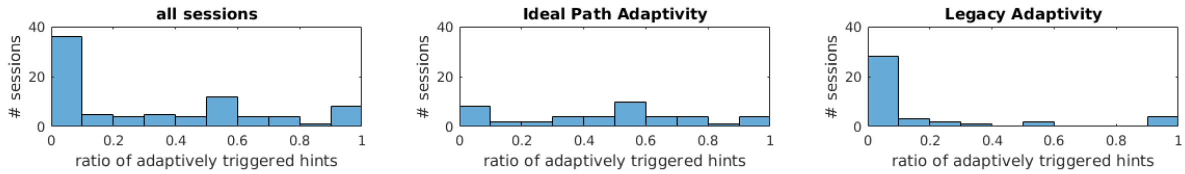


Figure 6.13: Ratio of hints triggered automatically by the ELAI adaptivity.

Comparison of Legacy and Ideal Path Based Adaptivity

In order to examine the contribution of the Ideal Path Score to suitable adaptivity decision (hypothesis H3) I compared my developed Ideal Path based adaptivity system with a version of the game with the original adaptivity mechanisms (termed here as "Legacy"). This "Legacy" adaptivity was designed by [Bie16] and is also the foundation of my Ideal Path based adaptivity, which extends and improves it. As described in section 6.2 presenting the experimental setup, participants were assigned to either play with the "Legacy" or the "Ideal Path" system. In the resulting dataset there are 43 sessions each played with "Legacy" and with "Ideal Path" adaptivity.

In a nutshell the evaluation shows that Ideal Path Scores clearly improved the adaptivity mechanisms of adaptive help messages and skill level classifications compared to the original adaptivity and therefore contribute to the questions of when and what should be adapted.

Adaptive Help

As figure 6.13 shows, adaptively triggered hints were much more common with the "Ideal Path" adaptivity. This indicates that despite a lack of clear correlation between the designed Ideal Path Score and measures like session duration or self-assessed attention and performance (see 6.4.3), the Ideal Path Score is successfull in practice.

A high ratio of adaptively triggered hints does not necessarily indicate good adaptivity as this also includes cases with many unnecessary hints. On the other hand, the fact that only a very low percentage of hints is triggered automatically through adaptivity clearly indicates that the adaptivity mechanism is unable to estimate the players subjective need to get help. This is the case for the "Legacy" adaptivity system, which has a mean of only 16% adaptively triggered hints. The "Ideal Path" adaptivity system is much better at providing needed help with a mean of 46% hints triggered adaptively. Still no participant using the "Ideal Path" version of the game stated that he received clearly too many hints (6 or 7 on the scale of 0-7) in any of the session questionnaires (see figure 6.14).

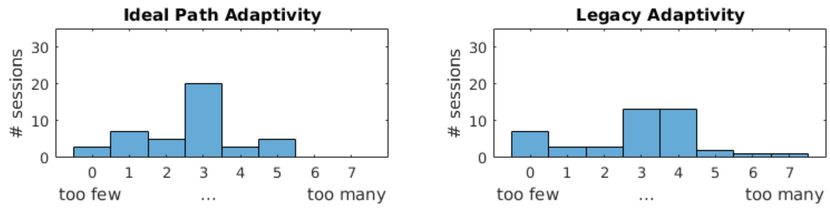


Figure 6.14: Participants' impression of amount of hints from the questionnaire for each game session.

Skill Level Estimation

In addition to triggering automatic hints, adaptivity in the SaFIRa game is responsible for estimation the player's skill level that then influences the difficulty level. The method of skill level computation is explained in detail in chapter 5.4. My assumption prior to the experiment was that the skill level of participants generally should increase with their experience in later game sessions (hypothesis H4.1) and that the mean Ideal Path Score of the session, which is used as a didactic factor in the skill level computation, would reflect this expected increase (hypothesis H4.2).

For comparability during this analysis, the normalization of the didactic factors (see 5.4) is calculated for all sessions with the same values. The minimum and maximum values in the overall dataset from the experiment are used rather than - as would be the case in a deployed system - only data available up to that point in time. I chose this approach for my analysis because the normalization values from my experiment would otherwise change very often. For a deployed system in practice on the other hand it is reasonable to assume that in its large dataset the extreme values used for the normalization would only change seldomly.

Because the skill levels were calculated during analysis after the experiment, the values of all variations (IP and Legacy) are available for all sessions regardless of the version the participant played during the experiment.

Figure 6.15 shows the calculated skill level for each session from the IP system and the Legacy system algorithm as well as the contributing didactic factors by themselves. The assumed trend of increasing skill level in subsequent game sessions is not apparent. However, the varying challenge between the five sessions played by each participant also need to be considered. To compare these influences figure 6.16 shows the subjective ratings of participants about how difficult each session appeared and how they rated their overall performance.

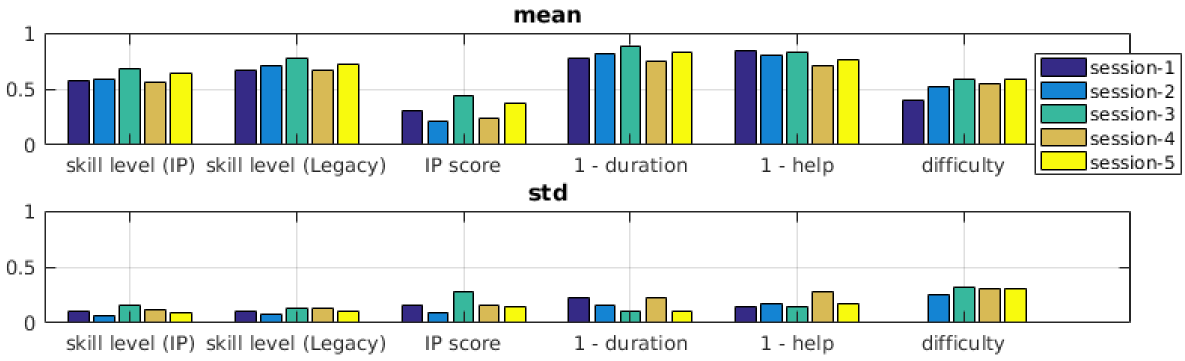


Figure 6.15: Means and standard deviations within session 1-5 for the computed skill level and didactic factors. Duration and help are inverted as higher values for them indicate lower skill.

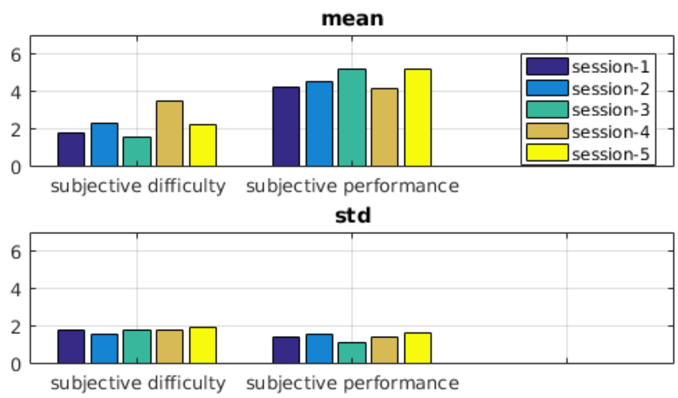


Figure 6.16: Subjective difficulty and performance ratings of participants by session. These ratings show that there are additional influencing factors for skill level estimation.

Looking at the plots, the participants' perception of their performance appears to be related to the duration factor, while perceived difficulty seems to be reflected by differences in mean Ideal Path Scores. The correlation coefficients shown in table 6.9 confirm that there are significant correlations of these ratings of performance or difficulty in particular with the "IP Score" and the "duration" didactic factors.

The mean Ideal Path Score of a session therefore is a useful indicator at least of the participants perception of performance and challenge. This leads to much higher correlations of the newly developed Ideal Path based skill level estimation with participants' perceived performance (r=0.30) and difficulty (r=-0.16) compared to the correlations of the Legacy system skill level with these (r=0.19 with performance, not significant with difficulty). However, these correlations are still rather low.

The interaction based "IP Move Factor" is contribution most to this correlation. While the original eye tracking score ("IP Gaze Factor") does not show correlations with perceived performance or difficulty the discretized version "IP Gaze discretized" actually shows better correlation with perceived performance than the help didactic factor (see table 6.9).

Correlation of the Ideal Path Score and Attention

The Ideal Path Score evaluated here was designed as described in chapter 4.1. The expectations (stated as hypotheses in section 6.1) were that the score correlates with the self-assessed ratings for attention (H1.1) and goal-orientedness (H1.2) and that the eye tracking based component contributes to this in particular (H1.3).

	Performance		Difficulty	
	r	p	r	p
skill level (IP)	0.29923	0.0051	-0.1617	0.1369
skill level (Legacy)	0.18889	0.0815	-0.0274	0.8025
duration	0.38949	0.0002	-0.3297	0.0019
help	0.17795	0.1012	-0.2902	0.0067
difficulty	-0.13922	0.2011	0.3872	0.0002
IP Score	0.38756	0.0002	-0.3434	0.0012
IP Move Factor	0.41118	0.00008	-0.3530	0.00085
IP Gaze Factor	0.00285	0.9792	-0.0725	0.5071
IP Move discretized	0.38391	0.0003	-0.3690	0.00047
IP Gaze discretized	0.21476	0.0471	-0.1148	0.2927

Table 6.9: Correlation of didactic factors with participants' subjective performance and difficulty ratings.

The fact that almost all participants rated their goal-orientedness and attention as very high limits this analysis however (see figure 6.8 in section 6.4.1). These limations are further discussed in section 6.5.4.

Correlations of the Ideal Path Score with the self-assessed attention and goal-orientedness ratings are not apparent in the boxplots of figure 6.17. The same is true for the gaze and interaction based factors comprising the overall Ideal Path Score. This analysis uses the mean score of the users session as the feature value.

Very low goal-directedness appears to be distinguishable in the boxplot but this has to be put into perspective by the extremely small sample size for this specific rating value. Table 6.10 shows Pearson's correlation coefficients and the corresponding p-values, which confirm that there is no significant correlation with subjective attention for any of the Ideal Path based factors. Regarding the subjective goal-orientedness there is a correlation especially with the Ideal Path Score and the interaction based "Move Factor", although Pearson's r only reaches 0.39 and 0.41 respectively.

P-values are testing the hypothesis of no correlation against the alternative that there is a nonzero correlation. A small p indicates a low probability that there is no correlation, often results with p-values below 0.05 are considered to show a statistically significant correlation. Regarding correlations with user attention none of the examined factors stay below that threshold, therefore indicating no statistically significant correlations.

As an additional feature the standard deviation of the Ideal Path Scores within a session may be helpful, with a correlation coefficient of -0.32 (see table 6.10). This shows that during sessions that are perceived as strongly goal-oriented the Ideal Path Scores, which are calculated for each move during the session, tend to vary less. This trend fits the general expectation that a highly goal-oriented user should perform more consistently than a less goal-oriented user.

Particularly the eye tracking based "IP Eye Factor" seems to provide no helpful information, questioning hypothesis H1.3, that eye tracking data improves the examined correlations. However, improvements to the eye tracking based score like discretization uncover a tendency of it to correlate with participants goal-orientedness (see "Gaze Factor discretized" in table 6.10). Further tuning of the eye tracking based features may still lead to an eye tracking factor that provides a significant contribution to an overall Ideal Path Score. A more dependable target variable than the subjective goal-orientedness rating could also support better evaluation.

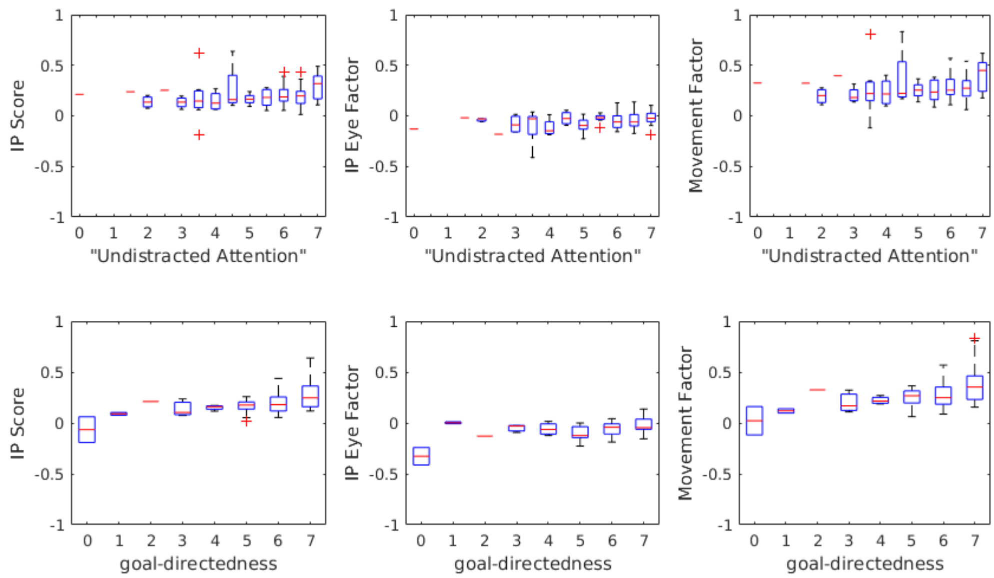


Figure 6.17: Correlations of mean Ideal Path Score of a session with attention / goal-orientedness.

	Attention		Goal-Orientedness	
	r	p	r	p
IP Score (Session Mean)	0.166	0.127	0.393	0.0002
IP Move Factor	0.158	0.146	0.395	0.0002
IP Eye Factor	0.111	0.309	0.138	0.206
Move Factor discretized	0.211	0.052	0.407	0.0001
Eye Factor discretized	0.086	0.429	0.288	0.007
IP Score (Session Std)	-0.034	0.756	-0.320	0.003

Table 6.10: Pearson's correlation coefficients of Ideal Path Score and its factors with participants' subjective attention and goal-orientedness ratings. Corresponding p-values indicating the (lack of) statistical significance are also reported.

Attention does not influence the session duration, that means the participants perceived attention did not influence his "performance" in the game, in the sense that attentive participants found the target faster. Figure 6.18 shows a scatter plot of session duration in relation to attention. The assumption that higher attention would lead to better performance (i.e. reaching the target faster) is refuted by this data. This is the case even after removing the influence of the difficulty level (the data shown contains only sessions with the medium difficulty level, which was the most common setting during the experiment). This raises the question whether any reliable inferences can be made with relation to this subjective attention rating at all.

Session duration is correlated with Ideal Path Score, i.e. there is a correlation between Ideal Path Score and the time taken to win a game session (see figure 6.19). Pearson's correlation coefficient of session duration with the session's mean Ideal Path Score is -0.48, with the Ideal Path Score "Move Factor" component even -0.51. Both correlations are statistically significant with p-values below $3 \cdot 10^{-6}$. The eye tracking component does not contribute to this correlation however.

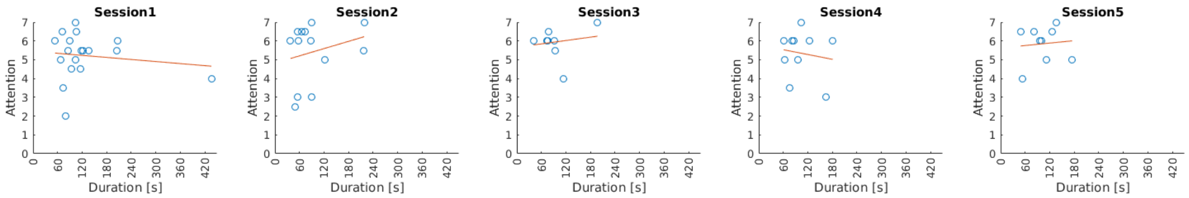


Figure 6.18: The time to complete a session (duration) does not correlate with the participants attention for the session. These plots only show sessions at medium difficulty settings to remove this other factor influencing the duration, the partial data plotted here consists of 60 sessions. The trends over the full dataset are similarly inconclusive.

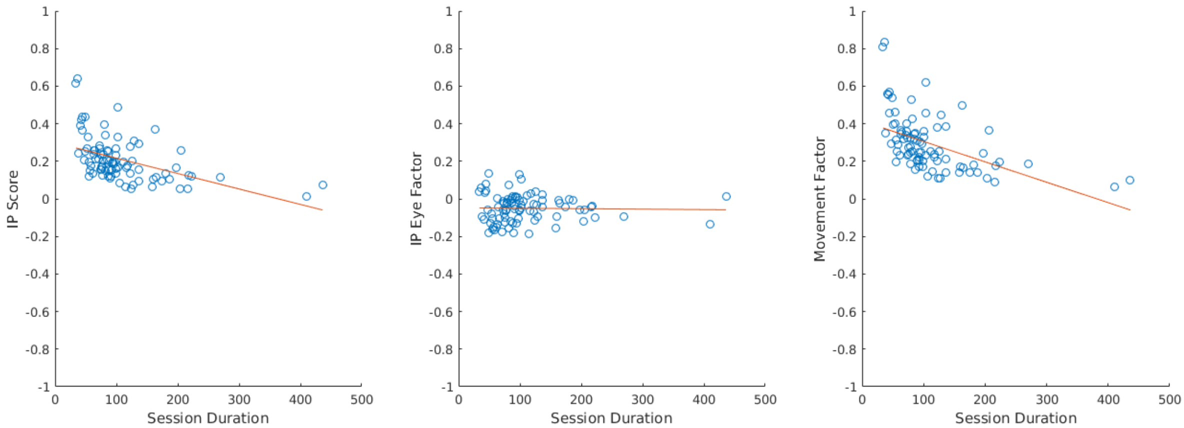


Figure 6.19: Relation between Ideal Path Score and session duration.

The main challenge during the analysis of the expected correlations was that the experiment did not provide observations of sessions with really low attention ratings. The small variations in the subjective attention (and goal-orientedness) ratings of participants did not show clear correlations with the developed Ideal Path Score, however, whether correlations exist for more extreme differences of the attention level cannot be answered with the conducted experiment.

This limitation is discussed in detail in section 6.5.4.

Generic Eye Tracking Features

In addition to the developed Ideal Path Score some generic eye tracking features were also evaluated based on the data collected from the experiment. These generic features were commonly used in related work (see chapter 2.3) and are examined here regarding their applicability in the context of my thesis.

My hypothesis regarding such generic eye tracking features (H2 as stated in section 6.1) was that they also correlate with the participants self-assessed attention.

Pearson's correlations coefficients of the participants ratings of attention and goal-orientedness with features like the mean fixation duration during a session, the fixation frequency or the changes of pupil diameter are shown in table 6.11. There is no correlation of participants ratings with these features, similar to the results regarding the Ideal Path Score presented in section 6.4.3. The problems limiting a more conclusive analysis are the same as stated there: The lack of data from participants with low attention prevents evaluation whether large differences in attention level can be detected by the features (see discussion in section 6.5.4).

	Attention		Goal-Orientedness		Performance		Difficulty	
	r	p	r	p	r	p	r	p
fixationDurationMean	-0.064	0.560	-0.098	0.368	-0.012	0.911	-0.199	0.067
fixationDurationStd	-0.022	0.839	-0.107	0.329	-0.010	0.930	-0.217	0.045
fixationFrequency	-0.040	0.716	-0.045	0.684	0.034	0.755	-0.238	0.027
pupilDiameterStd	-0.148	0.173	0.045	0.682	-0.070	0.525	0.112	0.304

Table 6.11: Pearson's correlation coefficients of generic eye tracking features with participants' ratings of subjective attention, goal-orientedness, perceived performance and perceived difficulty. The high p-values indicate the lack of statistically significant correlation apart from cases regarding Difficulty, which barely stay below the common significance threshold of 0.05.

Combination of generic and Ideal Path Score features is an evident approach to improve the expressiveness of the generic eye tracking features. Related work reported that such generic eye tracking features like the rate of fixations by themselves could not classify learning performance but improved classification in combination with content specific features [Bon13].

The test this approach, regression trees were trained and evaluated using leave-one-out cross-validation on different selections of variables including and excluding generic eye tracking features. Table 6.12 shows the resulting mean squared errors. While adding generic features to the original Ideal Path Score "Move Factor" and "Eye Factor" improved results slightly, adding generic features actually deteriorated results from a tree based on the improved, discretized Ideal Path factors. No clear conclusion about the benefits of generic eye tracking features can be derived from these results.

Overall, mean squared errors of the regression trees are extremely high, questioning any conclusions made about minor improvements. This leads back to the general problems regarding the observed attention ratings as discussed in section 6.5.4.

Feature Set	Predicted Attention
	Mean Squared Error
IP Move Factor + IP Gaze Factor	2.82
IP Move Factor + IP Gaze Factor + Generic Features	2.69
IP Move Factor (discrete) + IP Gaze Factor (discrete)	2.68
IP Move Factor (discrete) + IP Gaze Factor (discrete) + Generic Features	2.92

Table 6.12: Prediction errors regarding the attention rating by regression trees trained on the given variables. The "Generic Features" are mean fixation duration, standard deviation of fixation duration, fixation frequency, standard deviation of pupil diameter calculated for each session.

Discussion

As shown above (section 6.4), parts of the evaluation's results, particularly the analysis regarding attention, were inconclusive. The factors influencing these results, illustrated in figure 6.20, are therefore revisited here in the discussion.

Problems, limitations and possible improvements for future work are discussed regarding the framework and implementation of eye tracking and adaptivity (section 6.5.1), the concepts of the developed Ideal Path Score that was evaluated (section 6.5.2), the game SaFIRa used for the experiment (section 6.5.3) and the experimental setup (section 6.5.4).

The main challenges were caused by the complexity of the game, limiting inferences because specific aspects were often influenced by many parameters, and the lack of samples with low atten-

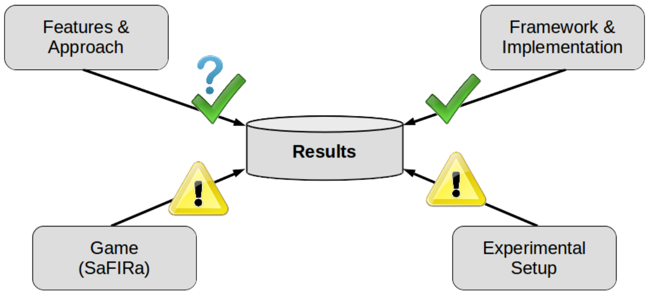


Figure 6.20: Overview of the discussion of components influencing the experiment's results.

tion levels, restricting the search for correlations to small variations of attention that are apparently insignificant.

Framework and Implementation

The implemented framework served as a suitable foundation for the experiment. Deployment to the lab system with a different eye tracking device worked seamlessly. The computation of adaptivity parameters by the framework, which integrated game and ELAI server, did not reveal any problems during the experiment.

Features and Concept

The results overall have to be interpreted with some care because of the limitations of the experimental setup (see 6.5.4) and the game "SaFIRa" used for the evaluation (see 6.5.3).

Adaptivity decisions were clearly improved by the developed approach of Ideal Path Scores compared to the original adaptivity of SaFIRa. The results show that the interaction based factor contributed more to the expressiveness of the Ideal Path Score, suggesting that inferences from interactions are at least easier to "get right" compared to inferences from eye tracking data. On the other hand, the discretization of the gaze factor showed clear improvements over its original version, so tuning of the specific feature proofs essential.

The game specific definition of the Ideal Path Score is an important factor and a considerable challenge. In the course of the experiment it turned out that the primary actions in SaFIRa are different than anticipated. Rather than searching for the relevant object, navigation through the city took up most of the playing time. The designed interpretation mechanisms for players' eye movements were not successfull because these were tuned to the expected rather than the actual gameplay. This demonstrates the importance and difficulty to define features that are exactly concerted with the game they are applied to.

The normalization mechanism for Didactic Factors could possibly trouble the adaptive skill level classification. Considering the large variations of the challenges and task durations observed between the 5 different game sessions every participant played, normalization based on global maximum and minimum values is questionnable. The Didactic Factors, among them the session duration and the number of hints displayed, are currently normalized using the overall maximum and minimum values in the database (see 5.4). This means a single, very easy or very hard session can skew the skill level calculation for all following players. A factor like the distance to the target at the start of the session could control for this effect and might lead to better skill level estimates. This issue was not examined any closer during the evaluation, however.

Correlations of user attention with the Ideal Path Score were not apparent. The total lack of such a trend for the gaze factor questions the assumption that eye movements can be a powerful tool to estimate attention despite the limitations posed by the experimental setup of this evaluation. Eye tracking information may be more suitable to assess player performance or to implement adaptivity mechanisms that are more directly based on specific eye movements, like noticing a displayed message.

Other sensors or classifiers could improve results by providing a more comprehensive model of the user. Facial expressions or body posture come to mind and could be extracted from the additional data recorded during the experiment using webcam and Kinect. The use of machine learning techniques to interpret the collected data is also an interesting possibility if enough data can be collected to train such systems.

Limitations of the Game "SaFIRa" and Possible Improvements

The game SaFIRa was chosen for the experiment of this thesis because it was readily available and modifiable. The game already contained ELAI integrated adaptivity mechanisms on which my work could improve upon. Also, SaFIRa is a full game - albeit of limited scope - rather than a simplistic setup solely for an experiment. This allowed to test adaptivity in a realistic context.

However, during the course of the experiments it became apparent that SaFIRa has a number of shortcomings as well as fuzzy game objectives. This turned out to be an important limitation for the analysis to make clear inferences because it led to varying attitudes and strategies of the participants and many degrees of freedom in the experiment. Analysis was difficult then because of a multitude of parameters incluenced the examined observations, limiting clear inferences.

The assumed learning objective for SaFIRa by [Bie16] is that players learn to distinguish (and name) different tank types that appear in the game. [Bie16] reports in his evaluation that 100 % of participants correctly recognized the target tank T42 in a post-test, while only 58 % recognized the other queried tank type. During my experiment I included the identical question and was able to reproduce the result that the target tank T42 is recognized by 100 % of participants. However, only 40 % of participants recognized the distractor tank type, that was not in the focus of the game. Looking at the gameplay in addition to these results, it is highly questionable whether the game is suited to teach different tank types because most time and effort is spent navigating through the city according to the virtual assistance's guidance rather than actually scanning and identifying vehicles.

A clearly defined gameplay objective was also missing for SaFIRa in its current state. While the lack of a coherent learning objective did not impede my experiment, the lack of an unambiguous gameplay objective left the participants playing style unrestrained. The players generally tried to find the target tank as fast as possible but the game does not present any score for a completed mission. Therefore the player does not receive feedback and - as requesting help is not penalized - has no general incentive to use as few help as possible. In this case, whether the player relies on many hints or challenges himself to win without them remains the choice of the player. Indeed, one participant of the experiment stated that he would have requested less help if there would have been a penalty for requesting help.

Adaptive hints are useful according to most participants. Some participants stated that they had already forgotten how to actively request hints or that they felt supported by this automatic measure. But one participant also felt that the automatic hints should give him more time to search on his own before intervening. To provide feedback and assurance to players moving in the right direction, it should also be considered to display occasional hints adaptively in such situations. Many participants actively requested hints to confirm that they were on the right track.

Other minor problems of the game are listed below. Some of these glitches confused or annoyed participants and probably influenced in-game behaviour, other problems and suggestions by participants are just listed here as the experiment turned out to also be an effective playtest.

- Distances stated in hints were scaled incorrectly, which confused players trying to orient themselves in the city.
- Cardinal directions in hints were confusing because a target just slightly North of the player leads to a message stating "cardinal direction North-West" rather than "West".
- Game boundaries are not intuitively marked, most players reached the limits of the accessible map and tried in vain to move further.
- The player does not receive any cardinal direction of the target at the start of a session and therefore needs to immediately request a hint or start moving in an arbitrary direction.
- Some players won the game "by accident" as they passed the target tank without even recognizing it, which is treated by the game like successfully finding the target.
- Hints of the highest difficulty level (not stating the cardinal direction) make the game much harder and require a very different strategy and more hints.
- The mechanism to request help, clicking on the virtual assistant and then again on the appearing questionmark, is not intuitive and often confuses players despite the explanation in the introduction.
- The button to show the next page of the introduction is easy to overlook, causing players to unintentionally skip the rest of the introduction.
- The design of the introduction page showing the three different tank types suggests that the player can choose the target tank type. If implemented, this could be a good measure to have players focus on different tank types.
- The low resolution of the underlying map makes street names hard to read.
- Where the avatar can move (i.e. what lines of the map represent legit streets) is not always clear.
- The text describing keyboard controls at the bottom left of the screen is often hard to read due to the background.
- Vehicles are sometimes placed partially "inside" buildings.

Limitations of the Experimental Setup

The biggest challenge for the experiment and analysis was posed by the use of the game SaFIRa in its current state. The game-related issues are discussed in the separate subsection 6.5.3.

The high number of degrees of freedom within the game made analysis of specific aspects like gameplay performance very difficult because these were often influenced by many factors. For example, session durations (i.e. the time taken by the player to find the target) depends on the randomized starting positions, the difficulty setting of the game and multiple aspects of the playing style of the particular player like his affinity to request hints. This multitude of parameters, some not objectively measurable, limited comparability across participants and game sessions. Examination of clear correlations between only two isolated parameters were therefore not possible.

The number of participants of the experiment had to be limited within the scope of this thesis due to time constraints. In connection with the challenges regarding the degrees of freedom, the limited size of the collected dataset may be responsible for some results lacking statistical significance. A larger dataset can also provide the base for other techniques like advanced machine learning to discover patterns in the data.

The question of correlations with participants' attention turned out to be particularly complex and needs to be examined in an experimental setup exclusively tuned to question. If possible, this should also be done with a larger number of participants, which could provide data for advanced machine learning and data mining techniques.

The questionnaire based approach to survey a subjective cognitive state - in this case the level of attention - is common in games research according to the extensive review of [Boy12] and was also used in related work like [Jen08]. From the experiences of the experiment the the designed questionnaire should be improved regarding the question whether there were too many or too few hints. This should be phrased more clearly as some participants related this question only to automatically shown hints, others to all hints including the actively requested ones.

No observations of low attention could be recorded during the experiment. The vast majority of participants rated their attention as very high, as presented in section 6.4.1. This limited the analysis of correlations with user attention. While the small differences in attention apparently had no effects on the playing performance or the Ideal Path Scores, comparison between sessions played with very different attention levels were not possible because such cases were not observed with this experimental setup. The question about such correlations therefore remains inconclusive.

The fact that most participants apparantly showed high attention is not surprising. Due to the special environment of an experiment where participants were observed during play, this was to be expected. By using an eight point scale I successfully tried to avoid an outcome where all participants would rate their attention at the exactly same level.

The variations between participants' ratings of their attention lacked correlations with measured variables like Ideal Path Score or even session duration. The subjective nature of participants' self-assessment ratings may be one explanation for the lack of correlations. As there were only small differences between the participants generally high ratings of their attention, these variations might mostly be subjective rather than objectively measurable differences in the "real" level of attention. With the collected data no conclusions can be made about more extreme differences in attention because such cases were not observed during this experiment.

One way to provoke differences in attention during future experiments is a setup that artificially provides distractions during some phases of the experiment. Another approach can be the use of a particular type of game that includes relaxing and demanding game phases which can be exactly identified.

7

Conclusion and Outlook

Overall, this work provides insights on how eye tracking can be integrated into the adaptive cycle of e-learning adaptivity. It highlights the challenges of meaningful assessment of eye movements and reports important experiences about the experimental setup to evaluate such an approach.

Adaptive serious games attempt to automatically tune their gameplay and challenges to the individual player in order to provide conditions that allow the player a flow experience as well as more effective learning. To make appropriate adaptivity decisions, such adaptive systems need accurate models of player's skills and current state. This thesis contributes to this overarching goal by developing and evaluating an approach to model player attention and performance based on players' interactions within the game and their eye movements recorded by an eye tracker.

Within the scope of this thesis I implemented a framework which facilitates integration of eye tracking into the adaptive cycle, designed an "Ideal Path Score" to assess attention and game performance, conducted a user study to evaluate the developed approach and analyzed details of the collected data.

Correlations with players' attention could not be discovered for the developed Ideal Path Score or its interaction and eye tracking based components. During the experiment participants only reported high or very high attention, therefore preventing comparisons of observations with largely different levels of attention during analysis of the results. To achieve conclusive results whether correlations between eye movements, Ideal Path Score and attention exist, further examination is necessary.

The tendencies shown in the evaluation of this thesis, however, cast some doubt on the initial assumption that eye tracking is a powerful tool to specifically estimate user attention.

Adaptivity mechanisms were improved by the developed Ideal Path Score. Compared to the original adaptivity mechanisms this new approach was able to assess a player's need for help much more accurately. The successful application of the Ideal Path Score towards the core objective - improved adaptivity - demonstrates that this is a promising approach.

However, it remains questionnable how much the eye tracking information contributed to this success. Rating eye movements in a meaningful, game specific way turned out to be challenging because of the complexity of gameplay and different player strategies even in a simple game like the one used in the experiment.

Based on the observations and the collected data, further improvements to the definition of the Ideal Path Score can be achieved. The potential for improvements was already demonstrated by an alternative, discretized definition of the "Gaze Factor", the component of the Ideal Path Score rating eye movements.

Future Work

This thesis provides a base for further analysis of user attention and more specialized adaptivity mechanisms based on eye movements.

Further study of user attention is needed to conclusively answer questions about correlations of eye movements with users' attention. The experience from the conducted experiment for this thesis can provide valuable input for an improved experimental setup. Adding phases of controlled distraction could provoke periods of low attention on the game, making observation of a larger range of attention levels possible. Also, the degrees of freedom in the evaluated game, like changing difficulty levels, should be limited as far as possible to control other influences on eye movements. To receive more significant results a larger number of players should participate in such an experiment.

Other sensors could help to build a more comprehensive model of the player's state or attention. During the conducted experiment additional data from a webcam and a Kinect sensor were collected. These were not analysed within the scope of this thesis, however. Features like facial expressions or posture should be considered and further examined regarding their use to improve attention estimation and adaptivity decisions.

Use of more advanced classification techniques from machine learning could provide a good way to combine features from such different sources if enough training data can be collected.

Transfering the approach to other games would put the results of this work on a broader foundation. Time constraints limited the implementation and evaluation of my approach to only one serious game in this thesis.

On a technical level, transfering the developed approach to other games should not pose difficulties as the developed framework was designed with a focus on extensibility and portability. Experiences how the details of the valuation of specific interactions and eye movements that make up the Ideal Path Score need to be adapted to different games would provide important insights about the overall approach. How suitable eye movements are to assess players' attention and game performance possibly also varies for different game genres.

Specialized adaptivity based on eye movements could provide valuable extensions to the more general adaptivity controlled by abstracted classifications of the player.

For example, [Bie16] reports that in the initial evaluation of the game SaFIRa some players did not notice adaptively updated help messages. Eye tracking information can reliable detect when a player looks at such a message. A simple adaptivity component could therefore continuously monitor whether the player notices the adaptively displayed help messages and more prominently highlight a message if necessary.

In such situations eye tracking enables very specific but potentially powerful adaptivity mechanisms.

Bibliography

[Alb07] ALBERT, Dietrich; HOCKEMEYER, Cord; RUST, Michael D Kickmeier; PEIRCE, Neil und CONLAN, Owen: Microadaptivity within Complex Learning Situations – a Personalized Approach based on Competence Structures and Problem Spaces. *Proceedings of the International Conference on Computers in Education ICCE 2007* (2007)

[Ame07] AMERSHI, Saleema und CONATI, Cristina: Unsupervised and supervised machine learning in user modeling for intelligent learning environments. *Proceedings of the 12th international conference on Intelligent user interfaces - IUI '07* (2007): S. 72

[Bar04] BARRIOS, Victor Manuel García; GÜTL, Christian; PREIS, Alexandra M; ANDREWS, Keith; PIVEC, Maja; MÖDRITSCHER, Felix und TRUMMER, Christian: AdELE: A Framework for Adaptive E-Learning through Eye Tracking. *October* (2004), Bd. 4(October): S. 1–8

[Bea10] BEATSON, Oliver: Challenge vs skill (2010), URL `https://en.wikipedia.org/wiki/File:Challenge_vs_skill.svg`

[Bed05] BEDNARIK, Roman: Potentials of Eye-Movement Tracking in Adaptive Systems. *Proceedings of the 4th Workshop on Empirical Evaluation of Adaptive Systems, held at the 10th International Conference on User Modeling UM2005* (2005): S. 1–8

[Bie16] BIEGEMEIER, Christian: *Web-basierte Schnittstelle zur Analyse und Adaption von Serious Games*, Master thesis, KIT; Fraunhofer IOSB (2016)

[Blo12] BLOG, Cygon's: Plugin Architecture (2012), URL `http://blog.nuclex-games.com/tutorials/cxx/plugin-architecture/`

[Boe09] BOEKER, Martin; ANDEL, Peter; SEIDL, Maximilian; STREICHER, Alexander; SCHNEEVOIGT, Timm; DERN, Peter und FRANKENSCHMIDT, Alexander: Uro-Island I - Game Based E-Learning in der Urologie. *GMS Medizinische Informatik, Biometrie und Epidemiologie* (2009), Bd. 5(April): S. 1–7

[Bon13] BONDAREVA, Daria; CONATI, Cristina und FEYZI-BEHNAGH, Reza: Inferring Learning from Gaze Data during Interaction with an Environment to Support Self-Regulated Learning (2013), (1): S. 229–238

[Boy12] BOYLE, Elizabeth A.; CONNOLLY, Thomas M.; HAINEY, Thomas und BOYLE, James M.: Engagement in digital entertainment games: A systematic review. *Computers in Human Behavior* (2012), Bd. 28(3): S. 771–780

[Cal08] CALVI, C.; PORTA, Marco und SACCHI, D.: e5Learning, an E-Learning Environment Based on Eye Tracking, in: *8th IEEE International Conference on Advanced Learning Technologies*, 1, IEEE, S. 376–380

[Can12] CANTONI, Virginio; PEREZ, Calet Jimenez; PORTA, Marco und RICOTTI, Stefania: Exploiting eye tracking in advanced e-learning systems, in: *Proceedings of the 13th International Conference on Computer Systems and Technologies - CompSysTech '12*, ACM Press, New York, New York, USA, S. 376

[Che06] CHEN, Hsiang: Flow on the net-detecting Web users' positive affects and their flow states. *Computers in Human Behavior* (2006), Bd. 22(2): S. 221–233

[Che07] CHEN, Jenova: Flow in games (and everything else). *Communications of the ACM* (2007), Bd. 50(4): S. 31

[Chi08] CHIANG, Yu T.; CHENG, Chao Yang und LIN, S. S J: The effects of digital games on undergraduate players' flow experiences and affect. *Proceedings - 2nd IEEE International Conference on Digital Game and Intelligent Toy Enhanced Learning, DIGITEL 2008* (2008): S. 157–159

[Con] CONATI, Cristina; MERTEN, Christina; AMERSHI, Saleema und MULDNER, Kasia: Using Eye-Tracking Data for High-Level User Modeling in Adaptive Interfaces: S. 1614–1617

[Con02] CONLAN, Owen; DAGGER, Declan und WADE, Vincent: Towards a Standards-based Approach to e-Learning Personalization using Reusable Learning Objects. *Proceedings* (2002), Bd. 2002: S. 210–217

[Con05] CONATI, Cristina; MERTEN, C; MULDNER, K und TERNES, D: Exploring eye tracking to increase bandwidth in user modeling. *User Modeling 2005, Proceedings* (2005), Bd. 3538: S. 357–366

[Con07] CONATI, Cristina und MERTEN, Christina: Eye-tracking for user modeling in exploratory learning environments: An empirical evaluation. *Knowledge-Based Systems* (2007), Bd. 20(6): S. 557–574

[Con12] CONNOLLY, Thomas M.; BOYLE, Elizabeth A.; MACARTHUR, Ewan; HAINEY, Thomas und BOYLE, James M.: A systematic literature review of empirical evidence on computer games and serious games. *Computers & Education* (2012), Bd. 59(2): S. 661–686

[Con13] CONATI, Cristina; JAQUES, Natasha und MUIR, Mary: Understanding attention to adaptive hints in educational games: An eye-tracking study. *International Journal of Artificial Intelligence in Education* (2013), Bd. 23(1-4): S. 136–161

[Cor13] CORNE, Ben: *Making Existing Educational Games Adaptive Using AOP*, Dissertation (2013)

[Csí92] CSÍKSZENTMIHÁLYI, M: The Flow Experience and its Significance for Human Psychology. *Optimal experience: psychological studies of flow in consciousness* (1992)

[Csi97] CSIKSZENTMIHALYI, Mihaly: *Finding flow: The psychology of engagement with everyday life*, Basic Books (1997)

[Dec94] DECI, Edward L.; EGHRARI, Haleh; PATRICK, Brian C. und LEONE, Dean R.: Facilitating Internalization: The Self-Determination Theory Perspective. *Journal of Personality* (1994), Bd. 62(1): S. 119–142

[Duc07] DUCHOWSKI, Andrew: Eye tracking methodology: Theory and practice. *Eye Tracking Methodology: Theory and Practice* (2007): S. 1–328

[FAA] Cognition: Wickens' Model, URL `http://www.hf.faa.gov/Webtraining/Cognition/CogFinal008.htm`

[Fog09] FOGG, Bj: A behavior model for persuasive design. *Proceedings of the 4th International Conference on Persuasive Technology - Persuasive '09* (2009): S. 40:1—-40:7

[Fow03] FOWLER, Martin; FOWLER, Martin; RICE, David und RICE, David: *Patterns of Enterprise Application Architecture*, A Martin Fowler signature book, Addison-Wesley (2003)

[Fre04] FREEMAN, Eric; ROBSON, Elisabeth; BATES, Bert und SIERRA, Kathy: *Head first design patterns*, " O'Reilly Media, Inc." (2004)

[Fu09] FU, Fong Ling; SU, Rong Chang und YU, Sheng Chin: EGameFlow: A scale to measure learners' enjoyment of e-learning games. *Computers and Education* (2009), Bd. 52(1): S. 101–112

[Fuk05] FUKUDA, Kyosuke; STERN, John A.; BROWN, Timothy B. und RUSSO, Michael B.: Cognition, blinks, eye-movements, and pupillary movements during performance of a running memory task. *Aviation Space and Environmental Medicine* (2005), Bd. 76(7 II): S. 75–85

[Gal01] GALLEY, Niels: Physiologische Grundlagen, Meßmethoden und Indikatorfunktion der okulomotorischen Aktivität. *Enzyklopädie der Psychologie* (2001), Bd. 4: S. 237–315

[Ghe10] GHERGULESCU, Ioana und MUNTEAN, Cristina: MoGAME: Motivation based Game Level Adaptation Mechanism. *the 10th Annual Irish Learning . . .* (2010)

[Gid13] GIDLÖF, Kerstin; WALLIN, Annika; DEWHURST, Richard und HOLMQVIST, Kenneth: Using Eye Tracking to Trace a Cognitive Process : Gaze Behaviour During Decision Making in a Natural Environment. *Journal of Eye Movement Research* (2013), Bd. 6(1): S. 1–14

[Gol99] GOLDBERG, Joseph H. und KOTVAL, Xerxes P.: Computer interface evaluation using eye movements: Methods and constructs. *International Journal of Industrial Ergonomics* (1999), Bd. 24(6): S. 631–645

[Güt04] GÜTL, Christian; PIVEC, Maja; TRUMMER, Christian; GARC, Victor M; PRIPFL, Jürgen und UMGEHER, Martin: AdeLE (Adaptive e-Learning with Eye-Tracking): Theoretical Background, System Architecture and Application Scenarios. *Iknow '04* (2004), Bd. II(2005)(2): S. 1–18

[Hen14] HENNING, Peter A.; HEBERLE, Florian; STREICHER, Alexander; ZIELINSKI, Andrea; SWERTZ, Christian; BOCK, Jürgen und ZANDER, Stefan: Personalized web learning: Merging open educational resources into adaptive courses for higher education. *CEUR Workshop Proceedings* (2014), Bd. 1181: S. 55–62

[Hol12] HOLMQVIST, Kenneth und MULVEY, Fiona: Eye tracker data quality: What it is and how to measure it (2012), Bd. 1(212): S. 45–52

[Hol16] HOLLENBERG, Stefan: *Durchführung von Befragungen mittels Fragebogen*, Springer Fachmedien Wiesbaden, Wiesbaden (2016), S. 23–28

[Hyr00] HYRSKYKARI, Aulikki; PAIVI, Majaranta; ANTTI, Aaltonen und RAIHA, Kari-Jouko: Design Issues of iDict: A Gaze-Assisted Translation Aid. *Eye Tracking Research and Applications Symposium* (2000): S. 9–14

[Jac03] JACOB, Robert J K und KARN, Keith S.: Eye Tracking in Human-Computer Interaction and Usability Research. Ready to Deliver the Promises. *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research* (2003): S. 531–553

[Jen08] JENNETT, Charlene; COX, Anna L.; CAIRNS, Paul; DHOPAREE, Samira; EPPS, Andrew; TIJS, Tim und WALTON, Alison: Measuring and defining the experience of immersion in games. *International Journal of Human Computer Studies* (2008), Bd. 66(9): S. 641–661

[Kah73] KAHNEMAN, Daniel: *Attention and effort*, Prentice Hall (1973)

[Kar11] KARDAN, Samad und CONATI, Cristina: Exploring Gaze Data for Determining User Learning with an Interactive Simulation (2011), (1)

[KR06] KICKMEIER-RUST, Michael; SCHWARZ, Daniel; ALBERT, Dietrich; VERPOORTEN, Dominique; CASTAIGNE, Jean-Loup und BOPP, Matthias: The ELEKTRA project: Towards a new learning experience. *M3 - Interdisciplinary Aspects on Digital Media & Education. Proceedings of the 2nd Symposium of the WG HCI & UE of the Austrian Computer Society.* (2006), Bd. 3(4): S. 19–48

[KR10] KICKMEIER-RUST, M. D. und ALBERT, D.: Micro-adaptivity: Protecting immersion in didactically adaptive digital educational games. *Journal of Computer Assisted Learning* (2010), Bd. 26(2): S. 95–105

[KR11] KICKMEIER-RUST, Michael D.; HILLEMANN, Eva und ALBERT, Dietrich: Tracking the UFO's paths: Using eye-tracking for the evaluation of serious games. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (2011), Bd. 6773 LNCS(PART 1): S. 315–324

[Lop11] LOPES, Ricardo und BIDARRA, Rafael: A semantic generation framework for enabling adaptive game worlds. *Proceedings of the 8th International Conference on Advances in Computer Entertainment Technology* (2011), (ii): S. 6:1—-6:8

[Mar96] MARSHALL, Martin N: Sampling for qualitative research Sample size. *Family Practice* (1996), Bd. 13(6): S. 522–525

[Meh11] MEHIGAN, Tracey J.; BARRY, Mary; KEHOE, Aidan und PITT, Ian: Using eye tracking technology to identify visual and verbal learners, in: *Proceedings - IEEE International Conference on Multimedia and Expo*, IEEE, S. 1–6

[Mer06] MERTEN, Christina und CONATI, Cristina: Eye-tracking to model and adapt to user metacognition in intelligent learning environments. *Proceedings of the 11th international conference on Intelligent user interfaces - IUI '06* (2006): S. 39

[Möd04] MÖDRITSCHER, Felix; GARCÍA-BARRIOS, Victor Manuel und GÜTL, Christian: Enhancement of SCORM to support adaptive E within the Scope of the Research Project AdeLE. *World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education* (2004): S. 2499–2505

[Nac08] NACKE, Lennart und LINDLEY, Craig A.: Flow and immersion in first-person shooters., in: *Proceedings of the 2008 Conference on Future Play Research, Play, Share - Future Play '08*, ACM Press, New York, New York, USA, S. 81

[Nak02] NAKAMURA, Jeanne und CSIKSZENTMIHALYI, Mihaly: The Concept of Flow Optimal Experience and Its Role in Development, in: *Handbook of positive psychology* (2002), S. 89–105

[Ora] ORACLE: Creating Extensible Applications, URL `https://docs.oracle.com/javase/tutorial/ext/basics/spi.html`

[Orv08] ORVIS, Karin A.; HORN, Daniel B. und BELANICH, James: The roles of task difficulty and prior videogame experience on performance and motivation in instructional videogames. *Computers in Human Behavior* (2008), Bd. 24(5): S. 2415–2433

[OSG10] OSGI: OSGi Architecture (2010), URL https://www.osgi.org/developer/architecture/

[Osh03] OSHEROVE, Roy: Creating a Plug-In Framework (2003), URL https://msdn.microsoft.com/en-us/library/ms972962.aspx

[Pal10] PALINKO, Oskar; KUN, Andrew L.; SHYROKOV, Alexander und HEEMAN, Peter: Estimating cognitive load using remote eye tracking in a driving simulator, in: *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*, ACM Press, New York, New York, USA, S. 141–144

[Pei08] PEIRCE, Neil; CONLAN, Owen und WADE, Vincent: Adaptive educational games: Providing non-invasive personalised learning experiences. *Proceedings - 2nd IEEE International Conference on Digital Game and Intelligent Toy Enhanced Learning, DIGITEL 2008* (2008): S. 28–35

[Piv05] PIVEC, Maja; PRIPFL, Jürgen; GÜTL, Christian; GARCÍA-BARRIOS, Victor Manuel; MÖDRITSCHER, Felix und TRUMMER, Christian: AdeLE first prototype: experiences made. *International Conference on Knowledge Management (I-Know)* (2005): S. 673–681

[Piv06] PIVEC, Maja; TRUMMER, Christian und PRIPFL, Jürgen: Eye-tracking adaptable e-learning and content authoring support. *Informatica (Ljubljana)* (2006), Bd. 30(1): S. 83–86

[Por08] PORTA, Marco: Implementing eye-based user-aware e-learning, in: *Proceeding of the twenty-sixth annual CHI conference extended abstracts on Human factors in computing systems - CHI '08*, ACM Press, New York, New York, USA, S. 3087

[Pre07] PRENSKY, Marc: *Digital game-based learning*, Paragon House, St. Paul, Minn. (2007)

[Pri08] PRIKKEL, John: Example Plug-In Framework (2008), URL https://en.wikipedia.org/wiki/Plug-in_(computing)#/media/File:Plug-InExample.png

[Rol05] ROLL, I., BAKER, R. S., ALEVEN, V., MCLAREN, B. M., & KOEDINGER, K. R: *User Modeling and User Profiling in Adaptive E-learning Systems*, Dissertation (2005)

[Rol13] ROLLER, Wolfgang; BERGER, Anton und SZENTES, Daniel: Technology based training for radar image interpreters. *2013 6th International Conference on Recent Advances in Space Technologies (RAST)* (2013): S. 1173–1177

[Rus] RUSTICI: Experience API: Layer 2: Record Any Learning Experience (Informal Learning), URL https://experienceapi.com/layer-2-record-any-learning-experience-informal-learning/

[Sch11] SCHÜTZ, Alexander C; BRAUN, Doris I und GEGENFURTNER, Karl R: Eye movements and perception: A selective review. *Journal of vision* (2011), Bd. 11(5): S. 1–30

[Shi77] SHIFFRIN, Richard M und SCHNEIDER, Walter: Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological review* (1977), Bd. 84(2): S. 127

[Shu03] SHUTE, Valerie und TOWLE, Brendon: Adaptive E-Learning (2003), Bd. 38(2): S. 105–114

[Shu12] SHUTE, Vj und ZAPATA-RIVERA, D: Adaptive educational systems. *Adaptive technologies for training ...* (2012), (1): S. 1–35

[Str13] STREICHER, Alexander; HEBERLE, Florian und BARGEL, Bela: Specification of the Learning Progress Model (2013)

[Str15a] STREICHER, Alexander und ROLLER, Wolfgang: Interoperable Adaptivität für Serious Games und Simulationen in der Bildauswertung. *Trends in Neuroergonomics, 11. Berliner Werkstatt Mensch-Maschine-Systeme* (2015): S. 293–298

[Str15b] STREICHER, Alexander und ROLLER, Wolfgang: Interoperable Adaptivität für Serious Games und Simulationen in der Bildauswertung. *Trends in Neuroergonomics, 11. Berliner Werkstatt Mensch-Maschine-Systeme* (2015), Bd. 7(1): S. 293–298

[Str15c] STREICHER, Alexander und ROLLER, Wolfgang: Towards an Interoperable Adaptive Tutoring Agent for Simulations and Serious Games. *International Conference on Theory and Practice in Modern Computing, MCCSIS 2015* (2015): S. 194–197

[Str16] STREICHER, Alexander und SMEDDINCK, Jan D.: Personalized and Adaptive Serious Games, in: Ralf Dörner; Stefan Göbel; Michael Kickmeier-Rust; Maic Masuch und Katharina Zweig (Herausgeber) *Entertainment Computing and Serious Games: International GI-Dagstuhl Seminar 15283, Dagstuhl Castle, Germany, July 5-10, 2015, Revised Selected Papers*, Springer International Publishing (2016), S. 332–377

[Swe05] SWEETSER, Penelope und WYETH, Peta: GameFlow. *Computers in Entertainment* (2005), Bd. 3(3): S. 3

[Swe13] SWERTZ, Christian; SCHMÖLZ, Alexander; FORSTNER, Alexandra; DAMBIER, Nathalie; HEBERLE, Florian; HENNING, Peter; STREICHER, Alexander; BURGHART, Catherine; BOCK, Jürgen; BADII, Atta; FUENTE, Luis De; PARODI, Elisabetta; THIEMERT, Daniel; GAL, Eran; RONEN, Michaela und ZANDER, Stefan: Lernpfadmodellierung mit der Webdidaktik für die adaptive Erweiterung von Lernmanagementsystemen. *DeLFI 2013 – Die 11. e-Learning Fachtagung Informatik* (2013): S. 277–280

[Tak00] TAKAHASHI, K; NAKAYAMA, M und SHIMIZU, Y: The response of eye-movement and pupil size to audio instruction while viewing a moving target. *Proceedings of the Symposium of Eye Tracking Research {&} Applications* (2000): S. 131–138

[Uen04] UENO, a und UCHIKAWA, Y: Relation between human alertness, velocity wave profile of saccade, and performance of visual activities. *Engineering in Medicine and Biology Society* (2004), Bd. 2: S. 933–935

[Wan05] WANG, Hua; CHIGNELL, Mark und ISHIZUKA, Mitsuru: Empathic tutoring software agents using real-time eye tracking, in: *Eye Tracking Research and Applications Symposium (ETRA)*, Bd. 2005, ACM Press, New York, New York, USA, S. 73–78

[Wei09] WEI, Haolin; MOLDOVAN, Arghir-Nicolae und MUNTEAN, Cristina: Sensing Learner Interest Through Eye Tracking (2009)

[Wet14] WETZEL, Stefanie; SPIEL, Katharina und BERTEL, Sven: Dynamically adapting an AI game engine based on players' eye movements and strategies. *Proceedings of the 2014 ACM SIGCHI symposium on Engineering interactive computing systems - EICS '14* (2014): S. 3–12

[Wic00] WICKENS, Christopher D und HOLLANDS, Justin G: *Engineering psychology and human performance*, Prentice Hall (2000)

[Wik14] Wikiversity: Self-efficacy and motivation: How does self-efficacy enhance motivation? (2014), URL https://en.wikiversity.org/wiki/Motivation_and_emotion/Book/2014/Self-efficacy_and_motivation

[xAP] xAPI Specification, URL https://github.com/adlnet/xAPI-Spec/blob/a752217060b83a2e15dfab69f8c257cd86a888e6/xAPI.md

List of Figures

List of Tables

Appendix A: Questionnaires of Evaluation

Attention Questionnaire

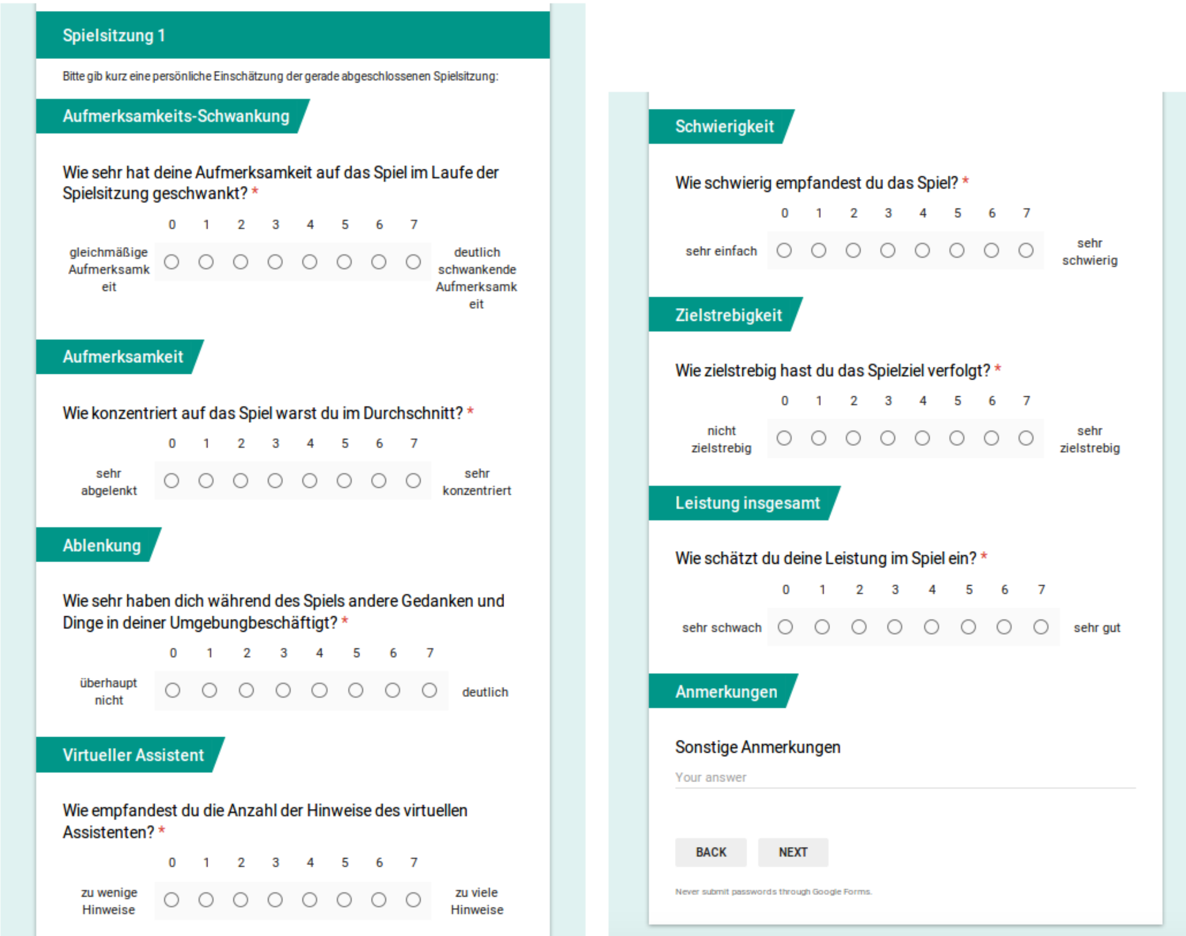
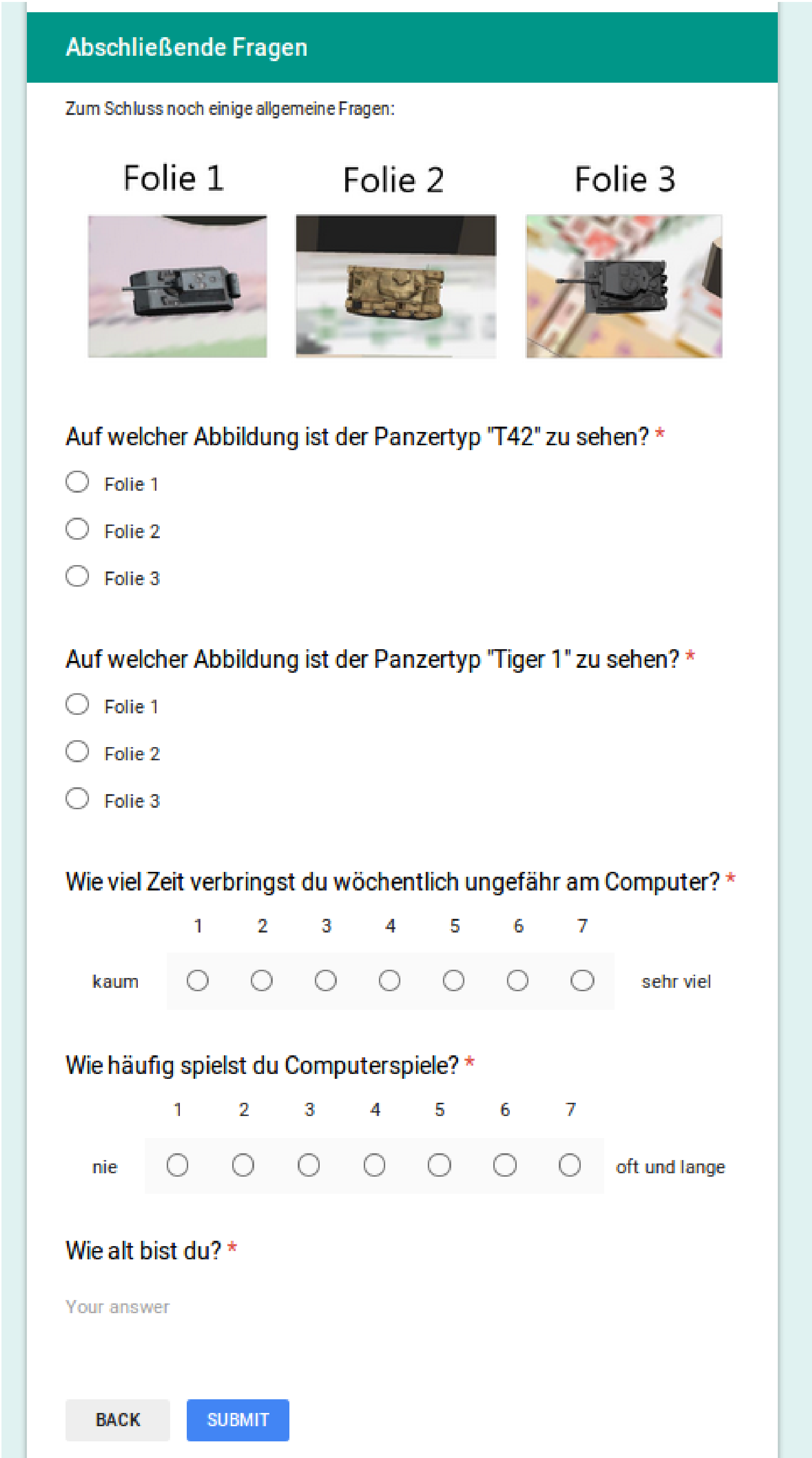


Figure 1: Original Questionnaire of the Experiment on Attention

Concluding Questionnaire



Abschließende Fragen

Zum Schluss noch einige allgemeine Fragen:

Folie 1 Folie 2 Folie 3

Auf welcher Abbildung ist der Panzertyp "T42" zu sehen? *

- ○ Folie 1
- ○ Folie 2
- ○ Folie 3

Auf welcher Abbildung ist der Panzertyp "Tiger 1" zu sehen? *

- ○ Folie 1
- ○ Folie 2
- ○ Folie 3

Wie viel Zeit verbringst du wöchentlich ungefähr am Computer? *

	1	2	3	4	5	6	7	
kaum	○	○	○	○	○	○	○	sehr viel

Wie häufig spielst du Computerspiele? *

	1	2	3	4	5	6	7	
nie	○	○	○	○	○	○	○	oft und lange

Wie alt bist du? *

Your answer

BACK SUBMIT

Figure 2: Original Questionnaire of the Experiment on Learning and Demographics