Entertainment Computing 31 (2019) 100307

Contents lists available at ScienceDirect

Entertainment Computing

journal homepage: www.elsevier.com/locate/entcom

Modeling and adjusting in-game difficulty based on facial expression analysis

Paris Mavromoustakos Blom[a,*], Sander Bakkes[b], Pieter Spronck[a]

[a] Tilburg University, Warandelaan 2, 5037 AB Tilburg, the Netherlands
[b] Utrecht University, Domplein 29, 3512 JE Utrecht, the Netherlands

ARTICLE INFO

Keywords:
Facial expression analysis
Game personalisation
Game difficulty adaptation
Dynamic difficulty adjustment

ABSTRACT

In this paper we introduce Facial Expression Analysis (FEA) both as a means of predicting in-game difficulty and as a modeling mechanism, based on which we develop in-game difficulty adjustment algorithms for single player arcade games. Our main contribution is the implementation of an online and unobtrusive game personalisation system. On the basis of FEA, our system is able to adapt the difficulty level of the game to the individual player, without interruptions, during actual gameplay.

Specifically, we study (a) how *perceived* in-game difficulty can be measured through facial expression analysis, and (b) how facial expression data can model player behavior and predict their affective state.

Experimental findings reveal that different in-game difficulty settings can be correlated to distinct player emotions (revealed in facial expressions). Furthermore, a model based on facial expression analysis is successfully applied to calculate an appropriate difficulty setting, tailored to the individual player. From these results, we may conclude that efficient game personalisation is achievable through FEA.

1. Introduction

In this paper, we study the correlation between players' emotional state (as observed through FEA) and in-game difficulty setting, in an attempt to validate FEA as an approach towards game personalisation. Furthermore, we research how FEA can be leveraged to provide an efficient mechanism for online and unobtrusive personalised games, where the game experience is continuously tailored to fit the individual player. Through FEA, we aim to extract a reliable and objective description of players' affective state, which will be applied in order to create personalised game spaces (levels).

The goal of this research is to generate personalised game levels based explicitly on FEA data, extracted during actual gameplay. We consider this to be a challenging task, given the diversity among players in terms of emotional expressiveness, gender, age, and in-game capabilities. An efficient game personalisation mechanism based on FEA is not only a novelty in the field of gaming AI, but may also set the foundations for further research on game personalisation through computer vision. Given the commercial popularity of webcams, FEA provides an accessible and unobtrusive input channel to apply player affect modeling methods and subsequently, Dynamic Difficulty Adjustment (DDA). We believe that this study could provide the basis for a generalisable player affect model, implemented across multiple games.

In order to examine whether personalised games can be implemented through FEA, we have divided our work into two parts:

The first part investigates the correlation between in-game difficulty and player emotional state, as described by the players' facial expressions. Through machine learning techniques, it aims to show that players' facial expressions are indicative of perceived in-game difficulty. Specifically, through a classification task, we are able to predict the difficulty level of the game currently played with high accuracy, by analysing players' facial features.

The second part, expands on the first part's findings, and introduces FEA as a technique towards measuring player engagement and dynamically adjusting in-game difficulty. Since it was shown that player facial expressions are an indicator of current game difficulty in the first part, we propose a FEA-based game personalisation system which aims at improving individual player in-game experience, while at the same time maximising player affection levels towards the game itself.

2. Related work

In this section, we present relevant research in the fields of Game

* Corresponding author at: Department of Cognitive Sciences and Artificial Intelligence Tilburg University, Dante 334, Warandelaan 2, 5037 AB Tilburg, the Netherlands.

E-mail addresses: p.mavromoustakosblom@uvt.nl (P.M. Blom), s.c.j.bakkes@uu.nl (S. Bakkes), p.spronck@uvt.nl (P. Spronck).

https://doi.org/10.1016/j.entcom.2019.100307
Received 21 March 2019; Accepted 13 June 2019
Available online 10 July 2019

Personalisation, Dynamic Difficulty Adjustment (DDA), and Affective Feedback. Our research is based on findings in these fields, each of which contributes towards implementing our game adaptation mechanism.

2.1. Game personalisation

Game personalisation is motivated by a significantly increased involvement and extensive cognitive elaboration when players are exposed to content of personal relevance [26]; it has been shown that they will exhibit stronger emotional reactions [9]. A positive effect on player satisfaction has been found, i.e., game personalisation raises player loyalty and enjoyment, which in turn can steer the gaming experience towards a (commercial) success [40]. The perspective of AI researchers to increase the engagement and enjoyment of players is one that is consistent with the perspective of game designers [28], i.e., personalisation methods are regarded as instrumental for achieving industry ambitions [21]. Tailoring the game experience to the individual player particularly benefits from the use of player models, and requires components that use these models to adapt parts of the game [4].

Our research follows the emerging trend of employing AI methods for adapting the game environment itself (as opposed to, more typically, adapting the behaviour of the game characters) [4]. In our investigation, we choose to focus on *personalising the game space* to the individual player with respect to experienced challenge. We define the "game space" as the representation of the game mechanisms through which the user interacts with the game.

2.2. Dynamic Difficulty Adjustment (DDA)

DDA is a technique for continuously adapting a game so that it maintains a certain level of difficulty [43,38,36]. In most cases, it aims at adapting the difficulty setting of a game to fit the skills and/or game experience of the individual player. DDA can be implemented using various approaches, from 'simple' heuristics to 'complicated' machine learning methods.

Various applications of DDA in games are found in literature. Hunicke and Chapman [14] introduce a probabilistic method for representing and reasoning about uncertainty in games. Their main goal is to keep the player engaged for the appropriate amount of time (also referred as keeping players in the "flow area", as described in Csikszentmihalyi's theory of Flow [8]). Zook and Riedl [46] follow a temporal data-driven approach, focused on modeling and predicting player performance instead of in-game difficulty. This approach is based on the assumption that in-game difficulty is a subjective measurement. Xue et al. [43] use a probabilistic graph to model player progression. In-game difficulty is dynamically adjusted in order to maximise a player's stay time in the progression graph. Jennings-Teats et al. [16] introduce a statistical approach, where a multilayer perceptron algorithm was built in order to rank generated level segments as a model of difficulty.

A term closely related to DDA is Challenge Balancing. Challenge balancing concerns automatically adapting the challenge that a game poses to the skills of a player [22,37]. It aims at achieving a 'balanced game', i.e., a game wherein the player is neither challenged too little, nor challenged too much. In most games, the only implementation of challenge balancing is provided by a *difficulty setting*, i.e., a discrete parameter that determines how challenging the game will be. However, as the challenge provided by a game is typically *multi-faceted*, it is hard for the player to estimate reliably the challenge level that is appropriate. Furthermore, generally only a limited set of discrete difficulty settings is available (e.g., easy, normal, hard). This entails that the available settings are not fine-tuned to be appropriate for each player. As such, researchers have developed advanced techniques for balancing the challenge level of games. Hunicke and Chapman [14] explored challenge balancing by controlling the strength of opponent characters (i.e., controlling the opponent character's health, accuracy, and employed weapons). Spronck et al. [37] investigated methods for automatically adjusting weights assigned to possible game scripts. Knowledge on the specific effect of game adaptations can be employed for maintaining a challenge level [3], and may be incorporated to steer the procedural generation of game content [42].

In our research, we take the distinct focus of balancing the game's challenge level by dynamically adapting the difficulty of the *content* that is placed within the game environment. We also focus on procedural content generation for tailoring the player experience. Our distinct focus in this matter, is to assess *online* and *unobtrusively* which game adaptations are required for optimising the individual player's experience while the game is being played, to have assessments on how the experienced player challenge impacts the procedural process (*cf.* Bakkes et al. [2]).

2.3. Affective feedback

Following the principles of affective computing as described by Picard et al. [27], researchers have been investigating the use of affective signals within Human-Computer Interaction (HCI) systems. More specifically, affective signals such as heart rate, skin conductance and facial expressions have been employed in order to build player models and create personalised gaming experiences.

Chanel et al. [7] introduce an approach based on emotion recognition, to model Tetris players with respect to three affective states: Boredom, anxiety and engagement. They have achieved a classification accuracy of 53.33%, using various features extracted from physiological sensors and questionnaires. Liu et al. [19] propose real-time DDA through player anxiety estimation, using wearable biofeedback sensors. In the present work, we employ Facial Expression Analysis as a method of applying real-time and unobtrusive DDA.

Facial expression analysis is a mature domain in computer vision with techniques that boast a high level of accuracy and robustness [11,31,18] without the requirement of expensive hardware [6]. For example, Buenaposada et al. have reported a 89% recognition accuracy in video sequences in unconstrained environments with strong changes in illumination and face locations.

Zaman and Shrimpton-Smith [45] evaluated an automated facial expressions analysis system to infer emotions that users had whilst performing common computer usage tasks. They generally reported a high level of correlation between the system's findings and human expert analyses. Tan et al. [39] performed a feasibility study in using facial expression analysis to evaluate player experiences. They concluded that typical game experiences yield a good variety of facial expressions (other than neutral) with variances of expression for individual participants being generally rich.

We postulate that the affective signals provided by FEA, can be utilised for efficient and accurate game personalisation. Since user facial expressions can be indicators of emotion [11], we investigate whether in-game difficulty can have an impact on player affect, as expressed through the face. In our research, we will leverage FEA in an attempt to predict the perceived and actual in-game difficulty, before proceeding to implement our game personalisation algorithms.

3. Methodology

In this section, we will give an outline of the methodology used in this research, as shown in Table 1.

This research is divided into two parts. The first part explores whether FEA during gameplay can lead to accurate determination of the perceived in-game difficulty. If it can, this will enable successful personalisation of games using facial expression analysis. This is the main topic of the second part, which focuses on the personalisation procedure, proposing various algorithms aiming towards reaching optimal player affection levels.

In the first part, users are requested to score three different versions

Table 1
Outline of the methodologies used in this research.

Methodological Part	Domain	Research Goals	Facial Expression Recognition Toolbox
First Part	PACMAN	Measuring in-game difficulty based on FEA	CERT
Second Part: Heuristics	INFINITE MARIO BROS.	Game personalisation based on Heuristics	INSIGHT
Second Part: Modeling	INFINITE MARIO BROS.	Game personalisation based on a trained Model	INSIGHT

of the same game in terms of difficulty. During gameplay, we detect player emotions and track facial Action Units (AU), which are fed as input to classification algorithms that aim to predict the perceived in-game difficulty.

The second part introduces two methods of translating probabilistic estimates of player emotions into in-game difficulty adaptations. A heuristic algorithm that adapts in-game difficulty based on player emotion estimates, and a model of player affection levels, are trained on emotion & head pose measurements. The goal is to build an effective game personalisation system, which should maximise player affection (as measured through FEA), while simultaneously converging to an optimal in-game difficulty level, for the individual player. The results are based on player feedback, through pairwise tests that compare the dynamic (adaptable) version to the static version of the same video game.

Although we acknowledge that facial expressions are not always strictly correlated to user affective state, we have chosen to use the term “emotion” to refer to one of the seven basic emotions as described by Ekman & Friesen [12]. We consider these emotions (as detected by the INSIGHT facial expression recognition SDK [35]) an adequate descriptor of player affective state in this study’s setting.

4. Measuring perceived in-game difficulty based on facial expression analysis

In the first part of our research, we aim to justify the use of FEA as an appropriate approach towards game personalisation. By predicting the perceived in-game difficulty, we show that player facial expressions can be correlated to different versions of the same game. Building a system that can accurately predict in-game difficulty solely by ‘looking’ at the player’s facial expressions, means that we will be able to dynamically adjust in-game difficulty to provide personalised game experiences. Below, we present our methods for predicting perceived in-game difficulty, by analysing players’ facial expressions. More details about this specific part of the study can be found in [20].

4.1. Experimental setup

In this experiment, the aim is to predict the *perceived* in-game difficulty through FEA. Players are asked to play different versions of the same video game, while their facial expressions are being recorded.

4.1.1. Pacman

Pacman is a widely popular arcade game, considered one of the classic games in its genre. The goal of the game is to navigate through a maze while accumulating points by ‘eating’ dots. Each ‘stage’ in the Pacman game is finished when all dots are eaten by the player.

Players participating in this experiment are asked to complete three different versions of the Pacman game. For each version the participant was instructed to play the game with the intention to collect all dots as fast as possible.

Similar to several other studies (Aponte, Levieux, and Natkin [47]; Li et al. [18]; Girouard et al. [48]) this study used a version of Pacman which the author altered to make the game more suitable for the experiment. The alteration revolved around the speed by which the player was allowed to move within the maze featured in the game. Since this research pertains data obtained from facial expressions, it was deemed important to eliminate factors that might negatively influence the experience of the game. Whereas ‘dying’ is a natural element of the game, the standard Pacman game only grants the player a limited number of lives. In our experiment, the player was granted an unlimited number of lives in order to avoid unnecessary frustration for some levels [50].

To familiarise themselves with the game, players were first presented with a short practice session. When participants were comfortable with the game controls, they were directed to the experiment.

4.1.2. Facial expression recognition & hardware setup

For the experiment, two laptops were used simultaneously. One laptop was used to run the game; another was used to record the participants’ faces during their play sessions. The following method was employed: participants were asked to sit in front of the first laptop on which a webcam was mounted. The camera itself was plugged into the second laptop that was running the recording software.

During gameplay, CERT (Computer Expression Recognition Toolbox) was used to extract facial expression data. CERT is able to track basic expressions of anger, disgust, fear, joy, sadness, surprise, contempt, a continuous measure of head pose (yaw, pitch, and roll), as well as 30 facial action units (AU’s) from the Facial Action Coding System [1]. The Facial Action Coding System (FACS) is designed by Ekman and Friesen and is used to “measure all facial behaviour, not just actions that might presumably be related to emotion” [11]. Ekman and Friesen describe action units to be anatomically separate and visually distinguishable facial movements. For an extensive survey on automatic facial expression analysis, including FACS, we refer readers to Fasel & Luettin [13].

4.1.3. Data collection

The data was processed as follows: For every participant, the average over all the frames per AU was calculated. The resulting value was subtracted from all individual values to account for the “neutral” expression, thereby correcting against a baseline. In the next step, the altered values were compared with a threshold: According to Grafsgaard et al. [49] a value (with a corrected baseline) above 0.25 indicates the presence of the emotion. The resulting values were binary, indicating either the presence or absence of a specific expression. Finally, after each session participants were asked to rate the difficulty of the game that they just played on a scale of 1 (very easy) to 7 (very difficult). This value is the “perceived difficulty”.

These binary values were used in a classification task where the goal was to predict the in-game difficulty level. For this task, the data mining program “Orange” [10] was used. The data was processed using 10-fold cross validation in a leave-one-out setting employing a classification tree, a Naïve Bayes classifier, and ZeroR as a baseline method.

In total, the recordings of participants that were collected provided 117,304 instances of frame-data (vectors consisting of AU values). For each participant, these vectors of frame data were subsequently labelled with the specific version each vector referred to and the perceived difficulty score that was assigned to this specific version by each participant. The perceived game version was selected as the target class during the classification tasks.

4.2. Results

The data for this experiment were obtained from 38 (N = 38)

Table 2
Perceived difficulty scoring by players for each version of Pacman on a 7-point scale. Each experiment setting indicates the order in which game versions were played.

	Perceived Difficulty		
Experiment setting	version A	version B	version C
1. A-B-C	2.8	3.5	3.8
2. A-C-B	4.2	5.0	4.0
3. B-A-C	3.0	5.0	3.4
4. B-C-A	3.0	4.0	3.2
5. C-A-B	3.0	4.2	3.2
6. C-B-A	3.0	3.2	4.2

participants. The average age of the participants was 35.1 years (SD = 14.54). Of these participants, 24 were recruited in the University library; the other 14 were selected from the of the authors' social circle. The data of 6 participants was left out of this research because of recording failures. Of the remaining 32 participants, 18 were male, 14 female.

Participants played through three different versions of Pacman. In version 'A', the speed with which Pacman moved through the maze was slower than that of its "enemies". In version 'C' Pacman's speed was higher than the enemies' speed, while in version 'B', Pacman's speed was left as originally designed.

We have provided an overview of the perceived difficulty scorings for each version, spread out over the different settings of the experiment in Table 2. For each setting of the experiment, the playing sequence of game versions is changed. Note that the scores presented here are subjective scores. For instance, in setting 5 of the experiment, participants first played version C, followed by versions A and B. In setting 2 of the experiment, the order of play was A-C-B. We see that the perceived difficulty of each version of the game is different for each experiment setting. For instance, whereas version A in setting 1 of the experiment was awarded an average difficulty score of 2.8, the same version was awarded a difficulty score of 4.2 in setting 5 and 6 of the experiment.

Table 3 shows that a classification tree provided the most accurate results, scoring an accuracy of 77% when predicting the perceived version of a game, without taking the actual version into account. A Naïve Bayes classifier provided an accuracy of 53%. While these results pertain to a combination of all levels, in Table 3 we also observe how the classification accuracy increases up to 90% when the actual game version is taken into account. This increase can be explained by the fact that the classifier uses a smaller training set, which is likely to have a more dominant majority class.

4.3. Discussion

Regarding perceived difficulty scores, Table 2 might seem counterintuitive and raise some questions. We expected that participants would rate Pacman version 'C' at a lower difficulty score than the other versions, since Pacman's moving speed was higher than the speed of its enemies. Version 'A' was expected to be rated as the most difficult one, given that Pacman was unable to "shake off" trailing ghosts. However, we observed that it required more game-playing skill to keep up with the fastest version ('C'), and only experienced players would rate that version at a lower difficulty.

Table 3
Prediction accuracy of perceived game version using facial expression analysis.

Method	Combined	version A	version B	version C
Classification Tree	77%	90%	87%	89%
Naïve Bayes	53%	63%	70%	74%
ZeroR	44%	32%	45%	55%

For all classification tasks, Naïve Bayes underperformed when compared to classification via a decision tree. Naïve Bayes classifiers assume that features are independent, while in our case, even though facial expressions can be isolated, the dynamics of the human face make it that facial expressions are often correlated or even dependent. For instance, to express disgust, a person has to use both the nose (AU 4) and the upper-brow (AU 9/10) to make the appropriate expression. Feature independence might cause Naïve Bayes classifiers to mistakenly assume that the AUs are independent of each other, whereas in actuality they are found in conjunction.

The results of the present study point in the direction where it is possible to accurately predict the perceived difficulty of a game. We were able to predict the perceived game version with an accuracy of 77%. However, when currently played game version was included as a classifier feature, prediction accuracy was increased to 88.7%. Considering the high accuracy of our predictions of the current difficulty of the game, it is implied that it is possible to use this data to make effective adjustments to the game.

4.4. Conclusion

In this preliminary study, we focused on confirming whether FEA can provide a sufficient tool towards personalised gaming. We have shown that in-game difficulty can be accurately measured solely through FEA. The ability to predict in-game difficulty through player facial expressions allows us to assume that varying in-game difficulty settings can trigger different emotions from the player, and consequently, indicate that facial expressions are an adequate descriptor of player affection level.

Based on these findings, we proceeded to implement an online and unobtrusive game personalisation mechanism, using FEA. In our main study, in-game difficulty will be dynamically adapted based on players' affective state, measured during gameplay.

5. Adapting in-game difficulty based on facial expression analysis

Having shown that player facial expressions can be correlated to in-game difficulty settings in Section 4, we will proceed to implement game personalisation algorithms. This section is focused on translating FEA data into actual in-game difficulty adjustments, creating personalised game spaces tailored to the individual player.

By introducing game personalisation methods we aim to achieve and retain a high level of user engagement. Based on the study in [8], our goal is to dynamically adjust in-game difficulty to keep players engaged for as long as possible.

This study is divided into two distinct sections: The initial (heuristic) approach describes heuristic methods of translating player emotions (as measured through FEA) into in-game difficulty adaptations, using a Gradient Ascent Optimisation (GAO) algorithm. The extended (modeling) approach is an extension of the former, by introducing head pose measurements (pitch, roll and yaw) as well as classification algorithms in order to create a probabilistic model of players' affective state. The model will ultimately be used to predict optimal in-game difficulty adjustments during gameplay.

5.1. Experimental setup

5.1.1. INFINITE MARIO BROS

We consider a typical video game: INFINITE MARIO BROS.[25]; an open-source clone of the classic video game SUPER MARIO BROS. It can be regarded an archetypic platform game; despite its relatively straightforward appearance it provides a diverse and challenging gameplay experience. We built upon a version of INFINITE MARIO BROS. that has been extended to procedurally generate entire Mario levels. These extensions have been made by Shaker et al. [32–34], Pedersen et al. [23,24], and

Togelius et al. [41].

We have made two further enhancements to the 2011 Mario AI Championship game engine of INFINITE MARIO BROS. First, we enhanced the engine such that it is able to procedurally generate *segments* of Mario levels while the game is in progress (Fig. 1). One game segment has a width of 112 game objects, and generally takes a player approximately 20–30 s to complete. This enhancement enables feedback on the observed player experience to rapidly impact the procedural process that generates the upcoming level segments. The upcoming level segments are generated seamlessly, such that no screen tears occur when the user is transitioning from one segment to the next (i.e., before the next segment can be observed a short 'gap' block is injected in the game space).

Our second enhancement to the game engine, is that within every segment we can now inject short *chunks* of specific game content. We enabled the game engine to generate five different types of chunks, (1) a straight chunk, containing enemies and jumpable blocks, (2) a hill chunk, also containing enemies, (3) a chunk with tubes, containing enemy plants, (4) a jump, and (5) a chunk with cannons. Each chunk can have six distinct implementations, determined by a per-chunk integer parameter $\in [0, 5]$. The challenge level of the chunk monotonically increases with the parameter value (e.g., a hill parameter value of 0 entails a chunk with no hills and no enemies, while a value of N, $N > 0$ entails N procedurally-generated hills with N relatively difficult enemies). Our enhanced engine has the desired property that the generated chunks are largely independent from each other, i.e., only in rare cases will one chunk be able to affect player behaviour in the surrounding chunks (e.g., a cannon bullet following the player to the next chunk). To benefit playability and level aesthetics, the order in which the chunks are encountered is randomised for each new segment. The five chunks (each 16 game objects in length) are preceded and succeeded by a flat, neutral chunk (also 16 game objects in length), to allow the player to prepare for the next game segment. To avoid player emotion estimates being "carried" from one segment to the next, a brief straight section has been added between segments, where we expect player facial expressions to fade out.

In both studies conducted (Sections 5.2 and 5.3), participants were asked to compare an adaptive to the static (baseline) version of SUPER MARIO BROS. The experiments were conducted in a home setting, with stable lighting conditions. In the static game, each chunk's difficulty is pre-defined and persistent throughout the entirety of the session. All chunks were assigned the same difficulty level $D \in [0\ldots5]$, namely $D = 1$ for the 'easy' setting, $D = 3$ for the 'normal' setting and $D = 5$ for the 'hard' setting. Difficulty level 1 was chosen over difficulty level 0 as the 'easy' setting, as difficulty level 0 removes all obstacles from the game and makes it impossible for the player to fail. We consider the static version to be a valid baseline, since the quality and quantity of obstacles is predefined for each of the difficulty levels. This type of game resembles the classic SUPER MARIO BROS. game, which we assume was designed to maximise entertainment.

5.1.2. INSIGHT facial expression recognition toolkit

In our approach, player emotions are tracked with the INSIGHT facial expression recognition SDK [35] during the entire game session, yet are taken into account independently for each chunk. We used INSIGHT rather than CERT in this part of the study for licensing reasons. That is, player expressions measured in chunk *c* (e.g., a chunk with cannons as content), will in our approach only affect the challenge level of that particular chunk type. As a result, online personalisation is achieved at the *content level* of the game. This characteristic allows the online personalisation to specifically tailor the challenge level of a certain content type, to the measured affective state of the player when interacting with this content.

INSIGHT classifies facial expressions at approximately 15 frames per second. For each frame, it outputs a probability distribution over seven distinct emotions, namely (1) neutralness, (2) happiness, (3) disgust, (4) anger, (5) fear, (6) sadness, and (7) surprise. Player emotions are being recorded over the course of a game session and used as input for the game adaptation mechanism, in the form of a vector $[e_1, e_2, \ldots, e_7]$, where e_n represents the probability estimate of a player expressing emotion n. Data is enriched by timestamps for each tracked frame, in order to allow the game engine to correlate between game frames and INSIGHT tracked frames. Depending on the progress of the player through the Mario game, a game chunk is typically interacted with for 2–10 s, resulting in a total of 30–150 classifications for each game chunk separately. The resulting probability distributions are averaged at the end of each chunk, into an estimate of a players's emotional state; it is an estimate that is relatively insensitive to classification noise of the facial expression system (which may occur in individual frames).

5.1.3. Facial expression tracking

There are two events at which assessments of the player's affective state are used to adapt the game; namely (1) when the next level segment needs to be generated, and (2) when the game resets due to player death. To this end, we take into consideration not only player assessments made during actual play of the game, but also in between ingame deaths of the player – as we observed that during this observational period many game players express high emotional activity. Furthermore, we particularly consider that most game players tend to maintain a relatively neutral facial expression during gameplay, with most emotional activity occurring when players experience an in-game death.

The described experimental setup is employed in two distinct experiments, described in the following sections.

5.2. Heuristic Approach: Game Personalisation through Facial Expression Analysis[1]

The heuristic approach uses a heuristic method of adjusting in-game difficulty based on classifications of player emotions. The main challenge in this respect is making accurate assessments of the player's expressions while he is playing the game, and mapping these assessments into challenge levels that are appropriate for the observed player. To that end, we employ a Gradient Ascent Optimisation (GAO) technique. GAO aims at optimising the challenge levels for each content type in the game (i.e., for each chunk type) such that human interactions with the content yield affective states that we consider desirable (i.e., happiness), while not yielding affective stances that we do not consider desirable (i.e., anger). In the present setup, we consider anger an undesirable emotion, and happiness a desirable emotion.

In this experiment, we investigate how participants experience the personalised game under actual game playing conditions, in comparison with a realistic (baseline) static game. To this end, in accordance with procedures employed by Shaker et al. [32]), we query for *pairwise preferences* (i.e., "is system *A* preferred over system *B*?"), a methodology with numerous advantages over rating-based questionnaires (e.g., no significant *order of reporting* effects) [44]. We perform pairwise tests of a static system *s*, with a fixed difficulty level, and a personalised system ***p***. The experiment follows a *within-subjects design* composed of two conditions (C1 and C2), consisting of a series of three sequentially performed pairwise tests, in randomised order. The pairwise tests compare the static system vs. the personalised system, both starting at identical challenge levels. Table 4 gives an overview of the resulting experimental conditions, with the initial challenge level of a system indicated between brackets.

The experiment is performed on ten participants, aged between 23 and 28 years, recruited at the University of Amsterdam. To minimise user fatigue impacting the experimental results, each of the six game-playing sessions is ended after a maximum of 4 level segments (i.e.,

[1] This section builds upon experiments that have been published in [5].

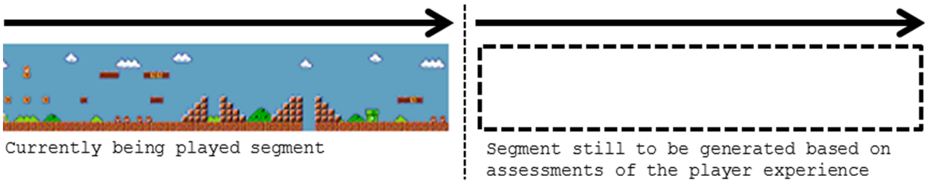


Fig. 1. Our enhanced version of INFINITE MARIO BROS. During gameplay it generates short *new level segments* of specific content *on-the-fly*, on the basis of classifications of the facial expression.

Table 4
Pairwise tests of the static system versus the personalised system. The initial challenge level is indicated between brackets.

Initial challenge level	Condition 1	Condition 2
Easy	s(easy) vs. p(easy)	p(easy) vs. s(easy)
Normal	s(normal) vs. p(normal)	p(normal) vs. s(normal)
Hard	s(hard) vs. p(hard)	p(hard) vs. s(hard)

approximately three minutes of play). After completing a pair of two games, we query the participants's preference through a 4-alternative forced choice (4-AFC) questionnaire protocol (e.g., *s* is preferred to *p*, *p* is preferred to *s*, both are preferred equally, or neither is preferred). The question presented to the participant is: "For which game did you find the challenge level more preferable?".

Fig. 2 lists the pairwise preferences as reported by the participants. The results show that when both gaming systems are set to an initial challenge level of 'easy', a significant majority (Z-test: $p = 0.037$) of participants prefers the personalised system over the static system (seven over three participants). Furthermore, we observe that when both gaming systems are set to an initial challenge level of 'normal', a significant majority (Z-test: $p = 0.037$) of participants prefers the personalised system over the static system (also seven over three participants). However, when both gaming systems are set to an initial challenge level of 'hard', only four participants prefer the personalised system over the static system (three participants), with the remaining three participants preferring neither.

Overall, we detect a preference towards the personalised system in 18 out of 30 tests, while only in 9 out of 30 tests the static system is preferred.

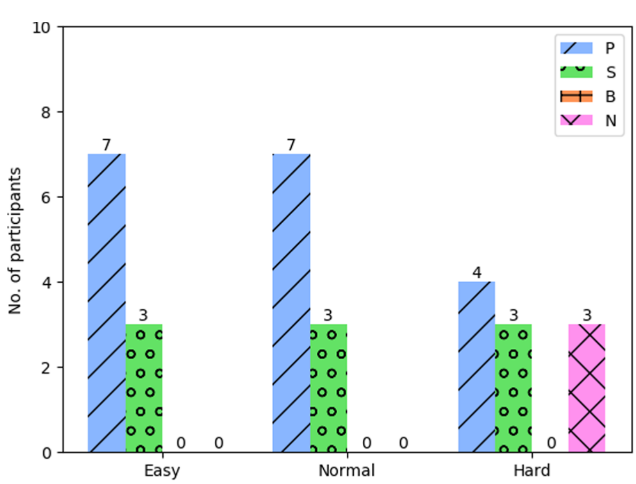


Fig. 2. Pairwise preferences of participants, per initial challenge level. The legend is a as follows, 'P' indicates a preference for the personalised system, 'S' indicates a preference for the static system, 'B' indicates that both are preferred equally, and 'N' indicates that neither is preferred; both are equally unpreferred.

5.3. Modeling approach: game personalisation through facial expression & head pose estimation analysis

This approach discusses dynamic difficulty adjustment through a trained player model, which is based on facial expression & head pose estimation analysis. Similarly to the heuristic approach, our aim is to translate player emotions into in-game difficulty adjustments. However, instead of using heuristics, we employ a classification task to predict player challenge level (the perceived chunk-specific difficulty level).

5.3.1. Model feature selection

Although neutralness, anger and happiness were the most frequently expressed emotions, all emotions classified by INSIGHT will be employed as features in our model. In that way, in cases where happiness, anger and neutralness are not the 'dominant' observed emotions, we are still able to extract information about player engagement.

Furthermore, to be able to measure user engagement beyond facial reactions, we have decided to track head pitch, roll & yaw along with the emotions vector. We have observed that players tend to change their posture with respect to the computer screen during the course of a game, either impreceptively or suddenly, while expressing certain emotions. Such movements can negatively impact the quality of the emotion estimations made by INSIGHT. For example, by tilting their head downward, players may be mistakenly classified as 'angry' even though their expression is neutral. By adding head pitch, roll & yaw to our system we aim to correlate emotion measurements with head pose measurements, in an attempt to 'explain' sudden bursts of emotions.

Lastly, we have decided to add both current in-game difficulty and an estimation of player challenge level to the feature set. Current difficulty level tracking can help discriminate spontaneous from consistent emotional activity, assuming that harder difficulty levels can cause persistent frustration, while lower difficulty levels tend to be encountered with higher average neutralness by players. Furthermore, in Section 4.2 we have shown that incorporating the actual difficulty level in the classifier substantially improves predictions on perceived difficulty. Challenge level is represented by a user-feedback based 5-point Likert scale for each game chunk. It is an integer value in the span [1,2, …,5], with 1 meaning 'too easy', 5 meaning 'too challenging' and 3 representing 'optimal challenge level'.

5.3.2. Model training

We have chosen to predict player challenge level using a Random Forest Classifier (RFC), because it enables us to retrieve a probability distribution over all possible output classes given unknown input. Also, its computational efficiency is considered to be appropriate for online adaptation. The RFC parameters used were $depth = unlimited$, $numberOfIterations = 100$, $bagSize = 100$.

In order to train the RFC, we introduced players to the personalised game, and asked them to finish 10 segments of INFINITE MARIO BROS. at each difficulty level ([1,2,…,5]). After completing each segment, users manually determined a Likert value (1–5) for each chunk separately through an in-game self-report prompt. This user feedback-based method has been employed in previous studies, such as Shaker et al. [32], and derives from IJsselsteijn's Game Experience Questionnaire (GEQ) [15]. In total, we have created a training set of approximately 1250 instances (25 participants), each labeled with a Likert estimate. We have selected both female and male players with varying skill

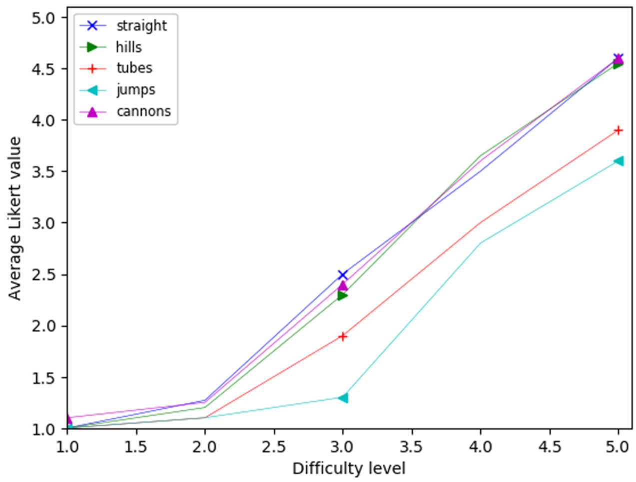


Fig. 3. Average chunk specific Likert preference during system training.

levels, to train the RFC, so as to include as much variety in our data as possible during the training phase.

In Fig. 3, the average Likert value determined by the players participating in the training phase is illustrated. One can observe a tendency of the average user to consider the hardest 'Jump' chunk possible (5) as the approximately optimal challenge level, while most of the chunks' optimal challenge level (3) seems to lie between difficulty levels of 3 and 4.

Player challenge level, described via the Likert estimate, is our model's target class, during the testing phase. When unknown instances are acquired by players, the expected output of the RFC is a probability distribution over the Likert estimates for each chunk type. The classification task is performed at the end of each segment, or right after in-game death. The Likert probability distribution estimated online by the RFC will immediately be used to perform game adaptations as described below.

Algorithm 1. Game Personalisation using a Random Forest Classifier

procedure RFCPERSONALISATION(I_n) ▷Unknown Instance
  $P_{L_i} \leftarrow classifyInstance(I_n)$
  for $i \leftarrow$ *each: chunk* **do**
    $E_{likert} = L[i]{*}P_{L_i}$
    $E_{normalised} = E_{likert}{*}1.5 - 4.5$
    $round(E_{normalised})$
    $newDifficulty = previousDifficulty - E_{normalised}$
  return *newDifficulty*

5.3.3. Game personalisation

Algorithm 1 describes the game personalisation procedure implemented in this approach. Feeding the observed instance (I_n) into the RFC classifier, we obtain a probability distribution P_{L_i} over all possible Likert classes $Li \in [1, 2, 3, 4, 5]$. We then calculate an overall Likert estimate value $E_{likert} = L[i]{*}P_{L_i}$. Using this value, game adaptations will take place, by adjusting the game difficulty for the next game segment, for each chunk individually. However, before actually calculating the next game difficulty setting, we normalise E_{likert} to calculate $E_{normalised} = E_{likert}{*}1.5 - 4.5$. This normalisation factor adjusts the minimum and maximum increase/decrease applied onto game difficulty to lie in the span of integers [-3,-2,…,3]. By doing this, we avoid increasing/decreasing game difficulty by extreme values (-5, -4, $+4$, $+5$) when the Likert estimate is close to its limits (1 or 5), an adaptation which we consider too steep to take in one single step of the personalisation algorithm.

5.4. Modeling approach: experiments & Results

In this section we will discuss the experiments and results retrieved, regarding the modeling approach.

5.4.1. Online personalisation - pairwise tests

In order to assess the modeling approach we have ran a set of pairwise tests, in which participants were asked to complete three segments of two different versions of INFINITE MARIO BROS.: a static (baseline) version versus a personalised version. After completing both tasks in three distinct difficulty levels (easy, normal, hard), the participants were asked to answer the following questions:

- For which game did you find the challenge level more appropriate?
- Which game did you find more challenging?
- Which game did you find more immersive?
- Which game did you find more frustrating?

Note that participants were not aware of the currently played game difficulty, and that the sequence in which different difficulty levels were played was randomised.

A total of 25 players (20 male, 5 female) compose the test set in this experiment. These participants have been recruited at the University of Amsterdam and the authors' social circle, and are not included in the model training data. The answers available to them again follow the 4-AFC protocol, as in the heuristic approach's experimenting, meaning that participants could choose P over S or S over P, both equally preferred or both equally unpreferred.

5.4.2. Pairwise tests - challenge level

As Fig. 4 illustrates, a majority of the participants show preference for the personalised version of the game when starting at 'easy' and 'normal' difficulty settings. However, preference levels tend to drop as the starting difficulty setting gets harder. Staring at 'hard' difficulty settings, even though the personalised version was not preferred by the majority of the participants, it was still more preferable than the static version.

As in the heuristic approach, players seem to favour a personalised gaming experience rather than a static one. However, even though the majority is statistically significant, one can observe a portion of the participants choosing neither of the two versions, when starting at hard game difficulty. This implies that our system is adapting game difficulty

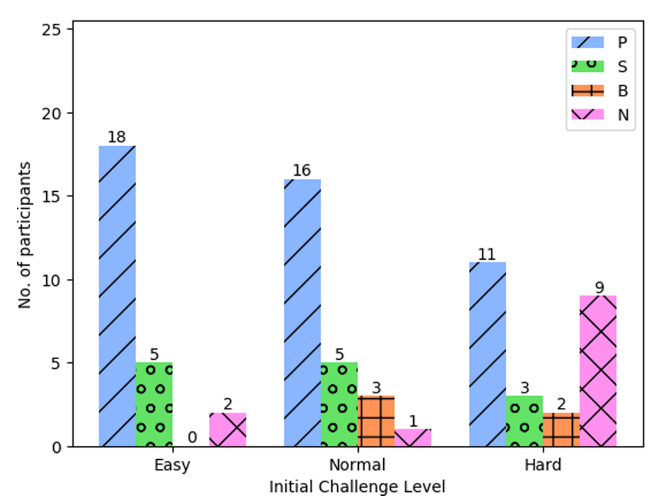


Fig. 4. Pairwise preferences of participants on the modeling approach, per initial challenge level. The legend is as follows, 'P' indicates a preference for the personalised system, 'S' indicates a preference for the static system, 'B' indicates that both are preferred equally, and 'N' indicates that neither is preferred; both are equally unpreferred.

Table 5
Participants' preferences starting at 'easy', 'normal' and 'hard' game difficulty settings. The legend is as follows, 'P' indicates a preference for the personalised system, 'S' indicates a preference for the static system, 'B' indicates that both are preferred equally, and 'N' indicates that neither is preferred; both are equally unpreferred.

Initial Difficulty		Most Challenging	Most Immersive	Most Frustrating
Easy	**P**	95%	100%	75%
	S	5%	0%	25%
	N	0%	0%	0%
	B	0%	0%	0%
Normal	**P**	71%	57%	67%
	S	29%	23%	33%
	N	0%	4%	0%
	B	0%	16%	0%
Hard	**P**	55%	39%	0%
	S	33%	39%	83%
	N	0%	11%	0%
	B	12%	11%	17%

efficiently at easy and normal starting difficulty settings, but drops in performance during hard difficulty game sessions. A possible explanation is that user anger – which should be the main adaptation factor in this case – may not be a sufficient descriptor, or the system should be fine-tuned so as to bring sharper in-game adaptations when anger is detected.

A significance test (Z-test) has been run on the above results, with the preference of the majority for the personalised version being significant in all starting conditions: $p = .0001$ in the 'easy' condition, $p = .0008$ in the 'normal' condition and $p = .005$ in the 'hard' condition.

5.4.3. Pairwise tests - demographics

In order to be able to explain the high percentage of participants choosing neither of the two versions when starting at 'hard' game difficulty levels, we have analysed the demographic information they have provided us with.

Table 5 illustrates the answers retrieved by the participants on the secondary questions. As can be observed, when starting at easy or normal game difficulty settings, the personalised version is consistently considered to be the most challenging, immersive but also frustrating version. However, when starting at hard game difficulty settings, the personalised version is equally preferred to the static one as 'most immersive', while still considered to be the most challenging one. An important observation in this case is that the static version dominates over the personalised one as 'most frustrating' with a percentage of 83% over 0%. We believe this phenomenon refers to the fact that the personalised version most of the time adapts to the player's frustration and decreases game difficulty, whereas the static version is not designed to adapt, thus, maximises potential player frustration. However, studies like [29] have shown that video games can still be considered engaging while being frustrating. Thus, we may interpret the frustration detected right after in-game death as a possible motivational factor, as long as the participants do not abandon the game.

Regarding players abandoning the game before finishing the experiment, we found that those who abandoned, on average, played considerably less hours per week from those who finished the experiment (6.5 v. 23 h spent gaming per week). As a consequence, we may state that our system is partially dependent on player skill level, although game abandonment may occur for reasons other than lack of player skill. Furthermore, players with a higher average of hours spent gaming per week, generally required less effort to familiarise with the game's controls, if not already familiar.

5.5. Discussion

Below, we will analyse the findings of our main study, particularly focusing on in-game difficulty adaptation, as well as in-game difficulty convergence to an appropriate level, for the individual player.

It is clear that in both versions of our game personalisation system as described in Sections 5.2 and 5.3, a significant majority of the participants has consistently preferred the personalised version over the classical (static) version of the game, starting at 'easy' and 'normal' difficulty settings.

Putting the two different approaches in comparison, in the modeling approach (5.3) we observe improved system convergence, i.e. the number of game segments necessary in order to reach an 'optimal' in-game difficulty setting is reduced in the modeling approach (see Fig. 5a–e).

5.5.1. Adapting to the individual player

In general, we have observed smoother adaptations of in-game difficulty in the modeling version of our system. As Fig. 3 shows, users who participated in the model's training set have set the optimal difficulty settings (Likert value of 3) between difficulty levels 3 and 4. The trained model can consistently correlate different facial expressions (which may include head movement) to a particular user challenge level, leading to smoother in-game adaptations, while the same user behavior would lead to steep adaptations in the heuristic system.

For example, in Fig. 5a–e we illustrate how the heuristic and modeling version of our system adapt the game to the same player, when starting at hard in-game difficulty settings in a four segment game session. We observe that the modeling version favours smoother game difficulty adaptations, but does not decrease difficulty below level 3. However, the first system allows steeper adaptations, which can lead to lower in-game difficulty (between 1 and 2) which might be preferred by this particular user.

5.5.2. Difficulty setting convergence

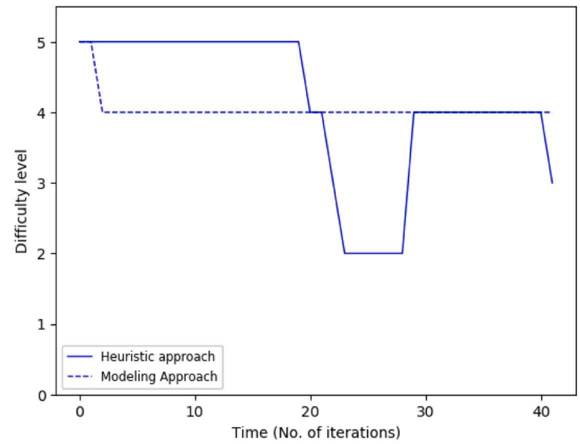
We have also examined whether (and how) in-game difficulty settings converge to the appropriate levels for the individual player. Generally, players tend to gain game skills throughout a game session. Our system should be able to immediately adapt to fast changes in player skill.

Looking at Fig. 5a–e, it is obvious how the modeling version of our system converges to the appropriate game difficulty setup after approximately 12 iterations of the algorithm, whereas the first system has not managed to converge to the optimal setup in the same time. The latter means that either the user's emotions are not yet stable through consecutive segments, or the user's neutral expression levels are still high. However, we may state that the difficulty settings determined by the first system by the end of the session are aiming towards the setup which the modeling version has converged to.

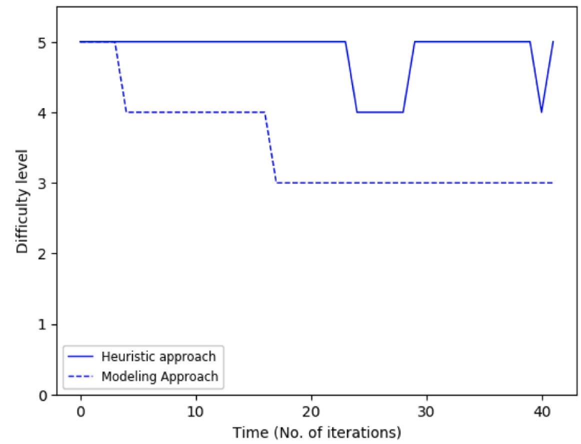
6. Conclusion

In the present study, we attempted to predict the difficulty of games using facial expression data of players. As shown, we were able to predict with relatively high accuracy (72%) the actual difficulty level that is being played based on the facial expressions of players. We were able to predict with 77% accuracy the perceived difficulty of a game when multiple levels are taken into account at the same time, whereas taking these levels into account individually resulted in 88% accuracy.

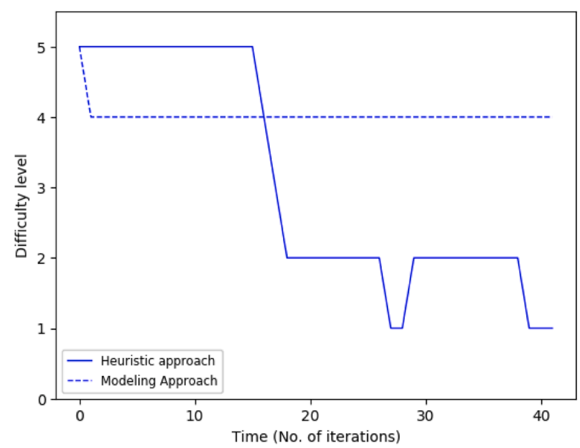
Regarding the game personalisation methods presented, user studies that validated these in the actual video game INFINITE MARIO BROS. revealed that game personalisation through heuristic methods can provide an effective basis for converging to an appropriate affective state for the individual player. Furthermore, building a model of player affective state made accurate in-game difficulty adaptation possible even through noisy emotion estimations (due to head movement), while it



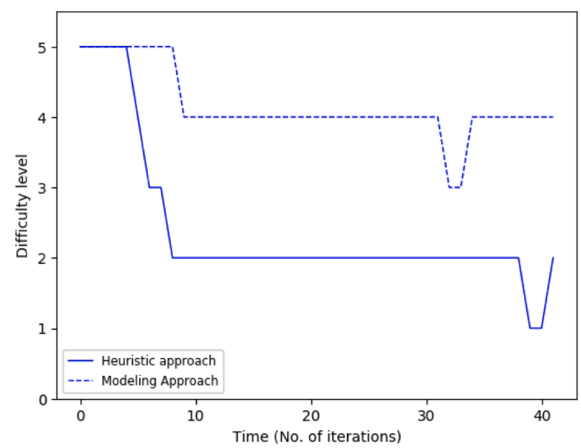
(a) Straight chunk difficulty adaptation



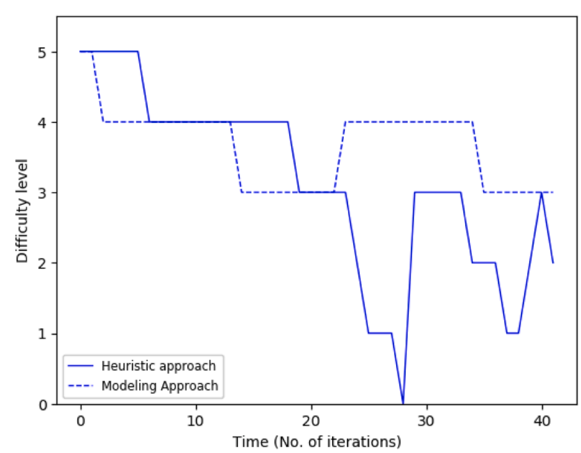
(b) Hills chunk difficulty adaptation



(c) Tubes chunk difficulty adaptation



(d) Jump chunk difficulty adaptation



(e) Cannons chunk difficulty adaptation

Fig. 5. Comparison of difficulty adaptation (per chunk) in the heuristic and modeling approach.

also achieved faster convergence compared to the heuristic method. As such, we may draw the overall conclusion that online and unobtrusive game personalisation is feasible by solely using facial expression analysis, while head pose detection can contribute to an even more effective game adaptation mechanism.

For future work, we aim towards improving the accuracy of our model (described in Section 5.3). This could be achieved by simply enriching the training set with observations from more players, regardless of age, gender or skill level. The present study's small participant pool may negatively affect the integrity of the statistical methods used[17]. Furthermore, we would like to investigate how adding more features to our classifier may further enhance player modeling. Both multi-modal modeling and multi-objective learning are fields of currently conducted research [30].

References

[1] The computer expression recognition toolbox (n.d.), March 2015. [Online] Available: http://mplab.ucsd.edu/ marni/Projects/CERT.html.
[2] Sander Bakkes, Shimon Whiteson, Guangliang Li, George Visniuc, Efstathios Charitos, Norbert Heijne, and Arjen Swellengrebel. Challenge balancing for personalised game spaces, 2014.
[3] Sander C.J. Bakkes, Pieter H.M. Spronck, H. Jaap Van den Herik, Rapid and reliable adaptation of video game AI, IEEE Trans. Comput. Intell. AI in Games 1 (2) (2009) 93–104.
[4] Sander C.J. Bakkes, Chek Tien Tan, Yusuf Pisan, Personalised gaming, J.: Creat. Technol. 3 (2012).
[5] Paris Mavromoustakos Blom, Sander Bakkes, Chek Tien Tan, Shimon Whiteson, Diederik M. Roijers, Roberto Valenti, Theo Gevers, et al., Towards personalised gaming via facial expression recognition, AIIDE, 2014.
[6] Nigel Bosch, Sidney K. D'mello, Jaclyn Ocumpaugh, Ryan S. Baker, Valerie Shute, Using video to automatically detect learner affect in computer-enabled classrooms, ACM Trans. Interact. Intell. Syst. (TiiS) 6 (2) (2016) 17.
[7] Guillaume Chanel, Cyril Rebetez, Mireille Bétrancourt, Thierry Pun, Boredom, engagement and anxiety as indicators for adaptation to difficulty in games, Proceedings of the 12th International Conference on Entertainment and Media in the Ubiquitous Era, ACM, 2008, pp. 13–17.
[8] M. Csikszentmihalyi, Flow: The Psychology of Optimal Experience: Steps Toward Enhancing the Quality of Life, Harper Collins Publishers, 1991.
[9] W.K. Darley, J.S. Lim, The effect of consumers' emotional reactions on behavioral intention: the moderating role of personal relevance and self-monitoring, Psychol. Market. 9 (4) (1992) 329–346.
[10] Janez Demšar, Tomaž Curk, Aleš Erjavec, Črt Gorup, Tomaž Hočevar, Mitar Milutinovič, Martin Možina, Matija Polajnar, Marko Toplak, Anže Starič, Miha Štajdohar, Lan Umek, Lan Žagar, Jure Žbontar, Marinka Žitnik, Blaž Zupan, Orange: data mining toolbox in python, J. Mach. Learn. Res. 14 (2013) 2349–2353.
[11] P. Ekman, V.W. Friesen, S. Ancoli, Facial signs of emotional experience, J. Personal. Soc. Psychol. 39 (1980) 6.
[12] Paul Ekman, Wallace V. Friesen, Constants across cultures in the face and emotion, J. Personal. Soc. Psychol. 17 (2) (1971) 124.
[13] Beat Fasel, Juergen Luettin, Automatic facial expression analysis: a survey, Pattern Recogn. 36 (1) (2003) 259–275.
[14] Robin Hunicke, Vernell Chapman, AI for dynamic difficulty adjustment in games, AAAI Workshop on Challenges in Game Artificial Intelligence, 2004, pp. 91–96.
[15] W.A. IJsselsteijn, Y.A.W. De Kort, Karolien Poels. The game experience questionnaire, 2013.
[16] Martin Jennings-Teats, Gillian Smith, Noah Wardrip-Fruin, Polymorph: dynamic difficulty adjustment through level generation, Proceedings of the 2010 Workshop on Procedural Content Generation in Games, PCGames '10, ACM, New York, NY, USA, 2010, pp. 11:1–11:4.
[17] Russell V. Lenth, Some practical guidelines for effective sample size determination, Am. Stat. 55 (3) (2001) 187–193.
[18] X. Li, S. He, Y. Dong, Q. Liu, X. Liu, Y. Fu, Z. Shi, W. Huang, To create DDA by the Approach of ANN from UCT-Created Data, in: 2010 International Conference on Computer Application and System Modeling (ICCASM), 2010, October, Vol. 8, pp. V8-475. IEEE.
[19] Changchun Liu, Pramila Agrawal, Nilanjan Sarkar, Shuo Chen, Dynamic difficulty adjustment in computer games through real-time anxiety-based affective feedback, Int. J. Human-Comput. Interact. 25 (6) (2009) 506–529.
[20] Stefan Methorst, Telltales: What's in a face. Master Thesis, Tilburg University, The Netherlands, 2015. [Online] Available: <http://arno.uvt.nl/show.cgi?fid=138346>.
[21] Peter D. Molyneux, The Future of Game AI - Lecture, Imperial College London, London, UK, 2006.
[22] J.K. Olesen, G.N. Yannakakis, J. Hallam. Real-time challenge balance in an RTS game using rtNEAT. In CIG'08, 2008, pp. 87–94.
[23] Chris Pedersen, Julian Togelius, Georgios N. Yannakakis, Modeling player experience in super mario bros, CIG 2009, 2009, pp. 132–139.
[24] C. Pedersen, J. Togelius, G.N. Yannakakis, Modeling player experience for content creation, IEEE Trans. Comput. Intell. AI Games 1 (2) (2009) 121–133.
[25] Markus Persson, Infinite mario, 2009. [Online] Available: <http://www.mojang.com/notch/mario>.
[26] R.E. Petty, J.T. Cacioppo, Issue involvement can increase or decrease persuasion by enhancing message-relevant cognitive responses, J. Personal. Soc. Psychol. 37 (10) (1979) 1915.
[27] R.W. Picard, Affective Computing, The MIT press, 2000.
[28] M.O. Riedl, Scalable personalization of interactive experiences through creative automation, Comput. Entertain. 8 (4) (2010) 26.
[29] Melvin Roest, Sander Bakkes, Engaging casual games that frustrate you. FDG, 2015.
[30] Diederik M. Roijers, Peter Vamplew, Shimon Whiteson, Richard Dazeley, A survey of multi-objective sequential decision-making, J. Artif. Intell. Res. 48 (2013) 67–113.
[31] R.M. Ryan, C.S. Rigby, A. Przybylski, The motivational pull of video games: a self-determination theory approach, Motivat. Emotion 30 (4) (2006) 344–360.
[32] Noor Shaker, Stylianos Asteriadis, Georgios N. Yannakakis, Kostas Karpouzis, A game-based corpus for analysing the interplay between game context and player experience, in: Affective Computing and Intelligent Interaction, 2011, pp. 547–556.
[33] Noor Shaker, Georgios N. Yannakakis, Julian Togelius, Digging deeper into platform game level design: session size and sequential features, in: Applications of Evolutionary Computation, 2012, pp. 275–284.
[34] N. Shaker, G.N. Yannakakis, J. Togelius, Crowdsourcing the aesthetics of platform games, IEEE Trans. Comput. Intell. AI Games 5 (3) (2013) 276–290.
[35] Sightcorp, Insight SDK, 2014. [Online] Available: <http://sightcorp.com/>.
[36] Mirna Paula Silva, Victor do Nascimento Silva, Luiz Chaimowicz, Dynamic difficulty adjustment on moba games, Entertain. Comput. 18 (2017) 103–123.
[37] Pieter H.M. Spronck, Ida G. Sprinkhuizen-Kuyper, Eric O. Postma, Difficulty scaling of game AI, in: GAMEON'2004, 2004, pp. 33–37.
[38] Adi Stein, Yair Yotam, Rami Puzis, Guy Shani, Meirav Taieb-Maimon, Eeg-triggered dynamic difficulty adjustment for multiplayer games, Entertain. Comput. 25 (2018) 14–25.
[39] Chek Tien Tan, Daniel Rosser, Sander Bakkes, Yusuf Pisan, A feasibility study in using facial expressions analysis to evaluate player experiences, Proceedings of The 8th Australasian Conference on Interactive Entertainment: Playing the System, ACM, 2012, p. 5.
[40] C.I. Teng, Customization, immersion satisfaction, and online gamer loyalty, Comput. Human Behav. 26 (6) (2010) 1547–1554.
[41] Julian Togelius, Sergey Karakovskiy, Robin Baumgarten, The 2009 Mario AI Competition, CEC 2010, 2010, pp. 1–8.
[42] Julian Togelius, Georgios N. Yannakakis, Kenneth O. Stanley, Cameron Browne, Search-based procedural content generation: a taxonomy and survey, IEEE Trans. Comput. Intell. AI Games 3 (3) (2011) 172–186.
[43] Su Xue, Meng Wu, John Kolen, Navid Aghdaie, Kazi A. Zaman, Dynamic difficulty adjustment for maximized engagement in digital games, in: Proceedings of the 26th International Conference on World Wide Web Companion, WWW '17 Companion, Republic and Canton of Geneva, Switzerland, 2017. International World Wide Web Conferences Steering Committee, pp. 465–471.
[44] Georgios N. Yannakakis, John Hallam, Ranking vs. preference: a comparative study of self-reporting, in: Affective Computing and Intelligent Interaction, 2011, pp. 437–446.
[45] Bieke Zaman, Tara Shrimpton-Smith, The facereader: measuring instant fun of use, Proceedings of the 4th Nordic conference on Human-computer interaction: changing roles, ACM, 2006, pp. 457–460.
[46] Alexander E. Zook, Mark O. Riedl, A temporal data-driven player model for dynamic difficulty adjustment, Proceedings of the Eighth AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment, AIIDE'12, AAAI Press, 2012, pp. 93–98.
[47] M.V. Aponte, G. Levieux, S. Natkin, Scaling the level of difficulty in single player video games, International Conference on Entertainment Computing, Springer, Berlin, Heidelberg, 2009, pp. 24–35.
[48] A. Girouard, E.T. Solovey, L.M. Hirshfield, K. Chauncey, A. Sassaroli, S. Fantini, R.J. Jacob, Distinguishing difficulty levels with non-invasive brain activity measurements, *IFIP Conference on Human-Computer Interaction*, Springer, Berlin, Heidelberg, 2009, August, pp. 440–452.
[49] J. Grafsgaard, J.B. Wiggins, K.E. Boyer, E.N. Wiebe, J. Lester, Automatically recognizing facial expression: Predicting engagement and frustration, in: Educational Data Mining, 2013, July.
[50] S.J. Kirsh, J.R. Mounts, Violent video game play impacts facial emotion recognition, Aggress. Behav.: Official Journal of the International Society for Research on Aggression 33 (4) (2007) 353–358.