UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

Towards Personalised Gaming via Facial Expression Recognition

Blom, P.M.; Bakkes, S.; Tan, C.T.; Whiteson, S.; Roijers, D.; Valenti, R.; Gevers, T.

Publication date
2014
Document Version
Final published version
Published in
Proceedings of the Tenth Artificial Intelligence and Interactive Digital Entertainment Conference (AIIDE 2014)

Link to publication

Citation for published version (APA):
Blom, P. M., Bakkes, S., Tan, C. T., Whiteson, S., Roijers, D., Valenti, R., & Gevers, T. (2014). Towards Personalised Gaming via Facial Expression Recognition. In I. Horswill, & A. Jhala (Eds.), *Proceedings of the Tenth Artificial Intelligence and Interactive Digital Entertainment Conference (AIIDE 2014)* (pp. 30-36). AAAI Press. http://www.aaai.org/Press/Proceedings/aiide14.php

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)

Download date:14 Nov 2026

Towards Personalised Gaming via Facial Expression Recognition

Paris Mavromoustakos Blom[1], Sander Bakkes[1], Chek Tien Tan[2], Shimon Whiteson[1], Diederik Roijers[1], Roberto Valenti[1], and Theo Gevers[1]
[1]University of Amsterdam, Intelligent Systems Laboratory, Amsterdam, The Netherlands
[2]University of Technology, Games Studio, Sydney, Australia

Abstract

In this paper we propose an approach for personalising the space in which a game is played (i.e., levels) dependent on classifications of the user's facial expression – to the end of tailoring the affective game experience to the individual user. Our approach is aimed at online game personalisation, i.e., the game experience is personalised *during actual play* of the game. A key insight of this paper is that game personalisation techniques can leverage novel computer vision-based techniques to *unobtrusively* infer player experiences automatically based on *facial expression analysis*. Specifically, to the end of tailoring the affective game experience to the individual user, in this paper we (1) leverage the proven INSIGHT facial expression recognition SDK as a model of the user's affective state (Sightcorp 2014), and (2) employ this model for guiding the online game personalisation process. User studies that validate the game personalisation approach in the actual video game INFINITE MARIO BROS. reveal that it provides an effective basis for converging to an appropriate affective state for the individual human player.

Introduction

Ideally, artificial intelligence (AI) in games provides satisfactory and effective game experiences for players regardless of gender, age, capabilities, or experience (Charles et al. 2005); it allows for the creation of personalised games, where the game experience is continuously tailored to fit the individual player. Indeed, we are now at a point where modern computer technology, simulation, and AI have opened up the possibility that more can be done with regard to *on-demand* and *just-in-time personalisation* (Riedl 2010). However, achieving the ambition of creating personalised games requires the development of novel techniques for assessing online and unobtrusively which game adaptations are required for optimizing the individual player's experience.

The goal of this research is to online generate game spaces (i.e. levels) such that the spaces optimise player challenge for the individual player. A major challenge to this end, is that in online gameplay only implicit feedback on the appropriateness of the personalisation actions is available, i.e., the AI can only observe the player interacting with the game, while not being provided with labels on the player experience. Still, methods for tailoring the affective game experience to the individual user require an indication on how appropriate the provided experience is to the player. However, explicitly asking for player feedback during gameplay, is usually too intrusive, and would gravely affect the game experience. It is thus of the essence to use as much implicit feedback as possible, to obtain an as accurate as possible model of the player experience.

A key insight of this paper is that game personalisation techniques can leverage novel computer vision-based techniques to *unobtrusively* infer player experiences automatically based on *facial expression analysis*. Specifically, to the end of tailoring the affective game experience to the individual user, in this paper we (1) leverage the established INSIGHT facial expression recognition SDK as a model of the user's affective state (Sightcorp 2014), and (2) employ this model for guiding the online game personalisation process.

As such, we consider challenge to be a cognitive state that might incorporate affective patterns that could be expressed through the face. We focus purely on attaining an appropriate challenge level through the online learning from affective signals; a relatively challenging task. This operates by adjusting procedural parameters that control the intended challenge level -per content type- within the game. This provides expressiveness to tailor the intended challenge level to specific users (by adapting specific content in a distinct manner). Specifically, we will control the *intended* challenge level based on measured affective states; we do not make assumptions on the relationship of affect and challenge.

Game Personalisation

Game personalisation is motivated by a significantly increased involvement and extensive cognitive elaboration when subjects are exposed to content of personal relevance (Petty and Cacioppo 1979); they will exhibit stronger emotional reactions (Darley and Lim 1992). Particularly, a positive effect on player satisfaction is indicated, i.e., game personalisation raises player loyalty and enjoyment, which in turn can steer the gaming experience towards a (commercial) success (Teng 2010). Indeed, the perspective of AI researchers to increase the engagement and enjoyment of the player is one that is consistent with the perspective of

Copyright © 2014, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

game designers (Riedl 2010), i.e., personalisation methods are regarded as instrumental for achieving industry ambitions (Molyneux 2006). 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 part of the game (Bakkes, Tan, and Pisan 2012).

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) (Bakkes et al. 2014). In our investigation, we choose to focus on *personalising the game space* to the individual player with respect to experienced challenge. Related work with regard to this scope is discussed next.

Player Experience Analysis

We build on the novel perspective that computer vision techniques can automatically infer gameplay experience metrics (Tan and Pisan 2012; Tan et al. 2012), a field broadly categorized into qualitative and quantitative methods.

Qualitative methods involve the collection and analysis of subjective data for games; this often includes direct observations, interviews and think-aloud protocols. These methods are most common amongst game practitioners and usually require formal playtest sessions in artificial play environments (Tan et al. 2012). Although these methods have been shown to usually reflect accurate states, they have several shortcomings. Firstly, they might inhibit true play experiences, as players might not be totally at ease when someone is watching or questioning them. Players might not be able to properly self-articulate their play experiences concurrently during gameplay and might not remember important details when post interviews are performed. Secondly, the sessions also often require a lot of time and resources to conduct and analyze. Hence there is a need for more efficient, accurate and versatile (ability to conduct in non-laboratory settings) ways to perform player experience analysis.

These reasons have driven much research towards quantitative methods that work on objective data. Quantitative methods have the potential to represent true player experiences in the game and are able to continuously capture a more diverse body of information. Common approaches include telemetry and psychophysiology.

Telemetry primarily deals with the logging of player in-game interactions to build player models, and several studies have been performed (Zammitto, Seif El-Nasr, and Newton 2010; Medler, John, and Lane 2011; Moura, Seif El-Nasr, and Shaw 2011; Gagne, Seif El-Nasr, and Shaw 2011). The advantage of Telemetry over qualitative methods is that it is non-disruptive and that it can continuously capture objective gameplay statistics in non-laboratory settings. However, the data is limited to the in-game actions available to the player and events in the game world. Hence these "virtual observations" do not capture full experiences and might not even represent the true experiences of the player in real life. For example, a player might take a long time to clear a level, but he might be having a high level of arousal in real life, having fun exploring the level, or simply be stimulated by the aesthetics.

Psychophysiology is the other main branch of quantitative player experience research, which consists of methods to infer psychological states from physiological measurements, that commonly include electrodermal activity (EDA), electromyography (EMG), electrocardiogram (ECG), electroencephalography (EEG), body temperature and pupil dilations. Current work (Mandryk, Atkins, and Inkpen 2006; Nacke and Lindley 2008; Yannakakis and Hallam 2009; Nacke, Grimshaw, and Lindley 2010; Zammitto, Seif El-Nasr, and Newton 2010; Drachen et al. 2010) mostly involve inferring emotional valence and arousal by employing a combination of the measurements. Amongst them, EDA and EMG seems to be most popular as they correspond accurately to emotional dimensions of arousal and valence respectively (Russell 1980). Similar to telemetry, physiological measurements are able to capture player experiences continuously in real-time. In addition, physiological data represent the real life experiences of the player. Unfortunately, most current approaches deal with expensive specialised equipment that are obtrusive, which are usually only viable in controlled laboratory settings. As such, we propose to investigate using a video-based approach to capture data in way that is more efficient, versatile, and does not affect natural gameplay.

Facial Expression Recognition

The first step in any facial expressions analysis system is to recognize facial expressions; being a fairly mature domain in computer vision with techniques that boast a high level of accuracy and robustness (Bartlett et al. 1999; Michel and El Kaliouby 2003; Buenaposada, Muñoz, and Baumela 2007; McDuff et al. 2011). For example, Buenaposada et al. 2007 have reported an 89% recognition accuracy in video sequences in unconstrained environments with strong changes in illumination and face locations.

In terms of using it for analysis of user experiences, there has been a limited number of works performed in non-game applications (Branco 2006; Zaman and Shrimpton-Smith 2006). Branco 2006 showed some encouraging results evaluating positive and negative expressions of users of an online shopping website. Zaman and Shrimpton-Smith 2006 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. In other domains, general emotion detection based on facial expression recognition (Ghijsen 2004; Baltrusaitis et al. 2011) have also shown promising results.

In our research, we take the distinct focus of balancing the game's challenge level by adapting the *content* that is placed within the game environment dependent on facial expression analysis. Particularly, we focus on procedural content generation (cf. Togelius et al. 2011; Yannakakis 2011) for tailoring the player experience. Our distinct focus in this matter, is to assess *online* and *unobtrusively* which game adaptations are required for optimizing the individual player's experience while the game is being played, so as to have assessments on the experienced player challenge impact the procedural process (*cf.* Bakkes et al. 2014).

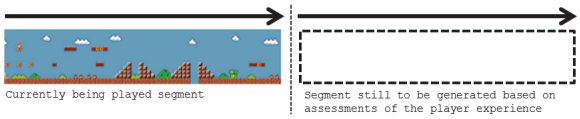


Figure 1: Our enhanced version of INFINITE MARIO BROS. During gameplay it generates short *new level segments* of specific content *on-the-fly*, on the basis of classifications of the facial expression.

Domain description

We consider a typical video game: INFINITE MARIO BROS. (Persson 2009); an open-source clone of the classic video game SUPER MARIO BROS. It can be regarded an archetypal platform game; despite its relatively straightforward appearance it provides a diverse and challenging gameplay experience. We build 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.* (2011; 2012; 2013), Pedersen *et al.* (2009b; 2009a), and Togelius *et al.* (2010).

We have made two further enhancements to the 2011 Mario AI Championship game engine of INFINITE MARIO BROS. First, it now is able to procedurally generate *segments* of Mario levels while the game is in progress (Figure 1). Second, we can now inject short *chunks* of specific game content: (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, stemming from a per-chunk parameter value $\in [0, 5]$. The challenge level of the chunk monotonically increases with the parameter value. In online gameplay, the only action that the personalisation algorithm can take is to output a vector of five integers (chunk parameters) $\in [0, 5]$ to the procedural process which in turn generates the next level segment. While the action space is relatively modest in size, its resulting expressiveness ranges from overly easy to exasperatedly hard level segments.

Approach

The goal of our approach is to online generate game spaces (i.e. levels) such that the spaces optimise player challenge for the individual player. To this end, a key insight is that game personalisation techniques can leverage novel computer vision-based techniques to *unobtrusively* infer player experiences automatically based on *facial expression analysis*. We perform emotion tracking, with the established INSIGHT facial expression recognition SDK (Sightcorp 2014), and gradient ascent optimisation of the individual game experience.

Emotion tracking

In our approach, player emotions are tracked with the INSIGHT facial expression recognition SDK (Sightcorp 2014) through the duration of a game session, yet are taken into account real-time and are chunk specific. As such we are not measuring, e.g., general happiness, but instead can map (parameters that generated) specific game content to specific

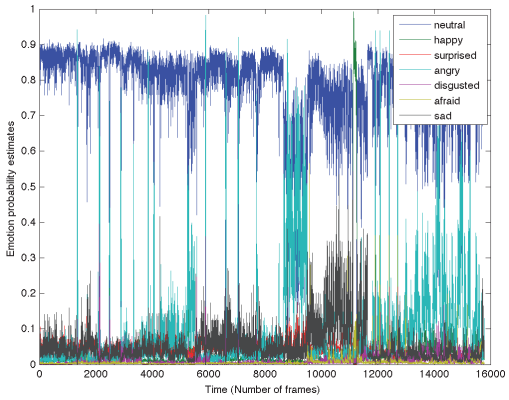


Figure 2: Classifications of the facial expressions of one human participant, over the course of a ten-minute game playing session. We observe that the dominant affective stance is 'neutral'.

affective states. We hereby assume that the classification probability of an affective stance indicates how strongly it is expressed by the player.

INSIGHT classifies facial expressions at approximately 15 frames per second. For each frame, it outputs a probability distribution over seven distinct emotions, namely (1) neutrality, (2) happiness, (3) disgust, (4) anger, (5) fear, (6) sadness, and (7) surprise. Depending on the progress of the player through the Mario game, a game chunk is typically interacted with for 2 to 10 seconds, resulting in a total of 30 to 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 stance; it is an estimate that is relatively insensitive to classification noise of the facial expression system (which may occur in individual frames). INSIGHT has an average accuracy of 93.2% over all classified emotions (Sightcorp 2014).

There are two events at which assessments on 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 in-game deaths of the human player – as we observed that during this observational period many game players express high emotional activity. Furthermore, we particularly consider that – following our experience with the target domain – most game players tend to maintain a relatively neutral facial expression during gameplay, with most emotional 'bursts' occurring when human players experience an in-game death. Figure 2 supports this intuition; it illustrates that 'neutral' is the dominant affective stance, as measured for one player over the course of a game play session of approximately ten minutes, with bursts of anger, happiness, and sadness being measured as well.

During actual play of the first segment of each game session, we calculate the variance in classification of each emotional stance e, from which we derive a factor $\alpha = 1 - var(e_1)$. This factor α is employed as a baseline for the gradient ascent algorithm; it aims at assessing the "emotional

Algorithm 1 Facial Expression-based Gradient Ascent Optimisation

procedure GAOPTIMIZE(e_t, e_{t-1}) ▷ Emotion vectors of current and previous segment
  $\alpha \leftarrow 5 * (1 - Var(e_1))$ ▷ Calculate α, scale to action space
  for $each : chunk$ **do**
    if playerDies(t) **then**
      $\phi = round(5 * \alpha * e_t[Anger])$
      $chunk.decreaseChallengeLevel(\phi)$
    else if segmentFinished(t) **then**
      if $e_t[Neutral] <= 0.8 * \alpha$ **then**
        $chunk.decreaseChallengeLevel(1)$
      else
        $\epsilon \leftarrow argmax_e |e_t - e_{t-1}|$
        $nextAction \leftarrow round(\epsilon * \alpha)$
        if $e \in \{angry, neutral\}$ **then**
          $nextAction \leftarrow -nextAction$
        $nextChallengeLevel \leftarrow previousChallengeLevel + nextAction$
        return $newChallengeLevel$

expressiveness" of each individual player. In our approach, the lower the variance in affective states (i.e., an observed player is not very expressive), the larger the gameplay adaptations to specific chunks when emotional bursts actually do occur (in either direction of intended challenge), considering that $var(e_1)$ is higher when a user shows high emotionality levels. Factor α is then multiplied to thresholds defined in the gradient ascent optimiser.

Gradient Ascent Optimisation

To map classifications of the human player's facial expressions to appropriate in-game challenge levels, we employ a Gradient Ascent Optimisation (GAO) technique. It is employed for optimising the challenge levels for each content type in the game (i.e., for each chunk) such that human interactions with the content yield affective stances that have positive valence (i.e., happiness), while minimising affective stances that have negative valence (i.e., neutrality and anger).

Our implementation of GAO is relatively straightforward (Algorithm 1). After a game segment has been completed by the human player, the probability-distribution vector of the measured emotional stances are retrieved for each individual chunk. The emotions taken into consideration for the present experiments are (1) neutrality, (2) happiness, and (3) anger; our preliminary trials with the Mario game suggested that these emotions were most likely to be expressed by human players (*cf.* Figure 2).

At each iteration of GAO, when a game segment is finished, the emotion vectors – of each individual chunk – of the recently played (finished) segment (S_t), plus the previously completed segment S_{t-1}, are fed into the algorithm. For each emotion that is taken into consideration (neutrality, happiness and anger), the difference between its current (e_t) and previous iteration (e_{t-1}) value is obtained. Next, the maximum of the three differences is determined, namely $argmax(e_t - e_{t-1})$. Since the three emotions are equally weighted, the maximum value calculated could be considered as the "most significant" change in emotional status of the user between two game segments. This value can be considered the desired challenge level of the next segment, S_{t+1}. Since emotions are probabilistic estimates, their difference between two segments follows $D_e \in [-1...1]$. In order to determine the next segment's challenge level per chunk, D_e has to be scaled up to action space of the employed procedural level generator of the Mario game, namely $[0...5]$. Thus, action $a_t = round(5D_e), a_t \in [-5...5]$ is calculated and defines the change in challenge level that will be presented in the next segment, where negative values define a drop in challenge and positive values define a respective increase. To summarize, the challenge level of a chunk in the next segment will be: $d_{S_{t+1}} = d_{S_t} + at$. This calculation will be individually applied to all chunks within a game segment. In practice, the algorithm will increase the game challenge level if the probability estimate of an emotion is higher in timestep t compared to $t-1$. However, we condition on which emotion is the one defining a_t, for the reason that an increase in "negative" emotions (neutrality and anger) should generate a decrease in game difficulty. That is why in these cases, we consider a_t to be $-a_t$.

In order to tailor GAO to the specific target domain, heuristic values are introduced in special occasions; all heuristic values follow from experimentation. Generally, as mentioned, users tend to show highly neutral expressions during gameplay, especially in gameplay settings of low challenge level. In order to prevent "stalling" the game at a certain challenge level due to lack of expressed emotionality, we introduce a heuristic threshold $\tau = 0.8\alpha$. The threshold is derived from our observations on player behaviour in the Mario game (Figure 2). If, by the end of a game segment, the level of neutrality of a player during a chunk was higher than the theshold τ, the level generator will force an increase in challenge by a unit measure (+1) in the next segment's respective chunk. This heuristic corresponds to the insight that the possibility of failure (and the positive affect that is provided by overcoming an obstacle) is an important factor to an appropriate game experience (Juul 2013).

On the other hand, lasting, excessively high challenge levels may impose an unpleasant experience on game players. In order to avoid player abandonment resulting from an inappropriately high challenge level, a second heuristic is applied onto emotions observed during in-game death. A threshold $\phi = 5\alpha \times \epsilon_{anger}$ is introduced regarding the anger measurement during death. The chunk in which death happened will instantly drop by $round(\phi)$ units of challenge level in an attempt to reduce player anger and boost player progress in the game. Note that $\epsilon_{anger} \in \{0...1\}$ is multiplied by 5 in order to directly map emotion probability scale into game challenge scale.

Experiments

Here we discuss the experiments that validate our approach in the actual video game INFINITE MARIO BROS.

Online personalisation – Pilot study

In the pilot study, we analyse the personalisation system's performance by observing one human participant interact with the system under controlled experimental conditions. The participant is placed in a room with stable lighting conditions, and is instructed to interact with the personalised Mario game as she would at home, while attempting to refrain from blocking the face (e.g., by moving a hand through the hair,

drinking coffee, etc.). The participant will interact with the game for ten minutes, starting at an initial challenge level of 'easy' (all parameter values being '1'). Our hypothesis is that when facial expressions can be classified accurately, our online personalisation method will converge to a challenge level that yields an appropriate affective state for the user.

Figure 3 illustrates the obtained results. For all chunks (Figure 3b – 3f), we observe the general trend where the algorithm decreases the per chunk challenge levels (Figure 3f) in the face of user anger, and increases the challenge levels in the face of user neutrality or happiness. Thereby, the online personalisation method operates as expected. For instance, Figure 3a reveals that the challenge level for the cannons chunk (Figure 3f) is initially increased because of high neutral levels. However, later in the game, high anger levels cause a drop in the challenge level. When lastly the angry emotion disappears, the challenge level becomes stable as well. Furthermore, in Figure 3b we observe that the online personalisation method appears stable in the face of classification noise. That is, after approximately 1000 classified frames, the human player suddenly expresses a 'mix' of emotions; denoting, in practise, that the player is talking or moving too much. As expected, the associated challenge level (see Figure 3a) remains stable in the face of this noise from the facial expression classifier.

Online personalisation – Pairwise tests

In this experiment, we investigate how human participants experience the personalised game under actual game playing conditions, in comparison with a realistic (baseline) static game.[1] To this end, in accordance with procedures employed by Shaker *et al.* 2011, 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) (Yannakakis and Hallam 2011). 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 randomised conditions (first s then p, or inversely), each condition consisting of a series of three sequentially performed pairwise tests, in randomized order. A pairwise test compares the static system vs. the personalised system, both starting at one of the three available challenge levels (easy, normal, or hard).

The experiment is performed by ten human participants. To minimise user fatigue impacting the experimental results, each of the three game-playing session is ended after a maximum of 4 level segments (i.e., 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, neither is preferred; both are equally unpreferred). The question presented to the participant is: "For which game did you find the challenge level more appropriate?".

[1]The static levels are built from chunks of a predetermined challenge level, in random order of occurrence for each new segment, so as to ensure variation and playability. Given this randomisation, the experimental trials are sufficiently short to prevent players from easily noticing possible chunk repetitions.

Table 1: Pairwise preferences of participants, per initial challenge level. The legenda is a 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.

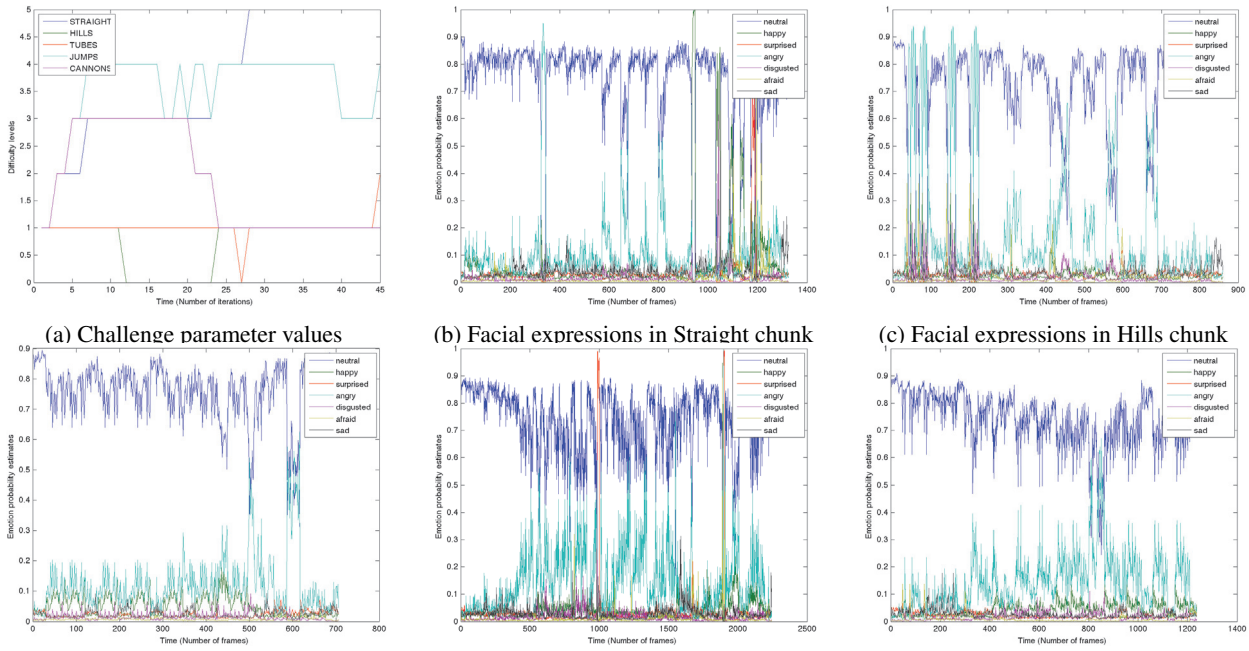
Participant	Easy	Normal	Hard
1	P	P	S
2	P	P	N
3	S	P	P
4	P	P	N
5	P	P	P
6	P	P	S
7	P	S	S
8	S	S	N
9	P	S	P
10	S	P	P
Totals	**70% P**	**70% P**	**40% P**
	30% S	30% S	30% S
	0% B	0% B	0% B
	0% N	0% N	30% N

Table 1 lists the pairwise preferences as reported by the human participants. The results reveals that when both gaming systems are set to an initial challenge level of 'easy', a significant majority ($p = 0.037$) of human participants prefers the personalised system over the static system (70% over 30%). Furthermore, we observe that when both gaming systems are set to an initial challenge level of 'normal', a significant majority ($p = 0.037$) of human participants prefers the personalised system over the static system (also 70% over 30%). When both gaming systems are set to an initial challenge level of 'hard', a narrow majority 40% of the human participants prefers the personalised system over the static system (30%), with the remaining 30% of the participants preferring neither; both are equally unpreferred.

From these results we may conclude that, generally, a majority of human participants prefers the personalised system over the static system. In the case the initial challenge level is 'easy' or 'normal', it concerns a significant majority. In the case the initial challenge level is 'hard', it concerns a narrow majority. A discussion on this latter phenomenon is provided next.

Discussion

The pairwise tests revealed that when a gaming system was initialised with a 'hard' challenge level, 30% of the participants preferred neither the static nor the personalised gaming system; both were equally unpreferred. Our data shows that these participants abandoned both the personalised and the static system, presumably because the employed challenge level was consistently too hard. While such user abandonment might be expected in the static system, one would however expect the personalised system to be able to adapt to these circumstances.



(a) Challenge parameter values (b) Facial expressions in Straight chunk (c) Facial expressions in Hills chunk

(d) Facial expressions in Tubes chunk (e) Facial expressions in Jumps chunk (f) Facial expressions in Cannons chunk

Figure 3: Pilot study. Learned challenge parameter values (Figure 3a) given the measured facial expressions per chunk (Figure 3b – 3f).

Indeed, our personalised system generally *does* decrease the challenge level when it measures the human player being angry. However, in these particular cases, no such measurements were made by the facial expression recognition SDK. We observed that the anger (frustration) of the human participants was not expressed in terms of facial expression, but in terms of hand gesturing, verbal actions, or head movements that prevented facial expressions from being assessed accurately. While this characteristic of the facial expression recognition SDK is outside of our control, we believe that more accurately assessments on player anger can nevertheless be obtained by simultaneously tracking additional features such as gaze and head movement.

Conclusion

In this paper we proposed an approach for personalising the space in which a game is played (i.e., levels) dependent on classifications of the user's facial expression – to the end of tailoring the affective game experience to the individual user. Our approach is aimed at online game personalisation (i.e., the game experience is personalised *during actual play* of the game). A key insight of this paper is that game personalisation techniques can leverage novel computer vision-based techniques to *unobtrusively* infer player experiences automatically based on *facial expression analysis*. Specifically, to the end of tailoring the affective game experience to the individual user, in this paper we (1) leveraged the established INSIGHT facial expression recognition SDK as a model of the user's affective state (Sightcorp 2014), and (2) employed this model for guiding the online game personalisation process.

The pilot study that tested the online personalisation method indicated that the method operates as expected – it decreases specific challenge levels in the face of user anger, and increases specific challenge levels in the face of user neutrality or happiness – and appears stable in the face of classification noise. The pairwise tests across ten human participants revealed that a significant majority of human participants prefers the personalised system over the static system, except in cases when anger (frustration) of the human participants was not expressed in terms of facial expression, but in terms of hand gesturing, verbal actions, or head movements that prevented facial expressions from being assessed accurately. From these results, we may conclude that the developed online personalisation method provides an effective basis for converging to an appropriate affective state for the individual human player.

For future work we will investigate how online game personalisation dependent on a player's facial expressions, can be made more accurate by tracking additional features such as gaze and head movement, and combining it with alternative (multi-objective) feedback models.

References

Bakkes, S. C. J.; Whiteson, S. A.; Li, G.; Visniuc, G.; Charitos, E.; Heijne, N.; and Swellengrebel, A. 2014. Challenge balancing for personalised game spaces.

Bakkes, S. C. J.; Tan, C. T.; and Pisan, Y. 2012. Personalised gaming. *Journal: Creative Technologies* 3.

Baltrusaitis, T.; McDuff, D.; Banda, N.; Mahmoud, M.; Kaliouby, R. E.; Robinson, P.; and Picard, R. 2011. Real-time inference of mental states from facial expressions and upper body gestures. *Face and Gesture 2011* 909–914.

Bartlett, M. S.; Hager, J. C.; Ekman, P.; and Sejnowski, T. J. 1999. Measuring facial expressions by computer image analysis. *Psychophysiology* 36(2):253–63.

Branco, P. 2006. *Computer-based facial expression analysis for assessing user experience*. Ph.D. Dissertation, Universidade do Minho.

Buenaposada, J. M.; Muñoz, E.; and Baumela, L. 2007. Recognising facial expressions in video sequences. *Pattern Analysis and Applications* 11(1):101–116.

Charles, D.; McNeill, M.; McAlister, M.; Black, M.; Moore, A.; Stringer, K.; Kucklich, J.; and Kerr, A. 2005. Player-centered design: Player modeling and adaptive digital games. In *DiGRA*, 285–298.

Darley, W., and Lim, J. 1992. The effect of consumers' emotional reactions on behavioral intention: The moderating role of personal relevance and self-monitoring. *Psychology and Marketing* 9(4):329–346.

Drachen, A.; Nacke, L.; Yannakakis, G.; and Pedersen, A. 2010. Correlation between heart rate, electrodermal activity and player experience in First-Person Shooter games. In *Proceedings of the 5th ACM SIGGRAPH Symposium on Video Games*, 49–54. ACM.

Gagne, A.; Seif El-Nasr, M.; and Shaw, C. 2011. A Deeper Look at the use of Telemetry for Analysis of Player Behavior in RTS Games. In *International Conference on Entertainment Computing ICEC*.

Ghijsen, M. 2004. Facial Expression Analysis for Human Computer Interaction. *Affective Computing, IEEE Transactions on* 2(3):147–161.

Juul, J. 2013. *The art of failure: An essay on the pain of playing video games*. Mit Press.

Mandryk, R. L.; Atkins, M. S.; and Inkpen, K. M. 2006. A continuous and objective evaluation of emotional experience with interactive play environments. *Proceedings of the SIGCHI conference on Human Factors in computing systems - CHI '06* 1027.

McDuff, D.; el Kaliouby, R.; Kassam, K.; and Picard, R. 2011. Acume: A new visualization tool for understanding facial expression and gesture data. In *Face and Gesture 2011*, 591–596. IEEE.

Medler, B.; John, M.; and Lane, J. 2011. Data cracker: developing a visual game analytic tool for analyzing online gameplay. In *Proceedings of the 2011 annual conference on Human factors in computing systems*, 2365–2374. ACM.

Michel, P., and El Kaliouby, R. 2003. Real time facial expression recognition in video using support vector machines. *Proceedings of the 5th international conference on Multimodal interfaces - ICMI '03* 258.

Molyneux, P. D. 2006. The future of game AI - Lecture. Imperial College London, London, UK. October 4, 2006.

Moura, D.; Seif El-Nasr, M.; and Shaw, C. D. 2011. Visualizing and Understanding Players Behavior in Video Games : Discovering Patterns and Supporting Aggregation and Comparison. In *SIGGRAPH*, 2–7.

Nacke, L., and Lindley, C. 2008. Flow and immersion in first-person shooters: measuring the player's gameplay experience. *Proceedings of the 2008 Conference on Future* 81–88.

Nacke, L. E.; Grimshaw, M. N.; and Lindley, C. a. 2010. More than a feeling: Measurement of sonic user experience and psychophysiology in a first-person shooter game. *Interacting with Computers* 22(5):336–343.

Pedersen, C.; Togelius, J.; and Yannakakis, G. N. 2009a. Modeling Player Experience for Content Creation. *IEEE Transactions on Computational Intelligence and AI in Games* 1(2):121–133.

Pedersen, C.; Togelius, J.; and Yannakakis, G. N. 2009b. Modeling player experience in super mario bros. In *CIG 2009*, 132–139.

Persson, M. 2009. Infinite mario. [Online] Available: http://www.mojang.com/notch/mario.

Petty, R., and Cacioppo, J. 1979. Issue involvement can increase or decrease persuasion by enhancing message-relevant cognitive responses. *Journal of personality and social psychology* 37(10):1915.

Riedl, M. 2010. Scalable personalization of interactive experiences through creative automation. *Computers in Entertainment* 8(4):26.

Russell, J. A. 1980. A circumplex model of affect. *Journal of Personality and Social Psychology* 39(6):1161–1178.

Shaker, N.; Asteriadis, S.; Yannakakis, G. N.; and Karpouzis, K. 2011. A game-based corpus for analysing the interplay between game context and player experience. In *Affective Computing and Intelligent Interaction*, 547–556.

Shaker, N.; Yannakakis, G. N.; and Togelius, J. 2012. Digging deeper into platform game level design: session size and sequential features. In *Applications of Evolutionary Computation*, 275–284.

Shaker, N.; Yannakakis, G. N.; and Togelius, J. 2013. Crowdsourcing the aesthetics of platform games. *IEEE Transactions on Computational Intelligence and AI in Games* 5(3):276–290.

Sightcorp. 2014. Insight SDK. [Online] Available: http://sightcorp.com/.

Tan, C. T., and Pisan, Y. 2012. Towards Automated Player Experience Detection With Computer Vision Techniques. In *CHI Workshop on Game User Experience*.

Tan, C. T.; Rosser, D.; Bakkes, S.; and Pisan, Y. 2012. A feasibility study in using facial expressions analysis to evaluate player experiences. In *Proceedings of The 8th Australasian Conference on Interactive Entertainment: Playing the System*, 5. ACM.

Teng, C. 2010. Customization, immersion satisfaction, and online gamer loyalty. *Computers in Human Behavior* 26(6):1547–1554.

Togelius, J.; Yannakakis, G. N.; Stanley, K. O.; and Browne, C. 2011. Search-based procedural content generation: A taxonomy and survey. *IEEE Transactions on Computational Intelligence and AI in Games* 3(3):172–186.

Togelius, J.; Karakovskiy, S.; and Baumgarten, R. 2010. The 2009 Mario AI Competition. In *CEC 2010*, 1–8.

Yannakakis, G. N., and Hallam, J. 2009. Real-time game adaptation for optimizing player satisfaction. *IEEE Transactions on Computational Intelligence and AI in Games* 1(2):121–133.

Yannakakis, G. N., and Hallam, J. 2011. Ranking vs. preference: a comparative study of self-reporting. In *Affective Computing and Intelligent Interaction*, 437–446.

Yannakakis, G. 2011. Experience-driven procedural content generation. *Affective Computing, IEEE* 1–16.

Zaman, B., and Shrimpton-Smith, T. 2006. The FaceReader: Measuring instant fun of use. In *Proceedings of the 4th Nordic conference on Human-computer interaction: changing roles*, number October in -, 457–460. ACM.

Zammitto, V.; Seif El-Nasr, M.; and Newton, P. 2010. Exploring Quantitative Methods for Evaluating Sports Games. *Signals* 8–11.