# Affect Recognition using Psychophysiological Correlates in High Intensity VR Exergaming

**Soumya C. Barathi, Michael Proulx, Eamonn O'Neill, Christof Lutteroth***
REal & Virtual Environments Augmentation Labs (REVEAL) University of Bath, UK

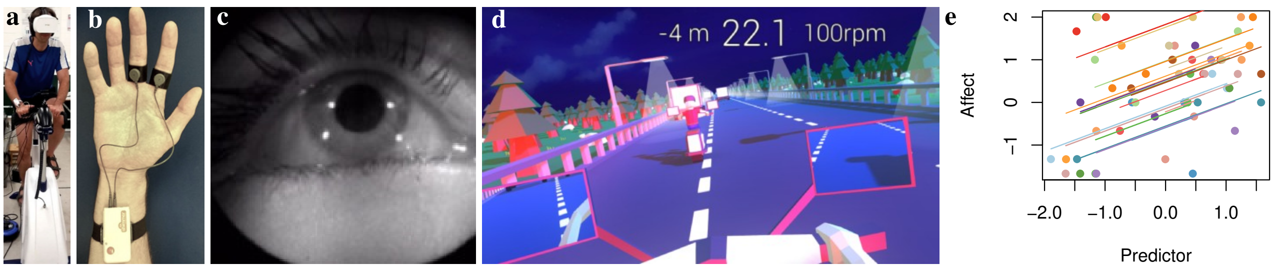


Figure 1: (a) Players exert themselves on an exercycle, (b) wearing a skin conductivity monitor and (c) an eye tracking enabled head-mounted display, while (d) playing a high intensity racing game. (e) Affect is predicted from sensor measurements using regression models (here: $sign(Fixations) \times Conductivity + Pupil$) with individual regression lines to represent each participant.

## ABSTRACT
User experience estimation of VR exergame players by recognising their affective state could enable us to personalise and optimise their experience. Affect recognition based on psychophysiological measurements has been successful for moderate intensity activities. High intensity VR exergames pose challenges as the effects of exercise and VR headsets interfere with those measurements. We present two experiments that investigate the use of different sensors for affect recognition in a VR exergame. The first experiment compares the impact of physical exertion and gamification on psychophysiological measurements during rest, conventional exercise, VR exergaming, and sedentary VR gaming. The second experiment compares underwhelming, overwhelming and optimal VR exergaming scenarios. We identify gaze fixations, eye blinks, pupil diameter and skin conductivity as psychophysiological measures suitable for affect recognition in VR exergaming and analyse their utility in determining affective valence and arousal. Our findings provide guidelines for researchers of affective VR exergames.

### Author Keywords
VR exergaming, affect recognition, psychophysiological correlates, high intensity exercise

*Correspondence: c.lutteroth@bath.ac.uk


*CHI '20, April 25–30, 2020, Honolulu, HI, USA.*

ACM ISBN 978-1-4503-6708-0/20/04 ...$15.00.
http://dx.doi.org/10.1145/3313831.3376596

### CCS Concepts
•**Human-centered computing** → **Human computer interaction (HCI);**

## INTRODUCTION
Physical inactivity is a serious health challenge resulting in global mortality of 3.2 million deaths and 69 million disability-adjusted life years [11]. Many adults fail to do the advised 150 minutes of moderate intensity weekly exercise [90], with the most frequently mentioned obstacles being lack of time and motivation [38]. Exergaming – gaming that involves physical exercise – is a promising motivational intervention to incentivise exercise, improving exercise adherence and enjoyability over conventional exercise by diverting focus from cognitive and physiological states while nearing or exceeding the ventilatory threshold [77, 94, 111]. Furthermore, virtual reality (VR) promotes attentional distraction from aversive bodily sensations such as panting and muscular pain [8]. A number of studies have combined immersive VR and exergaming to improve exercise performance, enjoyability and motivation [9, 17, 37, 60]. Some of these studies have investigated high intensity exercise protocols, which are around twice as time efficient as moderate intensity protocols and arguably more beneficial [44, 70]. Thus, high intensity VR exergaming could help tackle the main problems of exercise adherance, motivation and lack of time.

Affective state is a highly significant predictive value of usability in exergaming. Dynamic adaptation of game content according to the player's affective state would most likely result in better adherence and motivation [13]. Previous studies have been successful in recognising affect in moderate intensity non-VR exergames [78–80, 115]. However, affective recognition in *high intensity VR exergaming* has not yet been achieved. It presents unique challenges as the increased physical exertion and the VR headset interfere with common psychophysiological measurements. For example, it is difficult to recognise facial expression while wearing a VR head-mounted display (HMD), and facial expressions are affected by excessive panting and flared nostrils as a result of high intensity exercise. Studies generally rely on tedious self-reported user experience measures or questionnaires at the end of the experiment, which are useful to get overall feedback but lack the capacity to measure genuine and unadulterated reactions to events and experiences in the game [51]. Furthermore, they are redundant feedback in a sense, as they cannot be used to dynamically adapt the VR exergame during the session. The player experience is extremely fragile and can be disrupted by verbal enquiries of user experience as cognitive effort is necessary to express emotional experience in words [51].

Affectively adaptive videogames have been extensively studied; they use various neuropsychophysiological correlates such as blood pressure, heart rate variability (HRV), electroencephalography (EEG), facial electromyography (EMG), and galvanic skin response (GSR) to recognise affect [7,83,87–89]. Applying these measures in the context of an exergame imposes many challenges, as variations in some psychophysiological parameters may not be an autonomic response but due to physical activity. For example, perspiration can affect skin conductivity, and raised heart rate can affect heart rate variability. EMG signals can be distorted from exaggerated breaths because the oral cavity tends to become wider during exhaustion to facilitate higher intake of oxygen. Another problem of affect recognition is that common measures such as heart rate and skin conductance primarily reflect arousal and do only a limited job at best of indicating emotional valence, i.e. at distinguishing whether the affect is positive or negative [51].

The aim of this project is to identify psychophysiological markers of positive and negative affect that are suitable for use in the context of VR exergaming. We conducted two experiments to achieve our aim. The first experiment compares the impact of physical exertion and gamification on psychophysiological measurements during rest, conventional exercise, VR exergaming, and sedentary VR gaming. We used an eye tracker and skin conductivity sensor to test the suitability of different psychophysiological measures in determining affective responses. We also used validated questionnaires to determine user experience at the end of each condition and correlated them with the measures. In the second experiment, we recorded players' affective responses during 'underwhelming', 'overwhelming' and 'optimal' VR exergaming scenarios. We used experience sampling to elicit a ground truth of the affective state, and then correlated the psychophysiological measures with the ground truth as well as validated questionnaires of user experience. The measures that correlate significantly are thereby identified as predictors of affective state in high intensity VR exergaming.

In summary, we investigated the following research questions:

**RQ1** How do affective responses differ between sedentary gaming, exergaming, conventional exercise, and rest?

**RQ2** Which psychophysiological sensors are most suited for determining affect in high intensity VR exergaming?

**RQ3** What are the psychophysiological correlates of positive and negative affect in high intensity VR exergaming?

We make the following contributions:

1. A qualitative and quantitative analysis of differences between player's affective responses to VR gaming, VR exergaming, conventional exercise, and rest.
2. A novel method to measure positive and negative affect in VR exergaming.
3. An empirical study investigating the psychophysiological correlates of positive and negative affect.

To the best of our knowledge, our study is the first of its kind to investigate the use of pupillometry and skin conductivity to recognise affective state in high intensity VR exergaming.

## RELATED WORK

**Exercise Adherence:** Despite various awareness programs to communicate the physiological and psychological benefits of exercise, 40-65% of people enrolling for exercise programs typically drop out within 3-6 months [3,4,34]. Lack of time and motivation are the most commonly cited barriers [38]. Intrinsic motivation, i.e. motivation to perform an activity because it is enjoyable by itself without the need for other external rewards, plays a crucial role in adherence to exercise [69,99,112]. Therefore, VR exergames that improve intrinsic motivation could improve adherence. If they also reduce the time requirement of exercise, they could further improve exercise adherence. This can be achieved by using high intensity exercise protocols [90], making high intensity VR exergames a promising method to improve exercise adherence.

**Optimising Player Experience:** Behaviourist theories state that re-occurrence of a behaviour is likely when it is accompanied by a positive consequence whereas punishing consequence results in its termination [3,104]. Therefore, maximising enjoyability of exercise and minimising unpleasant effects such as fatigue and discomfort could promote adherence [3]. According to optimal experience theories [29,31] flow – an optimal and enjoyable experience – can occur when a person's skills match the challenge of a task. When the task is more challenging than a person's skills it leads to anxiety, whereas if a person is more skilled than the challenge level of the task it leads to boredom. Therefore, it is important to adapt the challenge level to enable optimal player experience. Exergames that can determine player experience and dynamically adapt its exercise and game intensities to optimise immersion [21] and enjoyment [1,40,99] would also improve exercise adherence.

**Relationship between Immersion and Flow** Immersion is defined as a psychological state in which one perceives being enveloped by an environment providing a continuous stream of stimuli and experiences [113]. Brown and Cairns identify three distinct levels of immersion which are engagement, engrossment and total immersion [21]. Engagement is the response of a user to an interaction that captures, preserves and stimulates their attention, especially when they are intrinsically motivated [62]. This is described to be the lowest level of involvement and to lower the barriers to enter this level, the gamer needs to invest time, effort, and attention which increase for more immersive experiences [21]. However, the experience of being engaged in an activity lacks the emotional attachment observed in the deeper levels of immersion. In the second level of immersion, engrossment, game features combine to the extent that emotions are directly affected by the game and the controls become invisible [63]. Total immersion is defined as shutting off from the real world so much so that the game is all that matters [63]. A positive correlation was found between immersion and appeal, implying that high immersion may lead to high appeal, or vice versa [26]. Flow is an optimal intrinsically enjoyable, subjectively effortless psychological state and can lead to peak performance in sports. Flow overlaps with immersion and both experiences have many mutual properties: concentration, distorting time perception, a balance between the player's skills and the game's challenge, and loss of self-awareness [63]. Seah et al. state that immersion is considered to be a precursor of flow [102]; immersion is an engaging positive user experience that could potentially distract the user from the physical exertion caused by exercising, thus making an exergame more enjoyable [35, 61, 85].

**User Experience** (UX) is defined as perceptions and responses resulting from the use of a system [33]. It is assessed by using various measures of involvement such as engagement, flow presence and immersion, and encapsulates the user's affect, preferences and behaviour during use [33, 63]. Evaluating a user's affective state facilitates profiling of experiences [63].

**Positive Affect**, i.e. enjoyment, plays an important role in motivating people to continue playing a game [108]. Immersion in games is pivotal to enjoyment [63] and media technology development pursues immersion as a route to enjoyment [50]. Immersion, presence, flow, psychological absorption and dissociation are a progression of ever-deeper indicators of game involvement [20] and a strong and direct determinant of enjoyment and self-reported performance in games [72]. They play a critical role in game enjoyment, which in turn increases intrinsic motivation and promotes adherence. All of them could lead to positive affect and improve player experience.

**Negative Affect:** According to the Experience Fluctuation Model the state of flow is achieved when there is balance between challenges and skills, and both challenges and skills are greater than their weekly average [73, 74]. The model also provides a detailed characterization of negative affective states such as anxiety, worry, apathy and boredom, all defined as imbalances between challenges and skills. For example, according to the model anxiety occurs when the challenge level is greater and skill level is lower than the weekly average.

**Inducing Positive Affect:** The dual flow model for enjoyment in exergames [106] consists of two dimensions: attractiveness, which is a psychological model balancing the player's perceived skill with perceived challenge; and effectiveness of the exercise, the physiological counterpart of flow balancing fitness and the intensity or challenge of the exercise [103]. Exergames must pose an adequate intensity, duration and density for optimal adaptation [49]. Furthermore, they must be attractive and adaptive to provide a low barrier to entry for newcomers while providing enough challenge for more experienced users in order to keep them interested [49]. We designed our exergame based on these models to induce positive affect.

**Affective Gaming:** The term 'affective ludology' refers to the scientific measurement of emotional and cognitive aspects of player experience while interacting with games [81]. Affective games dynamically adapt challenge and game content, and offer assistance according to the player's emotional state [46]. In order to do so, they need to recognise the player's affect. Arousal correlates with the pressure used to press buttons on a gamepad [107]; however, this is due to stress rather than enjoyment. Facial EMG has been used to determine emotional valence (positive or negative) during gaming [51, 82], and EEG has also shown to provide reliable measurements of affective player experience [7, 83]. Several studies use multimodal approaches including HR monitors, EMG and GSR [87–89].

**Affective Exergaming:** Measures such as GSR, HRV, EEG and facial EMG can be used to detect user experiences such as flow [82], immersion [83], and arousal [87–89]. However, many physiological measures are open to corruption [47] and are likely to get affected by exercise perspiration, panting and movement. Previous studies have been successful in detecting affect in moderate intensity non-VR exergames using facial expressions, GSR, temperature, respiration and movement [78–80, 115]. Facial expression recognition in VR is difficult because a player's HMD covers half her face. Moreover, high intensity exercise protocols are also a lot more physically demanding compared to moderate intensity exercise [95], so lead to higher perspiration, temperature, respiration and motion changes affecting physiological measurements. In order to recognise affect in VR exergames, we need to identify psychophysiological measures that are suitably robust.

**Valence and Arousal** of affect both need to be measured to conclude the emotional state. Previous studies that were able to achieve this were using fMRI [2], facial EMG [23] , EEG [25] and electrocardiography (ECG) [84]. However, most of these sensors are too sensitive to be used in an exergaming environment due to motion artefacts, and therefore we are not considering EMG, EEG and ECG in our studies. GSR measures the changes in electrical conductance of the skin as a result of increased sympathetic nervous system activity, indicating affect arousal but not valence. Recent attempts to use it also for valence through feature extraction [5] are likely to fail for high intensity VR exergaming because of extensive movement and perspiration. On the positive side, the eccrine glands on palms and soles are more sensitive to affect than exertion-induced perspiration [18, 30, 39] and affective responses typically precede the appearance of sweat, so GSR may be suitable for measuring affect arousal in VR exergames. Also pupil dilation, blink rate, and eye movements are potential measures of affect. Increases in pupil size reflect arousal associated with increased sympathetic activity [19, 55] and have been proposed for use in affective computing [92]. Blink rate is negatively correlated with dopamine activity which could reveal valence [64, 67, 68], with increased attention leading to a reduction in blink frequency [105]. Gaze fixations are often directed towards the object of one's thoughts [36, 48, 109], suggesting that they could also be markers of affective experiences.

Based on these considerations, we decided to investigate the potential of skin conductance, pupil dilation, blink rate and gaze fixations for affect recognition in VR exergaming.

**Experience Sampling** is a well established method for measuring experiences immediately as they are perceived in a particular moment [52], decreasing the chance of failing to remember and the inclination to select answers based on social desirability [116]. Recently received information is saved in short term memory, which is ephemeral in nature and has a limited storage capacity [96]. It moves into long term memory only upon rehearsal or meaningful association, with a risk of experiences getting lost in translation [41]. As a result, experience sampling is preferrable to post-experience questionnaires for measuring instantaneous affective response, therefore we used it to collect ground truth data about affect. To prevent the process of experience sampling from interrupting delicate player experiences such as immersion and flow, we integrated experience sampling directly into a VR exergame.

## EXPERIMENT I: GAMING AND EXERCISE

In order to address RQ1 and identify suitable psychophysiological sensors (RQ2), we investigated the difference in affective responses to sedentary VR gaming, high intensity VR exergaming, and conventional high intensity exercise. The overall experimental design is summarised in Table 1. We used a within-participants design for the independent variables *Game* with levels game (G) and no game (N), and *Exercise* with levels exercise (E) and no exercise (N). Outcome variables were measured during each of five *Phases* P-I, P-II, P-III, P-IV and P-V, which were defined based on a high intensity interval training (HIIT) exercise protocol. The combination of no-exercise and no-game forms the baseline condition (B) in which players remain seated in all five phases. In the sedentary gaming condition (G), players play a VR game that is an exact replica of our VR exergame, but the forward motion in the game is simply controlled by a hand pedal. In the conventional exercise condition (E), players exercise without the game in all five phases according to the HIIT protocol. And in the VR exergaming condition (EG), players play our VR exergame, which gamifies the HIIT protocol.

The method of acceleration (exercycle vs. hand pedal) was a potential confounding factor in the VR game experience, therefore we devised a model for the hand pedal that yielded a similar experience of acceleration as the exergame. Similar to the exergame where acceleration is becoming increasingly difficult, the acceleration effects of the pedal were attenuated with increasing speed. Similar to the exergaming condition, players of the sedentary game were asked to maintain a virtual speed of “65 to 70 RPM” during the low intensity phases and go as fast as they can during the high intensity sprints. To avoid further confounding factors, participants wore a HMD in all conditions. The VR environment of conditions B and E was bare but used the average sky colour and the average ground colour of the G and EG conditions. Exercise prompts and information about timing and speed were shown not only in EG but also in E and G. We recorded the baseline B first to get resting skin conductivity and blink rate before any other activity. After recording the baseline B, the order of G, E and EG was counterbalanced. During the G condition, players held the hand pedal in their left hand and the skin conductivity sensor was attached to their right hand to minimise movement artefacts. In order to differentiate sweat from affective response from exercise sweat, we measured skin conductivity at the fingers, which primarily contain eccrine glands highly responsive to emotional stimuli. We consistently used the same exercise protocol and ensured participants rested between conditions to control for the effects of exercise on sweating.

### High Intensity VR Exergame

The VR exergame (Fig. 1d) is played by riding a stationary exercycle while wearing an HMD. The player cycles along a straight path with a speed proportional to cycling revolutions per minute (RPM). Players can move from side to side by leaning slightly sideways, with a lateral movement speed proportional to the roll angle of the HMD. The game alternates between relaxing low intensity phases and high intensity sprints. The HMD shows prompts about exercise intensity, RPM and time remaining in the current phase. During the low intensity phases trucks appear and the player is instructed to avoid them with lateral movements. Players start the game with a score of 100 points and lose 10 points for every collision. In the high intensity phases, player’s goal is to beat a competitor represented by an automated avatar. In the first sprint, the game lets the player be ahead of the competitor for the first 10 seconds, and in the remaining 20 seconds the competitor is sped up to be ahead of the player. The second sprint reverses this, with the competitor ahead for the first 10 seconds and the player ahead for the remaining 20 seconds, which ensures a close match and an engaging race. Points are awarded for the distance covered during the game.

We designed the game according to the dual flow [103] and the game flow models [106], with the aim of eliciting a positive flow experience based on the psychological challenge of avoiding trucks and the physiological challenge of exercise. We adjusted the challenges to the player’s abilities for an optimal balance, facilitating engagement, immersion and flow [32, 63]. We designed the game aesthetics to elicit a positive affective response [59, 86]. During the low intensity phases, the virtual environment is sunny and bright to invoke calmness and relax the player, with an upbeat music playing at 120 beats per minute (BPM). During the high intensity sprints, the environment transitions into a dark scene with police cars flashing emergency lights, aiming to motivate and energise the player, with the speed of the music increasing to 140 BPM. Related work suggests that these visual stimuli can increase the aesthetic satisfaction of the game [56, 114], and that the high-tempo music can enhance energy levels and induce bodily action [65] with ergogenic effects by drawing the attentional focus away from negative exercise-induced bodily sensations [12, 14, 22, 66].

**Exercise Protocol and Equipment:** The game implemented a high intensity interval training (HIIT) protocol [43], which is an effective, time-efficient alternative to conventional low or moderate intensity exercise [42]. A 60 second warm-up phase was followed by two 30 second sprints separated by a 90 second recovery phase, finishing with a 90 second cool-down

Table 1: Design overview for Experiment I.

| Condition | Exercise | Game | Phase | | | | |
|---|---|---|---|---|---|---|---|
| | | | P-I (60s) | P-II (30s) | P-III (90s) | P-IV (30s) | P-V (90s) |
| Baseline (B) | No Exercise (N) | No Game (N) | Rest | Rest | Rest | Rest | Rest |
| Sedentary Gaming (G) | No Exercise (N) | Game (G) | Game | Game | Game | Game | Game |
| Conventional Exercise (E) | Exercise (E) | No Game (N) | Warm Up | Sprint 1 | Recovery | Sprint 2 | Cool down |
| Exergaming (EG) | Exercise (E) | Game (G) | Warm Up | Sprint 1 | Recovery | Sprint 2 | Cool down |

phase. During the warm-up, recovery and cool-down, players cycled at a low intensity of 65-70 RPM with 12 Nm resistance. During the sprints, players cycled as fast as possible with a resistance of 0.4 Nm $kg^{-1}$ initially based on body mass and adjusted to suit the player during a familiarisation phase. We used a Lode Excalibur Sport exercise bike and an FOVE HMD, connected to a PC running Unity with an Intel Xeon E5 2680 processor and two NVIDIA Titan X graphics cards.

### Outcome Variables

We collected psychophysiological measurements that are known to be associated with affect, considering their averages over the five phases for each condition. We measured blink rate in blinks per minute (*Blinks*) with the eye gaze tracker built into the FOVE HMD, recording pupillometry data with FOVE's Unity plugin at approx. 160Hz and counting blinks as periods with zero pupil diameter. We measured the tonic skin conductance (*Conductivity*) in microsiemens (µS) at 128 Hz using the Shimmer3 Consensys GSR development kit. Furthermore, we recorded the average power output (*Power*) in Watts during the sprint phases in conditions E and EG as a measure of physical performance, using the serial port interface provided by the Lode Excalibur Sport exercise bike.

We collected ground truth data for affect based on the Intrinsic Motivation Inventory (IMI) [98], which was validated in numerous sports science studies [27, 75]. We used the main Interest/Enjoyment subscale (*IMI Enjoy*) with a scoring range from 1 to 7, with 7 being the highest intrinsic motivation score.

### Procedure

We recruited 18 participants (14 male, 4 female, age 19-32, mean 23$\pm$3), who were students and employees of the University of Bath. We screened them with the Physical Activity Readiness Questionnaire (PAR-Q) [110] and excluded participants with health risks or a resting blood pressure greater than 140/90 mmHg. The remaining participants were informed about all experimental conditions and asked to complete a demographics questionnaire. After initialising sprint resistance based on body mass, participants went through a familiarisation phase which allowed them to experience the VR exergame. Participants then performed each of the four conditions: B, G, E and EG. They were instructed to work "very hard" during the sprints and to maintain 65-70 RPM during the low intensity phases for both EG and E. After conditions G, E and EG, participants completed the IMI and left qualitative feedback about their experience. Participants had a break of approx. 10 minutes between the gameplay rounds to avoid fatigue. At the end of the experiment participants were asked to comment on

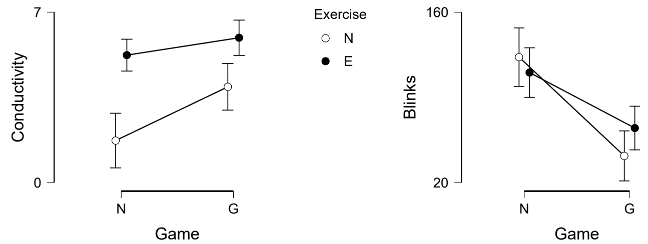


Figure 2: Conductivity (left) and Blinks (right) in the four conditions of Experiment I.

their experience during the different conditions. Each session took approx. 120 minutes.

### Hypotheses

Research suggests *Blinks* is lower and *Conductivity* is higher for higher levels of enjoyment [30, 105]; so we hypothesise:

**H1** EG is more enjoyable than E and G as measured by *IMI Enjoy*.
**H2** *Conductivity* correlates positively with enjoyment as measured by *IMI Enjoy* in conditions E and EG.
**H3** *Blinks* correlates negatively with enjoyment as measured by *IMI Enjoy* in conditions E and EG.

### Results

The results are summarised in Tables 2 and 4. The assumptions of analysis of variance (ANOVA) were met, so we analysed the data with repeated-measures ANOVAs using the $\omega^2$ measure for effect sizes [91], and two-tailed t-tests with Holm correction for pairwise comparisons. According to power analyses, the ANOVAs were able to detect medium-sized effects (Cohen's f=0.286) and the t-tests were able to detect large effects (Cohen's d=0.701) at $\alpha = 0.05$ with a power of 0.8. They allow us to better understand the uncertainty in the test results. Because of our within-participants design, we calculated and tested repeated measures correlations using the rmcorr R package [15] instead of simple Pearson correlations. This accounts for the individual differences between participants in both psychophysiological and self reported measurements, and increases the statistical power as no aggregation of measurements is necessary when testing intra-individual hypotheses [6]. The level of significance used for all tests was $\alpha = .05$. Plots show 95% confidence intervals of the means.

**Manipulation Check:** A one-way repeated-measures ANOVA was conducted on E, G and EG for *IMI Enjoy*. The main effect was significant ($F(2,34) = 26.11, p < .001^{***}$)

Table 2: Demographics and results for Experiment I (avg.±std. dev.).

| Demographics | Variable | B | G | E | EG |
|---|---|---|---|---|---|
| n=18 | Power | - | - | 289.52 ± 80.92 | 329.74 ± 61.16 |
| m=14; f=4 | IMI Enjoy | - | 3.60± 1.42 | 3.60 ± 0.99 | 5.67 ± 0.84 |
| age=23 ± 3 | Blinks | 123.06±113.28 | 41.94± 58.26 | 110.39 ± 112.78 | 64.83 ± 74.55 |
| | Conductivity | 1.73±0.99 | 3.93± 3.39 | 5.24 ± 3.16 | 5.95 ± 3.47 |

with a large effect size ($\omega^2 = 0.427$). The differences between E and EG ($t(17) = 6.778, p < .001^{***}$), and G and EG ($t(17) = 6.328, p < .001^{***}$) were significant. The difference between G and E was not significant. The results show that participants enjoyed the exergame more than conventional exercise and sedentary gaming, as intended, so we accept H1.

**Psychophysiological Differences:** A two-way repeated-measures ANOVA was conducted on *Game* (N vs. G) and *Exercise* (N vs. E) for *Conductivity* (Fig. 2 left). The effect of *Game* was significant ($F(1,17) = 11.994, p = .003^{**}$) with a medium effect size ($\omega^2 = 0.065$). The effect of *Exercise* was also significant ($F(1,17) = 33.132, p < .001^{***}$) with a large effect size ($\omega^2 = 0.203$). The interaction effect of exercise and game was significant ($F(1,17) = 4.694, p = .045^{*}$) with a small effect size ($\omega^2 = 0.016$). The difference in *Conductivity* between E and EG was significant ($t(17) = 2.370, p = .015^{*}$). A two-way repeated-measures ANOVA was conducted on *Game* (N vs. G) and *Exercise* (N vs. E) for *Blinks* (Figure 2 right). The effect of *Game* was significant ($F(1,17) = 26.168, p < .001^{***}$) with a medium effect size ($\omega^2 = 0.108$). The effect of *Exercise* was not significant indicating that exercise did not affect *Blinks* much. The interaction effect of exercise and game was significant with ($F(1,17) = 6.247, p = .023^{*}$) and a small effect size ($\omega^2 = 0.009$). The difference in *Blinks* between E and EG was significant ($t(17) = 3.607, p = .001^{**}$).

**Psychophysiological Correlates:** A repeated-measures correlation analysis for the conditions E and EG showed that *Conductivity* was significantly positively correlated with *IMI Enjoy* ($r(17) = 0.418, p = .038^{*}$), so we accept H2. *Blinks* was significantly negatively correlated with *IMI Enjoy* ($r(17) = -0.574, p = .005^{**}$), so we accept H3.

**Qualitative Feedback:** Comments indicate that players enjoyed being in an aesthetically appealing VR environment and experiencing the gamification of high intensity exercise. Typical comments in the VR exergaming condition were "challenging, thrilling, immersive", "combines best of both worlds, makes exercising way more enjoyable". Some comments such as "could push myself harder" suggest that, in addition to being entertained by the VR exergaming experience, participants also felt more motivated to exercise.

### Discussion

The manipulation check showed that the VR exergame was successful in eliciting positive affect (H2). The significant effects of *Game* and *Exercise* on *IMI Enjoy* indicate that they can both enhance positive affect (RQ1), and this is supported by the qualitative feedback. As a consequence, we can compare the conditions to find psychophysiological markers of positive affect. As we are interested to find markers specifically for VR exergames, the measurements in E and EG are most relevant, as participants were exercising at high intensity and using an HMD in both of these conditions.

The significant effect of exercise on *Conductivity* indicates that the sensor measurements are indeed affected by the exercise, increasing skin conductivity due to sweat. However, the effects of gamification on *Conductivity* while exercising were still significant, indicating that it is still sensitive enough as a marker of affect in a high intensity VR exergame. The correlation analysis confirms, as suggested in related work, that *Conductivity* is a linear predictor of positive affect, with more positive affective states leading to a higher skin conductivity.

The effect of exercise on *Blinks* was not significant, indicating that the sensor measurements are robust against the effects of exercise. There is a significant effect of gamification on *Blinks* both for sedentary activities (B vs. G) as well as while exercising (E vs. EG), making it a promising choice for affect recognition in general and for high intensity VR exergaming in particular. The correlation analysis confirms, as suggested in related work, that *Blinks* is a linear predictor of positive affect, with more positive affective states leading to a lower blink rate. Since *Blinks* appears to be a good choice for measuring affect in high intensity VR exergames, we widen our range of pupillometry measurements to also include gaze fixations and pupil dilation in Experiment II.

## EXPERIMENT II: CORRELATES OF AFFECT

In order to determine the suitability of different psychophysiological sensors (RQ2) and analyse the correlates of positive and negative affect in a high intensity VR exergame (RQ3), we designed three VR exergame scenarios to evoke the following three states: an *optimal* state (Opt) to induce positive affect, an *underwhelming* state (Under) to induce neutral affect, and an *overwhelming* state (Over) to induce negative affect. This is similar to Moller et al., who elicited a state of flow in players by tweaking the game intensity using overwhelming, underwhelming and optimal conditions [76]. However, our focus is not on flow as flow is often regarded as an extreme positive experience [63]. The aim of our experiment is to find the correlates of a broader spectrum of positive *and* negative affective states, therefore we used intrinsic motivation (*IMI Enjoy*) as a measure of affect instead of flow. The three states Opt, Under and Over form the levels of the independent variable *Game Scenario*, and we investigated the effects of Game Scenario on psychophysiological correlates using a counterbalanced within-participants design. We used a methodology extended from Experiment I, with the notable addition of experience sampling to record ground truth values of affective responses.

In addition to *Conductivity*, *Blinks* and *Power*, we measured pupil dilation and gaze fixations. The relationships of *Conductivity*, *Blinks* and pupil dilation with affect have been well documented [19, 30, 55, 105]. Gaze fixation is only a partial indicator of a person's thoughts as gaze may convey the object of a person's thoughts, which may not necessarily indicate affective state. Similarly, even though many studies explored the relationship between task performance and enjoyability [57, 71], better performance does not necessarily mean higher enjoyment. In summary, *Conductivity*, *Blinks*, and pupil dilation are more "conventional" predictors of affect, although they have not been explored yet for high intensity VR exergames, while gaze fixations and *Power* are more speculative in their relationship to affect.

### High Intensity VR Exergame

We modified the VR exergame from Experiment I to create VR exergaming environments for the optimal, underwhelming and overwhelming scenarios. To avoid confounding factors, the game mechanics, HIIT exercise protocol, equipment, ambient lighting and overall game environment in all three scenarios stayed the same. Because all conditions used the same exercise protocol, the noise due to movement artefacts when measuring *Conductivity* was similar and comparable.

The optimal scenario was designed to be engaging by presenting the player with the same appealing aesthetics and engaging gameplay elements as in Experiment I. The exercise intensity was designed to provide an optimal challenge with a positive competitive experience: in both sprints, the competitor was sped up to be ahead of the player for the first 10 seconds and slowed down to allow the player to get ahead in the remaining 20 seconds. The underwhelming scenario was designed to be devoid of stimuli and evoke boredom by using minimal aesthetics, without sound effects or music and no police cars with flashing lights. There was a complete absence of gameplay elements such as scoring points, dodging trucks and racing a competitor. The overwhelming condition was designed to elicit stress and frustration by using 'annoying', noisy sound effects such as blaring horns during the low intensity phases and wailing police sirens during the high intensity phases instead of music. The gameplay was extremely challenging, with the sprint competitor programmed to be always ahead of the player and the player losing 20 points instead of 10 when hitting a truck.

In order to make experience sampling of affect as unobtrusive as possible, we integrated it directly into the exergame. We used a colour coded experience sampling scale to measure affect valence and arousal as very negative (-2), negative (-1), neutral (0), positive (+1) and very positive (+2). The scale appeared 30 seconds after the start of the warm up phase and 5 seconds after each of the sprints, in all the three conditions. The scale stayed visible for 10 seconds and accepted input from two hand pedals attached to the exercycle handlebars. If players felt positive affect, they were instructed to click their right hand pedal once or twice depending on their arousal. Similarly, if they felt negative affect, they were instructed to click their left hand pedal once or twice. If their affective state was neutral, they were instructed to click both the pedals once.

### Outcome Variables and Data Analysis Approach

In addition to recording *Conductivity*, *Blinks* and *Power* to determine affective state, we recorded the total time of eye gaze fixations (*Fixations*) on visual components of the game: the competitor, the gap between the player and the competitor, the points, prompts, the displayed RPM and the timer. We used ray casting to detect the game components corresponding to a point of gaze. A low *Fixations* value indicates that the player was looking more at the peripheral VR environment or 'staring at nothing' instead of paying attention to the game. We also recorded a participant's pupil dilation (*Pupil*) during the warm up and in each of the two sprints, considering their average. Similar to Experiment I, we used the IMI Interest/Enjoyment subscale (*IMI Enjoy*) to measure intrinsic motivation. Lastly, the experience sampling method integrated in the exergame was used to collect ground truth values about the player's affective state; we consider the average of all values measured in a condition (*Affect*). We matched the sensor data and the ground truth by taking the average of the sensor data and of the experience sampling measures over a whole gameplay session.

*Conductivity*, *Fixations* and *Pupil* values are affected by systematic individual differences in eccrine activity [100, 101], fixation length [53, 54] and pupillary sensitivity [58], respectively. This experiment allowed us to collect enough data from each participant under VR exergaming conditions to compensate for these individual differences by normalising the variables using standard z-score transforms. *Fixations* and *Pupil* values were centred at the participant mean and scaled by dividing them by a participant's standard deviation. *Conductivity* values were not centred and only divided by the standard deviation, as they are only used as a predictor of affect arousal, see below.

We consider affect as a two dimensional model consisting of valence and arousal based on Russell's affect grid model [97], which describes affect along the dimensions of pleasure-displeasure (valence) and arousal-sleepiness (arousal). For this study we analyse valence as a ternary construct with three states: positive, negative and neutral, without a magnitude. Our experience sampling *Affect* scale with range $[-2, +2]$ accommodates both these dimensions: 1) arousal is represented by the absolute value with range $[0, 2]$, and 2) valence is represented by the sign (+, - or neutral). Although this approach has the advantage of differentiating neutral as well as positive and negative affect, it cannot distinguish between different affective states of the same valence, similar to other works on affect recognition using psychophysiological correlates [24, 41].

From related work we know that some psychophysiological variables are indicators of affect as a whole, or at least valence (positive, neutral or negative), while some variables are only indicators of arousal, with a valence in negative or positive direction. We formalise this by defining *Valence* as the sign of affect *sign(Affect)* with possible values -1, 0 and +1, and *Arousal* as the absolute value of affect |*Affect*| with non-negative values. Variables that predict the whole affective state, i.e. both valence and arousal, correlate directly with *Affect*. Variables that predict valence, i.e. whether the affect is negative, neutral or positive, are correlated with *sign(Affect)*. Variables that predict only arousal are correlated with |*Affect*|. These correlations can be tested using the repeated-measures correlation analysis from Experiment I. Since we treat valence as a ternary construct here, linear repeated-measures regression is more appropriate as a model than logistic regression.

### Procedure

We recruited 18 participants (14 male, 4 female, age 20-44, mean 26±5). The procedure of Experiment II was similar to that of Experiment I. We used the same screening procedure, exclusion criteria, questionnaires and exergame familiarisation phase. Additionally, participants practised answering the experience sampling scale by clicking the hand pedals, to make sure this would be easy during the experiment. Each experiment session took approximately 120 minutes.

### Hypotheses

We expect *Affect* to correlate with *IMI Enjoy* as the two concepts are highly related and both capable of measuring positive and negative affective response [33, 63]:

**H4** *Affect* correlates positively with enjoyment as measured by *IMI Enjoy*.

Related work and the results of Experiment I suggest that *Blinks*, *Pupil*, *Fixations* and *Power* are likely to be indicative of affect as a whole in the following manner:

**H5** *Blinks* correlates negatively with *Affect*.
**H6** *Pupil* correlates positively with *Affect*.
**H7** *Fixations* correlates positively with *Affect*.
**H8** *Power* correlates positively with *Affect*.

If *Blinks*, *Pupil*, *Fixations* and *Power* correlate with *Affect*, they will by implication also correlate in the same manner with valence to some degree; therefore we are also testing their corresponding correlations with *sign(Affect)*. Related work shows that *Conductivity* is a more prominent indicator of arousal rather than valence:

**H9** *Conductivity* correlates positively with |*Affect*|.

### Results

The results of Experiment II are shown in Tables 3 and 5. Similar to Experiment I, the assumptions of analysis of variance (ANOVA) were met, so we analysed the data with repeated-measures ANOVAs and two-tailed t-tests with Holm correction for pairwise comparisons. According to power analyses, the ANOVAs were able to detect medium-sized effects (Cohen's f=0.312) and the t-tests were able to detect large effects (Cohen's d=0.701) at $\alpha = 0.05$ with a power of 0.8. As for Experiment I, repeated-measures correlation with rmcorr was used to test correlations at a significant level of $\alpha = .05$.

**Manipulation Check:** A repeated-measures ANOVA was conducted on Game Scenario (Over, Under and Opt) for *IMI Enjoy*. The effect of Game Scenario was significant ($F(2,34) = 13.873, p < .001^{***}$) with a 'large' effect size ($\omega^2 = 0.261$). *IMI Enjoy* in Opt was significantly higher compared to Over ($t(17) = 2.480, p = .012^{*}$) with a 'medium' effect size (Cohen's d=0.585) and also significantly higher compared to Under ($t(17) = 5.497, p < .001^{***}$) with a 'large' effect size (Cohen's d=1.296). A repeated-measures ANOVA was conducted on Game Scenario (Over, Under and Opt) for *Affect*. The effect of Game Scenario was significant ($F(2,34) = 15.868, p < .001^{***}$) with a 'large' effect size ($\omega^2 = 0.284$). *Affect* in Opt was significantly higher compared to Over ($t(17) = 6.285, p < .001^{***}$) with a 'large' effect size (Cohen's d=1.481) and also significantly higher compared to Under ($t(17) = 4.216, p < .001^{***}$) with a 'large' effect size (Cohen's d=0.994). All this shows that Opt, Over and Under were successful in eliciting significantly different levels of affect, which is necessary in order to analyse the correlations of *Affect* with the psychophysiological variables.

**Validity of Affect:** A repeated-measures correlation was conducted on Game Scenario for *Affect* and *IMI Enjoy*. *Affect* was significantly, positively correlated with *IMI Enjoy* ($r(35) = 0.579, p < .001^{***}$), so we accept H4. This indicates that our experience sampling method (*Affect*) is a valid measure of affect in relation to *IMI Enjoy*.

**Correlates of Affect, Valence and Arousal:** *Blinks* was significantly negatively correlated with *Affect* ($r(35) = -0.374, p = .011^{*}$), so we accept H5. *Pupil* was significantly positively correlated with *Affect* ($r(35) = 0.346, p = .018^{*}$), so we accept H6. *Fixations* was significantly, positively correlated with *Affect* ($r(35) = 0.409, p = .006^{**}$), so we accept H7. *Power* was significantly positively correlated with *Affect* ($r(35) = 0.382; p = .010^{*}$), so we accept H8. *Blinks* was significantly negatively correlated with *sign(Affect)* ($r(35) = -0.460; p = .002^{**}$). *Pupil* was not significantly positively correlated with *sign(Affect)* ($r(35) = 0.270; p = .053$). *Fixations* was significantly, positively correlated with *sign(Affect)* ($r(35) = 0.512; p < .001^{***}$). *Power* was significantly positively correlated with *sign(Affect)* ($r(35) = 0.296; p = .038^{*}$). *Conductivity* was significantly positively correlated with |*Affect*| ($r(35) = 0.335; p = .021^{*}$), so we accept H9.

**Regression Analysis:** The combined linear effects of the psychophysiological variables on affect were analysed using multilevel linear regression models [28, 93] in R through the nlme package [16]. The psychophysiological variables were set up as fixed effects and participant number was set up as grouping factor. We tested the regression coefficients for significance with $\alpha = .05$, based on our hypotheses, reporting their 95% confidence intervals *CI*. We first analysed the effects of *Fixations*, *Pupil*, *Blinks* and *Power* on *Affect*: the effects of *Fixtions* ($B = 0.414, CI = [0.076, 0.752], t(49) = 2.463, p = .009^{**}$) and *Pupil* ($B = 0.322, CI = [-0.015, 0.660], t(49) = 1.920, p = .030^{*}$) were significant, and the effects of *Blinks* ($B = 0.002, CI = [-0.002, 0.005], t(49) = 0.821, p = .208$) and *Power* ($B = -0.001, CI = [-0.004, 0.002], t(49) = -0.663, p = .255$) were not significant. This indicates that *Fixations* and *Pupil* are the most important linear predictors of *Affect* ($R^2 = 0.246$), with *Blinks* and *Power* not improving the prediction significantly. We then analysed the effects of *Fixations*, *Blinks* and *Power* on *Valence*: the

Table 3: Demographics and results for Experiment II (avg.±std. dev.).

| Demographics | Variable | Under | Over | Opt |
|---|---|---|---|---|
| n=18 | Power | $300.85 \pm 116.76$ | $309.23 \pm 102.31$ | $332.88 \pm 113.95$ |
| m=14; f=4 | IMI Enjoy | $3.41 \pm 1.22$ | $4.45 \pm 1.22$ | $5.183 \pm 1.10$ |
| age=$26 \pm 5$ | Blinks | $103.67 \pm 91.34$ | $96.67 \pm 81.38$ | $86.83 \pm 86.136$ |
| | Pupil | $45.07 \pm 4.57$ | $42.33 \pm 5.31$ | $43.72 \pm 4.66$ |
| | Conductivity | $4.29 \pm 3.04$ | $4.40 \pm 2.73$ | $4.63 \pm 2.97$ |
| | Fixations | $3267.39 \pm 1432.98$ | $2777.44 \pm 1319.14$ | $3018.28 \pm 1613.41$ |

Table 4: Correlation coefficients of Experiment I.

| Variable | IMI Enjoy |
|---|---|
| Conductivity | H2 $^{*}$0.418 |
| Blinks | H3 $^{**}$-0.574 |

Table 5: Correlation coefficients of Experiment II.

| Variable | Affect | Valence | Arousal |
|---|---|---|---|
| Conductivity | -0.059 | -0.092 | H9 $^{*}$0.335 |
| Blinks | H5 $^{*}$-0.374 | $^{**}$-0.460 | 0.121 |
| Pupil | H6 $^{*}$0.346 | 0.270 | -0.211 |
| Fixations | H7 $^{**}$0.409 | $^{***}$0.512 | -0.115 |
| Power | H8 $^{*}$0.382 | $^{*}$0.296 | 0.116 |

effect of *Fixtions* ($B = 0.456, CI = [0.196, 0.716], t(50) = 3.521, p < .001^{***}$) was significant, and the effects of *Blinks* ($B < 0.001, CI = [-0.003, 0.003], t(50) = 0.005, p = .498$) and *Power* ($B = -0.001, CI = [-0.003, 0.001], t(50) = -1.013, p = .158$) were not significant. This indicates that *Fixations* is the most important linear predictor of *Valence* ($R^2 = 0.262$), with *Blinks* and *Power* not improving the prediction significantly. Combining our results about all predictors of *Valence*, *Arousal* and *Affect*, we then analysed the effects of $sign(Fixations) \times Conductivity$, *Pupil* and *Fixations* on *Affect*: the effects of *Pupil* ($B = 0.330, CI = [0.015, 0.646], t(50) = 2.106, p = .020^{*}$) were significant, and the effects of $sign(Fixations) \times Conductivity$ ($B = 0.466, CI = [-0.266, 1.199], t(50) = 1.279, p = .103$) and *Fixations* ($B = -0.018, CI = [-0.730, 0.694], t(50) = -0.050, p = .480$) were not significant. After removing *Fixations* from the model, the effects of both $sign(Fixations) \times Conductivity$ ($B = 0.449, CI = [0.130, 0.769], t(51) = 2.823, p = .003^{**}$) and *Pupil* ($B = 0.329, CI = [0.021, 0.637], t(51) = 2.146, p = .018^{*}$) were significant, indicating that this is a suitable model ($R^2 = 0.322$). Figure 1e shows a graph of model predictions vs. *Affect* with regression lines for each participant. The regression lines are scattered along the diagonal as each participant has her own baseline levels for the psychophysiological variables and the *Affect* measurements. The repeated-measures correlation and regression analyses are able to take these individual variations into account. Combining our results about more conventional predictors of *Valence*, *Arousal* and *Affect*, we finally analysed the effects of $sign(-Blinks) \times Conductivity$, *Pupil* and *Blinks* on *Affect*: the effects of $sign(-Blinks) \times Conductivity$ ($B = 0.548, CI = [0.018, 1.078], t(50) = 2.075, p = .021^{*}$), *Pupil* ($B = 0.379, CI = [0.054, 0.705], t(50) = 2.339, p = .011^{*}$) and *Blinks* ($B = 0.005, CI = [-0.0004, 0.011], t(50) = 1.863, p = .034^{*}$) were all significant, indicating that this is also a suitable model ($R^2 = 0.247$).

**Qualitative Feedback:** A recurring theme in Opt was positive engagement, e.g. "enticing", "motivating" and "stimulating". Participants also commented positively on the level of challenge in Opt, e.g. "the competition was fair which made it really fun". Comments on Under such as "dull", "super boring" and "repetitive" resonate with its purpose to underwhelm the player. Comments about Over, e.g. "disturbing", "annoying", "irritating", expressed that participants were clearly frustrated. They described the level of challenge as "disheartening" and the competitor as "too fast" and "winning by so much!".

## Discussion

The results indicate that affect in VR exergames can be measured by integrating an experience sampling scale directly into the exergame, mitigating the need for more intrusive measurements. The proposed scale correlates with the widely-used and well-validated IMI Interst/Enjoyment scale (H4). However, unlike IMI it has a straightforward neutral point that facilitates interpretation, and it can be quickly applied during the experience, making it easier to collect time series data about affect. The experiment demonstrates how affect can be manipulated in an exergame by design, based on game aesthetics and cognitive and physical challenge, making it easier for other researchers to collect new data sets on affect. This is also relevant for the calibration of affective exergames, where different affective responses need to be collected in order to determine personal parameters of affect recognition models. The main aim of our second experiment was to find the psychophysiological correlates of positive and negative affect (RQ3), and the hypothesised correlations (H5-H9) were all confirmed. This indicates that the identified correlates are valid for a range of affect measurements spanning positive as well as neutral and negative responses, and that they are robust enough to help predict affect in high intensity VR exergames.

While other works use machine learning approaches to recognise affect, these approaches often hide the underlying relationships between psychophysiological variables and recognised affect. By contrast, our analysis of linear relationships increases conceptual understanding and can be used as a basis to build more complex predictors modeling non-linear relationships. A main result is that measures related to the eye are promising predictors of affect in high intensity VR exergaming, and with the rise of eye-tracking enabled HMDs this is becoming increasingly relevant. In particular, the use of eye gaze fixations in this context is novel and promising, and there is potential to refine this approach by separately considering fixations on specific visual design elements of a game.

The correlation analysis confirms that not all psychophysiological measures are made equal. For example, skin conductivity is mainly a measure of arousal and blinks are more useful for predicting valence than arousal. The results also demonstrate that these measures can be combined to build stronger predictors, e.g. by estimating affect as a product of valence and arousal. The correlations shed some light on good and bad choices for sensor selection, giving system designers an idea of what can be expected from a particular psychophysiological measurement. For example, pupil dilation alone – although widely used in affective experiments – works only moderately well in a VR exergame. While predictors involving fixations can be comparatively strong, they generally need semantic information about the game, e.g. where important game elements are currently visible on the screen, which requires access to game internals. This would be difficult to obtain when recognising affect in a closed-source game, and could even be difficult if the source code is available. By comparison, "conventional" predictors involving blink rate, pupil dilation and skin conductivity are context agnostic, i.e. they can be used independently of the experience that is measured.

Predicting psychological states such as affect is generally difficult. There is usually no straightforward predictor that is highly correlated with the desired property. Reasonable predictors combine different measures to form an overall estimate, and also reduce variance by considering individual differences between users. For example, we used individual averages and standard deviations in order to compensate for individual differences in response sensitivity and model linear affect response for specific players. While these parameters can be obtained fairly easily and mostly automatically, they do require individual calibration; so an interesting direction of future work would be how affect recognition could be calibrated continuously as part of an exergaming experience.

**Limitations:** We took repeated measures in both experiments, which means that our results may have been affected by familiarisation and fatigue. We mitigated this with a familiarisation phase, breaks and counterbalancing. Furthermore, our participants were mainly in their 20s and mostly male, which may affect the generalisability of our results. Some participants were not strangers to the research team, some of them participated in both experiments, and they had varying previous VR experience. Our manipulation checks indicate that this did not impact the results noticeably as all conditions were successful in eliciting the desired affective responses that could then be correlated with psychophysiological measurements. All participants were naive to the goals of the experiments and all of them went through a standardised familiarisation phase in both experiments to mitigate the effects of training and previous VR experience. Focusing on an activity lowers blink rate and this is a potential limitation for Experiment 1. However, Experiment 2, which used the same game mechanics, exercise protocol, equipment, ambient lighting and overall game environment as the EG condition of Experiment 1, shows that blink rate was significantly correlated with *Affect*. Our findings are in line with previous studies showing that blink rate is negatively correlated with dopamine activity, which is associated with affect [64, 67, 68].

Our tested correlations have 'moderate' to 'large' effect sizes, so they are likely to be useful for similar populations. However, the accuracy of the linear model could be further investigated using longitudinal studies with bigger sample sizes, e.g. to shed light on the stability of predictors over time. Exercise-related personality traits such as competitiveness may have influenced the results [45], and they could be included as covariates in future research. To increase ecological validity, future studies could be conducted in real word conditions such as gyms to mitigate the influence of Hawthorne effects.

**Impact:** Our work paves the way for affectively adaptive high intensity VR exergames that can improve adherence. Our approach of determining the affective state is versatile and can be extended to apply in other contexts, e.g. for affectively adaptive conventional workout approaches.

## CONCLUSION

We identified psychophysiological measures suitable for affect recognition in high intensity VR exergaming and determined their relationship with affect. Building on this, we proposed and evaluated a novel, robust, multisensor approach for affect recognition, which will help exergame designers to scientifically measure, personalise and optimise the player experience. The results of our experiments can inform future VR exergaming studies as they help researchers to design and test affect predictors, and to formulate hypotheses about how sensor measurements relate to affect. Our data sets and analyses are available online [10].[1] In summary, we come to the following conclusions:

1. We identified gaze fixations, eye blinks, pupil diameter and skin conductivity as psychophysiological measures suitable for recognising affect in high intensity VR exergaming.
2. We presented an affective predictor with an optimal combination of the psychophysiological correlates of positive and negative affect in high intensity VR exergaming.

Our approach and findings can be used as a basis to design and build affectively adaptive high intensity VR exergames with great potential to improve exercise adherence.

## ACKNOWLEDGEMENTS

Soumya C. Barathi's research is funded by the UK's EPSRC Centre for Doctoral Training in Digital Entertainment (CDE), EP/L016540/1 and the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 665992. Eamonn O'Neill's and Christof Lutteroth's research is partly funded by CAMERA, the RCUK Centre for the Analysis of Motion, Entertainment Research and Applications, EP/M023281/1.

[1] https://doi.org/10.15125/BATH-00758

## REFERENCES


[1] Konstantinos Alexandris, Charalambos Tsorbatzoudis, and George Grouios. 2002. Perceived constraints on recreational sport participation: Investigating their relationship with intrinsic motivation, extrinsic motivation and amotivation. *Journal of Leisure Research* 34, 3 (2002), 233.

[2] Silke Anders, Martin Lotze, Michael Erb, Wolfgang Grodd, and Niels Birbaumer. 2004. Brain activity underlying emotional valence and arousal: A response-related fMRI study. *Human brain mapping* 23, 4 (2004), 200–209.

[3] James Annesi. 2003. Effects of a cognitive behavioral treatment package on exercise attendance and drop out in fitness centers. *European Journal of Sport Science* 3, 2 (2003), 1–16.

[4] James J Annesi. 1998. Effects of computer feedback on adherence to exercise. *Perceptual and Motor Skills* 87, 2 (1998), 723–730.

[5] Değer Ayata, Yusuf Yaslan, and Mustafa Kamaşak. 2016. Emotion recognition via random forest and galvanic skin response: Comparison of time based feature sets, window sizes and wavelet approaches. In *2016 Medical Technologies National Congress (TIPTEKNO)*. IEEE, 1–4.

[6] Jonathan Z Bakdash and Laura R Marusich. 2017. Repeated measures correlation. *Frontiers in Psychology* 8 (2017), 456.

[7] Fabrizio Balducci, Costantino Grana, and Rita Cucchiara. 2017. Affective level design for a role-playing videogame evaluated by a brain–computer interface and machine learning methods. *The Visual Computer* 33, 4 (2017), 413–427.

[8] Rosa M Banos, Patricia Escobar, Ausias Cebolla, Jaime Guixeres, Julio Alvarez Pitti, Juan Francisco Lisón, and Cristina Botella. 2016. Using virtual reality to distract overweight children from bodily sensations during exercise. *Cyberpsychology, Behavior, and Social Networking* 19, 2 (2016), 115–119.

[9] Soumya C Barathi, Daniel J Finnegan, Matthew Farrow, Alexander Whaley, Pippa Heath, Jude Buckley, Peter W Dowrick, Burkhard C Wuensche, James LJ Bilzon, Eamonn O'Neill, and others. 2018. Interactive Feedforward for Improving Performance and Maintaining Intrinsic Motivation in VR Exergaming. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 408.

[10] Soumya C. Barathi, Michael Proulx, Eamonn O'Neill, and Christof Lutteroth. 2020. Datasets and Analyses for "Affect Recognition using Psychophysiological Correlates in High Intensity VR Exergaming". University of Bath Research Data Archive, Bath, UK. `https://doi.org/10.15125/BATH-00758`.

[11] Philipe de Souto Barreto. 2013. Why are we failing to promote physical activity globally? *Bulletin of the World Health Organization* 91, 6 (2013), 390–390A.

[12] Martin J Barwood, Neil JV Weston, Richard Thelwell, and Jennifer Page. 2009. A motivational music and video intervention improves high-intensity exercise performance. *Journal of Sports Science & Medicine* 8, 3 (2009), 435.

[13] Antonis S Billis, Evdokimos I Konstantinidis, Aristea I Ladas, Magda N Tsolaki, Costas Pappas, and Panagiotis D Bamidis. 2011. Evaluating affective usability experiences of an exergaming platform for seniors. In *Biomedical Engineering, 2011 10th International Workshop on*. IEEE, 1–4.

[14] Daniel T Bishop, Costas I Karageorghis, and Georgios Loizou. 2007. A grounded theory of young tennis players' use of music to manipulate emotional state. *Journal of Sport and Exercise Psychology* 29, 5 (2007), 584–607.

[15] J Martin Bland and Douglas G Altman. 1995. Calculating correlation coefficients with repeated observations: Part 2—Correlation between subjects. *Bmj* 310, 6980 (1995), 633.

[16] Paul Bliese. 2006. Multilevel Modeling in R (2.2)–A Brief Introduction to R, the multilevel package and the nlme package. (2006).

[17] John Bolton, Mike Lambert, Denis Lirette, and Ben Unsworth. 2014. PaperDude: a virtual reality cycling exergame. In *Proc. CHI Extended Abstracts*. ACM, 475–478.

[18] W Boucsein. 2012. Electrodermal activity: Springer Science+ Business Media. (2012).

[19] Margaret M Bradley, Laura Miccoli, Miguel A Escrig, and Peter J Lang. 2008. The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology* 45, 4 (2008), 602–607.

[20] Jeanne H Brockmyer, Christine M Fox, Kathleen A Curtiss, Evan McBroom, Kimberly M Burkhart, and Jacquelyn N Pidruzny. 2009. The development of the Game Engagement Questionnaire: A measure of engagement in video game-playing. *Journal of Experimental Social Psychology* 45, 4 (2009), 624–634.

[21] Emily Brown and Paul Cairns. 2004. A grounded investigation of game immersion. In *Proc. CHI extended abstracts*. ACM, 1297–1300.

[22] Kimberly A Brownley, Robert G McMurray, and Anthony C Hackney. 1995. Effects of music on physiological and affective responses to graded treadmill exercise in trained and untrained runners. *International Journal of Psychophysiology* 19, 3 (1995), 193–201.

[23] John T Cacioppo, Richard E Petty, Mary E Losch, and Hai Sook Kim. 1986. Electromyographic activity over facial muscle regions can differentiate the valence and intensity of affective reactions. *Journal of Personality and Social Psychology* 50, 2 (1986), 260.

[24] John T Cacioppo, Louis G Tassinary, and Gary Berntson. 2007. *Handbook of psychophysiology*. Cambridge University Press.

[25] Hande Çelikkanat, Hiroki Moriya, Takeshi Ogawa, Jukka-Pekka Kauppi, Motoaki Kawanabe, and Aapo Hyvärinen. 2017. Decoding emotional valence from electroencephalographic rhythmic activity. In *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 4143–4146.

[26] Georgios Christou. 2014. The interplay between immersion and appeal in video games. *Computers in Human Behavior* 32 (2014), 92–100.

[27] Rachel B Clancy, Matthew P Herring, and Mark J Campbell. 2017. Motivation Measures in Sport: A Critical Review and Bibliometric Analysis. *Frontiers in Psychology* 8 (2017).

[28] Avital Cnaan, Nan M Laird, and Peter Slasor. 1997. Using the general linear mixed model to analyse unbalanced repeated measures and longitudinal data. *Statistics in Medicine* 16, 20 (1997), 2349–2380.

[29] Roddy Cowie, Ellen Douglas-Cowie, Nicolas Tsapatsoulis, George Votsis, Stefanos Kollias, Winfried Fellenz, and John G Taylor. 2001. Emotion recognition in human-computer interaction. *IEEE Signal Processing Magazine* 18, 1 (2001), 32–80.

[30] Hugo D Critchley. 2002. Electrodermal responses: what happens in the brain. *The Neuroscientist* 8, 2 (2002), 132–142.

[31] M Csikszentmihalyi. 1990. Flow. The Psychology of Optimal Experience. New York (HarperPerennial) 1990. (1990).

[32] Mihaly Csikszentmihalyi. 2014. *Flow and the Foundations of Positive Psychology*. Springer.

[33] ISO DIS. 2009. 9241-210: 2010. Ergonomics of human system interaction-Part 210: Human-centred design for interactive systems. *International Standardization Organization (ISO). Switzerland* (2009).

[34] Rod K Dishman. 1988. *Exercise adherence: Its impact on public health*. Human Kinetics.

[35] Stefan Engeser. 2012. *Advances in flow research*. Springer.

[36] Fernanda Ferreira, Jens Apel, and John M Henderson. 2008. Taking a new look at looking at nothing. *Trends in cognitive sciences* 12, 11 (2008), 405–410.

[37] Samantha Finkelstein, Andrea Nickel, Tiffany Barnes, and Evan A Suma. 2010. Astrojumper: Motivating children with autism to exercise using a VR game. In *CHI'10 Extended Abstracts on Human Factors in Computing Systems*. ACM, 4189–4194.

[38] Canadian Fitness and Lifestyle Research Institute. 2005. 2005 Physical Activity Monitor. (2005).

[39] Don C Fowles, Margaret J Christie, Robert Edelberg, William W Grings, David T Lykken, and Peter H Venables. 1981. Publication recommendations for electrodermal measurements. *Psychophysiology* 18, 3 (1981), 232–239.

[40] Christina M Frederick and Richard M Ryan. 1993. Differences in motivation for sport and exercise and their relations with participation and mental health. *Journal of Sport Behavior* 16, 3 (1993), 124.

[41] Tom Garner, Mark Grimshaw, and Debbie Abdel Nabi. 2010. A preliminary experiment to assess the fear value of preselected sound parameters in a survival horror game. In *Proceedings of the 5th Audio Mostly Conference: A Conference on Interaction with Sound*. ACM, 10.

[42] Martin J Gibala, Jenna B Gillen, and Michael E Percival. 2014. Physiological and health-related adaptations to low-volume interval training: influences of nutrition and sex. *Sports Medicine* 44, 2 (2014), 127–137.

[43] Martin J Gibala, Jonathan P Little, Maureen J MacDonald, and John A Hawley. 2012. Physiological adaptations to low-volume, high-intensity interval training in health and disease. *The Journal of Physiology* 590, 5 (2012), 1077–1084.

[44] Martin J Gibala and Sean L McGee. 2008. Metabolic adaptations to short-term high-intensity interval training: a little pain for a lot of gain? *Exercise and Sport Sciences Reviews* 36, 2 (2008), 58–63.

[45] Diane L Gill and Thomas E Deeter. 1988. Development of the sport orientation questionnaire. *Research Quarterly for Exercise and Sport* 59, 3 (1988), 191–202.

[46] Kiel Gilleade, Alan Dix, and Jen Allanson. 2005. Affective videogames and modes of affective gaming: assist me, challenge me, emote me. *DiGRA 2005: Changing Views–Worlds in Play.* (2005).

[47] Kiel M Gilleade and Alan Dix. 2004. Using frustration in the design of adaptive videogames. In *Proceedings of the 2004 ACM SIGCHI International Conference on Advances in computer entertainment technology*. ACM, 228–232.

[48] S Navid Hajimirza, Michael J Proulx, and Ebroul Izquierdo. 2012. Reading users' minds from their eyes: a method for implicit image annotation. *IEEE Transactions on Multimedia* 14, 3 (2012), 805–815.

[49] Sandro Hardy, Stefan Göbel, Michael Gutjahr, Josef Wiemeyer, and Ralf Steinmetz. 2012. Adaptation model for indoor exergames. *International Journal of Computer Science in Sport* 11, 1 (2012).

[50] Tilo Hartmann, Christoph Klimmt, and Peter Vorderer. 2010. Telepresence and media entertainment. *Immersed in media: Telepresence in everyday life* (2010), 137–157.

[51] Richard L Hazlett. 2006. Measuring emotional valence during interactive experiences: boys at video game play. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*. ACM, 1023–1026.

[52] Joel M Hektner, Jennifer A Schmidt, and Mihaly Csikszentmihalyi. 2007. *Experience sampling method: Measuring the quality of everyday life*. Sage.

[53] John M Henderson, Wonil Choi, Steven G Luke, and Joseph Schmidt. 2018. Neural correlates of individual differences in fixation duration during natural reading. *Quarterly Journal of Experimental Psychology* 71, 1 (2018), 314–323.

[54] John M Henderson and Steven G Luke. 2014. Stable individual differences in saccadic eye movements during reading, pseudoreading, scene viewing, and scene search. *Journal of Experimental Psychology: Human Perception and Performance* 40, 4 (2014), 1390.

[55] Eckhard H Hess and James M Polt. 1960. Pupil size as related to interest value of visual stimuli. *Science* 132, 3423 (1960), 349–350.

[56] Harrie F Hess and Jerry V Diller. 1969. Motivation for gambling as revealed in the marketing methods of the legitimate gambling industry. *Psychological Reports* 25, 1 (1969), 19–27.

[57] Morris B Holbrook, Robert W Chestnut, Terence A Oliva, and Eric A Greenleaf. 1984. Play as a consumption experience: The roles of emotions, performance, and personality in the enjoyment of games. *Journal of consumer research* 11, 2 (1984), 728–739.

[58] Sungpyo Hong, Joanna Narkiewicz, and Randy H Kardon. 2001. Comparison of pupil perimetry and visual perimetry in normal eyes: decibel sensitivity and variability. *Investigative ophthalmology & visual science* 42, 5 (2001), 957–965.

[59] Robin Hunicke, Marc LeBlanc, and Robert Zubek. 2004. MDA: A formal approach to game design and game research. In *Proc. AAAI Workshop on Challenges in Game AI*, Vol. 4. 1722.

[60] Christos Ioannou, Patrick Archard, Eamonn O'Neill, and Christof Lutteroth. 2019. Virtual Performance Augmentation in an Immersive Jump & Run Exergame. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 158.

[61] Susan A Jackson and Herbert W Marsh. 1996. Development and validation of a scale to measure optimal experience: The Flow State Scale. *Journal of Sport and Exercise Psychology* 18, 1 (1996), 17–35.

[62] Richard David Jacques. 1996. *The nature of engagement and its role in hypermedia evaluation and design.* Ph.D. Dissertation. South Bank University.

[63] Charlene Jennett, Anna L Cox, Paul Cairns, Samira Dhoparee, Andrew Epps, Tim Tijs, and Alison Walton. 2008. Measuring and defining the experience of immersion in games. *International journal of human-computer studies* 66, 9 (2008), 641–661.

[64] Jaime Kaminer, Alice S Powers, Kyle G Horn, Channing Hui, and Craig Evinger. 2011. Characterizing the spontaneous blink generator: an animal model. *Journal of Neuroscience* 31, 31 (2011), 11256–11267.

[65] Costas I Karageorghis, Leighton Jones, and Daniel C Low. 2006a. Relationship between exercise heart rate and music tempo preference. *Research quarterly for exercise and sport* 77, 2 (2006), 240–250.

[66] Costas I Karageorghis, David-Lee Priest, Peter C Terry, Nikos LD Chatzisarantis, and Andrew M Lane. 2006b. Redesign and initial validation of an instrument to assess the motivational qualities of music in exercise: The Brunel Music Rating Inventory-2. *Journal of sports sciences* 24, 8 (2006), 899–909.

[67] Craig N Karson. 1983. Spontaneous eye-blink rates and dopaminergic systems. *Brain* 106, 3 (1983), 643–653.

[68] Craig N Karson, Karen Faith Berman, Edward F Donnelly, Wallace B Mendelson, Joel E Kleinman, and Richard Jed Wyatt. 1981. Speaking, thinking, and blinking. *Psychiatry research* 5, 3 (1981), 243–246.

[69] Marcus Kilpatrick, Edward Hebert, and John Bartholomew. 2005. College students' motivation for physical activity: differentiating men's and women's motives for sport participation and exercise. *Journal of American college health* 54, 2 (2005), 87–94.

[70] Paul B Laursen and David G Jenkins. 2002. The scientific basis for high-intensity interval training. *Sports medicine* 32, 1 (2002), 53–73.

[71] Felissa K Lee, Kennon M Sheldon, and Daniel B Turban. 2003. Personality and the goal-striving process: The influence of achievement goal patterns, goal level, and mental focus on performance and enjoyment. *Journal of Applied Psychology* 88, 2 (2003), 256.

[72] Lili Liu, Rachael Ip, Anna Shum, and Christian Wagner. 2014. Learning effects of virtual game worlds: an empirical investigation of immersion, enjoyment and performance. (2014).

[73] Fausto Massimini and Massimo Carli. 1988. The systematic assessment of flow in daily experience. (1988).

[74] Fausto Massimini, Mihaly Csikszentmihalyi, and Massimo Carli. 1987. The monitoring of optimal experience: A tool for psychiatric rehabilitation. *Journal of Nervous and Mental disease* (1987).

[75] Edward McAuley, Terry Duncan, and Vance V Tammen. 1989. Psychometric properties of the Intrinsic Motivation Inventory in a competitive sport setting: A confirmatory factor analysis. *Research Quarterly for Exercise and Sport* 60, 1 (1989), 48–58.

[76] Arlen C Moller, Brian P Meier, and Robert D Wall. 2010. Developing an experimental induction of flow: Effortless action in the lab. *Effortless attention: A new perspective in the cognitive science of attention and action* (2010), 191–204.

[77] Javier Monedero, Elizabeth J Lyons, and Donal J O'Gorman. 2015. Interactive video game cycling leads to higher energy expenditure and is more enjoyable than conventional exercise in adults. *PloS One* 10, 3 (2015).

[78] Larissa Müller, Arne Bernin, Sobin Ghose, Wojtek Gozdzielewski, Qi Wang, Christos Grecos, Kai von Luck, and Florian Vogt. 2016. Physiological data analysis for an emotional provoking exergame. In *2016 IEEE Symposium Series on Computational Intelligence (SSCI)*. IEEE, 1–8.

[79] Larissa Müller, Arne Bernin, Andreas Kamenz, Sobin Ghose, Kai von Luck, Christos Grecos, Qi Wang, and Florian Vogt. 2017. Emotional journey for an emotion provoking cycling exergame. In *2017 IEEE 4th International Conference on Soft Computing & Machine Intelligence (ISCMI)*. IEEE, 104–108.

[80] Larissa Müller, Sebastian Zagaria, Arne Bernin, Abbes Amira, Naeem Ramzan, Christos Grecos, and Florian Vogt. 2015. Emotionbike: a study of provoking emotions in cycling exergames. In *International Conference on Entertainment Computing*. Springer, 155–168.

[81] Lennart Nacke. 2009. *Affective ludology: Scientific measurement of user experience in interactive entertainment*. Ph.D. Dissertation. Blekinge Institute of Technology.

[82] Lennart E Nacke and Craig A Lindley. 2010. Affective ludology, flow and immersion in a first-person shooter: Measurement of player experience. *arXiv preprint arXiv:1004.0248* (2010).

[83] Lennart E Nacke, Sophie Stellmach, and Craig A Lindley. 2011. Electroencephalographic assessment of player experience: A pilot study in affective ludology. *Simulation & Gaming* 42, 5 (2011), 632–655.

[84] Mohsen Naji, Mohammd Firoozabadi, and Parviz Azadfallah. 2014. Classification of music-induced emotions based on information fusion of forehead biosignals and electrocardiogram. *Cognitive Computation* 6, 2 (2014), 241–252.

[85] Jeanne Nakamura and Mihaly Csikszentmihalyi. 2014. The concept of flow. In *Flow and the Foundations of Positive Psychology*. Springer, 239–263.

[86] Simon Niedenthal. 2009. What we talk about when we talk about game aesthetics. (2009).

[87] Pedro A Nogueira, Rui Rodrigues, and Eugénio Oliveira. 2013a. Real-time psychophysiological emotional state estimation in digital gameplay scenarios. In *International Conference on Engineering Applications of Neural Networks*. Springer, 243–252.

[88] Pedro A Nogueira, Rui Rodrigues, Eugénio Oliveira, and Lennart E Nacke. 2013b. A hybrid approach at emotional state detection: Merging theoretical models of emotion with data-driven statistical classifiers. In *Proceedings of the 2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)-Volume 02*. IEEE Computer Society, 253–260.

[89] Pedro A Nogueira, Rui Rodrigues, Eugénio Oliveira, and Lennart E Nacke. 2013c. A regression-based method for lightweight emotional state detection in interactive environments. In *XVI Portuguese conference on artificial intelligence (EPIA)*.

[90] American College of Sports Medicine. 2017. *ACSM's guidelines for exercise testing and prescription*. Lippincott Williams & Wilkins.

[91] Stephen Olejnik and James Algina. 2003. Generalized eta and omega squared statistics: measures of effect size for some common research designs. *Psychological methods* 8, 4 (2003), 434.

[92] Timo Partala and Veikko Surakka. 2003. Pupil size variation as an indication of affective processing. *International journal of human-computer studies* 59, 1-2 (2003), 185–198.

[93] James L Peugh. 2010. A practical guide to multilevel modeling. *Journal of school psychology* 48, 1 (2010), 85–112.

[94] Ryan E Rhodes, Darren ER Warburton, and Shannon SD Bredin. 2009. Predicting the effect of interactive video bikes on exercise adherence: An efficacy trial. *Psychology, health & medicine* 14, 6 (2009), 631–640.

[95] Øivind Rognmo, Eva Hetland, Jan Helgerud, Jan Hoff, and Stig A Slørdahl. 2004. High intensity aerobic interval exercise is superior to moderate intensity exercise for increasing aerobic capacity in patients with coronary artery disease. *European Journal of Cardiovascular Prevention & Rehabilitation* 11, 3 (2004), 216–222.

[96] Gordon Rugg and Marian Petre. 2007. A Gentle Guide to Research Methods. 1st. (2007).

[97] James A Russell, Anna Weiss, and Gerald A Mendelsohn. 1989. Affect grid: a single-item scale of pleasure and arousal. *Journal of Personality and Social psychology* 57, 3 (1989), 493.

[98] Richard M Ryan. 1982. Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. *Journal of Personality and Social Psychology* 43, 3 (1982), 450–461.

[99] Richard M Ryan, Christina M Frederick, Deborah Lepes, Noel Rubio, and Kennon M Sheldon. 1997. Intrinsic motivation and exercise adherence. *International Journal of Sport Psychology* 28, 4 (1997), 335–354.

[100] Kenzo Sato and Richard L Dobson. 1970. Regional and individual variations in the function of the human eccrine sweat gland. *Journal of Investigative Dermatology* 54, 6 (1970), 443–449.

[101] K Sato and F Sato. 1983. Individual variations in structure and function of human eccrine sweat gland. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology* 245, 2 (1983), R203–R208.

[102] May-li Seah and Paul Cairns. 2008. From immersion to addiction in videogames. In *Proceedings of the 22nd British HCI Group Annual Conference on People and Computers: Culture, Creativity, Interaction-Volume 1*. British Computer Society, 55–63.

[103] Jeff Sinclair, Philip Hingston, and Martin Masek. 2009. Exergame development using the dual flow model. In *Proc. Australasian Conference on Interactive Entertainment*. ACM.

[104] Burrhus Frederic Skinner. 1953. *Science and human behavior*. Simon and Schuster.

[105] John A Stern, Larry C Walrath, and Robert Goldstein. 1984. The endogenous eyeblink. *Psychophysiology* 21, 1 (1984), 22–33.

[106] Penelope Sweetser and Peta Wyeth. 2005. GameFlow: a model for evaluating player enjoyment in games. *Computers in Entertainment (CIE)* 3, 3 (2005), 3–3.

[107] Jonathan Sykes and Simon Brown. 2003. Affective gaming: measuring emotion through the gamepad. In *CHI'03 extended abstracts on Human factors in computing systems*. ACM, 732–733.

[108] Ching-I Teng. 2010. Customization, immersion satisfaction, and online gamer loyalty. *Computers in Human Behavior* 26, 6 (2010), 1547–1554.

[109] Laura E Thomas and Alejandro Lleras. 2007. Moving eyes and moving thought: On the spatial compatibility between eye movements and cognition. *Psychonomic bulletin & review* 14, 4 (2007), 663–668.

[110] Scott Thomas, Jeff Reading, and Roy J Shephard. 1992. Revision of the physical activity readiness questionnaire (PAR-Q). *Canadian journal of sport sciences* (1992).

[111] Darren ER Warburton, Shannon SD Bredin, Leslie TL Horita, Dominik Zbogar, Jessica M Scott, Ben TA Esch, and Ryan E Rhodes. 2007. The health benefits of interactive video game exercise. *Applied Physiology, Nutrition, and Metabolism* 32, 4 (2007), 655–663.

[112] Philip M Wilson, Wendy M Rodgers, Chris M Blanchard, and Joanne Gessell. 2003. The relationship between psychological needs, self-determined motivation, exercise attitudes, and physical fitness. *Journal of Applied Social Psychology* 33, 11 (2003), 2373–2392.

[113] Bob G Witmer and Michael J Singer. 1998. Measuring presence in virtual environments: A presence questionnaire. *Presence* 7, 3 (1998), 225–240.

[114] Sandy Wolfson and Gill Case. 2000. The effects of sound and colour on responses to a computer game. *Interacting with computers* 13, 2 (2000), 183–192.

[115] Haris Zacharatos, Christos Gatzoulis, Yiorgos Chrysanthou, and Andreas Aristidou. 2013. Emotion recognition for exergames using laban movement analysis. In *Proceedings of Motion on Games*. ACM, 61–66.

[116] Jiri Zuzanek. 1999. Experience sampling method: Current and potential research applications. In *Workshop on Measurement of and Research on Time Use, National Research Council, Washington, DC*. 108.