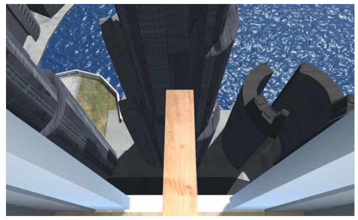
Characterizing Physiological Responses to Fear, Frustration, and Insight in Virtual Reality

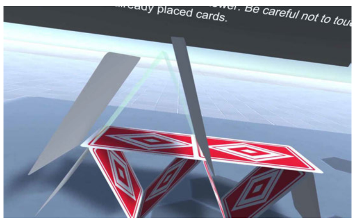
Tiffany Luong and Christian Holz

(a) Moments of fear (haunted house)



(b) Moments of fear (vertigo in a virtual pit)

(c) Moments of frustration (rigged archery)



(d) Moments of frustration (collapsing cards)

(e) Moments of insight (escape room)

Fig. 1: In a study with 24 participants, we elicited *transient psychological states* using: (a&b) moments of fear, (c&d) moments of frustration when accomplishing a task, and (e) moments of insight when solving a puzzle. During the study, we measured participants' cardiac, pulmonary, electrodermal, and pupillary activity to capture participants' physiological behavior and present an analysis to understand the *short-term responses* to events in VR.

Abstract—Physiological sensing often complements studies of human behavior in virtual reality (VR) to detect users' affective and cognitive states. Some psychological states, such as fear and frustration, can be particularly hard to differentiate from a physiological perspective as they are close in the arousal and valence emotional space. Moreover, it is largely unclear how users' physiological reactions are expressed in response to *transient psychological states* such as fear, frustration, and insight—especially since these are rich indicators for characterizing users' responses to dynamic systems but are hard to capture in highly interactive settings. We conducted a study ($N = 24$) to analyze participants' pulmonary, electrodermal, cardiac, and pupillary responses to moments of fear, frustration, and insight in immersive settings. Participants interacted in five VR environments, throughout which we measured their physiological reactions and analyzed the patterns. We also measured subjective fear and frustration using questionnaires. We found differences between fear and frustration pupillary, respiratory, and electrodermal responses, as well as between the pupillary changes that followed fear in a horror game and those that followed fear in a vertigo experiment. We present the relationships between fear levels, frustration levels, and their physiological responses. To detect these affective events and states, we introduce user-independent binary classification models that achieved an average micro F_1 score of 71 % for detecting fear in a horror game, 75 % for fear of vertigo, 76 % for frustration, and 75 % for insight, showing the promise for detecting these states from passive and objective signals.

Index Terms—Virtual Reality, Affective Computing, Emotions, Cognitive State, Physiological Measures

1 Introduction

Virtual reality (VR) immerses users into worlds—real or fictional—that can substantially differ from any of their regular environments. This inherent property of VR is useful for entertainment and discovery, such as gaming and virtual travel. More interestingly, researchers can also harness this property to expose users to situations that would not be feasible in real-life or even dangerous [9], all while observing their behavior, actions, and reactions. This feature provides an optimal training environment for participants to develop new skills, as they can fail without facing real-world consequences [17].

In addition to monitoring users' physical behavior, researchers have leveraged VR to study users' physiological behaviors, such as stress during training [17, 47], mental workload in manufacturing [7], or racial discrimination [2]. Studies of users' psychological states can help better understand human cognition and learning processes [24]. Beyond VR experiences [24], they can produce implications for the design of real-world procedures [9].

However, studies of psychological states have mostly focused on *persistent* states, which facilitates their detection and makes monitoring easier. Studying and detecting *transient psychological states* is considerably more challenging as they occur suddenly and are difficult to reliably elicit. Fear, frustration, and insight are particularly interesting in this regard because they can reveal findings that are orthogonal to those gained from persistent states. Moments of fear and frustration can have manifold of causes, but they always impair users' performances in VR tasks [46] and discourage engagement and future use [26]. Conversely, insight—colloquially called "Aha! moment" [25] or "Eureka moment" [30]—can have the opposite effect and is often identified as a form of creativity that can result in important innovations [30]. Past research has shown that users better remember solutions when accompanied by moments of insight than when missing this feeling of epiphany [11]. Identifying the reasons that lead to such "Aha! Moments" is non-trivial and has long been of interest [25], but there is a lack of studies on frustration and insight responses that allow users to freely move about as they would in regular VR settings.

• *The authors are with the Dept. of Computer Science at ETH Zürich, CH.*
• *E-mails: {firstname.lastname}@inf.ethz.ch.*

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

In this paper, we present a controlled study of users' physiological responses to contexts and events designed to elicit fear, frustration, and insight. The purpose of our study was to capture the physiological responses to these events and transient states, monitored through wearable sensors that record pulmonary activity (breathing belt), electrodermal activity (finger straps), cardiac activity (ear clip), and pupillary activity (eye tracker). Our wearable setup allowed us to expose participants to a wide variety of scenarios and tasks to be completed during regular interaction in VR. From our analysis of the physiological recordings, we extract common patterns following interventions in VR and discuss their potential for designing experiments and experiences in VR.

As shown in Figure 1, our study environment elicited a variety of psychological states through five VR scenarios: a) fear when navigating a haunted house in a horror game, induced by visual and auditory effects, b) fear of vertigo when experiencing heights at various levels while walking on a plank, c) frustration in a rigged archery game when arrows slip and balloons fail to pop, d) frustration in a tower of cards game when the tower collapses just before completion, and e) moments of insight when completing escape-room puzzles.

Our analysis shows that although fear and frustration share close proximity inside the emotional space [51], they elicit different physiological and behavioral responses. While in both cases, participants' heart rate variability (HRV) decreased and their pupil diameters increased, the temporal evolution of their pupil dynamics, respiratory rate, and phasic skin response greatly differed. Moments of insight also manifested themselves in cardiac and pupillary responses that differed from responses to frustration and fear. We further confirmed that physiological responses to fear (in response to imminent threat in a horror game) expressed different pupil variations than in fear of heights.

In a second step, we corroborate our analysis of these important multi-modal signals through learning-based detection models. Our user-independent classifiers achieve promising accuracies for detecting fear of horror (mean micro F_1 score of 71 %), fear of vertigo ($F_1 = 75\,\%$), frustration ($F_1 = 76\,\%$), and moments of insight ($F_1 = 75\,\%$).

Overall, our results demonstrate the link between physiological patterns that result from users' responses to transient moments of frustration, fear, and insight. In our discussion, we outline the implications of these indicators for future interactive systems, as they afford *objective* and *passive* monitoring of short moments that can influence user experience in VR applications using non-invasive eye-tracking, pulmonary and cardiac activity monitoring, and electrodermal activity tracking—passive input modalities that are emerging as part of recent VR headsets themselves [3, 34] as well as everyday glasses [23].

Taken together, we make the following contributions:

1. a controlled environment inside five VR scenarios to elicit *transient* psychological states, including fear in a horror scenario, fear of vertigo, and frustration, all in response to events unfolding in interactive scenes, and moments of insight while solving a puzzle,
2. an analysis of cardiovascular, pulmonary, and autonomic physiological responses and their potential for indicating these transient psychological states, and
3. a computational demonstration of detecting such transient psychological states using user-independent classifications models that take physiological signals from wearable sensors as input, operating in real-time in highly interactive VR scenarios.

2 Background

2.1 Affective and Cognitive States in VR

Emotions have often been explored in the context of VR [35, 49]. Kolakowska et al. [29] distinguished three ways to classify emotional models: the dimensional, discrete, and hybrid perspectives. From a dimensional perspective, emotions are considered continuous phenomena, usually within two or three dimensions [35]. The most popular emotional dimensions are valence (from negative to positive) and arousal (degree of mental activation) [35]. From a discrete perspective, emotions are considered distinct. For example, the Big Six distinguish Happiness, Sadness, Fear, Anger, Disgust, and Surprise [12]. The hybrid perspective combines discrete and dimensional viewpoints. For example, the Russell [51]'s circumplex model of affect replaces some discrete emotions in the dimensional space (arousal on x, valence on y). In this model, *afraid* and *frustrated* are situated in the same emotional space (negative valence, positive arousal; see supplementary materials).

In 1990, Hodges et al. [22] produced fear of height experiences in VR. The experiment has been frequently replicated since [38, 53], including in the well-known virtual pit experiment [38, 62]. While vertigo is often used to treat clinically phobic patients in VR, vertigo in VR was also shown to work on non-phobic users [38, 53]. Besides the fear of height [53], there are also other types of fear that have been used in entertainment (e.g., horror games [45]). Indeed, VR horror games have been highly anticipated also to promote movies such as *Paranormal Activity: The Ghost Dimension* or *The Conjuring 2* [33].

In contrast, frustration has rarely been the subject of VR-based investigations. Experiencing frustration during UI interaction can negatively influence productivity, user experience, reduce acceptance of a technology or of an interaction, and can drive users to seek alternative systems [32]. Paladines-Jaramillo et al. [43] adapted Rozensweig's picture test for VR where users are exposed to various situations to study their tolerance to frustration. They reported preliminary feedback and showed that the adapted test was promising. Another study found that interactive loading screens improved users' experience in VR compared to passive ones by measuring users' frustration, enjoyment, and perceived speed of loading UIs [19].

As for insight, it is a positive instant that occurs where one, for example, suddenly finds the solution to a problem, understands a joke, or realizes something about a situation or oneself. Insight has been shown to benefit creativity and learning retention [30, 60]. Outside VR, past work mainly relied on the *compound remote associates test* [30], where users see three words (e.g., food, forward, break) and form a compound or familiar two-word phrase with each of the three problem words (e.g., fast for fast-food, fast-forward, and breakfast) to elicit insight. Despite the many tools for creativity in VR (e.g., [27, 28, 39, 50]), there is only one study in which researchers tried to detect insight in a learning setting using electrodermal activity (EDA, also galvanic skin response GSR) [10].

2.2 Measuring Affective and Cognitive States

In VR, most studies rely on subjective methods such as questionnaires to measure participants' psychological states and experiences [52]. To measure fear and distress, past VR studies used the Subjective Units of Discomfort Scale (SUDS) [35, 53, 64] or custom Likert scale questionnaires to prompt participants about their experienced fear [33]. Frustration has been assessed using emotions questionnaires such as the PAD or the Self-Assessment Manikin [6, 14]. Others have reused items from questionnaires such as the NASA task load index for measuring workload [18] or the User Engagement Scale (UES), which measures users' engagement in an application [42]. Previous studies have also assessed frustration through customized Likert scales [19, 54] or asked participants to self-report moments of frustration through button presses [26]. Since moments of insight are difficult to trigger and identify, past studies have simply asked participants to self-report when they experienced a moment of insight [10, 60].

While questionnaires are well established, they are typically presented *post-hoc* after each study condition or after the whole experiment [59]. Questionnaires are thus restricted by human language and can be biased since answers result from what participants believe they have felt or perceived in *retrospect*. Questionnaires thus cannot capture the nuances in the modulation of users' psychological states during VR use, as self-report measures cannot be assessed at high frequency.

All these reasons have motivated researchers to use physiological signals to detect changes in the Autonomic Nervous Systems (ANS) and the Central Nervous Systems (CNS) to evaluate VR experiences [13]. Since physiological responses follow psychological processes, researchers have hypothesized that they could be linked back to psychological states through characteristic features [13]. Thus, the impact of perceived emotion and cognition on physiological changes in VR experiences has been extensively studied in the past [35, 38, 57]. In particular, fear in VR has been found to affect heart rate and skin conductance [45]. Fear and frustration are closely linked to arousal [51], which impacts cardiac activity [37], electroencephalogram (EEG) signals [37], and EDA [38]. Outside VR, frustration was shown to affect HRV [61], EDA [26, 48], pupillometry, and behavioral data [26, 61]. Frustration was studied in mostly static settings [26, 61] and moments of insight were mainly studied using brain-computer interfaces [60]. The latter studies showed that a moment of insight was associated with a burst in the γ-band power activity 300 ms before the button press signaling that a solution was derived [30]. In VR, Collins et al. [10] predicted "Aha! Moments" through EDA signals while participants solved a hypercube puzzle, achieving a prediction accuracy up to 98.81% (train and tested using pooled data from all users).

As mentioned above, fear in VR has been explored in clinical and entertainment settings. However, despite the common terminology of *fear*, the stimuli used to induce fear in these two different settings highly differ. Further, despite the fact that physiological sensors such as eye-trackers are increasingly directly embedded into VR headsets (e.g., HTC Vive Pro Eye, Varjo VR-3), the impact of different types of fear on the pupil diameter and on the respiration rate in VR has not been explored yet. These emerging signal modalities are promising as users tend to be highly active in immersive virtual environments, which introduces motion artifacts in epidermal and cardiac activity measurements. Frustration and insight have rarely been measured in VR. Yet, frustration as a state is particularly interesting to measure in the context of interaction design or cyberness—issues that affect VR technologies as they aim to gain wider adoption. However, it is unclear how frustration manifests itself in users' physiological responses or how it may be detected automatically inside interactive and immersive settings. Likewise, frustration and fear are closely situated in the arousal and valence emotional space [51]. Differentiating physiological responses to both moments is also relevant in this regard, particularly due to the frequent deliberate use of fear in VR experiments. Similarly, moments of insight have not been studied in representative VR conditions, which would require unencumbered use and thus wearable physiological sensors despite the growing interest in improving productivity, creativity, and learning retention using VR [30]. Therefore, the study in this paper fills this gap and investigates how fear, frustration, and insight differ in the physiological responses they elicit as well as how accurately they can be detected using wearable sensors.

3 Experiment

The goal of this experiment was to quantify the impact of acute fear, frustration, and insight events on the dynamics of physiological signals using non-invasive sensors. These insights facilitated passively recognizing such events through classification models in a second step.

3.1 Environments

In total, we designed and verified five VEs to elicit specific transient psychological states (Fig. 1). We piloted the experiment and iterated on the designs of the environments with 3 participants before the study to validate the stimuli. The controls and interactions were explained to the participants before each game. During the experiment, participants were totally immersed in the VE and when they needed guidance or help, the experimenter could display messages through appearing notes in the VE, not to break their presence [56].

VE1: In the **horror game**, participants investigated and had to escape from a haunted house. They could freely teleport in the virtual mansion to a variety of highlighted locations and interact with virtual objects using controllers. A variety of events occurred when participants teleported or interacted with virtual objects: a rat passed by in front of them, a zombie flashed by in front of them and turned off the light, slammed the door in front of them, or ran towards the participants. After finishing the game, participants reported their anxiety level using the SUDS [42]. When designing the environment, we paid special attention to audio-visual details to engage users (e.g., flickering lights, panicking respiration sounds, sound effects with interactions).

VE2: The **vertigo scene** replicated the virtual pit experiment [38] using an elevator platform at various heights (similar to Seinfeld et al.'s study [53]). Participants stood in a virtual elevator, from which a plank led outside into a city environment. For this environment, the experimenter placed a physical plank on the floor, calibrated to match the position and size of the virtual plank to provide haptic cues while walking. The elevator was completely opaque to avoid motion sickness during vertical motion, and sound effects were added (doors, elevator going up, wind). On each floor level, participants' task was to exit the elevator and walk on the plank as far as possible, look around, and return back into the elevator after. This brought up a questionnaire where they rated their anxiety level, after which the elevator advanced to the next floor. Participants walked on the plank at four elevations (in the order: 0 m, 50 m, 100 m, and 150 m).

VE3: In the **archery game**, participants used a bow and arrows to shoot balloons that spawned and slowly rose. The environment elicited punctual and acute moments of frustration. As participants increased their score of shot balloons, increasingly many balloons failed to pop or arrows 'accidentally' dropped before launch. The environment started with a training phase where participants repeatedly shot arrows at a bulls-eye to practice the interaction and then shot eight balloons. The game behaved normally, so participants could gain confidence in the interaction, and their own performance to score. However, the trial phase was, then, rigged. Participants' goal was to shoot 30 balloons. After participants had reached a score of 5/30, the game applied a 50 % chance for an arrow to fail popping a balloon (in this case, the balloon would just be pushed away) or drop down before launching. To increase pressure, a timer counted down from 120 s and another counter showed the number of remaining balloons that would spawn, which we limited to 40. If participants still reached a score of 29/30, all remaining shots were rigged until their trial timed out. A final "Game Over" was displayed in front of them at the end of the timer.

VE4: In the **cards game**, participants' task was to assemble a tower of cards from a deck of cards, which would collapse to elicit moments of frustration. Using the controllers, participants grabbed cards with either hand and carefully placed them to build the tower. Transparently shown guides helped snap cards into place to create a solid structure. Assembling the tower required nine steps for completion as shown in Figure 1. To make the game challenging, the tower would collapse when a card grabbed by participants accidentally collided with already placed cards. To ensure frustration near completion, the game ensured that the tower always collapsed when placing the last card. When the tower collapsed, the message "Oh no! The tower collapsed. Please try again." reinforced the frustration. In total, participants had three attempts. A counter displayed their progress. After the last attempt, a final "Game Over" message announced the end of the game.

VE5: In the **escape room game**, participants' task was to find a way out of a prison cell by discovering cues that would trigger moments of insight. To move around the cell, participants could use the controllers to freely teleport around. They needed to find two codes to open two doors in the room. The first code became legible when participants happened to be at the right location in the cell, looking towards the wall, such that the bars occluded parts of the wall painting and revealed the code. The second code was encoded as the sequence of flashes from the flickering exit sign above of the second door. Counting the number of flickers between breaks revealed the three-digit code.

We designed the puzzle to be challenging and ensured that participants would not find them right away through piloting. Before attempting the escape room, participants were instructed to pull the trigger button whenever they felt that they just had an insight or thought they knew how to advance. If they pulled it by accident, they were instructed to report this, so the experimenter could void the event. In addition, the experimenter labeled moments if participants verbally expressed statements such as "Aah!" or "Oh!" when gazing at the symbols or at the flickering exit sign.

The experiment followed a within-subject design. All users experienced all five VEs designed to elicit different psychological states. The order of the exposure to the environments was randomized to avoid order effects. Each environment was preceded by a baseline phase during which users were asked to relax (first baseline: 1 min, following baselines: 30 s) and was followed by a pause where the participants took off the headset.

3.2 Dependent variables and collected data

The variables we assessed were: events, self-report measures, and physiological measures.

Events: Each environment produced transient moments of *fear*, *frustration*, or *insight* through the events listed in Table 1. Our environment logged these events with timestamps for each participant. The vertigo scene logged the moment of farthest position on the plank as the event.

State | Label	VE	Events
Fear	VE1	Rat jumpscare, zombie jumpscare 1, zombie jumpscare 2, zombie jumpscare 3
	VE2	Elevator floor 50, elevator floor 100, elevator floor 150
Frustration	VE3	Missed shot, fallen arrow
	VE4	Tower collapsed following user input, tower collapsed under system control
Insight	VE5	Moments of insight experienced by the user

Table 1: Affective and cognitive stimuli logged in our experiment.

Self-reports: To assess *fear* in VE1 and VE2, participants filled out the Subjective Units of Discomfort Scale (SUDS) [64] as used to measure fear, discomfort, and distress in previous VR studies [53]. The scale measures changes related to anxiety levels from 0 (absolute relaxation) to 100 (worst anxiety experienced). In VE1, participants reported their discomfort level once at the end of the horror game inside VR. In VE2, participants reported their discomfort level each time they returned to the elevator on each floor.

To measure *frustration* in VE3 and VE4, participants filled out the User Engagement Scale (UES) [42] on a laptop after completing each scene. One item on the perceived usability subscale especially targets frustration: "I felt frustrated while using this application."

For capturing *insight* in VE5, participants report moments of insight by pulling the trigger on any of the VR controllers while in the game.

Physiological measure: Our apparatus collected participants' cardiac activity through a photoplethysmography sensor (PPG, capturing blood volume changes at 128 Hz to detect heartbeats and heart rate), respiratory activity through breathing (sampling a breathing belt at 20 Hz to obtain respiratory rate), autonomic response from skin conductance (EDA at 128 Hz, which measures emotional arousal), and pupillometry (from eye tracking sampled at 120 Hz). All signals were logged with Unix timestamps.

3.3 Physiological data processing

Table 2 lists the features we extracted from the physiological signals recorded during the study. We extracted them from a 30 s window centered on each labeled event.

Signal modality and extracted features
PPG: Beats per minute (bpm), mean inter-beat interval (ibi), median absolute deviation of intervals between heart beats (mad), standard deviation of intervals between heart beats (sdnn), root mean square of successive differences between neighboring heartbeat intervals (rmssd), standard deviation of successive differences between neighboring heartbeat intervals (sdsd), proportion of differences between successive heart beats greater than 50 ms and 20 ms (pnn50, pnn20), sd1, sd2, s, sd1/sd2
EDA: number of peaks, {mean, min, max, standard deviation} amplitude of the peaks, {mean, min, max, standard deviation} rise time to the peaks, and {mean, min, max, standard deviation} recovery time after the peaks
Pupillometry: *(averaged from both eyes)* Number of blinks, {mean, min, max, standard deviation} of the filtered pupil diameter
Respiration: Respiratory rate (bpm)

Table 2: Extracted physiological features from the different sensors used in this study.

PPG: Before feature extraction, we band-pass processed the raw signal (0.33–5 Hz, 900^{th} order FIR filter) to remove drift and higher-frequency artefacts. Table 2 shows the 13 time-domain features we extracted, mainly HRV features, including four non-linear measures derived from Poincaré plot analysis (sd1, sd2, s, and sd1/sd2) using Python's HeartPy package. To ensure proper signal quality, we inspected the PPG signal for each event and excluded events that showed unintelligible signals from the analysis.

EDA: For each window, we filtered the raw EDA signal using a high-pass filter (2^{nd} order Butterworth with a 1 Hz cutoff) and standardized the signal. The phasic component was obtained by applying a high-pass filter (2^{nd} order Butterworth with a .5 Hz cutoff), and the tonic component was obtained by subtracting the phasic part from the filtered signal. The tonic part represents slow changes (Skin Conductance Levels, SCL), while the phasic part corresponds to the rapid responses to a stimulus (Skin Conductance Responses, SCR) [5]. SCR are characterized by their peaks through features such as count, rise time, and recovery time. Table 2 shows the features we extracted using Python's neurokit package. Again, we discarded windows of EDA signals that exhibited one or more artifacts (i.e., steep drop or constant signal portion) through manual inspection.

Pupillometry: From the logged signals (each eye), we extracted the number of blinks and the pupil diameter, filtered by subtracting the pupil diameter affected by the brightness to the measured pupil diameter for each participant [8]. To ensure proper calculations, we recorded a baseline for each participant where the scene progressed from complete darkness to an average brightness level of 0.8 in 9 steps to determine the relationship between brightness and participants' pupil diameters [8]. For all further analysis, we assessed the scene's current brightness alongside the logged raw eye data.

We modeled the effect of scene brightness on pupil diameter through tanh (the relation between luminance and pupil diameter was previously established [40]): $pupild(b) = a_1 - a_2 * \tanh(a_3 * \log(b) + a_4)$, where $pupild$ is the pupil diameter and b is the VE brightness level. We computed the coefficients a_1, a_2, a_3, a_4 for each eye and participant using the data collected during the baseline (the supplementary material shows an example of fitting for one participant).

We extracted the number of blinks from raw pupil diameters using Hershman's approach [20]. All analyses in the remainder of this paper exclusively operate on filtered pupil diameters. We extracted features from the left and the right eyes separately and averaged them (Table 2).

Respiration: We derived the respiratory rate from the force sensor contained in the belt using standard low-pass filtering and averaging peak-to-peak intervals.

3.4 Apparatus

The experiment was conducted in a room of 5 m × 5 m. Strict safety and sanitary protocols were followed to ensure COVID-19 compliance according to the local regulations, including wiping and sanitizing the headset, controllers, and sensors before and after each participant, using masks and extra VR HMD shields.

Participants were equipped with an HTC Vive Pro Eye, a Vernier GoDirect respiration belt, and the Shimmer3 GSR+ sensors for PPG and EDA measurements (Fig. 2). The respiration belt was placed on the participant's rib cage above their clothes with the sensor at their chest. As shown in Figure 2, participants wore the Shimmer wristband on their non-dominant wrists, the EDA sensors on the middle phalanx of their middle and ring fingers, and had the PPG sensor clipped to their earlobes on the non-dominant hand's side. The eye-tracker data was collected through the VR headset cable, and the respiration belt and the shimmer sensors reported signals through Bluetooth. The computer running the VR applications integrated all signals and logged them with Unix timestamps. The wire of the headset passed through a ring suspended from the ceiling, length adjusted to match the participant's height. A physical shelf served as the plank (275 mm × 115 mm × 20 mm), which the experimenter placed on the floor and calibrated to the virtual pit in VE2 to match the position of the virtual plank (Fig. 2). The plank was removed for all other environments.

Participants used both HTC Vive controllers in all environments. We developed the frontend in Unity 3D. The experience ran on an Intel Core i7-9700K CPU 3.90 Hz computer with 32 GB of RAM, supported by an NVIDIA GeForce RTX 3070 GPU.

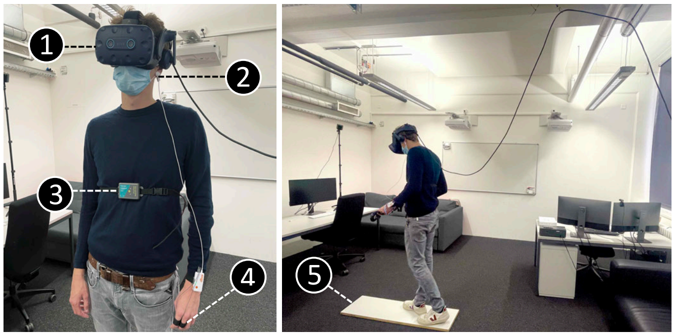


Fig. 2: Study apparatus. Participants wore a (1) VR headset that included sensors for pupillary activity, (2) a photoplethysmography ear clip to assess cardiac activity, (3) a respiration belt, and (4) electrodermal activity sensors mounted on the fingers. Participants performed tasks in an office space. (5) A physical plank was placed on the floor for the vertigo environment (VE2).

3.5 Experimental Procedure

The experiment consisted of five parts, lasting around 70 min total:

1) Written consent and instructions. Participants filled out a consent form and signed a COVID-19 statement prior to the experiment. The experimenter then introduced them to the experiment, the equipment that was involved, and the data we recorded (which was anonymized). Before starting, participants filled out questionnaires to assess their experience with VR and video games, their susceptibility to motion sickness (MSSQ [15]), demographic information, background, profile, trait anxiety (STAI-T [59]), and level of alertness [21].

2) Equipment setup. The experimenter equipped participants with the respiration belt and the Shimmer sensors and adjusted the length of the suspended headset cable to allow free motion and avoid breaks in presence [58]. Before each environment, participants received instructions about how to navigate the following environment and how they could use the controllers. For the vertigo scene, the experimenter placed and calibrated the plank. For the escape game, participants were told that they would receive hints in the scene if needed.

3) Baseline. Before interacting in each environment, participants were exposed to an empty baseline scene and instructed to stand still and relax for pupillometry calibration. Before the first environment, we recorded data for one minute, progressively switching the scene from dark to bright, whereas the subsequent baselines lasted for 30 s.

4) Experiment. Interaction in a scene immediately started following the baseline recording; participants did not take off the VR headset before. Once they had completed the tasks inside VE1, VE3, and VE4, they took off the headset and filled out questionnaires on a laptop. After the vertigo scene, the experimenter removed the plank.

5) Debriefing. After the final scene, participants were encouraged to report thoughts and feedback they may have using the laptop. A final debriefing informed them about the rigged archery game and tower of cards to elicit frustration.

3.6 Participants

The study was approved by the local committee responsible for the ethical conduct of studies. The inclusion criteria for participants were: the participants had to be 18–70 years old. They should neither have had any COVID-19 symptoms nor have been in contact with confirmed cases in the previous 14 days. They should not have known health-related problems, physical disability, clinical acrophobia, clinical anxiety, be clinically autistic, or take drugs. They had to understand and speak English. During recruiting, participants were informed that the VEs included a horror game, a vertigo experience, puzzle games, and that their physiological signals would be recorded. Participants were not informed about the environment's purpose of eliciting psychological states or that they may be rigged to elicit frustration.

24 healthy participants from the higher education institutions in our city volunteered to take part in the study (7 female, ages 21–33, $M = 27.1$, $SD = 3.1$). They originated from 15 distinct countries. Nine participants wore glasses, 20 subjects reported having had little exposure to VR technology ("none or occasionally"), and five reported that they regularly played video games. No participant reported elevated susceptibility to motion sickness on the MSSQ before the experiment. Likewise, no participant reported having experienced any symptoms of motion sickness during or following the study. Using the STAI-T, participants reported an average trait anxieties (tendency to be generally anxious) score of 45.88 ($SD = 3.78$). The scores range from 20 to 80 (higher scores correlate with greater anxiety). Finally, in their ratings on the Stanford Sleepiness Scale before the experiment [21], 20 participants rated their state of alertness as "Awake, but relaxed" and four as "Active, vital, and alert."

3.7 Results: Subjective ratings

We now summarize participants' subjective ratings (i.e., SUDS and frustration level), their qualitative feedback, and our observations during the user study. Figures 3 and 4 illustrate participants' self-reported measures from the questionnaires.

Elicitation of fear

Overall, participants reported a mean score of 46 of 100 ($SD = 26$) on the SUDS at the end of the horror game (Fig. 3). Six participants reported a score higher than 90, and two reported a score of 77. The scores confirm that the horror game induced mitigated reactions in participants. From our observations of the horror environment, fear elicited in participants varied in extent: some screamed, jumped, cursed in their native language, and some even asked the experimenter how long the game would still last as they considered stopping the experiment because it was too scary. Others just walked through the mansion and did not show any sign of fear.

In the vertigo scene, participants reported a mean SUDS score of 36.5 ($SD = 20.5$) across all floors above ground. One participant's average was above 80 and two more had an average higher than 60. The higher the elevator went, the more anxious participants rated themselves (Fig. 3). Again, we observed varied responses to the elevator experience. Some participants took no more than two steps on the plank and their knees were shaking. Others seemed at ease and just walked on the plank without hesitation.

Elicitation of frustration

For the archery and the tower of cards environments, participants' goal was to best complete the task. After the archery environment, they reported a median frustration rating of 3.5 of 5 (Fig. 4) on the UES PU subscale [42]. We observed that many participants openly expressed their frustration, shouting "Hey!", "What is going on?!", "Gosh!", "This is a disaster!", and "Oh no!", especially when arrows dropped and when the game-over screen appeared.

After the tower of cards environment, participants reported a median frustration score of 3 of 5 (Fig. 4). Across their three attempts, the tower collapsed 29 times due to their own collision with placed cards and 46 times because they had reached the last two cards and our system rigged the game. In the latter case, most expressed their frustration through "Come on!", "Oh, why??", "What the...", "No, what?"

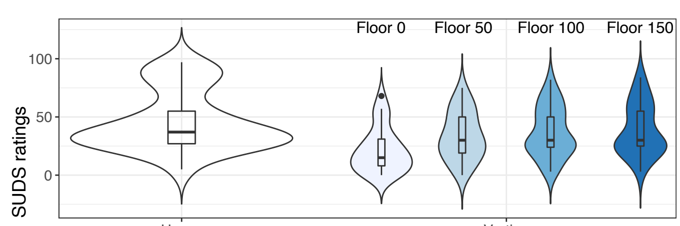


Fig. 3: Reported subjective anxiety levels in the horror (VE1) and vertigo (VE2) environments (Subjective Units of Discomfort Scale [64]).

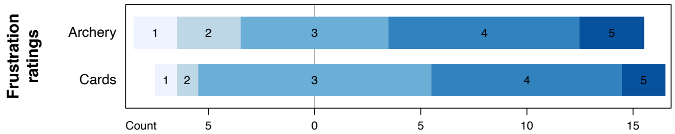


Fig. 4: Reported subjective frustration levels after the archery (VE3) and tower of cards (VE4) environments (User Engagement Scale [42]).

Elicitation of Insight

In total, participants triggered the insight button 44 times ($M = 1.8$ times per game-play, $SD = 1.7$). Of the 24 participants, 19 expressed at least once a moment of insight. Two did not need a hint from the experimenter for the first code (wall painting), while four did not need hints for the second (blinking exit sign). Afterward, most participants mentioned that the puzzles were particularly difficult, but found satisfaction when they figured them out.

4 Characterization of Physiological Responses

We now analyze the physiological responses to the transient moments of fear, frustration, and insight. We first compare signal behavior with participants' baselines. Then, we compare the differences in responses to fear and frustration and investigate the differences between fear in the horror game and fear of heights. Finally, we examine more finely the correlation between the reported anxiety level (i.e., SUDS score) in the fear environments, the frustration level (from the UES), and their physiological responses.

4.1 Impact of transient states on physiological behavior

Comparing participants' physiological responses to elicited psychological states to their baseline behavior, we first extracted instances of such moments above a certain magnitude. To analyze the effect of fear, frustration, and insight, we respectively only considered subsamples of participants who experienced fear, i.e., this resulted in all *fear* events from 14 participants who rated fear ≥ 40 of 100 on the SUDS in the horror or the vertigo environment, in all *frustration* events from 23 participants who rated frustration ≥ 3 of 5 in the archery or the cards game, and in all *insight* events from 19 participants who self-reported at least one moment of insight (Table 3). We chose the respective thresholds for fear and frustration as the rounded mean between the mean scores reported in the two related VEs.

Table 3 summarizes our results (please see the supplementary material for all values).

We summarize the procedure here for clarity and ran the same analysis in Sections 4.1, 4.2, and 4.3. We separately performed statistical tests by pair (i.e., fear vs. baseline, frustration vs. baseline, insight vs. baseline), considering affective event (i.e., baseline, fear, frustration, insight) a within-subject factor and participant a blocking factor in each test. For all dependent variables, we performed Shapiro-Wilk tests to verify if the variables followed a normal distribution. Because all tests showed positive results, we performed Wilcoxon signed-rank tests and report p values and effect sizes r.

The effect sizes were largest for pupillometry ($r_{max} = .72$) and respiration metrics ($r_{max} = .64$), followed by EDA ($r_{max} = .36$) and PPG metrics ($r_{max} = .25$). This suggests a stronger relationship between the states events and participants' eye behavior and respiratory activity.

Overall, our analysis shows that all events increased participants' bpm, pupil max diameter, and respiration rate. All events also impacted the variation of the pupil diameter (min, max, and/or SD of the pupil diameter) and the skin conductance response (SCR) (i.e., more peaks). While participants were asked to relax during the baseline, a few peaks were still detected in the SCR during the baseline.

Frustration and insight increased the mean pupil diameter. Fear decreased it. Both fear and insight increased the variation of the pupil diameter (i.e., lower min, higher max, higher SD), while frustration significantly decreased it (i.e., higher min, higher max lower SD) and decreased participants' blinks.

Fear and frustration both tended to increase HRV and tended to increase the SCR peaks amplitude.

Modality	Feature	B → Fe $N = 14$	B → Fr $N = 23$	B → I $N = 19$	Fe → Fr $N = 13$	Fe_V → Fe_H $N = 13$
PPG	bpm	↑***	↑ **	↑*	(↑)	(↓)
	HRV	↓ +	↓ **	(↑)		↓ +
EDA	N SCR peaks		(↑)		(↑)	
	SCR peaks amplitude	↑ *	↑ +	(↑)	(↑)	(↓)
	SCR peaks rise time		(↓)			(↑)
	SCR peaks recovery time	(↑)	(↓)	(↑)	(↓)	
Eyes	N Blinks	(↓)	↓ *		↓ *	↓ *
	Mean pupil diameter	↓ *	↑ ***	↑ ***	↑ **	↓ +
	Min pupil diameter	↓ +	↑ ***	(↓)	↑ ***	↓ *
	Max pupil diameter	↑ ****	↑ **	↑ ****	↓ ***	↑ ***
	SD pupil diameter	↑ *	↓ ****	↑ ****	↓ ***	↑ *
Respiratory	Respiration rate	↑ ****	↑ ****	↑ ***	(↑)	(↓)

Table 3: Effect of event type on physiological features. B=Baseline, Fe=Fear, Fe_V=Fear in the Vertigo environment, Fe_H=Fear in the Horror environment, Fr=Frustration, I=Insight events. Arrows indicate the relation from the first to the second type of event indicated in the columns. Example (*number of blinks* row, *Fe → Fr* column): ↓* means that the number of blinks is significantly lower for frustration events than the number of blinks for fear events. Parentheses indicate not-significant effects (Wilcoxon's $r \geq .1$). Significances: $^{+}p \leq .1$, $^{*}p \leq .1$, $^{**}p \leq .01$, $^{***}p \leq .001$, $^{****}p \leq .0001$. Exact values are given in the supplementary material.

4.2 Physiological responses: fear vs. frustration

Table 3 summarizes our analysis of the difference between physiological responses to fear and frustration. We followed our prior analysis method, considering event type (fear or frustration) a within-subject factor and participant a blocking factor. We only analyzed events from the 13 participants who reported both a high anxiety rating (SUDS score) and a high frustration rating in at least one of the environments.

Overall, moments of frustration caused participants to blink less, increased their mean pupil diameters, and reduced variation in pupillary activity compared to moments of fear. Fear decreased participants' mean pupil diameter and increased pupillary activity over a larger amplitude compared to moments of frustration.

We found no significant differences that related to participants' cardiac activity and EDA to elicited transient moments.

4.3 Physiological responses: fear/horror vs. fear/vertigo

To analyze the differences between participants' physiological responses to different types of fear across VE1 and VE2, we extracted all events from the 13 participants who reported a high level of anxiety in both environments. We considered environment (i.e., horror or vertigo environment) a within-subject factor and participant a blocking factor. Table 3 reports our results.

Overall, the difference between physiological responses to horror events and vertigo events manifested mostly in participants' pupillary responses. Following moments of fear in the horror game, they blinked significantly less, their pupil diameter was significantly lower, and their amplitudes varied more than following moments of fear of vertigo.

Moments of fear in the horror game also led to lower median HRV than when following moments of fear of vertigo ($r = .3$, $p = .08$).

4.4 Correlating subjective ratings and physio. responses

Above, we compared transient states to identify statistical differences between them. We now investigate the relationship between levels of fear, levels of frustration, and the physiological responses expressed by participants and analyze the correlations between them.

To account for variability across participants, we analyzed the percentage change of all metrics from each participant's baseline rather than absolute values with the exception of pupil diameter (because it is already adjusted per participant). We normalized all metrics following Wiederhold et al.'s adapted approach [63]: $percentage\ change = 100 \times [(value - baseline)/baseline]$. All events from all 24 participants were included in this analysis.

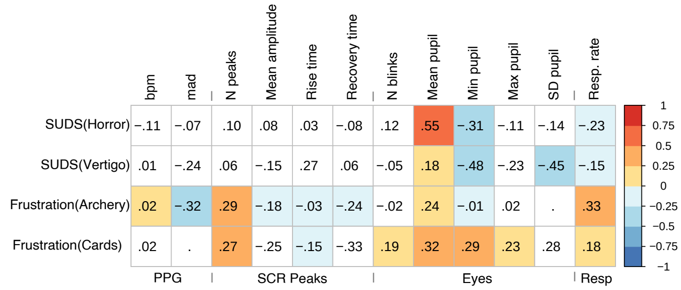


Fig. 5: Spearman's ρ correlations between the subjective ratings and the normalized physiological features. The discomfort level was assessed using the Subjective Units of Discomfort Scale (SUDS) [64] (horror and vertigo scenes). Frustration was reported through the User Engagement Scale (UES) short-form [42] (archery and cards scenes). Only significant correlations are colored ($p \leq .05$).

Using Spearman's ρ coefficient, we evaluated the zero-order correlation between subjective scores (i.e., SUDS and UES) and the percentage change of physiological features around fear and frustration events. Because we could not quantify the magnitude of insights, we limited this analysis to fear and frustration events. Figure 5 summarizes the correlations we obtained. For the sake of readability, we use "increase" and "decrease" to denote positive or negative correlations below.

In summary, the more anxious or frustrated participants were, the more their HRV decreased (significant for archery), and the more their mean pupil diameter increased (all $p \leq .05$, all $\rho \geq .018$).

Higher frustration levels significantly increased the number of peaks in the SCR (cards, archery), increased the respiration rate (cards, archery), decreased the SCR peaks amplitudes (archery), shortened the time it took for peaks to rise (cards, archery), and shortened the time it took for SCR peaks to recover (archery). Some of these correlations are not significant in the cards game, but the correlation coefficient all follow the same direction.

Higher fear levels significantly decreased the minimum pupil diameter and decreased the SD of the pupil diameter in the vertigo scene. Contrary to previous results shown in Table 3, higher fear levels also significantly decreased the respiration rate.

We also observed differences in trends between the frustration and fear responses. Higher frustration significantly increased the respiration rate in both the archery and cards came. On the contrary, higher fear levels significantly decreased the respiration rate in the horror game and vertigo environment. Increased frustration levels in the cards game led to increased min and max pupil diameters. Increasing fear levels led to decreased min and max pupil diameters. Increased frustration significantly shortened the SCR peaks rise time in the cards game. Fear, however, tended to elongate the rise time of SCR peaks in the vertigo scene (not significant).

Among other significant correlations, frustration level in the cards game is positively correlated with the number of blinks of participants while previous analyzes showed an overall decrease in the number of blinks due to frustration events.

4.5 Results summary

Table 4 lists the effects of fear, frustration, and insight on heart rate, HRV, SCR peaks, the number of blinks, pupillary activity, and respiration rate. These effects are illustrated based on the previous findings, favouring in the order: significant correlations $p \leq .05$, significant effects $p \leq .05$, non-significant correlations with Spearman's $\rho > .015$, non-significant effects significant effects with Wilcoxon's $r > .1$.

The SCR peaks duration account for the SCR peaks mean rise time and mean recovery time.

5 Detecting moments of fear, frustration, insight

Having analyzed the characteristic impact of affective and cognitive states on participants' physiology in an immersive environment, we now investigate the stability of the main features we identified in the

	Physiological responses							
	HRV	SCR peaks			num	pupil diameter		resp
		num	amplitude	duration	blinks	mean	sd	rate
Fear								
Horror	↓*			↑	↑	↑*	↑	↓*
Vertigo							↓*	
Frustration								
Archery	↓*	↑*	↓*	↓*		↑*	↓*	↑*
Cards					↑*			
Insight	↑		↑	↑		↑*	↑*	↑*

Table 4: Summary of effects of fear, frustration, and insight on the physiological responses. The arrows indicate how transient states affect features, based on the correlation analysis, then based on the statistical differences reported. * indicates that at least one of the effects or correlations was consistently statistically significant across the different analyses ($p \leq .05$).

previous section for automatically detecting such transient states. For this, we compare the performance of three learning-based techniques that we trained on the physiological features in Table 2 supervised by the labels in Table 1, both of which we collected in our study.

5.1 Data selection

We trained separate binary classifiers to determine the presence of fear (both and each horror and vertigo), frustration (both archery and cards), and insight in a person's continuous physiological signals (i.e., five classifiers). All signals were normalized by subtracting the features extracted from the baseline from each value, as recommended in previous work [34]. As described in Section 4.1, we extracted valid *fear* events from participants who had rated their anxiety level above 40 in the horror or the vertigo environment and we extracted valid *frustration* events from participants who had rated their frustration level above 3 in the archery or the cards game. We also extracted all *insight* events. In total, this amounted to 509 frustration samples, 88 fear in horror game samples, 42 fear of vertigo samples, and 43 insight samples.

For absence labels, we sourced a matching number of 'none' events for each type of event, balanced to match event occurrences across scenes and participants (e.g., for 509 frustration samples, we selected 509 'none' samples from the same VE). We picked random moments from related VEs (e.g., 'none' events from archery) where participants' physiological signals indicated activity, but with a minimum offset of 30 s to a proper event that occurred in the scene. We also padded each selected sample with a buffer of 10 s before and after to prevent selecting another sample from this region.

5.2 Classifiers

For each event type, we trained three supervised machine learning classifiers: 1) logistic regression, 2) linear support vector machine (SVM), and 3) a random forest (1000 estimators and a max depth of 5). SVMs and random forests have been frequently used in past VR studies to classify affective and cognitive states in VR [35] as well as for classifying physiological signals more generally [41]. In our case, linear SVMs outperformed those with RBF kernels. As a third classification method, we chose logistic regressions, because our ratings were ordinal data.

We trained each classifier using the features listed in Table 2, separately calculating the features from 21 sliding 30-second windows, starting in the period of $[-15..5]$ s (step size = 1 s), where the event occurred at 0 s. For comparison, we trained separate classifiers with all the features from subsets of all sensors. We tuned the hyperparameters using random search.

5.3 Results

To assess our classifiers, we used the micro F_1 score that represents the harmonic mean between precision and recall on the positive class (as we are more interested in positive classes). We evaluated our trained models using leave-one-subject-out cross-validation, holding out all data from one participant per fold for testing, and averaging all N

State	Algo.	Modalities	F_1 score
Fear—Horror	*log*	PPG, Eyes	71 %
Fear—Vertigo	*log*	PPG, EDA, Eyes	75 %
Fear	*lsvc*	PPG, EDA, Resp	66 %
Frustration	*lsvc*	PPG, Eyes, Resp	76 %
Insight	*rf*	PPG, EDA, Eyes, Resp	75 %

Table 5: Our classifiers' F_1 scores (following leave-one participant out evaluation). We filtered samples by SUDS and frustration score. Classifiers: *log* = Logistic Regression, *lsvc* = Linear Support Vector Classifier, *rf* = Random Forest.

models' F_1 scores. Table 5 lists the results of these user-independent models for the best-performing models. Please see the supplementary material for all sensor combinations and resulting F_1 scores.

As evident in the table, all respectively best trained models achieved an F_1 score of above 66 %, though using a different subset of sensors and thus features as input. This shows the promise of using passively collected physiological signals to detect the presence of transient moments of fear, frustration, or insight for future applications. Interestingly, apart from detecting moments of insights, all other best performing classifiers are linear.

6 Discussion

6.1 Physiological responses to fear

In this section, we consider fear ratings independent of the VE (horror game and vertigo environments).

Moments of fear led to participants' pupils varying over larger amplitudes and decreased their heart rate variability (HRV). These observations match the findings of previous studies on fear and anxiety [8]. Moments of fear also increased the peak amplitudes of electrodermal activity responses (SCR), which is consistent with past findings about the activation of the SCR due to stress and arousal [35]. Fear also correlated with pupil diameter increase and pupillary variation in our study. This corroborates past observations of increasing pupil diameter in response to negative stimuli [1].

Our correlation analysis also showed that the higher the reported anxiety level was, the lower their respiratory rate was. Prior studies have linked fear with an increase in respiratory rate [31]. However, our results are more consistent with Blatz's findings that imminent threat (sudden backward-tilting chair) decreases heart rate and respiratory rate [4]. This difference may arise from the fact that many previous studies have compared the physiological responses to an active condition with physiological behavior during rest. Our results also show this contrast: all five VEs increased participants' respiratory rates and heart rates compared to the baseline, likely a result of physical activity and not of elicited states. Another explanation for the decreasing respiratory rates we observed may be the nature of the stimuli for fear. Previous studies have analyzed the *overall* fear experienced during *long* conditions [16, 36, 45], whereas we focused on short moments and emotionally intense affective events.

Alternatively, the emotional outburst experienced by some participants following a jumpscare could have cut their breath short, thus decreasing their respiratory rate (and heart rate) for a longer period than for those who experienced less intense fear.

6.2 Physiological responses to frustration

In our study, frustrating events caused participants' HRV to decrease, their blink rate to drop, and their pupil diameters to increase. This is in line with prior findings (e.g., [26, 61]) that blink rate and respiratory rate yield interesting insights about the occurrence of a frustration event. The decrease in HRV and increase in SCR activity we observed can be explained by the stress or emotional arousal reaction to frustration events (e.g., when arrows dropped), supporting previously reported effects of negative emotions [31].

Regarding eye activity, blink rate decreased for frustration events compared to participants' baseline, but there was a positive correlation between frustration and the number of blinks in the cards game. An explanation for this observation is that the number of blinks can be considered a behavioral measure extracted from a physiological signal (rather than an autonomic response). Thus, it may simply be a personal reaction to frustration, much like blinking more in surprise or squinting eyes in confusion. Stimuli context may be another explanation for this. During archery, participants aimed to reach a score in a given time, likely increasing their focus on the task and thus decreasing blink rate—even though the game, at times, may have appeared not to work as it should have. This would have decreased the overall blink rate despite occurrences of frustration. Finally, the decrease in blinks could also be explained by engagement. Prior work linked decreasing blink rates to sustained attention, cognitive load, and increased engagement [55]. It could be that the more participants were engaged in the game, the more frustrated they became by the unfortunate events. This explanation is supported by the previously found circular relation between emotional intensity and presence, which itself is related to engagement [49].

6.3 Physiological responses to insight

Insight is particularly difficult to elicit in a controlled manner as it requires users to discover solutions. Unlike analytic problem-solving, users cannot readily explain the exact path they followed to reach the solution [60]. Our escape game worked well to elicit insight, as participants had to discover that codes were encoded in different ways (light flashes to counts). Since the environment offered opportunities for only two such insights, we have a limited amount of data points only and will need more data to strengthen effects significance.

Few prior studies have investigated moments of insight in a physiological context. Collins et al. [10] detected moments of insight using EDA responses in a user-*dependent* model, pooling all users in the training and testing dataset, and reached an accuracy of up to 98.81 %. Our results, however, showed no significant difference between the EDA responses elicited by insight and those during the baseline, which may be due to our limited sample size.

Insight increased participants' pupil diameter, led to variations over larger amplitudes (lower min, higher max) and stronger deviations from the mean (higher SD) compared to the baseline. Previous work found insight to improve memory and creativity [11] and our results are in line with previous results that found strong links between cognitive states (e.g., cognitive load, attention) and pupil diameter [34, 44].

6.4 Difference between fear and frustration

Past work mainly aimed to classify different classes of emotional arousal and valence [35]. However, identifying the differences between close emotions situated in the same emotional space such as fear and frustration is particularly difficult and has consequences for detecting these states. We designed two environments per emotion to evaluate the differences between fear and frustration across contexts.

We found significant differences between fear and frustration in blinks and pupillary activity, as well as differences in their correlation with respiratory rate. As above, blink rate can be explained by individuals' behavioral differences.

Fear events impacted participants' pupil diameter significantly more (lower min, higher max, higher SD) than frustration events, which on the contrary provoked a decrease in the SD of participants' pupil diameters. This difference can be explained by the higher arousal level induced by fear events compared to frustration events.

More interestingly, a higher anxiety level was correlated with an increase in the average pupil diameter and a *decrease* in the minimum pupil diameter. Comparatively, frustration level was correlated with an increase in the average pupil diameter and an *increase* in the minimum pupil diameter in the cards environment. Thus, higher anxiety provoked a constriction and dilation of the pupils, while higher frustration mainly provoked a dilation of the pupils.

Anxiety level due to fear events negatively correlated with participants' respiratory rate, as opposed to frustration level. While both fear and frustration are negative emotions, fear events can be associated with imminent threats, provoking users to be more guarded, contrarily to frustration. Past work has shown that more anxious participants tend to have more tense muscles [35]. This tension could also be associated with compression of their thoracic cavity, making them retain more their breath with increasing anxiety. Higher frustration levels, however, would make individuals more active, which would increase their respiratory rate. This restlessness can also be found in the shortening of their SCR peak activity.

Indeed, our correlation analysis consistently showed that higher frustration levels increased the number of SCR peaks, decreased the SCR peak amplitudes, the peak rise times, and the peak recovery times (compared with no correlation with fear). This held true in both, the archery and the cards game. Frustration can, thus, be associated with repetitive small activations of the SCR (multiple short and low-amplitude SCR peak activation).

6.5 Difference between fear/horror vs. fear/vertigo

Despite the fact that fear is a term used under the same denomination to refer to the emotion felt when one feels in danger, we found differences between fear in a vertigo scene and fear experienced during a VR horror game physiological responses.

The correlation coefficient between the fear level and the mean pupil diameter was much higher in the horror game than in the vertigo scene (see Fig. 5). Increasing anxiety levels also decreased participants' respiratory rate more in the horror game than in the vertigo environment. These differences can be explained by the overall more intense horror stimuli compared to the vertigo stimuli.

We also found that both the horror and the vertigo environment significantly increased the SD of participants' pupil diameter compared to the baseline. However, anxiety level was negatively correlated with the SD of the pupil diameter in the vertigo environment (Table 4). This difference may stem from the difference in the acuteness of the stimuli. Horror stimuli were overall more sudden than events in the vertigo scene, which may have elicited physiological changes of higher response amplitudes.

6.6 Recognition of short affective and cognitive events

We trained separate user-independent classifiers to determine the presence of fear, frustration, and insight and obtained a micro F_1 score above 66 % by using different subsets of sensors.

Interestingly, eye features proved important in all best-performing classifiers except for recognizing context-independent fear. Our analyses showed different trends between the overall impact of fear events on the physiological responses against the baseline (i.e., increase of the pupil SD) and the correlation between anxiety level reported in fear environments and the physiological responses (i.e., negative correlation between anxiety ratings and the pupil diameter SD). The fact that eye features seemed little relevant for the classification of transient psychological states when mixing different types of fear highlights the difficulty to classify states across contexts, especially as they differ in their physiological responses despite using the same terminology.

Overall, the performances of our models are comparable to prior work (e.g., Kapoor et al. [26] classified frustration with an accuracy of 79.2 %, but used more modalities such as face recognition and pressure exerted on a mouse.) It is important to note that we evaluated our recognition models using a user-independent approach and without any context-dependent variables (e.g., no task performances or input-related features). We also had few data points in the fear and insight environments. Considering these conditions, our models performed surprisingly well compared to those in the literature [41].

For comparison, we also trained user-dependent models. They achieved average F_1 scores of up to 85 % for fear environments and 91 % for frustration. The models for insight performed poorly (64 %), because of the small amount of data points available for training within each participant. This performance is also consistent with past user-dependent classifiers [10].

Our classifiers were primarily tuned in an attempt to investigate whether the modalities and features importance in the classification corroborate our previous analyses. We believe that our results indicate a promising direction for future work to further tune hyperparameters in addition to gathering more samples to bolster detection accuracy. For example, modulating the size of the window could be useful to improve the detection of fear, frustration, and insight events in real-time.

6.7 Limitations

Our study focused on transient psychological states that can impact user experience in VR but that are particularly short in duration. Despite being categorized in the same emotional space (i.e., negative valence, positive arousal), our results shed light on the differences between fear, frustration, and insight. However, we analyzed these differences in a limited number of contexts and additional states in this emotional space should be investigated, such as anger [41] and other types of fear. We also presented only one VE on a positively valenced state, because positive responses of short duration can be particularly difficult to elicit—as we showed in the case of insight. Further investigating the physiological responses to these additional states could help better understand the relations between affective and cognitive states.

The dataset sizes and statistical power of our analyses were limited ($N \in [13, 24]$). We accounted for individual variability by normalizing physiological responses in the correlation analysis and model training, but larger sample sizes and a more diverse population are needed for generalization and to detect potential interaction effects.

We analyzed the impact of transient states and their differences using wearable sensors in interactive settings. While the EDA and PPG sensors yielded interesting insights about the effects of transient states, their signals naturally suffered more from motion artifacts than the respiration and eye tracking sensors. The resulting need for more elaborate preprocessing currently limits their readiness for use in interactive contexts. Advancing to robust but non-invasive wearable sensors will be essential to improve the analysis of human cognition and behavior using physiological signals.

Finally, while our study afforded participants free movement and interaction, our physiological recordings have no baseline that is as insightful as EEG. Future efforts could build on our elicitation design to restrict head motion and use EEG to better understand the underlying processes associated with emotion and cognition.

7 Conclusion

In this paper, we have presented the results of a user study that elicited transient psychological states that are short in duration, such as fear, frustration, and insight. Specifically, we investigated the cardiac, epidermal, pulmonary, and pupillary responses to changes in these states. 24 participants were exposed to five virtual environments that comprised different fear, frustration, and insight events.

Despite the fact that fear and frustration are situated in the same emotional space (negative valence, positive arousal), we found differences in their physiological responses: anxiety level decreased respiration rate across different fear contexts, while respiration rate increased with frustration level. Higher fear level was linked to pupil dilation and constriction, and higher frustration level produced an overall dilation of the pupil. Frustration also triggered more changes in skin conductance responses, shortening the overall phasic skin peak amplitudes and duration. High-intensity states such as fear and insight were also associated with higher amplitudes of change in participants' pupil diameters. We also found significant differences in the pupillary behavior in response to fear events in a horror game and fear events in a vertigo environment.

In a second step, we demonstrated that the behavior of the features that we selected to characterize physiological responses to these transient moments was systematic across participants and event types. Using the events and participants' physiological signals, we trained binary classifiers for detecting moments of fear, frustration, and insight in an objective, unobtrusive, and real-time manner. Our user-independent models detect these transient events with F_1 scores above 66 %. The implications of both, our results as well as our data-driven detection of transient states show the potential to aid further understanding of emotions and cognitive states in immersive environments, including in productivity, creativity, and learning scenarios using VR.

References

[1] Areej Babiker, Ibrahima Faye, Kristin Prehn, and Aamir Malik. Machine learning to differentiate between positive and negative emotions using pupil diameter. *Frontiers in psychology*, 6:1921, 2015.

[2] Domna Banakou, Sameer Kishore, and Mel Slater. Virtually Being Einstein Results in an Improvement in Cognitive Task Performance and a Decrease in Age Bias. *Frontiers in Psychology*, 9, 2018. ISSN 1664-1078. Publisher: Frontiers.

[3] Guillermo Bernal, Tao Yang, Abhinandan Jain, and Pattie Maes. PhysioHMD: a conformable, modular toolkit for collecting physiological data from head-mounted displays. In *Proceedings of the 2018 ACM International Symposium on Wearable Computers*, pages 160–167, 2018.

[4] Wm E Blatz. The cardiac, respiratory, and electrical phenomena involved in the emotion of fear. *Journal of Experimental Psychology*, 8(2):109, 1925.

[5] Wolfram Boucsein. *Electrodermal Activity*. Springer-Verlag New York Inc., 2nd ed. edition, September 2011. ISBN 978-1-4614-1125-3.

[6] Margaret M Bradley and Peter J Lang. Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry*, 25(1):49–59, 1994. Number: 1 Publisher: Pergamon.

[7] Chin-Jung Chao, Sheng-Yu Wu, Yi-Jan Yau, Weng-Yan Feng, and Feng-Yi Tseng. Effects of three-dimensional virtual reality and traditional training methods on mental workload and training performance. *Human Factors and Ergonomics in Manufacturing & Service Industries*, 27(4):187–196, 2017. Publisher: Wiley Online Library.

[8] Hao Chen, Arindam Dey, Mark Billinghurst, and Robert W Lindeman. Exploring pupil dilation in emotional virtual reality environments. 2017.

[9] Luca Chittaro and Fabio Buttussi. Assessing knowledge retention of an immersive serious game vs. a traditional education method in aviation safety. *IEEE transactions on visualization and computer graphics*, 21(4):529–538, 2015. Number: 4 Publisher: IEEE.

[10] Jonny Collins, Holger Regenbrecht, Tobias Langlotz, Yekta Said Can, Cem Ersoy, and Russell Butson. Measuring Cognitive Load and Insight: A Methodology Exemplified in a Virtual Reality Learning Context. In *2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 351–362. IEEE, 2019.

[11] Amory H Danek and Jennifer Wiley. What causes the insight memory advantage? *Cognition*, 205:104411, 2020.

[12] Paul Ekman, E Richard Sorenson, and Wallace V Friesen. Pancultural elements in facial displays of emotion. *Science*, 164 (3875):86–88, 1969. Number: 3875 Publisher: American Association for the Advancement of Science.

[13] Stephen H Fairclough. Fundamentals of physiological computing. *Interacting with computers*, 21(1-2):133–145, 2008. Number: 1-2 Publisher: Oxford University Press Oxford, UK.

[14] Yuan Gao, Nadia Bianchi-Berthouze, and Hongying Meng. What does touch tell us about emotions in touchscreen-based gameplay? *ACM Transactions on Computer-Human Interaction (TOCHI)*, 19 (4):1–30, 2012.

[15] John F. Golding. Predicting individual differences in motion sickness susceptibility by questionnaire. *Personality and Individual Differences*, 41(2):237–248, July 2006. ISSN 0191-8869.

[16] Daniel Gromer, Max Reinke, Isabel Christner, and Paul Pauli. Causal interactive links between presence and fear in virtual reality height exposure. *Frontiers in psychology*, 10:141, 2019. Publisher: Frontiers.

[17] Kurinchi Selvan Gurusamy, Rajesh Aggarwal, Latha Palanivelu, and Brian R Davidson. Virtual reality training for surgical trainees in laparoscopic surgery. *Cochrane database of systematic reviews*, (1), 2009. Number: 1 Publisher: John Wiley & Sons, Ltd.

[18] Sandra G Hart and Lowell E Staveland. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology*, volume 52, pages 139–183. Elsevier, 1988.

[19] David Heidrich, Annika Wohlan, and Meike Schaller. Perceived speed, frustration and enjoyment of interactive and passive loading scenarios in virtual reality. In *International Conference on Human-Computer Interaction*, pages 343–355. Springer, 2020.

[20] Ronen Hershman, Avishai Henik, and Noga Cohen. A novel blink detection method based on pupillometry noise. *Behavior research methods*, 50(1):107–114, 2018.

[21] E Hoddes, V Zarcone, and W Dement. Stanford sleepiness scale. *Enzyklopädie der Schlafmedizin*, 1184, 1972.

[22] Larry F Hodges, Rob Kooper, Thomas C Meyer, Barbara O Rothbaum, Dan Opdyke, Johannes J. de Graaff, James S Williford, and Max M North. Virtual environments for treating the fear of heights. *IEEE computer*, 28(7):27–34, 1995.

[23] Christian Holz and Edward J. Wang. Glabella: Continuously sensing blood pressure behavior using an unobtrusive wearable device. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 1(3), sep 2017. doi: 10.1145/3132024. URL `https://doi.org/10.1145/3132024`.

[24] Camille Jeunet, Louis Albert, Ferran Argelaguet, and Anatole Lécuyer. "do you feel in control?": towards novel approaches to characterise, manipulate and measure the sense of agency in virtual environments. *IEEE transactions on visualization and computer graphics*, 24(4):1486–1495, 2018.

[25] Craig A Kaplan and Herbert A Simon. In search of insight. *Cognitive psychology*, 22(3):374–419, 1990.

[26] Ashish Kapoor, Winslow Burleson, and Rosalind W. Picard. Automatic prediction of frustration. *International Journal of Human-Computer Studies*, 65(8):724–736, August 2007. ISSN 1071-5819.

[27] D. Keefe, D. Feliz, T. Moscovich, D. Laidlaw, and J. LaViola. CavePainting: a fully immersive 3D artistic medium and interactive experience. In *I3D '01*, 2001.

[28] Pascal Knierim, Valentin Schwind, Anna Maria Feit, Florian Nieuwenhuizen, and Niels Henze. Physical keyboards in virtual reality: Analysis of typing performance and effects of avatar hands. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–9, 2018.

[29] Agata Kolakowska, Agnieszka Landowska, Mariusz Szwoch, Wioleta Szwoch, and Michal R. Wróbel. Modeling emotions for affect-aware applications. *Information Systems Development and Applications*, pages 55–69, 2015.

[30] John Kounios and Mark Beeman. The Aha! Moment: The Cognitive Neuroscience of Insight. *Current Directions in Psychological Science*, 18(4):210–216, 2009.

[31] Sylvia D. Kreibig. Autonomic nervous system activity in emotion: a review. *Biological Psychology*, 84(3):394–421, 2010. ISSN 1873-6246. Number: 3.

[32] Jonathan Lazar, Adam Jones, and Ben Shneiderman. Workplace user frustration with computers: An exploratory investigation of the causes and severity. *Behaviour & Information Technology*, 25 (03):239–251, 2006.

[33] Jih-Hsuan Tammy Lin. Fear in virtual reality (vr): Fear elements, coping reactions, immediate and next-day fright responses toward a survival horror zombie virtual reality game. *Computers in Human Behavior*, 72:350–361, 2017.

[34] Tiffany Luong, Nicolas Martin, Anais Raison, Ferran Argelaguet, Jean-Marc Diverrez, and Anatole Lécuyer. Towards Real-Time Recognition of Users' Mental Workload Using Integrated Physiological Sensors Into a VR HMD. In *IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 1–13, Virtual, Brazil, November 2020.

[35] Tiffany Luong, Anatole Lecuyer, Nicolas Martin, and Ferran Argelaguet. A Survey on Affective and Cognitive VR. *IEEE transactions on visualization and computer graphics*, PP, September 2021. ISSN 1941-0506.

[36] Eric Malbos, Ronald M Rapee, and Manolya Kavakli. Behavioral presence test in threatening virtual environments. *Presence: Teleoperators and Virtual Environments*, 21(3):268–280, 2012. Number: 3 Publisher: MIT Press.

[37] Javier Marín-Morales, Juan Luis Higuera-Trujillo, Alberto Greco, Jaime Guixeres, Carmen Llinares, Enzo Pasquale Scilingo, Mariano Alcañiz, and Gaetano Valenza. Affective computing in virtual reality: emotion recognition from brain and heartbeat dynamics using wearable sensors. *Scientific reports*, 8(1):1–15, 2018. Number: 1 Publisher: Nature Publishing Group.

[38] Michael Meehan, Sharif Razzaque, Brent Insko, Mary Whitton, and Frederick P Brooks. Review of four studies on the use of physiological reaction as a measure of presence in stressfulvirtual environments. *Applied psychophysiology and biofeedback*, 30(3): 239–258, 2005. Number: 3 Publisher: Springer.

[39] Manuel Meier, Paul Streli, Andreas Fender, and Christian Holz. Tapld: Rapid touch interaction in virtual reality using wearable sensing. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, pages 519–528. IEEE, 2021.

[40] Parry Moon and Domina Eberle Spencer. On the stiles-crawford effect. *JOSA*, 34(6):319–329, 1944.

[41] Domen Novak, Matjaž Mihelj, and Marko Munih. A survey of methods for data fusion and system adaptation using autonomic nervous system responses in physiological computing. *Interacting with Computers*, 24(3):154–172, May 2012. ISSN 0953-5438.

[42] Heather L. O'Brien, Paul Cairns, and Mark Hall. A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form. *International Journal of Human-Computer Studies*, 112:28–39, April 2018. ISSN 1071-5819.

[43] Fernando Paladines-Jaramillo, Veronica Egas-Reyes, Diego Ordonez-Camacho, Juan Salazar, Veronica Realpe, and Ivan Terceros. Using virtual reality to detect, assess, and treat frustration. In *Conference on Information and Communication Technologies of Ecuador*, pages 377–390. Springer, 2020.

[44] Oskar Palinko, Andrew L Kun, Alexander Shyrokov, and Peter Heeman. Estimating cognitive load using remote eye tracking in a driving simulator. In *Proceedings of the 2010 symposium on eye-tracking research & applications*, pages 141–144, 2010.

[45] Federica Pallavicini, Ambra Ferrari, Alessandro Pepe, Giacomo Garcea, Andrea Zanacchi, and Fabrizia Mantovani. Effectiveness of virtual reality survival horror games for the emotional elicitation: Preliminary insights using Resident Evil 7: Biohazard. *International Conference on Universal Access in Human-Computer Interaction*, pages 87–101, 2018. Publisher: Springer.

[46] Thomas D Parsons and Christopher G Courtney. Interactions between threat and executive control in a virtual reality stroop task. *IEEE Transactions on Affective Computing*, 9(1):66–75, 2016. Number: 1 Publisher: IEEE.

[47] Thomas D Parsons and James L Reinebold. Adaptive virtual environments for neuropsychological assessment in serious games. *IEEE Transactions on Consumer Electronics*, 58(2), 2012. Number: 2 Publisher: IEEE.

[48] David Portugal, Marios Belk, João Quintas, Eleni Christodoulou, and George Samaras. Identification of an individual's frustration in the work environment through a multi-sensor computer mouse. In *International Conference on Human Aspects of IT for the Aged Population*, pages 79–88. Springer, 2016.

[49] Giuseppe Riva, Fabrizia Mantovani, Claret Samantha Capideville, Alessandra Preziosa, Francesca Morganti, Daniela Villani, Andrea Gaggioli, Cristina Botella, and Mariano Alcañiz. Affective interactions using virtual reality: the link between presence and emotions. *CyberPsychology & Behavior*, 10(1):45–56, 2007. Number: 1 Publisher: Mary Ann Liebert.

[50] Hugo Romat, Andreas Fender, Manuel Meier, and Christian Holz. Flashpen: A high-fidelity and high-precision multi-surface pen for virtual reality. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, pages 306–315. IEEE, 2021.

[51] James A Russell. A circumplex model of affect. *Journal of personality and social psychology*, 39(6):1161, 1980. Number: 6 Publisher: American Psychological Association.

[52] Valentin Schwind, Pascal Knierim, Nico Haas, and Niels Henze. Using Presence Questionnaires in Virtual Reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–12. Association for Computing Machinery, New York, NY, USA, 2019.

[53] Sofia Seinfeld, Ilias Bergstrom, Ausias Pomes, Jorge Arroyo-Palacios, Francisco Vico, Mel Slater, and Maria V Sanchez-Vives. Influence of music on anxiety induced by fear of heights in virtual reality. *Frontiers in psychology*, 6:1969, 2016. Publisher: Frontiers.

[54] Paula R Selvidge, Barbara S Chaparro, and Gregory T Bender. The world wide wait: effects of delays on user performance. *International Journal of Industrial Ergonomics*, 29(1):15–20, 2002.

[55] Greg J Siegle, Naho Ichikawa, and Stuart Steinhauer. Blink before and after you think: Blinks occur prior to and following cognitive load indexed by pupillary responses. *Psychophysiology*, 45(5): 679–687, 2008.

[56] Mel Slater, John McCarthy, and Francesco Maringelli. The influence of body movement on subjective presence in virtual environments. *Human Factors*, 40(3):469–477, 1998. Number: 3 Publisher: Sage Publications Sage CA: Los Angeles, CA.

[57] Mel Slater, Andrea Brogni, and Anthony Steed. Physiological responses to breaks in presence: A pilot study. In *Presence 2003: The 6th annual international workshop on presence*, volume 157. Citeseer, 2003.

[58] Mel Slater, David-Paul Pertaub, Chris Barker, and David M Clark. An experimental study on fear of public speaking using a virtual environment. *CyberPsychology & Behavior*, 9(5):627–633, 2006. Number: 5 Publisher: Mary Ann Liebert.

[59] Charles D Spielberger. State-Trait anxiety inventory. *The Corsini encyclopedia of psychology*, pages 1–1, 2010. Publisher: Wiley Online Library.

[60] Giulia Sprugnoli, Simone Rossi, Alexandra Emmendorfer, Alessandro Rossi, Sook-Lei Liew, Elisa Tatti, Giorgio di Lorenzo, Alvaro Pascual-Leone, and Emiliano Santarnecchi. Neural correlates of eureka moment. *Intelligence*, 62:99–118, 2017.

[61] Brandon Taylor, Anind Dey, Daniel Siewiorek, and Asim Smailagic. Using physiological sensors to detect levels of user frustration induced by system delays. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '15, page 517–528, New York, NY, USA, 2015. Association for Computing Machinery. ISBN 9781450335744.

[62] Martin Usoh, Kevin Arthur, Mary C Whitton, Rui Bastos, Anthony Steed, Mel Slater, and Frederick P Brooks Jr. Walking¿ walking-in-place¿ flying, in virtual environments. In *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, pages 359–364, 1999.

[63] Brenda K Wiederhold, Dong P Jang, Sun I Kim, and Mark D Wiederhold. Physiological monitoring as an objective tool in virtual reality therapy. *CyberPsychology & Behavior*, 5(1):77–82, 2002.

[64] Joseph Wolpe. Psychotherapy by reciprocal inhibition. *Conditional reflex: a Pavlovian journal of research & therapy*, 3(4): 234–240, 1968. Publisher: Springer.