VRFS: Multimodal Dataset and Analysis of Physiological Responses in Virtual Reality and Flatscreen Gaming Environments

ANONYMOUS AUTHOR(S)

Recent research has focused on the effectiveness of Virtual Reality (VR) in games as a more immersive method of interaction. However, there is a lack of systematic comparison of physiological effects between VR and flatscreen (FS) gaming. This paper introduces a multimodal affective dataset named the **V**irtual **R**eality and **F**lat**S**creen Dataset (VRFS), which is the first multimodal dataset of emotional and physiological responses to commercially available games in VR and FS environments. To create the VRFS dataset, we first selected four games through a pilot study of 6 participants to cover all four quadrants of the valence-arousal space. Using these games, we recorded the physiological activity, including Blood Volume Pulse and Electrodermal Activity, and self-reported emotions of 33 participants in a user study. Our data analysis revealed that VR gaming elicited more pronounced emotions, higher arousal, increased cognitive load and stress, and lower dominance than FS gaming. The entire VRFS dataset, containing over 15 hours of multimodal data comparing FS and VR gaming across different games, is publicly available for research purposes. This dataset provides a valuable resource for further investigating the physiological and emotional effects of VR and FS gaming.

ACM Reference Format:
Anonymous Author(s). 2022. VRFS: Multimodal Dataset and Analysis of Physiological Responses in Virtual Reality and Flatscreen Gaming Environments . 1, 1 (December 2022), 21 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Recent advances in Virtual Reality (VR) technology have ushered in a new era of VR gaming. As a result, many interactive experiences traditionally available in the conventional Flatscreen (FS) space are now moving to VR. Many traditional desktop games, primarily the first-person shooter [70, 80] and horror-adventure type games [59, 75], have quickly been ported to VR. Therefore, at this time of pivotal change, there is a need to assess the potential usability gains that VR brings to help make quantitative decisions instead of speculative claims. It can be done by investigating the users' emotional responses in VR game-play and by doing a comparative analysis of FS gaming.

While the VR gaming experience represents a burgeoning area of research within human-computer interaction circles, there is a lack of systematic comparison of the physiological effects between VR and FS gaming. We found only a few studies that have explored *emotional challenges* in gaming [6, 51]. Emotions play a critical part in our lives, influencing our decision-making, perception, social interactions, learning, memory, and creativity [71, 81]. Thus understanding emotional responses helps researchers to design better user experiences and improve the usability, acceptability, and accessibility of technologies. Due to this, researchers have often used several physiological signals in affective computing to predict arousal, valence, or even specific emotions by detecting physiological changes in the human body [55].

There have been multiple explorations in the field of emotional datasets in the past decade. Emotional databases are essential in human-computer interaction (HCI) and games because they can provide valuable information on how

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
© 2022 Association for Computing Machinery.
Manuscript submitted to ACM

people feel and react in different situations. This information can be used to design and develop more effective and engaging user interfaces and create more immersive and emotionally compelling games. Emotional databases can be used to train machine learning algorithms and other systems to recognize and interpret emotions. These models can be further used to create more immersive and engaging experiences where the application responds to the player's emotions and provides personalized experiences based on their emotional state. The research done in this paper has the three following contributions. First, we present **V**irtual **R**eality and **F**lat**S**creen Dataset (VRFS), the first-ever systematic comparison of physiological effects between VR and FS gaming. Second, we identify the emotional and physiological effects of VR and FS gameplay via self-reporting questionnaires and physiological signals. Together, these measures provided a comprehensive subjective and objective outlook on the two gameplay methods tested in our user study. Finally, we provide a novel 15+ hours of the multimodal affective dataset in which we elicited the emotions of multiple participants using commercially available games in VR and FS settings. The dataset for this study can be found here - Google Drive link.

The rest of the paper is organized as follows. In the next section, we discuss some of the earlier work done in this domain. The related work section is followed by a description of our stimuli selection procedure, including the pilot study and the final list of representative games. After this, we describe our experimental methodology and present the data. The following section provides the results of the data analysis. Next, we discuss how the results compare to the hypotheses. Finally, we conclude by pointing toward future research directions.

2 RELATED WORK

2.1 Emotions in VR

Emotional analysis for immersion in VR has been much studied recently. One ideal scenario is driving in VR vs. driving in real life. Researchers have used driving simulation games, and wheel controllers to create an environment as close as possible to real driving and used self-reporting surveys for felt emotions and experience [12, 73]. These studies revealed that VR dissociates the user from the real world more than FS, thus making VR more immersive. The uneasiness and motion sickness caused by the current VR systems made these studies more challenging. Research with young adults on a driving simulator in VR and FS showed that VR elicits more positive emotions and the sense of immersion and flow is more remarkable in VR games than in FS [47]. Researchers have established that Electroencephalography (EEG) signals from the brain are a relevant metric for decoding emotional arousal [24]. One study tried to recreate traffic-light-based responses in VR and tracked the system's effectiveness with EEG signals from the brain [37]. They showed that in an average of eight subjects, traffic events could reach an 87% accuracy compared to real-world events in VR. Studies have also shown that VR can elicit specific emotions with predictable outcomes. Researchers have been able to create "anxious" and "relaxation" producing simulations in VR with a significant certainty of predicted emotion arousal [58]. With these results, researchers have looked into VR as a medium for required emotion elicitation as an upgrade to existing stimuli. Meuleman et al. [42] observe that emotional responses in a VR gaming study are clustered in two segments; joy and fear. Work has also been done to successfully create deep learning models to predict peak emotions in gameplay using several biosignals [56]. VREED [69], is one of the first datasets that provide behavioral (eye tracking) and physiological signals [Electrocardiogram: ECG and Galvanic Skin Response: GSR (also known as Electrodermal Activity: EDA)] data in addition to self-reporting for comparison of emotion elicitation in VR and FS 360° Video-Based Virtual Environments. Further, recognizing the emotions users feel allows researchers to control these scenarios live, making systems like this useful for emotional training for children with Autism. Similar research has been done to compare the immersiveness, as well as motion sickness in first and third-person points of view in VR [43] which suggests that although the first-person perspective is more immersive, it is also more prone to inducing sickness. The emotions induced and their strength varies significantly in FS and VR as factors such as screen size and worldview have been shown to significantly affect mental immersion, which in turn affects emotions felt by the users [61, 64]. Other factors such as *cybersickness* also come into play when exploring emotions in VR gaming [79]. Overall positive emotions such as *hope, courage, relaxation, and calmness* are more associated with gaming in VR [51]; however, it is essential to note that these emotions are also greatly affected by the genre of the game title. It is possible to induce particular emotions by designing gameplay in a particular manner [19].

2.2 Analysis of Emotions

Emotional responses can be assessed via three different methods: self-reports from the subject (collected, for example, through questionnaires or interviews), physiological changes in the subject's body (e.g., heart rate or skin conductivity), and directly observable behaviors (e.g., facial expressions or eye gaze) [39]. Among these methods, self-reports are often subject to several biases, and limitations [15], however, these are our greater source for the ground truth. To understand emotions, we must first have a way to quantify them. For this purpose, many researchers have created different methods of classifying emotions, such as the Tree of Emotions [49], the wheel of emotions [54], the Pleasure, Arousal Dominance scale (PAD), also known as the Valence Arousal Dominance (VAD) model. From the VAD model, the Circumplex Model of Affects (CMA) was derived, which is a two-dimensional cartesian model to represent emotional stimuli based only on arousal (Y-Axis) and valence (X-Axis) only [60]. CMA representation is easier to understand and use than the VAD model. Additionally, the CMA is based on a circular structure, which allows for a more intuitive representation of emotions.

Meanwhile, emotion recognition from physiological signals is expedient since it taps the pure, unaltered emotion in contrast to behavioral responses like facial expressions, which can be faked. Recent advancements in wearable technologies have shown a strong potential for hassle-free acquisition of physiological signals in a non-intrusive manner and have thus inspired us to investigate emotional responses in VR and FS gameplay using physiological signals. Physiological signals can get more information on how users interact with and respond to a system. Researchers have also utilized multiple sensors to help evaluate and monitor human physiological signals. Signals like EDA [38, 72], ECG [44, 48], and Electroencephalogram (EEG) [26, 31, 71] have been widely used. The use of these sensors is showcased in a wide variety of applications, including medical [21, 53], neuromarketing [14, 76], sports training [10, 62] and many more. Researchers have used physiological signals like electrodermal and cardiovascular activity to analyze and predict emotions [36]. They have used these findings to create guidelines for VR developers to create more immersive games. Classification and analysis of emotions have been important for researchers for a long time. From a tree-like structure that divided emotions into primary, secondary, and tertiary [49] to a size axis model proposed by Kort et al. [30]. This led to the development of the Valence Arousal and Dominance Model by Osgood, Suci, and Tannenbaum [46]. The use of physiological signals for emotional classification has increased over the past decade thanks to improved technology surrounding procuring such data. For instance, The recent Empatica E4 device has shown the ability to capture physiological data reliably in most cases [40].

2.3 Affective Datasets

Data gathering is essential for many studies; however, this process is extremely lengthy, expensive, and requires significant resources. To combat these, many studies utilize publicly available datasets instead. One prominent dataset is the International Affective Picture System (IAPS) [33] that provides a range of picture-based emotional stimuli and participant ratings. Another recent example is the expanded version of the International Affective Digitized Sounds (IADS-E) [77] dataset. It augments the previous International Affective Digitized Sounds (IADS) [9] dataset and provides a larger selection of audio-based emotional stimuli combined with the participant rating.

Instead of just the emotional ratings, many datasets also offer physiological as well as behavioral data of the participants. These datasets record the participants' behavioral (eye gaze movement, face recording) and physiological (EDA, Blood Volume Pulse: BVP) signals while they are exposed to emotional stimuli. For example, MANHOB-HCI [66] is an affective dataset consisting of the face videos, audio signals, eye-gaze data, and peripheral/central nervous system physiological signals of participants reacting to 20 different emotional videos. Another dataset called the Database for Emotion Analysis using Physiological signals (DEAP) [29] contains affective music videos, their reported ratings, and the recorded EEG data of the participants. Similarly, Database for Emotion Recognition through EEG and ECG Signals from Wireless Low-cost Off-the-Shelf Devices (DREAMER) [28] contains the ECG and EEG data from participants experiencing audio-visual stimuli. The MEG-Based Multimodal Database for Decoding Affective Physiological Responses (DECAF) [1] contains the recorded brain signals similar to DEAP. However, these were captured using a Magnetoencephalogram (MEG) that requires physical contact with the participant's scalp but facilitates a natural affective response.

There also exist multiple datasets for participants reacting to different stimuli, such as the Bio-Reactions and Faces for Emotion-based Personalization for AI Systems (BIRAFFE2) [32] that consists of accelerometer data, ECG, and EDA signals, participants' facial expression data, and a personality and game engagement questionnaires. In the field of VR, VREED [69] is one of the first datasets that provide behavioral (eye tracking) and physiological signals (ECG and EDA) data in addition to self-reporting for comparison of emotion elicitation in VR and FS 360° in video-based virtual environments. Another dataset, Affective Virtual Reality System (AVRS) [82], has rated Arousal, Valence, and Dominance of similar 360° VR environments using self-reporting. Further, datasets have rated arousal and valence and have analyzed correlations between head movement and self-reported values [35]. In addition to self-reporting, some datasets have also used physiological data collected via EEG for emotional classification [25, 67].

A common and significant limitation of most affective datasets (including DEAP [29], MAHNOB-HCI [65], etc.) is that only a few have explored a VR-based stimulus for evoking emotions. However, VR technology has proven to provide substantial immersive experiences. Additionally, even the limited datasets that analyzed VR stimulus do not allow systematic comparison of VR and FS responses as they all rely on a different stimulus, such as [35] that uses 360° videos, [25] uses pictures embedded in VR environments. Finally, these datasets primarily include either physiological or psychological responses, while both are equally important for comparative analysis, limiting the research's scope and conclusions. To that effect, we are introducing the VRFS dataset, which would allow for a comprehensive analysis of the physiological signals and psychological responses in both VR and FS to increase understanding of people playing games in VR and how to improve the experience for new users.

3 STIMULI SELECTION PROCEDURE

The selection of effective games is essential to elicit the appropriate affective responses; therefore, the selection process underwent multiple stages to select the best game. First, the researchers shortlisted commercially available games, which had both VR and FS variants available, based on online reviews and placed them on a CMA. After placing games on the CMA, 12 games were shortlisted such that each quadrant of the CMA had three games. This pilot study selects one game from each quadrant of the CMA. Table 1 elaborates on the game selection procedure.

Table 1. Stimuli Selection Procedure.

	Procedure
Focus groups	Three Researchers met over two 2-hour sessions to extensively discuss and experience the 12 games to find suitable games for the pilot trial.
Pilot trial	In a span of two weeks, six volunteers spent around 90 minutes each engaging in and rating the selected 12 games. Each volunteer played four games, such that each game was played twice. Self-Assessment Manikin (SAM) [arousal, valence, dominance] and Visual Analog Scale (VAS) [joy, anger, calmness, sadness, disgust, relaxation, happiness, anxiousness, fear, and dizziness] were recorded to select the game.

3.1 Stimuli Shortlisting

Using various game distribution platforms such as Steam and Microsoft Xbox, the researchers attempted to find games that offered the same gameplay in both FS and VR modes. A short list was curated according to the following inclusion criteria:

- Games that were available on the Microsoft Windows platform,
- Games that offered almost similar experiences both in FS and VR, apart from obvious medium differences such as resolution and controls,
- Games that were playable offline only or that had uniform and predictable play sessions (E.g. Racing games with fixed laps or simple task-based games) for all participants,
- Games from which a sub-section can be identified that can be replayed or recreated by participants, and
- Games that could run comfortably on the test system (consistent frame rate with no frame drops).

After shortlisting, the researchers found 21 games and selected 12 based on each game's position in the CMA. The aim was to have three games in each quadrant. These 12 games were used for the pilot trial and then narrowed down to 4, one from each quadrant.

3.2 Pilot Trial

Six volunteers (five males and one female) aged between 19 and 21 ($\mu = 20$, $\sigma = 0.74$) played and rated the selected 12 games. The volunteers played four games each in a randomized order. Each game was effectively played and rated twice. To optimize the resources, each volunteer only played the FS version of each game. Since the perceived category of emotions has not been found to differ significantly between VR and FS (e.g., [2]), we expected that the chosen games through such a pilot trial would elicit the desired emotions in VR gameplay as well. The following tools were used to judge arousal:

- Self-Assessment Manikin (SAM) is a widely used state measurement tool [8]. It has simple cartoon-like manikin icons that can be used to plot the CMA dimensions (arousal and valence). The valence scale ranges from 1="happy" to 5="sad" pictures of SAM. Meanwhile, the arousal scale ranges from 1="calm" to 5="excited" pictures of SAM.
- The Visual Analog Scale (VAS) [22] is a continuous slider-based scale of numbers ranging from 1-100, with two verbal descriptors at each end. Using VAS, volunteers rated how they felt while engaging in the games, using separate scales for joy, happiness, calmness, relaxation, anger, disgust, anxiousness, fear, and sadness.

After completing the pilot study, we had the SAM and VAS results for each game, using which we could see how each game performed and in which quadrant they lie in, as shown in Fig. 1. Arousal and valence ratings for each of the 12 games can be found in Table 2. SAM and VAS ratings have been collected to be used as the "ground truth" or reference point for the research, against which the physiological responses can be compared. Although we collected dominance as well through SAM, we found that it did not differ significantly between participants (H: 11.0, p-value: 0.44) according to the Kruskal Wallis test, so we excluded the dominance while performing the stimuli selection.

Table 2. Arousal, valence, and dominance of each game in the pilot study. The selected games are highlighted in bold.

Name	Arousal	Valence	r	$d\theta$
WAR THUNDER	6	4	0.7	2.8
HITMAN	4	**2**	1.58	27.8
RACEROOM	3	4	1.11	14.4
MINECRAFT	**8**	6	1.58	22.0
NO MAN'S SKY	7	5	1.0	45.0
THE FOREST	3	5	1.0	45.0
FOREWARNED	5	7	1.0	45.0
PROJECT CARS 2	6	3	1.11	20.1
MICROSOFT FLIGHT SIM	6	5	1.0	45.0
SUBNAUTICA	4	**9**	2.06	31.8
PHASMOPHOBIA	3	6	1.11	14.5
DIRT RALLY 2	**2**	3	1.80	7.7

In order to analyze which game should be chosen as the representative of their quadrant, we plotted (see Fig. 1) the distance r of that game from the origin as well as the angle $d\theta$ that they are farther from 45° as

$$r = \sqrt{Arousal^2 + Valence^2}, \text{ and}$$
$$d\theta = \left| \frac{\pi}{4} - \tan^{-1}\left(\frac{Arousal}{Valence}\right) \right|$$

Games with the highest r value, i.e., the largest distance from the origin and the least difference from 45°, would be the best representatives for that quadrant. Taking values farther away from the origin would show higher arousal or valence in that preferred quadrant. Values near the 45° line would ensure that both arousal and valence are almost equally high. The following games were chosen from each quadrant (see gameplay screenshots in Fig. 2).

- *Quadrant 1: High Arousal- High Valence (HVHA)* Dirt Rally 2.0 (Fig. 2(b)) is chosen as it has the minimum $d\theta$ value of 7.7 along with the maximum r value of 1.80. This is a rally-style racing game that offers a realistic driving experience.
- *Quadrant 2: High Arousal - Low Valence (LVHA)* Phasmophobia (Fig. 2(a)) is chosen in this quadrant due to its high r value of 1.11 and is the closest to 45° with a $d\theta$ value of 15.5. This is an investigative horror game where the users have to find clues to solve an objective.
- *Quadrant 3: Low Arousal - Low Valence (LVLA)* Minecraft (Fig. 2(c)) is the clear winner here since it has a higher r value of 1.58 with a minimum $d\theta$ value of 22. This is a block-based sandbox video game in which players gather and use blocks to create various objects.

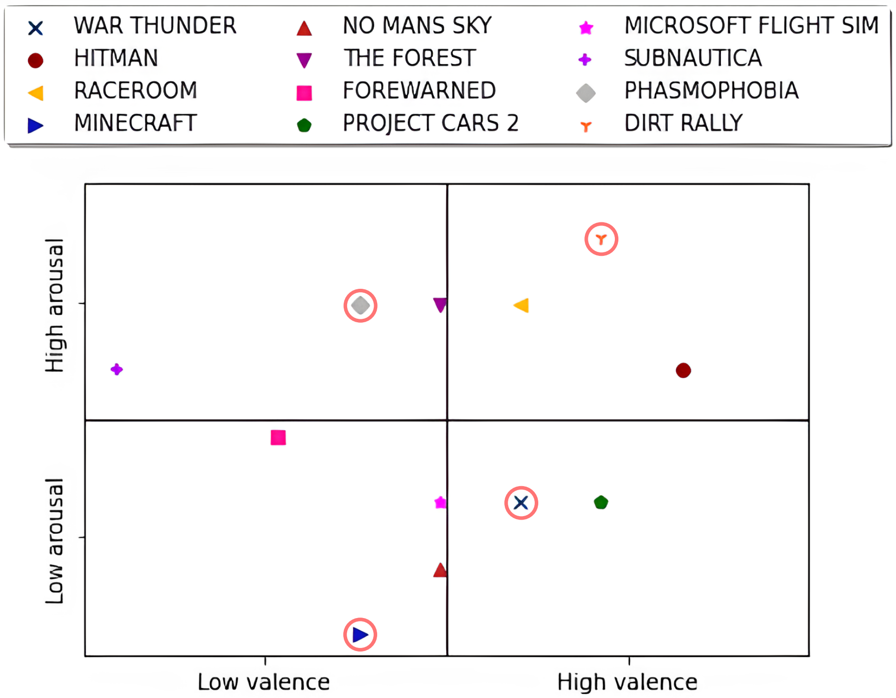


Fig. 1. Games plotted in their respective quadrants according to SAM results from the pilot study. Encircled games were selected for the final study.

- *Quadrant 4: Low Arousal - High Valence (HVLA)* We chose War Thunder (Fig. 2(d)) in this quadrant since it has a much lower $d\theta$ value of 2.8 compared to other games. This military-style aerial combat fighting simulator aims to shoot down enemy aircraft.

4 EXPERIMENTAL SETUP AND DATA COLLECTION

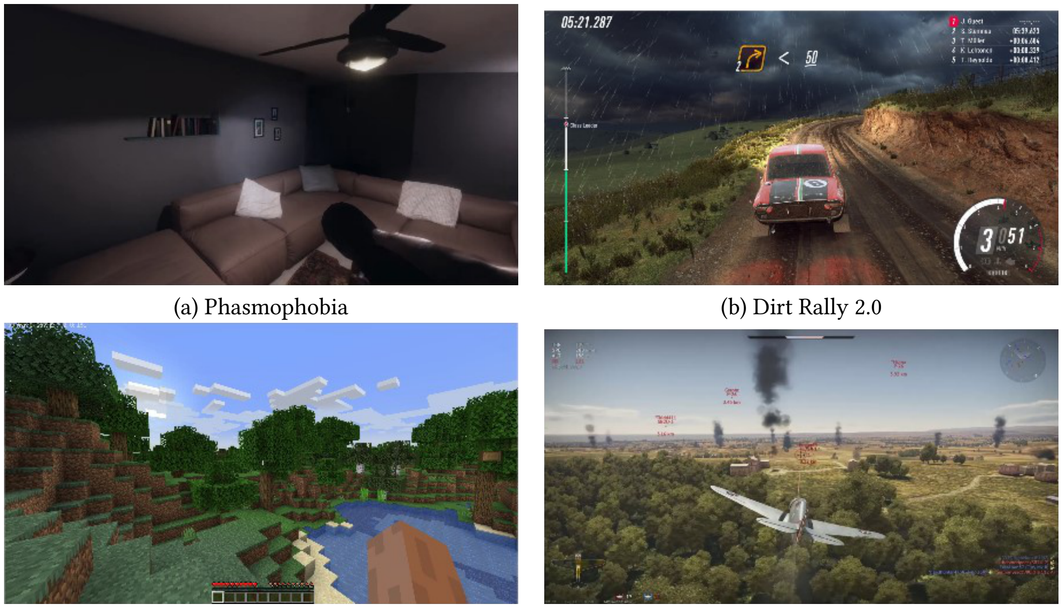
This experimental study aimed to gather physiological and self-reporting data from participants in different emotion-eliciting games played by them in both VR and FS mediums. Further, we want to analyze and identify these signals for evident patterns of arousal and immersiveness and report our findings to future researchers and developers.

4.1 Ethics

All participants signed a consent form before the study was conducted. The study was approved by our institute's Institutional Review Board (IRB). To preserve the privacy of all participants, raw physiological data collected during the study would not be available. Only pre-processed and anonymized data would be shared publicly.

4.2 Participant Screening Criteria

An invite was sent via various channels to students of our institute. The invite had a brief description of the experiment and a demographic questionnaire. The following volunteers were excluded from the study:



(a) Phasmophobia (b) Dirt Rally 2.0

(c) Minecraft (d) War Thunder

Fig. 2. In-game screenshots of selected four games.

- Individuals who reported having a seizure(s), migraines, or any other medical condition in the past that could increase the risk during the user study,
- Participants who reported having a prior history of motion sickness,
- Individuals who rated four or higher on a Likert scale on the question “How easily do you get motion sick or carsick?” where one was “Rarely get motion sick” and five was “Very often get motion sick”, and
- Due to the ongoing COVID-19 pandemic, participants who were COVID positive or reported symptoms including cough, fever, and fatigue.

Finally, 36 individuals (16 female and 20 male) aged between 19 and 22 years ($\mu = 20.8, \sigma = 0.71$) volunteered to participate in this study. 45.4% of the participants (n=15) reported having used VR previously, and none reported feeling any motion sickness during or after exposure to VR.

4.3 Psychological Measures

Participants completed the following self-reported measures both before and during the experiment:

- *A pre-exposure questionnaire*
  - Participants were asked to report any demographic information such as age, gender, ethnicity, dominant hand, and previous experience with VR, and
  - SAM and VAS measures were taken to form a baseline. Questions were modified to ask the participant how they felt at that moment.
- *A post-exposure questionnaire*
  - SAM and VAS measures were taken immediately after playing the game in both VR and FS, and

  - Using VAS, participants were asked to rate how dizzy they felt after playing the games in both VR and FS.

4.4 Physiological Measures

With the four selected games from the pilot study, we invited 33 participants to play these games in FS and VR. The details of the data we collected are explained in Table 3. We measured the participants' Blood Volume Pulse (BVP) and EDA. These signals have been shown to represent physiological state accurately [69] and can be measured non-intrusively with a wrist device. The device is worn similarly to a wristwatch and does not obstruct the gameplay experience of the participants. The BVP and EDA signals are explained in more detail below. These signals were recorded using an Empatica E4 device and synced with each other. The sampling rate for these is hard coded into the firmware and optimized to capture the frequency content of relevant signals.

4.4.1 Blood Volume Pulse. The BVP signal measures the changes in blood volume inside arteries and capillaries. A Photoplethysmography (PPG) sensor uses light from an LED to measure the refraction and estimate blood volume. The amount of light that returns to the PPG sensor is positively proportional to the blood volume. PPG gives the average value of blood in the tissues through which light has passed. Using this measure of the amount of blood passing over each pulse through a PPG, we can effectively calculate Heart Rate variability (HRV) using BVP [52]. We then extract the Low-Frequency and High-frequency features from the HRV. The LF band (0.04–0.15 Hz) is affected by breathing from 3 to 9 bpm. [27] while the HF or respiratory band (0.15 – 0.40 Hz) is influenced by breathing from 9 to 24 bpm [45]. The ratio of LF to HF power (LF/HF ratio) provides the ratio between the sympathetic nervous system (SNS) and parasympathetic nervous system (PNS) activity under controlled conditions. The BVP in Empatica E4 is calculated through the PPG sensor at 64 Hz.

4.4.2 Electrodermal Activity. Electrodermal activity (EDA) refers to the variation of the electrical properties of the skin in response to sweat secretion. Applying a low constant voltage, Skin Conductance (SC) change can be measured non-invasively. EDA has been measured with the Galvanic Skin Response sensor. EDA data consists of two components, Skin Conductance Level (SCL) [tonic component] and Skin Conductance Response (SCR) [phasic component]. SCL is a general measure of slow-moving psycho-physiological activation [63], while SCRs depict higher-frequency changes directly related to an external stimulus [20]. Typically, SCR and heart rate are the best discriminators for arousal detection [11]. The time series of SC can be characterized by a slowly varying tonic activity (i.e., SCL) and a fast varying phasic activity (i.e., SCR) [3]. EDA was sampled at a frequency of 4 Hz, with the data being measured in micro siemens (μS).

4.5 Apparatus and Setup

All apparatus was set up in the "Experiment Room," and only one participant and two researchers were allowed at one time to ensure necessary social distancing rules. Following are various hardware devices that were part of our experimental setup:

- *Flatscreen Display and Speakers:* A single HP 24-inch Monitor was used to play games on the FS. The monitor resolution was 1920x1080 pixels. The refresh rate was 60 Hz, and all games were played at 60 FPS to ensure a smooth experience. The monitor comes with in-built speakers, which were used to stream audio in FS mode.

Table 3. Details of experimental data collection.

Stimuli Selection Method	Based on the six pilot trial results
Participant Information and Pre-Exposure Data	33 Participants provided demographic information (i.e. sex, age), answered a health inclusion questionnaire
Post-Exposure Data	SAM (arousal, valence) and VAS
Recorded Physiological Signals	BVP and EDA
Study Design	2x4 mixed within-subject design (2 conditions (VR+FS) × 4 variables (4 Games))
Additional Materials	Self Reporting Questionnaire, and verbal instructions protocol

- *VR Headset:* An Oculus Rift S VR headset[1] was used. It comes with one headset and two controllers. The headset is wired to the computer, and both handheld controllers are wireless. Rift S uses a single fast-switch LCD panel with a resolution of 2560×1440 and an 80 Hz refresh rate. It has a field of view of 115 degrees. The Rift S has inbuilt speakers in the headset, positioned just above the ears, which were used for audio delivery in VR experiences. *Steering Wheel Controller:* For the driving game (Dirt Rally 2.0), we used a steering wheel controller Thrustmaster T300RS[2]. This is a steering wheel + 2-leg paddle device. This simulates a real-world driving experience. The wheel also has vibrational feedback.
- *Physiological Collection Device:* The Empatica E4[3] wristband was used to measure various physiological signals. It has sensors for Photoplethysmography (PPG) (measures BVP), a 3-axis accelerometer, an EDA sensor, and an infrared thermopile for skin temperature. It also has an internal clock for syncing and a physical event marking button to mark events in real time.

Both VR and FS versions of the games were played on a system with hardware configurations better or equivalent to the recommended specifications of all the games. It had an Intel Core-i7 8700K CPU and an NVidia GeForce GTX 1080Ti GPU. In addition, a 14-inch Windows FHD laptop was used to complete emotion self-reporting questionnaires (SAM and VAS).

4.6 Hypotheses

Overall, we expected VR games to be perceived as more immersive [50] and arousing. Due to the high physical nature of control in VR games, we expected the mentally and physically challenging nature to reflect in the physiological signals. The following hypotheses informed our study.

- **H1:** Analyzing variance in heart rate helps establish a relationship between physiological load and a higher cognitive load [23]. As immersion in VR is higher (implying a more significant burden) than in FS gaming [5], we expected a higher cognitive load to be observed while using VR gaming, which will be reflected as increased HR values.
- **H2:** A positive relationship is observed between emotional intensity and immersion [58]. Since immersion in VR is higher than in FS gaming, we expected that immersion induced through VR gameplay would enhance one's appraisal of their context compared to FS gameplay.

[1]Oculus Rift S VR headset: https://www.oculus.com/rift-s/
[2]Thrustmaster T300RS Force Feedback racing wheel: https://www.thrustmaster.com/en-gb/products/t300rs/
[3]Empatica E4 Wristband: https://www.empatica.com/research/e4/

- **H3:** LF/HF sympathovagal balance values increase during "sympathetic dominance" caused by emotional and physiological stress [17]. VR gaming is likely to induce emotional and physiological stress compared to FS gaming due to the higher cognitive load in VR gaming.
- **H4:** A higher cognitive load is observed with learning novel information [68]. VR controls and navigation are relatively new to the target users, requiring more concentration and physical effort. Accordingly, users are likely to feel less dominant during VR gaming compared to FS gaming.
- **H5:** VR gaming is likely to induce more arousal compared to FS gaming due to higher immersion in the VR environment, which will be reflected in increased SAM ratings. Since SCRs also have been found to be much more frequent when the individual is aroused [7], we expected an increased skin conductance (EDA) activity as well.

4.7 Experimental Procedure

We collect and analyze qualitative psychological responses and quantitative physiological measures in a 2x4 mixed within-subject design study.Each session lasted between 30-40 minutes, depending on the game. Each participant was asked to visit the lab twice and play the same game on the FS in one session and VR in the other. To avoid the order effects, this order was randomized to ensure equal participants in both categories, VR first and FS first. A verbal instructions protocol was used to ensure that instructions were held constant for all participants. Before the start of the session, participants were informed about the study itself but not about the purpose or the hypotheses. Participants were asked to sign the consent form and fill out the "participant demographics and pre-exposure" questionnaire. This was followed by a brief introduction of their specific game and the objective they had to complete. Then, the Empatica E4 wristband was attached to the participants and switched on. The LED signals of the E4 device were noted to ensure correct functioning. After the equipment test was complete, the participants were told to relax and not think of anything extreme or arousing to establish a baseline. This lasted for 5 minutes. In the case of the VR experiment, participants were introduced to the use of VR while the researchers helped fit the Rift S headset to the participant's comfort. Comfort in the use of hand controllers was confirmed by asking participants to navigate basic menus.

In the case of the FS, the participants were shown the basic controls of the keyboard and mouse. In Dirt Rally 2.0, the Thrustmaster wheel was used in both FS and VR. After this, the study continued with the participant playing the game, trying to complete the decided tasks. At the end of each game session, participants filled out the "Post-Exposure" questionnaire consisting of the SAM and VAS. At the end of both sessions, participants were briefed about the study objectives and thanked for their participation.

4.8 Dataset

We use the FLIRT module in python to extract various statistical and signal features for the EDA and HRV data. Accelerometer data is ignored in our analysis because it does not carry significant impressions useful for the final analysis. The trial was conducted on 36 participants; however, three of the data points had to be dropped due to participants' inability to complete the gameplay for various reasons. This led to a total of 33 participants, split between 4 games. Data is ordered by the game and subsequently by the player, each having a VR and FS folder. Inside each folder is the zip file containing the data along with the CSV files for the data. These are ACC, BVP, EDA, HR, IBI, and TEMP (see Table 4) . Each CSV file starts with the time in the beginning, followed by the frequency of the recording, and finally, the data collected over the recording.

Table 4. Number of Participants across different games in the dataset

	Dirt Rally 2 (HVHA)	Phasmophobia (LVHA)	Minecraft (LVLA)	War Thunder (HVLA)	Duration (minutes)
VR	9	8	7	9	μ= 13.7, σ = 2.76
FS	9	8	7	9	μ= 14.1, σ = 2.65

5 RESULTS

5.1 Validation of Stimulation

This analysis aims to confirm that the chosen games elicit the emotions assigned to them by the CMA quadrant. Table 5 provides the mean (μ) and standard deviation (σ) for the different games during FS gameplay, VR gameplay as well as VR+FS gameplay. We observe the Arousal and valence do fall in the needed quadrant; however, we still need to confirm statistical significance. We observe higher reported arousal for games rated as High Arousal (High Arousal High Valence, μ = 3.33; High Arousal Low Valence, μ = 2.75) and vice versa (Low Arousal High Valence, μ = 2.00; Low Arousal Low Valence, μ = 1.17) as evident by the combined values. Similarly, in the case of High Valence rated games, we observed a higher reported valence (High Arousal High Valence, μ = 2.88; Low Arousal High Valence, μ = 2.79), and vice versa (High Arousal Low Valence, μ = 2.12; Low Arousal Low Valence, μ = 2.28), as evident by the combined values. Similar trends have been observed in VR and FS gameplay individually.

Table 5. Arousal and Valence Ratings as per the CMA quadrant.

Intended CMA Quadrant	**Type**	**Arousal**		**Valence**		**r**	
		μ	σ	μ	σ	μ	σ
Low Arousal Low Valence	VR	1.00	0.94	2.00	1.25	2.43	1.25
	FS	1.33	0.47	2.56	0.96	2.91	1.00
	VR+FS	1.17	0.76	2.28	1.15	2.67	1.15
High Arousal Low Valence	VR	3.38	0.86	2.38	0.99	4.24	0.87
	FS	2.12	0.93	1.88	0.78	2.99	0.73
	VR+FS	2.75	1.09	2.12	0.93	3.62	1.02
Low Arousal High Valence	VR	2.86	0.83	3.00	0.76	4.23	0.72
	FS	1.14	0.64	2.57	0.90	2.91	0.82
	VR+FS	2.00	1.13	2.79	0.86	3.57	1.02
High Arousal High Valence	VR	3.44	0.68	3.22	1.23	4.88	0.63
	FS	3.22	0.42	2.53	0.67	3.98	0.51
	VR+FS	3.33	0.58	2.88	1.08	4.43	0.73

The Kruskal-Wallis test is conducted on the data given by Table 5 to confirm the significance of the stimulation ability of the games. In the case of both FS and VR, the reported arousal varied significantly (H: 30.67, p-value: 9.97e-7) across the games in all four CMA quadrants. We also performed the post hoc Dunn test; the results are presented in the table.

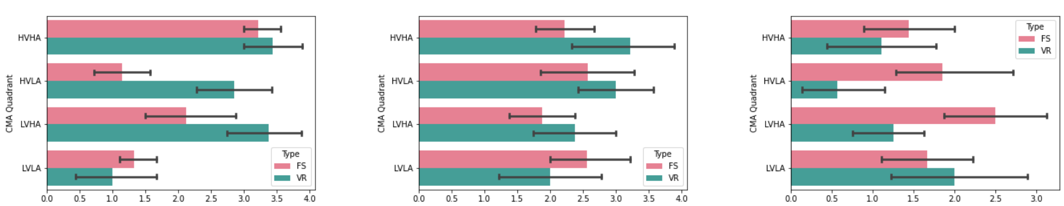
Table 6. Post-hoc result on comparing Arousal between four quadrants using Dunn test.

	HVHA	HVLA	LVHA	LVLA
HVHA	1.0	0.002	0.156	1.3e-7
HVLA	2.2e-3	1.0	0.1	0.006
LVHA	0.015	0.1	1.0	2.1e-4
LVLA	1.3e-7	0.06	2.2e-3	1.0

Similarly, in the case of FS. This indicates that the participants playing the games in the HVLA and HVHA category experience significantly higher Valence as compared to those playing games in the LVLA and LVH A category, along with those playing games in the HVHAand the LVHAcategories experiencing higher Arousal than their counterparts playing games in the HVLAand the LVLAcategories. Even though participants perceived the HVHAin FS as less arousing than in VR, the Arousal ratings in this quadrant were still significantly higher than the LVLAof the same. In summary, the participants experienced four distinct emotional states over the differing valence and arousal dimensions (HVHA, HVLA, LVHA, LVLA).

5.2 Psychological Results

5.2.1 Self Assessment Mannequin (SAM). We show plots of mean and standard deviation of the reported arousal, valence, and dominance of different games in Fig. 3. The data was found to be non-normal upon using the Shapiro test (arousal: W = 0.89, p-value = 0.003; valence: W = 0.88, p-value = 0.002; dominance: W = 0.90, p-value = 0.006). We observed a higher reported arousal in VR gameplay (μ = 2.6 σ = 1.3) than in FS gameplay (μ = 2.0 σ = 1.0) as shown in Fig. 3(a). We conducted the nonparametric Wilcoxon signed-rank test on the data and found the difference to be statistically significant (W = 66.5, p-value = 0.007). We also observed a higher value for Valence in VR gameplay (μ = 2.6 σ = 1.2) as compared to FS gameplay (μ = 2.3 σ = 0.9) as shown in Fig. 3(b). However, the Wilcoxon signed-rank test found this difference to be statistically non-significant (W = 75.5, p-value = 0.262). In contrast, we observed that dominance was lower in VR gameplay (μ = 1.3 σ = 1.1) compared to FS gameplay (μ = 1.8 σ = 1.0), and this difference was found to be statistically significant (W = 88.0, p-value = 0.020).



(a) (b) (c)

Fig. 3. Mean and Standard Deviation of (a) arousal, (b) valence, and (c) dominance of players in different games.

5.2.2 Visual Analogue Scale (VAS). We plot the mean and standard deviation recorded in VR and FS gameplay for the different emotions in Fig. 5 and reported emotions per CMA quadrant in Fig. 4. The data collected for the emotions were found to be non-normally distributed upon conducting the Shapiro-Wilks test (joy: W = 0.92, p-value = 0.016;

anger: W = 0.87, p-value = 0.001; calmness: W = 0.86, p-value = 0.001; sadness: W = 0.70, p-value = 0.000; disgust: W = 0.62, p-value = 0.000; relaxation: W = 0.88, p-value = 0.001; happiness: W = 0.92, p-value = 0.017; fear: W = 0.92, p-value = 0.015; anxiousness: W = 0.91, p-value = 0.008; dizziness: W = 0.91, p-value = 0.008;). Upon conducting the Wilcoxon signed-rank test, we observe the p-values shown in Table 7. We observe that joy, anger, happiness, and dizziness are reported to be higher during VR gameplay as compared to FS gameplay. Of these, only joy and dizziness have a statistically significant difference between the two gameplay (joy: W = 101.50, p-value = 0.004; dizziness: W = 8.50, p-value = 0.000) difference between the two gameplay. On the other hand, calmness, sadness, and relaxation, while higher during VR gameplay than in FS gameplay, are not different enough to be statistically significant.

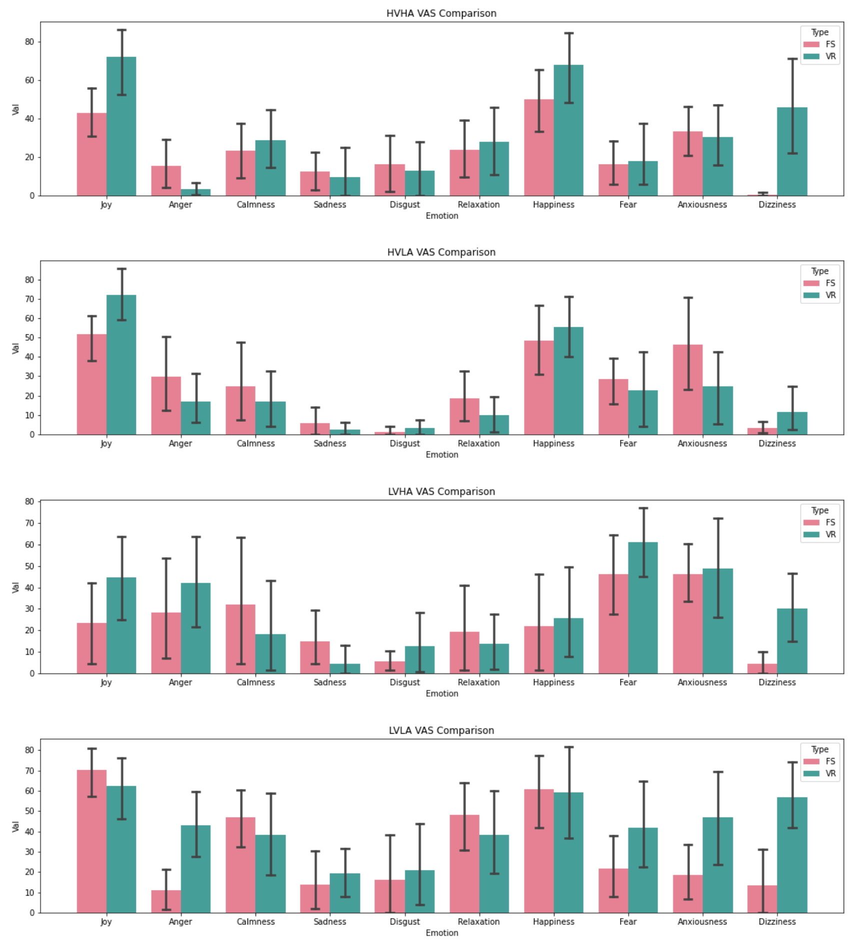


Fig. 4. Mean of different emotions between VR and FS across the different quadrants of CMA.

5.3 Physiological Results

5.3.1 Electrodermal Activity (EDA). The raw EDA data is pre-processed using the convex EDA algorithm [20] to obtain the tonic (SCL) and phasic (SCR) components. The cvxEDA model is a physiologically inspired model which describes

Table 7. p-values of different emotions between VR and FS

Emotion	VR		FS		p-value
	μ	σ	μ	σ	
Joy	63.45	26.01	48.39	26.62	0.0041
Anger	25.73	26.53	20.48	25.19	0.3198
Calmness	26.18	27.52	31.88	31.11	0.1389
Sadness	9.42	17.11	11.67	18.24	0.3007
Disgust	12.79	22.98	10.33	22.14	0.2176
Relaxation	23.30	26.88	28.24	27.16	0.1708
Happiness	53.55	33.28	46.64	30.38	0.1469
Fear	34.88	31.65	27.15	23.75	0.2096
Anxiousness	37.55	31.58	35.27	26.79	0.9478
Dizziness	37.27	32.27	5.48	14.76	0.0000

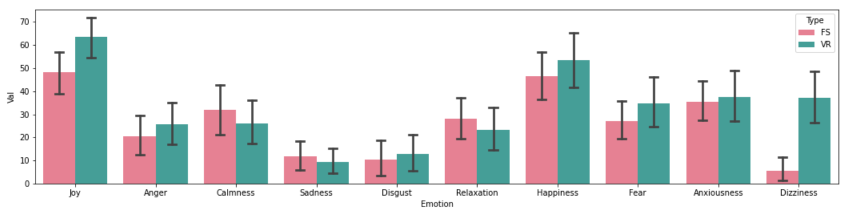


Fig. 5. Mean of different emotions upon playing games in VR and FS.

skin conductance as a combination of three factors: the phasic component, the tonic component, and additive white Gaussian noise. The proposed cvxEDA algorithm uses convex optimization to break down the EDA signal into its different components and provides a comprehensive explanation of EDA. At the same time, we collected baseline EDA data and found that it did not differ significantly between participants (H = 32.0, p-value = 0.47) according to the Kruskal Wallis test, so we did not perform the normalization of the EDA data employing the baseline data. From SCR, we obtain the statistical features such as mean, standard deviation, minimum, maximum, and percentiles which are shown in Fig. 6. These results were found to be non-normal upon conducting the Shapiro-Wilks test (W = 0.25, p-value = 0.00). While we observe lower SCR ratings during VR gameplay (μ = 0.35, σ = 0.75) than in FS gameplay (μ = 0.51, σ = 1.57), the difference between the two is statistically nonsignificant (W = 222.0, p-value = 0.295) according to the Wilcoxon Signed-rank test.

5.3.2 Heart Rate. We present the mean heart rate values for the participants during VR and FS gameplay, based on the game played in Fig. 7(a). The average heart rate of the players followed a normal distribution according to the Shapiro-Wilks test (W = 0.97, p-value = 0.46). The heart rate observed during VR gameplay (μ = 81.52, σ = 10.71) was higher than that observed during FS gameplay (μ = 77.27, σ = 9.92). We found the difference to be statistically significant (t-stat = -2.5, p-value = 0.016) upon conducting a paired t-test on the data. Using the heart rate and the Inter Beat Interval (IBI), we extracted the low-frequency (LF) and high-frequency (HF) features. We then calculated the LF/HF ratio using these two and sorted them based on the different emotions as shown Fig. 7(b).

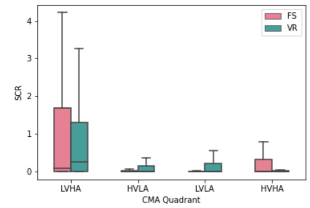
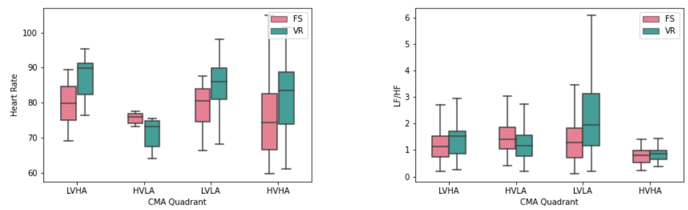


Fig. 6. Electrodermal Activity of Players to Different Games.



(a) (b)

Fig. 7. (a) Mean heart rates of participants based on the CMA quadrant. (b) Mean LF/HF features of the player in each quadrant for VR and FS.

We found that the mean LF/HF values were not normally distributed according to the Shapiro-Wilks test (W = 0.87, p-value = 0.005). We also observed higher LF/HF values during VR gameplay (μ = 1.43, σ = 0.63) compared to FS gameplay (μ = 1.37, σ = 0.60). Upon conducting the Wilcoxon Signed-rank test, we observed that the difference was statistically significant (W = 157, p-value = 0.894).

6 DISCUSSION

We present statistical validation of our five hypotheses presented above and provide inferences, followed by a discussion on the implications of our work and the limitations that we observed in our study.

6.1 Inferences

Our first hypothesis (H1) stated that participants would experience a higher cognitive load during VR gameplay than FS gameplay, reflecting an increased average heart rate. Our analysis observed a higher mean heart rate in the VR Game case. Using a paired t-test, we obtained a statistically significant result (t-stat = -2.5, p-value = 0.016), validating H1. We partially attribute this increase to the increased immersion felt during VR gameplay.

The second hypothesis (H2) proposes that increased immersion in VR gameplay will result in a greater elicitation of desired emotions. To evaluate this, we will refer to the radius of the SAM result in the CMA graph (Equation 1). By examining this data, we can see that emotional elicitation during VR gameplay (μ = 3.92, σ = 1.32) is higher than FS gameplay (μ = 3.22, σ = 0.91), and this difference is statistically significant (W = 89.5, p-value = 0.003) according to the paired t-test. Additionally, when we compare the games in the different quadrants of the CMA plot, we observe that emotional elicitation during VR gameplay is higher and statistically significant according to the Wilcoxon Signed Rank Test for all quadrants, except for the LVLA case, as shown in Table 6.

According to our third hypothesis (H3), we expected that the LF/HF Sympathovagal balance would increase in the case of VR due to the increase in cognitive load. In our study, we observed a significantly higher value of LF/HF in VR (μ = 1.43, σ = 0.63) as compared to FS (μ = 1.37, σ = 0.60) but the difference was not statistically significant using the Wilcoxon signed rank test (W: 157, p-value: 0.894). This could be because the LF/HF ratio does not accurately measure cardiac sympathovagal balance [4].

The fourth hypothesis (H4) suggests that emotions felt by users will be less dominant in virtual reality (VR) compared to flat screen (FS) displays. Our data support this hypothesis, as we found a lower dominance of emotions in VR gameplay (μ = 1.3 σ = 1.1) compared to FS gameplay (mean = 1.8, SD = 1.0). This difference was significant, as indicated by the Wilcoxon signed-rank test (W = 88.0, p = 0.02). It is possible that this effect was due to the fact that many participants were new to VR and had little to no experience with it, while they were familiar with using FS monitors. This lack of experience may have led to a decrease in emotional dominance in the unfamiliar VR environment.

Finally, according to our fifth hypothesis (H5), the participants should experience a higher level of arousal while in virtual reality (VR) compared to being in a full-screen (FS) environment. In our study, we observe a decreased SCR (VR: μ = 0.34, σ = 0.75; FS: μ = 0.51, σ = 1.61), the difference was not statistically significant. However, we did observe a statistically significant increase in the participants' arousal levels when playing in VR (μ = 2.6 σ = 1.3) compared to FS (μ = 2.0 σ = 1.0) (Wilcoxon test: W = 66.5, p-value = 0.007). This effect may be partially attributed to the greater immersion level and dynamic VR experiences.

6.2 Implications and Future Work

Our research can enable a better understanding of user engagement and emotional response in VR and FS gaming. Consequently, this may allow developers to design platform-specific experiences for VR and conventional FS platforms based on the intended emotions that are required to be enticed by the end users. Further, the findings of this research can also assist in designing and developing more engaging games, not only for entertainment but also for non-entertainment purposes such as rehabilitation [74] and in managing post-traumatic stress disorder [57]. Using insights from this study, developers can create tailored experiences that incite specific emotions by leveraging the strengths of VR. Multimodal signals can be used for this purpose to adapt gaming experiences based on emotions and their physiological response [13]. Creating an adaptive, personalized physical therapy game that adjusts to the user's mental and physical state in runtime may yield immense potential [78].

While the physiological effects of VR have been investigated in general, the emotional reactions are not well-known [34]. The proposed dataset in this research can be used to understand the emotional effects of gaming in VR and compare them with FS gaming. It can help monitor and mitigate the emotional consequences of gaming, devising a proper regulatory framework for gamification technology.

6.3 Limitations

One limitation of our study was the size and sample of the participants. Accordingly, a larger study is needed to produce a more extensive dataset that is more generalizable to a broader population. Our research focused exclusively on adults' emotional reactions, but it is very likely that younger players, including children, will experience more extreme emotional reactions [16]. Therefore, research should explore how different demographics, based on age, gender, etc., are susceptible to comparative effects. It would also be interesting to explore individual differences in psychological and physiological responses, for example, based on openness to experience [41], exposure beyond the novelty effect period

[18], etc. It will also be essential to gather longitudinal data, as this will enable a more robust understanding of how the observations of the current study change over time.

7 CONCLUSION

In this study, we present the first systematic comparison of VR and FS gameplay using both quantitative and qualitative measures. These measures included self-reported emotions and physiological effects such as EDA and BVP. We first conducted a pilot study with six volunteers to identify four games that elicited emotions in all four quadrants of the CMA. Further, in a user study with 33 participants, we found that the induced emotions were accurately perceived while playing the selected games. We observed significant differences in arousal and valence ratings across the four quadrants, indicating that participants experienced distinct emotional states over the valence and arousal dimensions (HVHA, HVLA, LVHA, LVLA). Overall, we found that VR gameplay led to more intense emotions, higher arousal, increased cognitive load and stress, and lower dominance compared to FS gameplay. The novel VRFS dataset from this study is made publicly available for other researchers to use for testing methods, hypotheses, and algorithms.

REFERENCES

[1] Mojtaba Khomami Abadi, Ramanathan Subramanian, Seyed Mostafa Kia, Paolo Avesani, Ioannis Patras, and Nicu Sebe. 2015. DECAF: MEG-Based Multimodal Database for Decoding Affective Physiological Responses. *IEEE Transactions on Affective Computing* 6, 3 (2015), 209–222. https://doi.org/10.1109/TAFFC.2015.2392932

[2] RT Al Alam and Nicola Dibben. 2021. A comparison of presence and emotion between immersive virtual reality and desktop displays for musical multimedia. In *Future Directions of Music Cognition 2021 Virtual Conference Proceedings*. Ohio State University Libraries.

[3] Mathias Benedek and Christian Kaernbach. 2010. A continuous measure of phasic electrodermal activity. *Journal of Neuroscience Methods* 190, 1 (2010), 80–91. https://doi.org/10.1016/j.jneumeth.2010.04.028

[4] George Billman. 2013. The LF/HF ratio does not accurately measure cardiac sympatho-vagal balance. *Frontiers in Physiology* 4 (2013). https://doi.org/10.3389/fphys.2013.00026

[5] Jim Blascovich and Jeremy Bailenson. 2011. *Infinite reality: Avatars, eternal life, new worlds, and the dawn of the virtual revolution.* William Morrow & Co.

[6] Julia Ayumi Bopp, Klaus Opwis, and Elisa D. Mekler. 2018. An Odd Kind of Pleasure: Differentiating Emotional Challenge in Digital Games. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3173574.3173615

[7] Wolfram Boucsein. 2012. *Electrodermal activity.* Springer Science & Business Media.

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

[9] Margaret M Bradley and Peter J Lang. 2007. The International Affective Digitized Sounds (; IADS-2): Affective ratings of sounds and instruction manual. *University of Florida, Gainesville, FL, Tech. Rep. B-3* (2007).

[10] Eglė Butkevičiūtė, Liepa Bikulčienė, Tatjana Sidekerskienė, Tomas Blažauskas, Rytis Maskeliūnas, Robertas Damaševičius, and Wei Wei. 2019. Removal of movement artefact for mobile EEG analysis in sports exercises. *IEEE Access* 7 (2019), 7206–7217.

[11] Yekta Said Can, Bert Arnrich, and Cem Ersoy. 2019. Stress detection in daily life scenarios using smart phones and wearable sensors: A survey. *Journal of biomedical informatics* 92 (2019), 103139.

[12] Shi Cao, Krithika Nandakumar, Raiju Babu, and Benjamin Thompson. 2020. Game play in virtual reality driving simulation involving head-mounted display and comparison to desktop display. *Virtual Reality* 24, 3 (01 Sep 2020), 503–513. https://doi.org/10.1007/s10055-019-00412-x

[13] Guillaume Chanel, Cyril Rebetez, Mireille Bétrancourt, and Thierry Pun. 2011. Emotion Assessment From Physiological Signals for Adaptation of Game Difficulty. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans* 41, 6 (2011), 1052–1063. https://doi.org/10.1109/TSMCA.2011.2116000

[14] Ubaldo Cuesta, Jose Ignacio Niño, and L Martínez-Martínez. 2018. Neuromarketing: analysis of packaging using Gsr, eye-tracking and facial expression. In *Papper presented at The European Conference on Media, Communication & Film*.

[15] Marion Devaux and Franco Sassi. 2016. Social disparities in hazardous alcohol use: self-report bias may lead to incorrect estimates. *The European Journal of Public Health* 26, 1 (2016), 129–134.

[16] Aimee Drolet, Patti Williams, and Loraine Lau-Gesk. 2007. Age-related differences in responses to affective vs. rational ads for hedonic vs. utilitarian products. *Marketing Letters* 18, 4 (2007), 211–221.

[17] Dwain L Eckberg. 1997. Sympathovagal balance: a critical appraisal. *Circulation* 96, 9 (1997), 3224–3232.

[18] Aviv Elor, Michael Powell, Evanjelin Mahmoodi, Mircea Teodorescu, and Sri Kurniawan. 2021. Gaming Beyond the Novelty Effect of Immersive Virtual Reality for Physical Rehabilitation. *IEEE Transactions on Games* 14, 1 (2021), 107–115.

[19] Thomas Friedrichs, Carolin Zschippig, Marc Herrlich, Benjamin Walther-Franks, Rainer Malaka, and Kerstin Schill. 2015. Simple games–complex emotions: Automated affect detection using physiological signals. In *International Conference on Entertainment Computing*. Springer, 375–382.

[20] Alberto Greco, Gaetano Valenza, Antonio Lanata, Enzo Pasquale Scilingo, and Luca Citi. 2015. cvxEDA: A convex optimization approach to electrodermal activity processing. *IEEE Transactions on Biomedical Engineering* 63, 4 (2015), 797–804.

[21] Alberto Greco, Gaetano Valenza, and Enzo Pasquale Scilingo. 2016. *Advances in Electrodermal activity processing with applications for mental health*. Springer.

[22] Gillian A Hawker, Samra Mian, Tetyana Kendzerska, and Melissa French. 2011. Measures of adult pain: Visual analog scale for pain (vas pain), numeric rating scale for pain (nrs pain), mcgill pain questionnaire (mpq), short-form mcgill pain questionnaire (sf-mpq), chronic pain grade scale (cpgs), short form-36 bodily pain scale (sf-36 bps), and measure of intermittent and constant osteoarthritis pain (icoap). *Arthritis care & research* 63, S11 (2011), S240–S252.

[23] Dennis Held, Sascha Meudt, and Friedhelm Schwenker. 2016. Bimodal recognition of cognitive load based on speech and physiological changes. In *IAPR Workshop on Multimodal Pattern Recognition of Social Signals in Human-Computer Interaction*. Springer, 12–23.

[24] Simon M Hofmann, Felix Klotzsche, Alberto Mariola, Vadim Nikulin, Arno Villringer, and Michael Gaebler. 2021. Decoding subjective emotional arousal from EEG during an immersive Virtual Reality experience. *Elife* 10 (2021), e64812.

[25] Marko Horvat, Marko Dobrinić, Matej Novosel, and Petar Jerčić. 2018. Assessing emotional responses induced in virtual reality using a consumer EEG headset: A preliminary report. In *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*. IEEE, 1006–1010.

[26] Amir Jalilifard, Ednaldo Brigante Pizzolato, and Md Kafiul Islam. 2016. Emotion classification using single-channel scalp-EEG recording. In *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. 845–849. https://doi.org/10.1109/EMBC.2016.7590833

[27] Markad V Kamath, Mari Watanabe, and Adrian Upton. 2012. Heart rate variability (HRV) signal analysis: clinical applications. (2012).

[28] Stamos Katsigiannis and Naeem Ramzan. 2018. DREAMER: A Database for Emotion Recognition Through EEG and ECG Signals From Wireless Low-cost Off-the-Shelf Devices. *IEEE Journal of Biomedical and Health Informatics* 22, 1 (2018), 98–107. https://doi.org/10.1109/JBHI.2017.2688239

[29] Sander Koelstra, Christian Muhl, Mohammad Soleymani, Jong-Seok Lee, Ashkan Yazdani, Touradj Ebrahimi, Thierry Pun, Anton Nijholt, and Ioannis Patras. 2012. DEAP: A Database for Emotion Analysis ;Using Physiological Signals. *IEEE Transactions on Affective Computing* 3, 1 (2012), 18–31. https://doi.org/10.1109/T-AFFC.2011.15

[30] Barry Kort, Rob Reilly, and Rosalind W Picard. 2001. An affective model of interplay between emotions and learning: Reengineering educational pedagogy-building a learning companion. In *Proceedings IEEE international conference on advanced learning technologies*. IEEE, 43–46.

[31] Ilkka Kosunen, Mikko Salminen, Simo Järvelä, Antti Ruonala, Niklas Ravaja, and Giulio Jacucci. 2016. RelaWorld: Neuroadaptive and Immersive Virtual Reality Meditation System. In *Proceedings of the 21st International Conference on Intelligent User Interfaces* (Sonoma, California, USA) *(IUI '16)*. Association for Computing Machinery, New York, NY, USA, 208–217. https://doi.org/10.1145/2856767.2856796

[32] Krzysztof Kutt, Dominika Drążyk, Laura Żuchowska, Maciej Szelążek, Szymon Bobek, and Grzegorz J Nalepa. 2022. BIRAFFE2, a multimodal dataset for emotion-based personalization in rich affective game environments. *Scientific Data* 9, 1 (2022), 1–15.

[33] Peter J Lang, Margaret M Bradley, Bruce N Cuthbert, et al. 2005. *International affective picture system (IAPS): Affective ratings of pictures and instruction manual*. NIMH, Center for the Study of Emotion & Attention Gainesville, FL.

[34] Raymond Lavoie, Kelley Main, Corey King, and Danielle King. 2021. Virtual experience, real consequences: the potential negative emotional consequences of virtual reality gameplay. *Virtual Reality* 25, 1 (2021), 69–81.

[35] Benjamin J Li, Jeremy N Bailenson, Adam Pines, Walter J Greenleaf, and Leanne M Williams. 2017. A public database of immersive VR videos with corresponding ratings of arousal, valence, and correlations between head movements and self report measures. *Frontiers in psychology* 8 (2017), 2116.

[36] Yi Li, Adel S. Elmaghraby, Ayman El-Baz, and Estate M. Sokhadze. 2015. Using physiological signal analysis to design affective VR games. In *2015 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*. 57–62. https://doi.org/10.1109/ISSPIT.2015.7394401

[37] Chin-Teng Lin, I-Fang Chung, Li-Wei Ko, Yu-Chieh Chen, Sheng-Fu Liang, and Jeng-Ren Duann. 2007. EEG-based assessment of driver cognitive responses in a dynamic virtual-reality driving environment. *IEEE Transactions on Biomedical Engineering* 54, 7 (2007), 1349–1352.

[38] Mary F Macedonio, Thomas D Parsons, Raymond A Digiuseppe, Brenda A Weiderhold, and Albert A Rizzo. 2007. Immersiveness and physiological arousal within panoramic video-based virtual reality. *Cyberpsychology & Behavior* 10, 4 (2007), 508–515.

[39] Iris B Mauss and Michael D Robinson. 2009. Measures of emotion: A review. *Cognition and emotion* 23, 2 (2009), 209–237.

[40] Cameron McCarthy, Nikhilesh Pradhan, Calum Redpath, and Andy Adler. 2016. Validation of the Empatica E4 wristband. In *2016 IEEE EMBS international student conference (ISC)*. IEEE, 1–4.

[41] Robert R McCrae and Oliver P John. 1992. An introduction to the five-factor model and its applications. *Journal of personality* 60, 2 (1992), 175–215.

[42] Ben Meuleman and David Rudrauf. 2021. Induction and Profiling of Strong Multi-Componential Emotions in Virtual Reality. *IEEE Transactions on Affective Computing* 12, 1 (Jan. 2021), 189–202. https://doi.org/10.1109/TAFFC.2018.2864730 Conference Name: IEEE Transactions on Affective Computing.

[43] Diego Monteiro, Hai-Ning Liang, Wenge Xu, Marvin Brucker, Vijayakumar Nanjappan, and Yong Yue. 2018. Evaluating enjoyment, presence, and emulator sickness in VR games based on first- and third- person viewing perspectives. *Computer Animation and Virtual Worlds* 29, 3-4 (2018), e1830.

https://doi.org/10.1002/cav.1830 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/cav.1830 e1830 cav.1830.

[44] Mimma Nardelli, Gaetano Valenza, Alberto Greco, Antonio Lanata, and Enzo Pasquale Scilingo. 2015. Recognizing Emotions Induced by Affective Sounds through Heart Rate Variability. *IEEE Transactions on Affective Computing* 6, 4 (2015), 385–394. https://doi.org/10.1109/TAFFC.2015.2432810

[45] Task Force of the European Society of Cardiology et al. 1996. Heart rate variability: standards of measurement, physiological interpretation and clinical use. *circulation* 93 (1996), 1043–1065.

[46] Charles Egerton Osgood, George J Suci, and Percy H Tannenbaum. 1957. *The measurement of meaning*. Number 47. University of Illinois press.

[47] Federica Pallavicini and Alessandro Pepe. 2019. Comparing player experience in video games played in virtual reality or on desktop displays: Immersion, flow, and positive emotions. In *Extended Abstracts of the Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts*. 195–210.

[48] Renato Panda, Ricardo Malheiro, and Rui Pedro Paiva. 2020. Novel Audio Features for Music Emotion Recognition. *IEEE Transactions on Affective Computing* 11, 4 (2020), 614–626. https://doi.org/10.1109/TAFFC.2018.2820691

[49] W Gerrod Parrott. 2001. *Emotions in social psychology: Essential readings*. psychology press.

[50] Randy Pausch, Dennis Proffitt, and George Williams. 1997. Quantifying immersion in virtual reality. In *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*. 13–18.

[51] Xiaolan Peng, Jin Huang, Alena Denisova, Hui Chen, Feng Tian, and Hongan Wang. 2020. A palette of deepened emotions: Exploring emotional challenge in virtual reality games. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–13.

[52] Erik Peper, Rick Harvey, Lin I-Mei, Hana Tylova, and Donald Moss. 2007. Is There More to Blood Volume Pulse Than Heart Rate Variability, Respiratory Sinus Arrhythmia, and Cardiorespiratory Synchrony?. *Biofeedback* 35, 2 (2007), 54 – 61. https://ezproxy.lib.ucalgary.ca/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=26393327&site=ehost-live

[53] Paweł Pławiak. 2018. Novel methodology of cardiac health recognition based on ECG signals and evolutionary-neural system. *Expert Systems with Applications* 92 (2018), 334–349.

[54] Robert Plutchik. 2001. The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American scientist* 89, 4 (2001), 344–350.

[55] Denise Quesnel, Steve Dipaola, and Bernhard E Riecke. 2017. Deep Learning for Classification of Peak Emotions within Virtual Reality Systems. (2017). https://doi.org/10.13140/RG.2.2.15443.27686 Publisher: Unpublished.

[56] Denise Quesnel, Steve DiPaola, Bernhard E Riecke, et al. 2017. Deep learning for classification of peak emotions within virtual reality systems. *1st Artificial Intelligence Meets Virtual and Augmented Worlds* (2017), 6–11.

[57] Greg M Reger, Kevin M Holloway, Colette Candy, Barbara O Rothbaum, JoAnn Difede, Albert A Rizzo, and Gregory A Gahm. 2011. Effectiveness of virtual reality exposure therapy for active duty soldiers in a military mental health clinic. *Journal of traumatic stress* 24, 1 (2011), 93–96.

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

[59] Katja Rogers, Giovanni Ribeiro, Rina R Wehbe, Michael Weber, and Lennart E Nacke. 2018. Vanishing importance: studying immersive effects of game audio perception on player experiences in virtual reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–13.

[60] James A Russell. 1980. A circumplex model of affect. *Journal of personality and social psychology* 39, 6 (1980), 1161.

[61] Steven Schmidt, Stefan Uhrig, and Domenic Reuschel. 2020. Investigating the Relationship of Mental Immersion and Physiological Measures during Cloud Gaming. In *2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*. 1–6. https://doi.org/10.1109/QoMEX48832.2020.9123133

[62] Francesco Sessa, Giovanni Messina, Anna Valenzano, Antonietta Messina, Monica Salerno, Gabriella Marsala, Giuseppe Bertozzi, Aurora Daniele, Vincenzo Monda, and Raffaele Russo. 2018. Sports training and adaptive changes. *Sport Sciences for Health* 14, 3 (2018), 705–708.

[63] Cornelia Setz, Bert Arnrich, Johannes Schumm, Roberto La Marca, Gerhard Tröster, and Ulrike Ehlert. 2009. Discriminating stress from cognitive load using a wearable EDA device. *IEEE Transactions on information technology in biomedicine* 14, 2 (2009), 410–417.

[64] William J Shelstad, Dustin C Smith, and Barbara S Chaparro. 2017. Gaming on the rift: How virtual reality affects game user satisfaction. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 61. SAGE Publications Sage CA: Los Angeles, CA, 2072–2076.

[65] Mohammad Soleymani, Jeroen Lichtenauer, Thierry Pun, and Maja Pantic. 2011. A multimodal database for affect recognition and implicit tagging. *IEEE transactions on affective computing* 3, 1 (2011), 42–55.

[66] Mohammad Soleymani, Jeroen Lichtenauer, Thierry Pun, and Maja Pantic. 2012. A Multimodal Database for Affect Recognition and Implicit Tagging. *IEEE Transactions on Affective Computing* 3, 1 (2012), 42–55. https://doi.org/10.1109/T-AFFC.2011.25

[67] Nazmi Sofian Suhaimi, James Mountstephens, and Jason Teo. 2022. A Dataset for Emotion Recognition Using Virtual Reality and EEG (DER-VREEG): Emotional State Classification Using Low-Cost Wearable VR-EEG Headsets. *Big Data and Cognitive Computing* 6, 1 (2022). https://doi.org/10.3390/bdcc6010016

[68] John Sweller. 2011. Cognitive load theory. In *Psychology of learning and motivation*. Vol. 55. Elsevier, 37–76.

[69] Luma Tabbaa, Ryan Searle, Saber Mirzaee Bafti, Md Moinul Hossain, Jittrapol Intarasisrisawat, Maxine Glancy, and Chee Siang Ang. 2022. VREED: Virtual Reality Emotion Recognition Dataset Using Eye Tracking & Physiological Measures. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 4 (Dec. 2022), 178:1–178:20. https://doi.org/10.1145/3495002

[70] Chek Tien Tan, Tuck Wah Leong, Songjia Shen, Christopher Dubravs, and Chen Si. 2015. Exploring gameplay experiences on the Oculus Rift. In *Proceedings of the 2015 annual symposium on computer-human interaction in play*. 253–263.

[71] Samarth Tripathi, Shrinivas Acharya, Ranti Dev Sharma, Sudhanshu Mittal, and Samit Bhattacharya. 2017. Using deep and convolutional neural networks for accurate emotion classification on deap dataset.. In *Twenty-ninth IAAI conference*.

[72] Gaetano Valenza, Antonio Lanata, and Enzo Pasquale Scilingo. 2012. The Role of Nonlinear Dynamics in Affective Valence and Arousal Recognition. *IEEE Transactions on Affective Computing* 3, 2 (2012), 237–249. https://doi.org/10.1109/T-AFFC.2011.30

[73] Marcel Walch, Julian Frommel, Katja Rogers, Felix Schüssel, Philipp Hock, David Dobbelstein, and Michael Weber. 2017. Evaluating VR Driving Simulation from a Player Experience Perspective. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI EA '17)*. Association for Computing Machinery, New York, NY, USA, 2982–2989. https://doi.org/10.1145/3027063.3053202

[74] Mark D Wiederhold, Megan Crisci, Vrajeshri Patel, Makoto Nonaka, and Brenda K Wiederhold. 2019. Physiological monitoring during augmented reality exercise confirms advantages to health and well-being. *Cyberpsychology, Behavior, and Social Networking* 22, 2 (2019), 122–126.

[75] Graham Wilson and Mark McGill. 2018. Violent video games in virtual reality: Re-evaluating the impact and rating of interactive experiences. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play*. 535–548.

[76] Mahendra Yadava, Pradeep Kumar, Rajkumar Saini, Partha Pratim Roy, and Debi Prosad Dogra. 2017. Analysis of EEG signals and its application to neuromarketing. *Multimedia Tools and Applications* 76, 18 (2017), 19087–19111.

[77] Wanlu Yang, Kai Makita, Takashi Nakao, Noriaki Kanayama, Maro G Machizawa, Takafumi Sasaoka, Ayako Sugata, Ryota Kobayashi, Ryosuke Hiramoto, Shigeto Yamawaki, et al. 2018. Affective auditory stimulus database: An expanded version of the International Affective Digitized Sounds (IADS-E). *Behavior Research Methods* 50, 4 (2018), 1415–1429.

[78] Georgios N Yannakakis. 2018. Enhancing health care via affective computing. (2018).

[79] Caglar Yildirim. 2019. Cybersickness during VR gaming undermines game enjoyment: A mediation model. *Displays* 59 (2019), 35–43. https://doi.org/10.1016/j.displa.2019.07.002

[80] Caglar Yildirim, Michael Carroll, Daniel Hufnal, Theodore Johnson, and Sylvia Pericles. 2018. Video game user experience: to VR, or not to VR?. In *2018 IEEE Games, Entertainment, Media Conference (GEM)*. IEEE, 1–9.

[81] Shanbin Zhang, Guangyuan Liu, and Xiangwei Lai. 2015. Classification of evoked emotions using an artificial neural network based on single, short-term physiological signals. *Journal of Advanced Computational Intelligence and Intelligent Informatics* 19, 1 (2015), 118–126.

[82] Wenzhuo Zhang, Lin Shu, Xiangmin Xu, and Dan Liao. 2017. Affective virtual reality system (AVRS): design and ratings of affective VR scenes. In *2017 International Conference on Virtual Reality and Visualization (ICVRV)*. IEEE, 311–314.