An Analysis of Physiological and Psychological Responses in Virtual Reality and Flat Screen Gaming

Ritik Vatsal*, Shrivatsa Mishra*, Rushil Thareja, Mrinmoy Chakrabarty, Ojaswa Sharma, Jainendra Shukla
Indraprastha Institute of Information Technology, Delhi, India

***Abstract*—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 robust analysis of the physiological effects between VR and flatscreen (FS) gaming. This paper introduces the first systematic comparison and analysis of emotional and physiological responses to commercially available games in VR and FS environments. To elicit these responses, 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 Virtual Reality and Flat Screen (VRFS) dataset, containing over 15 hours of multimodal data comparing FS and VR gaming across different games, is also made publicly available for research purposes. Our analysis provides valuable insights for further investigations into the physiological and emotional effects of VR and FS gaming. The dataset is available at - GitHub link.**

***Index Terms*—Multi-modal recognition, Physiological signals, Artificial, augmented, and virtual realities, Games.**

I. Introduction

EMOTIONS in Virtual Reality (VR) tend to be different than those in traditional Flatscreen (FS) gaming because VR provides a more immersive and sensory experience. In traditional FS gaming, players view the game on a two-dimensional screen, which can create a sense of detachment between the player and the game world. However, in VR, players are placed in a fully-realized three-dimensional environment and are able to interact physically with the game world, which creates a stronger sense of presence and immersion [1]. Although prior research has examined the impact of both VR [2]–[7], and FS gaming [8]–[10] on individuals' physiological signals, these investigations have been conducted in distinct settings, leading to different experimental conditions and outcomes. Through our study, the utilization of both measures in conjunction has allowed for the attainment of a comprehensive and holistic perspective on the two distinct gaming methodologies that were subject to testing. This approach has not only allowed us to pinpoint the design obstacles and possibilities that are pivotal to the advancement of the gaming experience but also to obtain a more thorough understanding of the subject matter.

The first two authors (*) contributed equally to this work.

Recent advances in VR technology have ushered in a new era of VR gaming. As a result, many interactive experiences traditionally available in the conventional FS space are now moving to VR. Many traditional desktop games, primarily the first-person shooter [2], [3] and horror-adventure type games [4], [5], 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.

Emotions play a critical part in our lives, influencing our decision-making, perception, social interactions, learning, memory, and creativity [11], [12]. Understanding emotional responses helps researchers to design better user experiences and improve the usability, acceptability, and accessibility of technologies. We found only a few studies that have explored *emotional challenges* in gaming [13], [14]. Researchers have often used several physiological signals to predict arousal, valence, or even specific emotions by detecting physiological changes in the human body [15]. Due to the longer history and greater accessibility, emotions in FS games have been extensively researched, exploring their importance for player engagement and enjoyment. Studies have investigated factors like game content, player experience, personality, and culture, [8], [9] and the use of physiological and behavioral measures has provided deeper insight into players' emotional responses [10].

Our primary novel contribution in this work is to present the first-ever systematic comparison of physiological effects between VR and FS gaming, identifying 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. As part of this study, we have curated a 15+ hour multimodal affective dataset, named **V**irtual **R**eality and **F**lat**S**creen (VRFS). This novel dataset has been made publicly accessible for research purposes. The dataset for this study can be found here - GitHub link.

II. Related work

This study utilizes a methodical approach of literature review that is composed of three separate sections to thoroughly investigate different facets of the research problem.

Section II-A scrutinizes diverse analytical techniques utilized in emotion research, including physiological measurements and self-report measures. The subsequent section II-B conducts a comprehensive evaluation of the existing literature on emotions in VR. Lastly, section II-C presents an extensive analysis of recent emotional datasets that are germane to the research and explicates how this study diverges from them, offering a distinctive contribution to the field.

A. Analysis of Emotions

Emotional responses can be assessed via three different methods: self-reports from the participant (collected, for example, through questionnaires or interviews), physiological changes in the participant's body (e.g., heart rate or skin conductivity), and directly observable behaviors (e.g., facial expressions or eye gaze) [16]. Among these methods, self-reports are often subject to several biases, and limitations [17], 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 [18], the wheel of emotions [19], 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 [20]. CMA representation is easier to understand and use than the VAD model. The CMA is based on a circular structure, which allows for a more intuitive representation of emotions.

Additionally, subjective measures like Self-Assessment Manikin (SAM) and Visual Analog Scale (VAS) are commonly employed to assess arousal and valence in VR [21], [22]. SAM and VAS offer a convenient and reliable way to quantify emotional responses, aiding researchers in understanding the emotional impact of VR environments and applications.

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 offer additional unique information regarding users' interactions with and responses to a system. Researchers have also utilized multiple sensors to help evaluate and monitor human physiological signals. Signals like EDA [23], [24], ECG [25], [26], and Electroencephalogram (EEG) [11], [27], [28] have been widely used. The use of these sensors is showcased in a wide variety of applications, including medical [29], [30], neuromarketing [31], [32], sports training [33], [34] and many more. Additionally, researchers have also proposed the use of these physiological signals as controllers in games to enhance connect among multiple players [35]. 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 [18] to a size axis model proposed by Kort et al. [37]. This led to the development of the Valence Arousal and Dominance Model by Osgood, Suci, and Tannenbaum [38].

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 [39].

B. Emotions in VR

Emotional analysis for immersion in VR has been much studied recently. Driving scenarios are commonly used in VR research as a means to study and understand human perception, due to the familiar and relatively realistic experience they offer. 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 [7], [40], [41]. 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 [42]. Researchers have established that EEG signals from the brain are a relevant metric for decoding emotional arousal [43]. One study tried to recreate traffic-light-based responses in VR and tracked the system's effectiveness with EEG signals from the brain [44]. They showed that in an average of eight participants, 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 [45]. With these results, researchers have looked into VR as a medium for required emotion elicitation as an upgrade to existing stimuli. Meuleman et al. [46] 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 [15]. VREED [47], 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 and motion sickness in first and third-person points of view in VR [48], 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 [49]–[51]. Other factors such as *cybersickness* also come into play when exploring emotions in VR gaming [52]. Overall positive emotions such as *hope, courage, relaxation, and calmness* are more associated with gaming in VR [14]; 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 [53].

C. 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) [54] 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) [55] dataset. It augments the previous International Affective Digitized Sounds (IADS) [56] 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 [57]. 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, MAHNOB-HCI [58] 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 providing a valuable resource for analyzing emotion recognition and affective computing. Another dataset called the Database for Emotion Analysis using Physiological signals (DEAP) [59] analyzes affective music videos, their reported ratings, and the recorded EEG data of the participants. Similarly, the Database for Emotion Recognition through EEG and ECG Signals from Wireless Low-cost Off-the-Shelf Devices (DREAMER) [60] analyzes the ECG and EEG data from participants experiencing audio-visual stimuli. The MEG-Based Multimodal Database for Decoding Affective Physiological Responses (DECAF) [61] 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) [62] 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 [47] is one of the first datasets that analyzes 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) [63], 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 [64]. In addition to self-reporting, some datasets have also used physiological data collected via EEG for emotional classification [65], [66].

The review outlined in section II-B establishes that there is a dearth of rigorous analysis regarding the physiological effects associated with VR and FS gaming. This deficiency in the study of physiological and psychological reactions to both VR and FS gaming can be attributed to the unavailability of standardized datasets that offer emotional responses for both types of gaming. Section II-C of the review highlights the significant limitation of many affective datasets, including DEAP [59] and MAHNOB-HCI [58], in that they do not utilize VR stimuli to elicit emotions, despite its well-known ability to provide immersive experiences. Additionally, existing datasets that analyze VR stimuli rely on different stimulus modalities, such as 360° videos [38] or pictures embedded in VR environments such as 360° videos [64] or pictures embedded in VR environments [66]. Most affective datasets primarily focus on either physiological or psychological responses, making systematic comparisons between VR and non-VR responses challenging and hampering the ability to draw comprehensive conclusions. This limitation emphasizes a gap in the field and presents an important opportunity for researchers to perform more comprehensive and standardized investigation that could improve the understanding of individuals playing games in VR and aid in enhancing the experience for new users. To this end, we have also made the entire VRFS dataset publicly accessible.

III. Stimuli Selection Procedure

The selection of effective games is essential to elicit the appropriate affective responses; therefore, our selection process underwent multiple stages to select the best games. We first shortlisted commercially available games that 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. Our game selection process is summarized below.

- **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.
– SAM [arousal, valence, dominance] and VAS [joy, anger, calmness, sadness, disgust, relaxation, happiness, anxiousness, fear, and dizziness] were recorded to select the game.

A. 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.

B. 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 (for example, see [67]), 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:

- SAM is a widely used state measurement tool [68]. 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.
- VAS [69] 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.

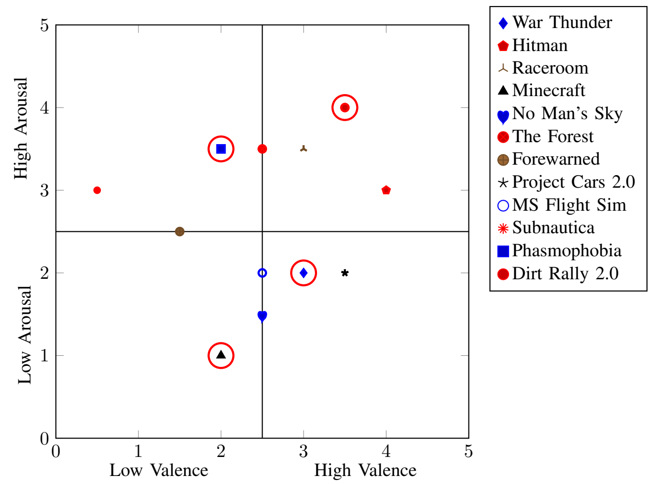

Fig. 1. Games plotted in respective quadrants according to arousal and valence results from pilot SAM study. Encircled games were selected for the final study.

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, as shown in Fig. 1. Arousal and valence ratings for each of the 12 games can be found in Table I. 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 I
REPORTED 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 2.0	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 2.0	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

$$r = \sqrt{Arousal^2 + Valence^2}, \text{ and}$$

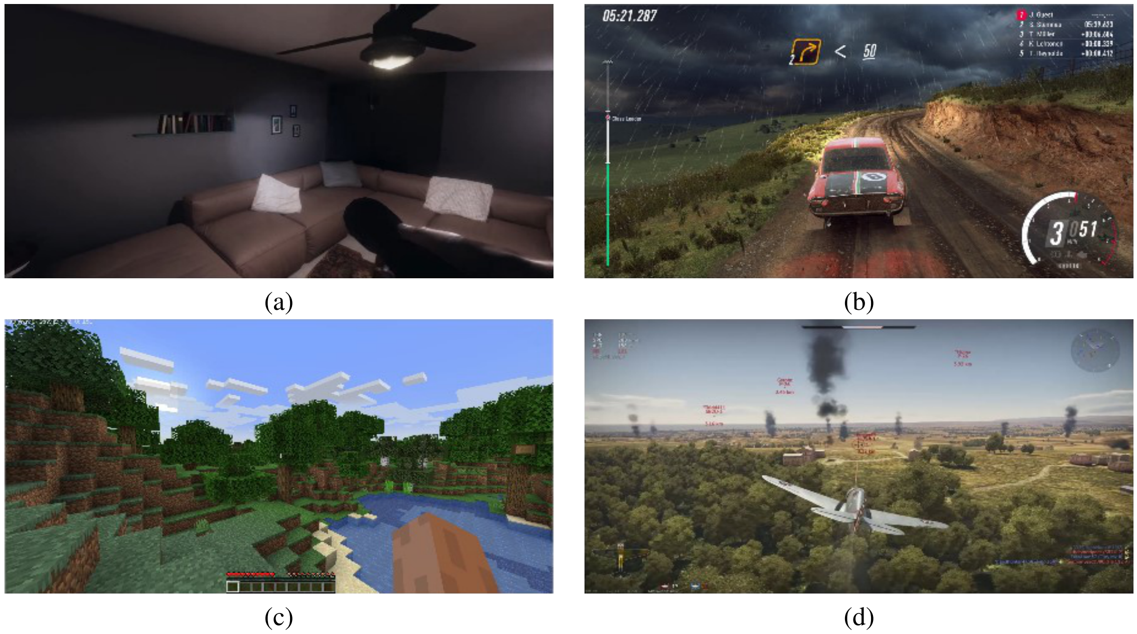

Fig. 2. In-game screenshots of selected four games selected games. (a) Phasmophobia: High Arousal Low Valence; (b) Dirt Rally 2.0: High Arousal High Valence; (c) Minecraft: Low Arousal Low Valence; (d) War Thunder: Low Arousal High Valence

$$d\theta = \left| \frac{\pi}{4} - \tan^{-1}\left(\frac{Arousal}{Valence}\right) \right|$$

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.
- *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.

IV. 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.

A. Ethics

The study was approved by the Institutional Review Board (IRB) of Indraprastha Institute of Information Technology, Delhi. All participants were provided with an overview of the study and a written consent was taken before the study was conducted. 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.

B. 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:

- 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.

C. 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.

D. 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 II. We measured the participants' BVP and EDA. These signals have been shown to represent physiological state accurately [47] 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 (BVP - 64 Hz, EDA - 4 Hz).

1) BVP: 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. The BVP in Empatica E4 is processed through the PPG sensor at 64 Hz. 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 [70]. We then extract the Low-frequency (LF) and High-frequency (HF) features from the HRV. The LF band (0.04–0.15 Hz) is affected by breathing from 3 to 9 bpm. [71] while the HF or respiratory band (0.15 – 0.40 Hz) is influenced by breathing from 9 to 24 bpm [72]. The ratio of LF to HF power (LF/HF ratio) has been proposed as an indicator of sympathovagal balance, reflecting an increase in values during states of "sympathetic dominance" induced by emotional and physiological stress [73]. Similar to previous research [74], we have employed the LF/HF ratio within the context of an ambulatory condition, which encompassed participants having unrestricted movement (with the exception of keeping the off-hand restricted) and accounting for the variability in breathing patterns.

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 was sampled at a frequency of 4 Hz, with the data being measured in micro siemens (μS). 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 [75], while SCRs depict higher-frequency changes directly related to an external stimulus [76]. Typically, SCR and heart rate are the best discriminators for arousal detection [77]. 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) [78].

TABLE II
DETAILS OF EXPERIMENTAL DATA COLLECTION.

Stimuli Selection Method	Based on the six pilot trial results
Participants 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-participant study (2 conditions (VR+FS) $\times$ 4 variables (4 Games))
Additional Materials	Self Reporting Questionnaire, and verbal instructions protocol

E. Apparatus and Setup

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

- *FS 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.
- *VR Headset:* An Oculus Rift S VR headset[1] was used for the VR experience. It comes with one headset and

[1] Oculus Rift S VR headset: https://www.oculus.com/rift-s/

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 are 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 and a 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 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).

F. Hypotheses

Overall, we expected VR games to be perceived as more immersive [79] 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.

1) **H1:** Previous research has suggested a positive correlation between heart rate and cognitive load [77], [80], [81]. Moreover, studies have shown higher immersion and increased cognitive load in virtual reality (VR) gaming compared to traditional first-person shooter (FS) gaming [82]. Consequently, we anticipated an elevated heart rate during VR gaming.
2) **H2:** A positive relationship is observed between emotional intensity and immersion [45]. 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.
3) **H3:** Due to the higher cognitive load involved, VR gaming induces greater emotional and physiological stress compared to FS gaming [83]. Additionally, emotional and physiological stress often leads to an increase in LF/HF sympathovagal balance values, indicating sympathetic dominance [73]. Hence, we anticipated an elevation in LF/HF ratio during VR gaming.
4) **H4:** A higher cognitive load is observed with learning novel information [84]. 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.
5) **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 [85], we expected an increased skin conductance (EDA) activity as well.
6) **H6:** Past literature establishes a correlation among valence, arousal, and dominance [86]. In the present context, we postulate a heightened prominence of this correlation within Virtual Reality (VR) gaming as compared to Full-Screen (FS) gaming.

G. Experimental Procedure

We collect and analyze qualitative psychological responses and quantitative physiological measures in a 2x4 mixed-participant 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 counterbalanced 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. The session ended when the participant completed the tasks for that specific game. 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.

[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/

H. Dataset

We use the FLIRT module [87] in Python to extract various statistical and signal features for the EDA and HRV data. FLIRT module preprocessed the data for artifact removal and noise filtering using the extended Kalman filter (EKF) for EDA data and the Malik [88], Kamarth [89] and Acar [90] rule for the heart rate. 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(see Table III). We provide anonymized ACC, BVP, EDA, HR, IBI, and TEMP data in the form of CSV files for both VR and FS. Each CSV file starts with a timestamp, followed by the frequency of the recording and the data collected over the recording. We have also added the script we used to get the results of this study.

TABLE III
NUMBER OF PARTICIPANTS AND AVERAGE TIME TAKEN FOR 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

V. RESULTS

A. Validation of Stimulation

Through our analysis, we intend to confirm that the chosen games indeed elicit the emotions assigned to them by the CMA quadrant. Table IV provides the mean (μ) and standard deviation (σ) for different games during FS gameplay, VR gameplay, as well as both types of gameplays taken together (VR+FS). We observe that the arousal and valence fall in the expected quadrant; however, further statistical significance analysis is required to confirm this. Considering both modalities taken together (VR+FS), we observe that higher reported arousal is evident for games rated as high arousal. (HVHA Median: 4.0 (4.0 - 5.0); LVHA Mdn: 3.5 (3.0 - 5.0). Similarly, low arousal is reported for games rated as Low arousal (HVLA Mdn: 3.0 (2.0 - 3.8)), LVLA Mdn: 2.0 (2.0 - 2.8)). This trend continues for valence as well with High valence rated games, generating a higher reported valence (HVHA Mdn: 4.0 (3.0 - 4.8); HVLA Mdn: 4.0 (3.0 - 4.0)) and vice-versa (LVHA Mdn: 3.0 (2.0 - 4.0); LVLA Mdn: 3.0 (2.2 - 4.0)). Taken independently, both the VR and FS modalities confer the above trends.

The Kruskal-Wallis ANOVA test is conducted on the data shown in Table IV 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(3): 30.67, p-value: 9.97e-7) across the games in all four CMA quadrants. The Kruskal-Wallis test result was followed up with posthoc Dunn's test; the results are presented in the table. 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 LVHA category, along with those playing games in the HVHA and the LVHA categories experiencing higher arousal than their counterparts playing games in the HVLA and the LVLA categories. Even though participants perceived the HVHA in FS as less arousing than in VR, the arousal ratings in this quadrant were still significantly higher than the LVLA of the same. In summary, the participants experienced four distinct emotional states over the differing valence and arousal dimensions (HVHA, HVLA, LVHA, LVLA) through a statically significant difference in reported arousal and valence.

TABLE IV
REPORTED AROUSAL AND VALENCE RATINGS ACCORDING TO THE CMA QUADRANT.

Intended CMA Quadrant	**Type**	**Arousal**		**Valence**		**r**	
		μ	σ	μ	σ	μ	σ
Low Valence Low Arousal (LVLA)	VR	2.00	0.94	3.00	1.25	3.72	1.25
	FS	2.33	0.47	3.56	0.96	4.27	0.98
	VR+FS	2.17	0.76	3.28	1.15	4.00	1.16
Low Valence High Arousal (LVHA)	VR	4.38	0.86	3.38	0.99	5.61	0.87
	FS	3.12	0.93	2.88	0.78	4.36	0.70
	VR+FS	3.75	1.09	3.12	0.93	4.99	1.01
High Valence Low Arousal (HVLA)	VR	3.86	0.83	4.00	0.76	5.62	0.72
	FS	2.14	0.64	3.57	0.90	4.24	0.79
	VR+FS	3.00	1.13	3.79	0.86	4.93	1.02
High Valence High Arousal (HVHA)	VR	4.44	0.68	4.22	1.23	6.25	0.69
	FS	4.22	0.42	3.22	0.79	5.36	0.53
	VR+FS	4.33	0.58	3.72	1.15	5.81	0.76

TABLE V
BONFERRONI-CORRECTED P-VALUES FOR POSTHOC DUNN'S TESTS ON SAM DATA OF ALL FOUR QUADRANTS

	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

B. Psychological Results

1) SAM: We show plots of the 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-Wilks test (for all: W>0.89, p-value<0.00009). Upon comparing the arousal between VR and FS gameplay using nonparametric Wilcoxon's signed-rank test, we found that the reported arousal in VR (Mdn: 4.00 (3.00 - 5.00)) was higher than FS (Mdn: 3.00 (2.00 - 4.00)), which was statistically significant (W(33) = 66.5, p-value = 0.007, Cliff's delta = -0.25). We also observed a higher value for valence in VR gameplay (Mdn: 4.00 (3.00 - 5.00)) as compared to FS gameplay (Mdn: 3.00 (3.00 - 4.00))

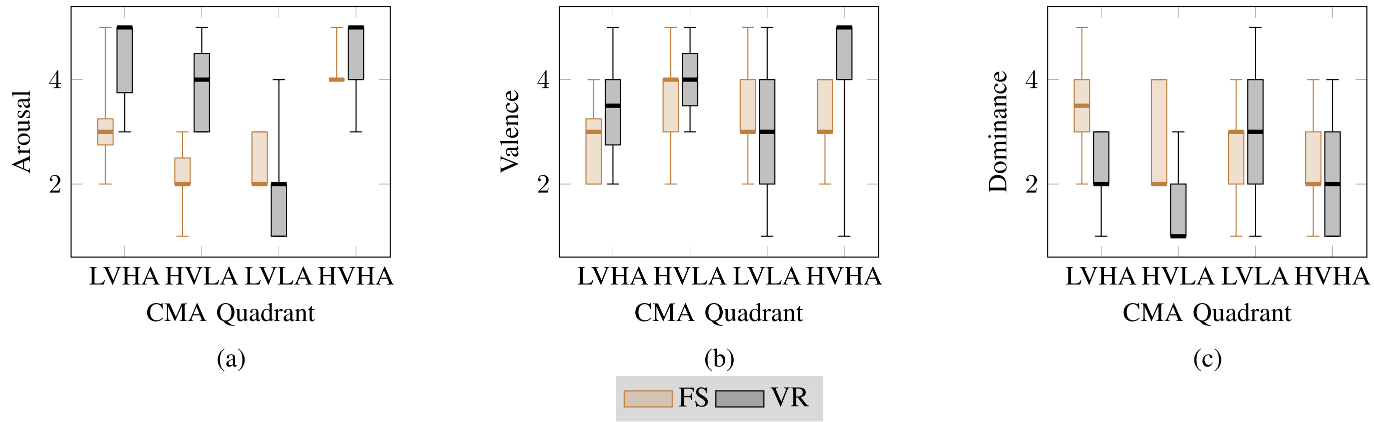

Fig. 3. Reported values of (a) arousal, (b) valence, and (c) dominance of players using SAM in different games in both VR and FS.

as shown in Fig. 3(b). However, the Wilcoxon signed-rank test found this difference statistically non-significant (W(33) = 75.5, p-value = 0.262, Cliff's delta = 0.27). In contrast, we observed that dominance was lower in VR gameplay (Mdn: 2.00 (1.00 - 3.00)) compared to FS gameplay (Mdn: 3.00 (2.00 - 4.00)), and this difference was found to be statistically significant (W(33) = 88.0, p-value = 0.02, Cliff's delta = 0.53). To ascertain the relationship between the reported arousal, valence, and dominance, a non-parametric Spearman's rank correlation was done, where we found that the dominance was weakly correlated to both arousal(ρ(64) = -0.27, p = 0.029) and valence(ρ(64) = -0.27, p = 0.031). Upon further inspection, we found a moderate negative correlation between dominance and valence for high-rated dominance cases (dominance $\geq$ 4, ρ(14) = -0.42, p = 0.12), which was not significant; however, there was also a strong negative correlation between the same for low-rated dominance cases (dominance $\leq$ 2, ρ(32) = -0.53, p = 0.01), which was statistically significant.

2) VAS: We plotted the mean and standard deviation recorded in VR and FS gameplay for the different emotions in Fig. 4 and reported emotions per CMA quadrant. The data collected for the emotions were found to be non-normally distributed upon conducting the Shapiro-Wilks test (p-value $<$ 0.02). Upon conducting the Wilcoxon signed-rank test, we observe the p-values as shown in Table VI, with the significant values marked in bold. 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(33) = 101; p-value = 0.004, Cliff's delta = 0.51; dizziness: W(33) = 8.5, p-value = 0.0, Cliff's delta = 0.53) 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.

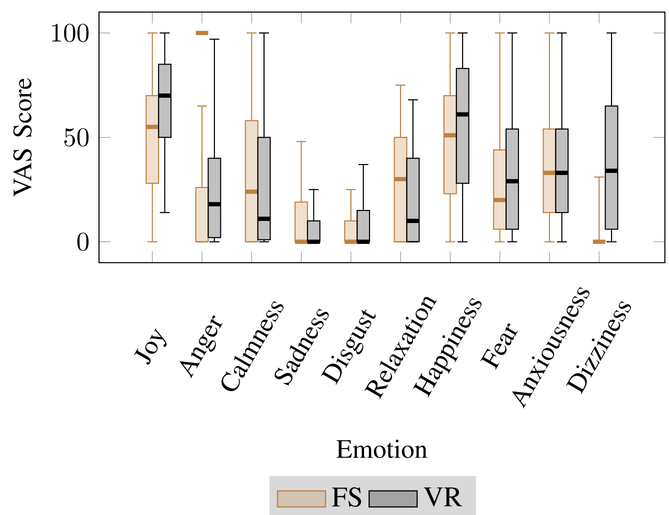

Fig. 4. Reported values of different emotions upon playing games in VR and FS using VAS.

TABLE VI
A COMPARISON OF SELF REPORTED LEVELS OF DIFFERENT EMOTIONS ON THE VAS TEST BETWEEN VR AND FS GAMEPLAY. EMOTIONS WITH STATISTICALLY SIGNIFICANT DIFFERENCES (P-VALUE $<$ 0.05) BETWEEN VR AND FS GAMEPLAY ARE SHOWN IN BOLD.

Emotion	**VR**		**FS**		***p*-value**	**Cliff's delta**
	μ	σ	μ	σ		
Joy	**63.45**	**26.01**	**48.39**	**26.62**	**0.0041**	**0.51**
Anger	25.73	26.53	20.48	25.19	0.3198	0.46
Calmness	26.18	27.52	31.88	31.11	0.1389	0.43
Sadness	9.42	17.11	11.67	18.24	0.3007	0.26
Disgust	12.79	22.98	10.33	22.14	0.2176	0.37
Relaxation	23.30	26.88	28.24	27.16	0.1708	0.35
Happiness	53.55	33.28	46.64	30.38	0.1469	0.43
Fear	34.88	31.65	27.15	23.75	0.2096	0.47
Anxiousness	37.55	31.58	35.27	26.79	0.9478	0.49
Dizziness	**37.27**	**32.27**	**5.48**	**14.76**	**0.0000**	**0.53**

C. Physiological Results

1) EDA: The raw EDA data is pre-processed using the convex EDA algorithm (cvxEDA) [76] to obtain the tonic (SCL) and phasic (SCR) components. The cvxEDA model is a physiologically inspired model which describes skin conductance as a combination of three factors: the phasic component, the tonic component, and additive white Gaussian noise. The cvxEDA algorithm uses convex optimization to break down the EDA signal into its different components and provides a comprehensive explanation of EDA. We normalized the data with the baseline using Z-score normalization. From SCR, we obtain the statistical features such as mean, standard deviation, minimum, maximum, and percentiles shown in Fig. 5. These results were found to be non-normal upon conducting the Shapiro-Wilks test (p-value = 0.00). While we observe higher SCR ratings during VR gameplay (Mdn: 0.45 (-0.03-3.54)) than in FS gameplay (Mdn: 0.02 (0.00-1.68)), the difference between the two is statistically nonsignificant (W(33) = 221, p-value = 0.288, Cliff's delta = 0.44) according to the Wilcoxon Signed-rank test.

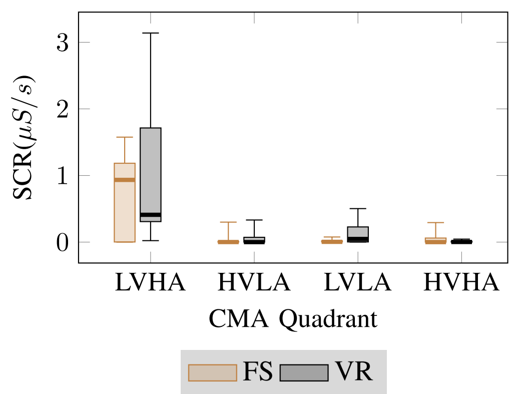

Fig. 5. SCR of players in VR and FS to different dames using EDA data.

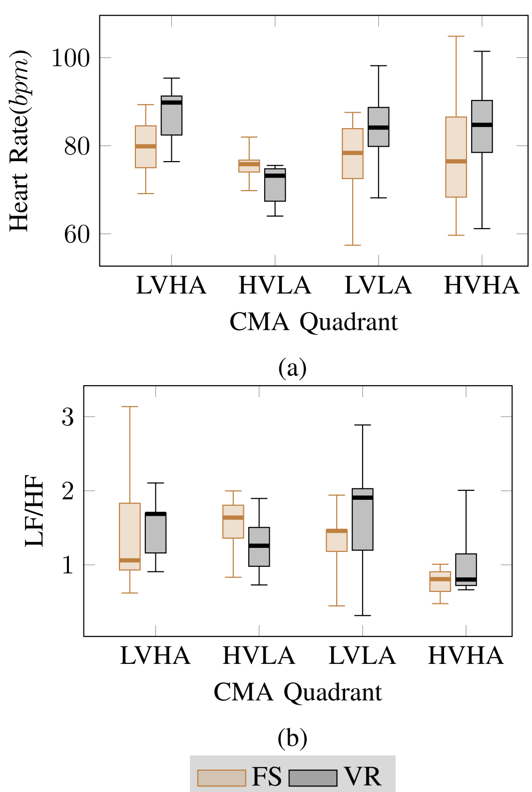

Fig. 6. (a) Observed heart rates of participants playing different games in VR and FS. (b) Observed LF/HF features of participants playing different games in VR and FS.

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. 6(a). The heart rate was normalized w.r.t. the baseline for each player. 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(32) = -2.5, p-value = 0.016, Cohen's d = 0.41) upon conducting a paired t-test on the data after confirming normality. Using the heart rate and the Inter Beat Interval (IBI), we extracted the LF and HF features. We then calculated the LF/HF ratio using these two and sorted them based on the different emotions as shown Fig. 6(b).

We found that the mean LF/HF values were also normally distributed according to the Shapiro-Wilks test (FS: p-value=0.21; VR: p-value=0.59). 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 not statistically significant (t(24) = -0.33, p-value = 0.74, Cohen's d = -0.09).

VI. Discussion

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

A. Inferences

We expected an elevated heart rate during VR gameplay according to our first hypothesis (H1) due to increased cognitive load in VR. Our analysis observed a higher mean heart rate in the VR scenario. A paired t-test indicates a statistically significant result (t(32) = -2.5, p-value = 0.016, Cohen's d = 0.41), validating H1. Therefore, our results align with those of Milinska et al. [81], who explain this phenomenon as a consequence of increased immersion which in turn leads to an additional cognitive burden on the user.

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 (arousal) during VR gameplay (Mdn: 4.00 (3.00 - 5.00)) is higher than FS gameplay (Mdn: 3.00 (2.00 - 4.00)), and this difference is statistically significant (W(33) = 75.5, p-value = 0.262, Cliff's delta(33) = 0.27) according to the Wilcoxon Signed Rank test. Additionally, when we compare the games in the different quadrants of the CMA plot, we observed 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 V.

According to our third hypothesis (H3), we expected that the LF/HF Sympathovagal balance would increase in the case of VR due to the increased emotional and physiological stress. 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).

Still, the difference was not statistically significant using a t-test (t(24) = -0.33, p-value = 0.74, Cohen's d = -0.09). This is most likely because the LF/HF ratio does not accurately measure cardiac sympathovagal balance [91].

The fourth hypothesis (H4) suggests that emotions felt by users will be less dominant in VR compared to FS displays. Our experimental data backs this hypothesis; the dominance of emotions has been found to be lesser in VR gameplay (Mdn: 2.00 (1.00 - 3.00)) than in FS gameplay (Mdn: 3.00 (2.00 - 4.00)). This difference was significant, as indicated by the Wilcoxon signed-rank test (W(33) = 88.0, p-value = 0.02, Cliff's delta = 0.54). 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.

According to our fifth hypothesis (H5), the participants should experience a higher level of arousal while in VR compared to being in a FS environment. Our study observed an increased SCR (VR Mdn: 0.45 (-0.03-3.54); FS Mdn: 0.02 (0-1.68)); the difference was not statistically significant. However, we did observe a statistically significant increase in the participants' arousal levels when playing in VR (Mdn: 4.00 (3.00-4.00)) compared to FS (Mdn: 3.00 (2.00-4.00)) (W(33) = 66.5, p-value = 0.007, Cliff's delta = 0.25). This effect may be partially attributed to the greater immersion, intensity, variability, and dynamic nature of VR experiences. We observe similar results with 3D VR eliciting a larger emotional response as compared to 2D VR [92].

In line with our sixth hypothesis (H6), we found evidence of a correlation between dominance and arousal as well as dominance and valence. This supports previous research that reported a quadratic relationship between arousal and dominance [86]. In addition, our results revealed a higher correlation between valence and dominance in VR ($\rho(33)$ = 0.34, p = 0.047) compared to FS games ($\rho(33)$ = 0.06, p = 0.73). We observe a similar trend between arousal and dominance (VR: $\rho(33)$ = 0.31, p = 0.07; FS: $\rho(33)$ = 0.002, p = 0.98). However, there is no correlation between arousal and valence, which might be due to the variance in the gameplay itself, and more research is needed about this effect.

B. Implications and Future Work

Our research provides valuable insights into user engagement and emotional response in both VR and FS gaming. These insights can be leveraged by developers to design platform-specific experiences that elicit intended emotions in end-users. Our findings can also inform the development of more engaging games for both entertainment and non-entertainment purposes, such as rehabilitation [93] and managing post-traumatic stress disorder [94].

By incorporating multimodal signals, developers can create adaptive gaming experiences that respond to users' emotions and physiological responses, for example, an adaptive physical therapy game that adjusts to the user's mental and physical state in real-time has immense potential [95], [96].

Our dataset can also be used by developers to design more engaging and immersive video games. Our findings indicate that emotional response is less dominant in VR compared to FS displays, suggesting that developers can explore ways to improve emotional engagement in VR games. By incorporating our findings into their game design, developers can enhance the emotional experience of players and improve the overall user experience.

While the physiological effects of VR have been investigated [97], the emotional reactions are not well-known. Our curated dataset, VRFS, can be used to further understand the emotional effects of gaming in VR and compare them with FS gaming. This knowledge can inform the development of a regulatory framework for gamification technology and help monitor and mitigate the emotional consequences of gaming.

C. 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 [98]. Therefore, research should explore how different demographics, based on age, gender, etc., are susceptible to comparative effects. Another limitation of this study is that participants were asked to keep their off-hand as still as possible while playing the game to prevent motion artifacts. This may have reduced comfort for some participants, and it is possible that more natural data could have been collected if participants had been allowed to move their off-hand freely. The imposed restriction on off-hand movement during data collection introduces a partially ambulatory context, which could potentially impact the interpretation of LF/HF ratio. Therefore, it is important to consider the results within this specific context, and further research may be necessary to validate the use of LF/HF ratio in partially ambulatory conditions. It would also be interesting to explore individual differences in psychological and physiological responses, for example, based on openness to experience [99], exposure beyond the novelty effect period [100], 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.

VII. Conclusion

In this study, we have conducted a comprehensive analysis comparing the emotional and physiological responses of VR and FS gaming. Through a pilot study involving six participants, we selected four games that covered all four quadrants of the valence-arousal space to elicit emotional responses. We then recorded the physiological activity, including blood volume pulse and electrodermal activity, and self-reported emotions of 33 participants during a user study. We observed that the emotional responses induced during the gameplay of the identified games were correctly perceived by the participants. We also found a significant difference in the arousal and valence ratings reported in the four quadrants, indicating that the participants experienced four distinct emotional states over the valence and arousal dimensions (HVHA, HVLA, LVHA, LVLA). These results suggest that the identified games successfully elicited a range of emotional responses in both VR and FS gaming environments. Moreover, our analysis revealed that VR gaming led to more pronounced emotional responses, higher levels of arousal, increased cognitive load and stress, and lower levels of dominance compared to FS gaming. These findings provide valuable insights into the differential emotional and physiological effects of VR and FS gaming and have important implications for future research in this area. Additionally, we have made available the curated VRFS dataset, which contains over 15 hours of multimodal data comparing FS and VR gaming across different games. We hope that this dataset will be useful for other researchers to test their methods, hypotheses, and algorithms on this novel dataset.

Acknowledgments

This study is partially supported by the Center For Design and New Media (a TCS Foundation Initiative supported by Tata Consultancy Services) and the Infosys Centre for Artificial Intelligence at IIIT Delhi. We are grateful to all the participants for their cooperation during the study.

References

[1] M. Slater, V. Linakis, M. Usoh, and R. Kooper, "Immersion, presence, and performance in virtual environments: An experiment with tri-dimensional chess," *ACM Virtual Reality Software and Technology (VRST)*, 06 1999.

[2] C. T. Tan, T. W. Leong, S. Shen, C. Dubravs, and C. Si, "Exploring gameplay experiences on the oculus rift," in *Proceedings of the 2015 annual symposium on computer-human interaction in play*, 2015, pp. 253–263.

[3] C. Yildirim, M. Carroll, D. Hufnal, T. Johnson, and S. Pericles, "Video game user experience: to vr, or not to vr?" in *2018 IEEE Games, Entertainment, Media Conference (GEM)*. IEEE, 2018, pp. 1–9.

[4] K. Rogers, G. Ribeiro, R. R. Wehbe, M. Weber, and L. E. Nacke, "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*, 2018, pp. 1–13.

[5] G. Wilson and M. McGill, "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*, 2018, pp. 535–548.

[6] M. Granato, D. Gadia, D. Maggiorini, and L. A. Ripamonti, "An empirical study of players' emotions in vr racing games based on a dataset of physiological data," *Multimedia Tools and Applications*, vol. 79, pp. 33 657–33 686, 2020.

[7] A. Dey, T. Piumsomboon, Y. Lee, and M. Billinghurst, "Effects of sharing physiological states of players in a collaborative virtual reality gameplay," in *Proceedings of the 2017 CHI conference on human factors in computing systems*, 2017, pp. 4045–4056.

[8] G. N. Yannakakis and A. Paiva, "Emotion in games," *Handbook on affective computing*, vol. 2014, pp. 459–471, 2014.

[9] P. Vorderer, C. Klimmt, and U. Ritterfeld, "Enjoyment: At the heart of media entertainment," *Communication theory*, vol. 14, no. 4, pp. 388–408, 2004.

[10] C. A. Anderson, A. Shibuya, N. Ihori, E. L. Swing, B. J. Bushman, A. Sakamoto, H. R. Rothstein, and M. Saleem, "Violent video game effects on aggression, empathy, and prosocial behavior in eastern and western countries: a meta-analytic review." *Psychological bulletin*, vol. 136, no. 2, p. 151, 2010.

[11] S. Tripathi, S. Acharya, R. D. Sharma, S. Mittal, and S. Bhattacharya, "Using deep and convolutional neural networks for accurate emotion classification on deap dataset." in *Twenty-ninth IAAI conference*, 2017.

[12] S. Zhang, G. Liu, and X. Lai, "Classification of evoked emotions using an artificial neural network based on single, short-term physiological signals," *Journal of Advanced Computational Intelligence and Intelligent Informatics*, vol. 19, no. 1, pp. 118–126, 2015.

[13] J. A. Bopp, K. Opwis, and E. D. Mekler, "An odd kind of pleasure: Differentiating emotional challenge in digital games," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, ser. CHI '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 1–12.

[14] X. Peng, J. Huang, A. Denisova, H. Chen, F. Tian, and H. Wang, "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*, 2020, pp. 1–13.

[15] D. Quesnel, S. DiPaola, and B. Riecke, "Deep learning for classification of peak emotions within virtual reality systems," *International SERIES on Information Systems and Management in Creative eMedia (CreMedia)*, no. 2017/2, pp. 6–11, 2018.

[16] I. B. Mauss and M. D. Robinson, "Measures of emotion: A review," *Cognition and emotion*, vol. 23, no. 2, pp. 209–237, 2009.

[17] M. Devaux and F. Sassi, "Social disparities in hazardous alcohol use: self-report bias may lead to incorrect estimates," *The European Journal of Public Health*, vol. 26, no. 1, pp. 129–134, 2016.

[18] W. G. Parrott, *Emotions in social psychology: Essential readings*. psychology press, 2001.

[19] R. Plutchik, "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*, vol. 89, no. 4, pp. 344–350, 2001.

[20] J. A. Russell, "A circumplex model of affect." *Journal of personality and social psychology*, vol. 39, no. 6, p. 1161, 1980.

[21] T. Xie, M. Cao, and Z. Pan, "Applying self-assessment manikin (sam) to evaluate the affective arousal effects of vr games," in *Proceedings of the 2020 3rd International Conference on Image and Graphics Processing*, ser. ICIGP '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 134–138.

[22] "Using virtual reality to control preoperative anxiety in ambulatory surgery patients: A pilot study in maxillofacial and plastic surgery," *Journal of Stomatology, Oral and Maxillofacial Surgery*, vol. 119, no. 4, pp. 257–261, 2018, 54th SFSCMFCO Congress.

[23] M. F. Macedonio, T. D. Parsons, R. A. Digiuseppe, B. A. Weiderhold, and A. A. Rizzo, "Immersiveness and physiological arousal within panoramic video-based virtual reality," *Cyberpsychology & Behavior*, vol. 10, no. 4, pp. 508–515, 2007.

[24] G. Valenza, A. Lanata, and E. P. Scilingo, "The role of nonlinear dynamics in affective valence and arousal recognition," *IEEE Transactions on Affective Computing*, vol. 3, no. 2, pp. 237–249, 2012.

[25] M. Nardelli, G. Valenza, A. Greco, A. Lanata, and E. P. Scilingo, "Recognizing emotions induced by affective sounds through heart rate variability," *IEEE Transactions on Affective Computing*, vol. 6, no. 4, pp. 385–394, 2015.

[26] R. Panda, R. Malheiro, and R. P. Paiva, "Novel audio features for music emotion recognition," *IEEE Transactions on Affective Computing*, vol. 11, no. 4, pp. 614–626, 2020.

[27] A. Jalilifard, E. B. Pizzolato, and M. K. Islam, "Emotion classification using single-channel scalp-eeg recording," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2016, pp. 845–849.

[28] I. Kosunen, M. Salminen, S. Järvelä, A. Ruonala, N. Ravaja, and G. Jacucci, "Relaworld: Neuroadaptive and immersive virtual reality meditation system," in *Proceedings of the 21st International Conference on Intelligent User Interfaces*, ser. IUI '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 208–217.

[29] P. Pławiak, "Novel methodology of cardiac health recognition based on ecg signals and evolutionary-neural system," *Expert Systems with Applications*, vol. 92, pp. 334–349, 2018.

[30] A. Greco, G. Valenza, and E. P. Scilingo, *Advances in Electrodermal activity processing with applications for mental health*. Springer, 2016.

[31] M. Yadava, P. Kumar, R. Saini, P. P. Roy, and D. Prosad Dogra, "Analysis of eeg signals and its application to neuromarketing," *Multimedia Tools and Applications*, vol. 76, no. 18, pp. 19 087–19 111, 2017.

[32] U. Cuesta, J. I. Niño, and L. Martínez-Martínez, "Neuromarketing: analysis of packaging using gsr, eye-tracking and facial expression," in *Papper presented at The European Conference on Media, Communication & Film*, 2018.

[33] F. Sessa, G. Messina, A. Valenzano, A. Messina, M. Salerno, G. Marsala, G. Bertozzi, A. Daniele, V. Monda, and R. Russo, "Sports

training and adaptive changes," *Sport Sciences for Health*, vol. 14, no. 3, pp. 705–708, 2018.

[34] E. Butkevičiūtė, L. Bikulčienė, T. Sidekerskienė, T. Blažauskas, R. Maskeliūnas, R. Damaševičius, and W. Wei, "Removal of movement artefact for mobile eeg analysis in sports exercises," *IEEE Access*, vol. 7, pp. 7206–7217, 2019.

[35] R. B. Robinson, E. Reid, A. E. Depping, R. Mandryk, J. C. Fey, and K. Isbister, "'in the same boat',: A game of mirroring emotions for enhancing social play," in *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, ser. CHI EA '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 1–4.

[36] Y. Li, A. S. Elmaghraby, A. El-Baz, and E. M. Sokhadze, "Using physiological signal analysis to design affective vr games," in *2015 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, 2015, pp. 57–62.

[37] B. Kort, R. Reilly, and R. W. Picard, "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, 2001, pp. 43–46.

[38] C. E. Osgood, G. J. Suci, and P. H. Tannenbaum, *The measurement of meaning*. University of Illinois press, 1957, no. 47.

[39] C. McCarthy, N. Pradhan, C. Redpath, and A. Adler, "Validation of the empatica e4 wristband," in *2016 IEEE EMBS international student conference (ISC)*. IEEE, 2016, pp. 1–4.

[40] M. Walch, J. Frommel, K. Rogers, F. Schüssel, P. Hock, D. Dobbelstein, and M. Weber, "Evaluating vr driving simulation from a player experience perspective," in *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, ser. CHI EA '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 2982–2989.

[41] S. Cao, K. Nandakumar, R. Babu, and B. Thompson, "Game play in virtual reality driving simulation involving head-mounted display and comparison to desktop display," *Virtual Reality*, vol. 24, no. 3, pp. 503–513, Sep 2020.

[42] F. Pallavicini and A. Pepe, "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*, 2019, pp. 195–210.

[43] S. M. Hofmann, F. Klotzsche, A. Mariola, V. Nikulin, A. Villringer, and M. Gaebler, "Decoding subjective emotional arousal from eeg during an immersive virtual reality experience," *Elife*, vol. 10, p. e64812, 2021.

[44] C.-T. Lin, I.-F. Chung, L.-W. Ko, Y.-C. Chen, S.-F. Liang, and J.-R. Duann, "Eeg-based assessment of driver cognitive responses in a dynamic virtual-reality driving environment," *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 7, pp. 1349–1352, 2007.

[45] G. Riva, F. Mantovani, C. S. Capideville, A. Preziosa, F. Morganti, D. Villani, A. Gaggioli, C. Botella, and M. Alcañiz, "Affective interactions using virtual reality: the link between presence and emotions," *Cyberpsychology & behavior*, vol. 10, no. 1, pp. 45–56, 2007.

[46] B. Meuleman and D. Rudrauf, "Induction and Profiling of Strong Multi-Componential Emotions in Virtual Reality," *IEEE Transactions on Affective Computing*, vol. 12, no. 1, pp. 189–202, Jan. 2021, conference Name: IEEE Transactions on Affective Computing.

[47] L. Tabbaa, R. Searle, S. M. Bafti, M. M. Hossain, J. Intarasisrisawat, M. Glancy, and C. S. Ang, "VREED: Virtual Reality Emotion Recognition Dataset Using Eye Tracking & Physiological Measures," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 5, no. 4, pp. 178:1–178:20, Dec. 2022.

[48] D. Monteiro, H.-N. Liang, W. Xu, M. Brucker, V. Nanjappan, and Y. Yue, "Evaluating enjoyment, presence, and emulator sickness in vr games based on first- and third- person viewing perspectives," *Computer Animation and Virtual Worlds*, vol. 29, no. 3-4, p. e1830, 2018.

[49] S. Schmidt, S. Uhrig, and D. Reuschel, "Investigating the relationship of mental immersion and physiological measures during cloud gaming," in *2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*, 2020, pp. 1–6.

[50] W. J. Shelstad, D. C. Smith, and B. S. Chaparro, "Gaming on the rift: How virtual reality affects game user satisfaction," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 61, no. 1. SAGE Publications Sage CA: Los Angeles, CA, 2017, pp. 2072–2076.

[51] S. Karaosmanoglu, K. Rogers, D. Wolf, E. Rukzio, F. Steinicke, and L. E. Nacke, "Feels like team spirit: Biometric and strategic interdependence in asymmetric multiplayer vr games," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, ser. CHI '21. New York, NY, USA: Association for Computing Machinery, 2021.

[52] C. Yildirim, "Cybersickness during vr gaming undermines game enjoyment: A mediation model," *Displays*, vol. 59, pp. 35–43, 2019.

[53] T. Friedrichs, C. Zschippig, M. Herrlich, B. Walther-Franks, R. Malaka, and K. Schill, "Simple games–complex emotions: Automated affect detection using physiological signals," in *International Conference on Entertainment Computing*. Springer, 2015, pp. 375–382.

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

[55] W. Yang, K. Makita, T. Nakao, N. Kanayama, M. G. Machizawa, T. Sasaoka, A. Sugata, R. Kobayashi, R. Hiramoto, S. Yamawaki *et al.*, "Affective auditory stimulus database: An expanded version of the international affective digitized sounds (iads-e)," *Behavior Research Methods*, vol. 50, no. 4, pp. 1415–1429, 2018.

[56] M. M. Bradley and P. J. Lang, "The international affective digitized sounds (; iads-2): Affective ratings of sounds and instruction manual," *University of Florida, Gainesville, FL, Tech. Rep. B-3*, 2007.

[57] R. Robinson, K. Wiley, A. Rezaeivahdati, M. Klarkowski, and R. L. Mandryk, ""let's get physiological, physiological!": A systematic review of affective gaming," in *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*, ser. CHI PLAY '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 132–147.

[58] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, "A multimodal database for affect recognition and implicit tagging," *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 42–55, 2012.

[59] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "Deap: A database for emotion analysis ;using physiological signals," *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 18–31, 2012.

[60] S. Katsigiannis and N. Ramzan, "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*, vol. 22, no. 1, pp. 98–107, 2018.

[61] M. K. Abadi, R. Subramanian, S. M. Kia, P. Avesani, I. Patras, and N. Sebe, "Decaf: Meg-based multimodal database for decoding affective physiological responses," *IEEE Transactions on Affective Computing*, vol. 6, no. 3, pp. 209–222, 2015.

[62] K. Kutt, D. Drążyk, L. Żuchowska, M. Szelążek, S. Bobek, and G. J. Nalepa, "Biraffe2, a multimodal dataset for emotion-based personalization in rich affective game environments," *Scientific Data*, vol. 9, no. 1, pp. 1–15, 2022.

[63] W. Zhang, L. Shu, X. Xu, and D. Liao, "Affective virtual reality system (avrs): design and ratings of affective vr scenes," in *2017 International Conference on Virtual Reality and Visualization (ICVRV)*. IEEE, 2017, pp. 311–314.

[64] B. J. Li, J. N. Bailenson, A. Pines, W. J. Greenleaf, and L. M. Williams, "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*, vol. 8, p. 2116, 2017.

[65] N. S. Suhaimi, J. Mountstephens, and J. Teo, "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*, vol. 6, no. 1, 2022.

[66] M. Horvat, M. Dobrinić, M. Novosel, and P. Jerčić, "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, 2018, pp. 1006–1010.

[67] R. Al Alam and N. Dibben, "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, 2021.

[68] M. M. Bradley and P. J. Lang, "Measuring emotion: the self-assessment manikin and the semantic differential," *Journal of behavior therapy and experimental psychiatry*, vol. 25, no. 1, pp. 49–59, 1994.

[69] G. A. Hawker, S. Mian, T. Kendzerska, and M. French, "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*, vol. 63, no. S11, pp. S240–S252, 2011.

[70] E. Peper, R. Harvey, L. I-Mei, H. Tylova, and D. Moss, "Is there more to blood volume pulse than heart rate variability, respiratory sinus arrhythmia, and cardiorespiratory synchrony?." *Biofeedback*, vol. 35, no. 2, pp. 54 – 61, 2007.

[71] M. V. Kamath, M. Watanabe, and A. Upton, "Heart rate variability (hrv) signal analysis: clinical applications," 2012.

[72] T. F. of the European Society of Cardiology *et al.*, "Heart rate variability: standards of measurement, physiological interpretation and clinical use," *circulation*, vol. 93, pp. 1043–1065, 1996.

[73] D. L. Eckberg, "Sympathovagal balance: a critical appraisal," *Circulation*, vol. 96, no. 9, pp. 3224–3232, 1997.

[74] F. Shaffer, R. McCraty, and C. L. Zerr, "A healthy heart is not a metronome: an integrative review of the heart's anatomy and heart rate variability," *Frontiers in psychology*, vol. 5, p. 1040, 2014.

[75] C. Setz, B. Arnrich, J. Schumm, R. La Marca, G. Tröster, and U. Ehlert, "Discriminating stress from cognitive load using a wearable eda device," *IEEE Transactions on information technology in biomedicine*, vol. 14, no. 2, pp. 410–417, 2009.

[76] A. Greco, G. Valenza, A. Lanata, E. P. Scilingo, and L. Citi, "cvxeda: A convex optimization approach to electrodermal activity processing," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 4, pp. 797–804, 2015.

[77] Y. S. Can, B. Arnrich, and C. Ersoy, "Stress detection in daily life scenarios using smart phones and wearable sensors: A survey," *Journal of biomedical informatics*, vol. 92, p. 103139, 2019.

[78] M. Benedek and C. Kaernbach, "A continuous measure of phasic electrodermal activity," *Journal of Neuroscience Methods*, vol. 190, no. 1, pp. 80–91, 2010.

[79] R. Pausch, D. Proffitt, and G. Williams, "Quantifying immersion in virtual reality," in *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, 1997, pp. 13–18.

[80] D. Held, S. Meudt, and F. Schwenker, "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, 2016, pp. 12–23.

[81] M. Malińska, K. Zużewicz, J. Bugajska, and A. Grabowski, "Heart rate variability (hrv) during virtual reality immersion," *International Journal of Occupational Safety and Ergonomics*, vol. 21, no. 1, pp. 47–54, 2015.

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

[83] D. Conway, I. Dick, Z. Li, Y. Wang, and F. Chen, "The effect of stress on cognitive load measurement," in *Human-Computer Interaction–INTERACT 2013: 14th IFIP TC 13 International Conference, Cape Town, South Africa, September 2-6, 2013, Proceedings, Part IV 14*. Springer, 2013, pp. 659–666.

[84] J. Sweller, "Cognitive load theory," in *Psychology of learning and motivation*. Elsevier, 2011, vol. 55, pp. 37–76.

[85] W. Boucsein, *Electrodermal activity*. Springer Science & Business Media, 2012.

[86] A. B. Warriner, V. Kuperman, and M. Brysbaert, "Norms of valence, arousal, and dominance for 13,915 english lemmas," *Behavior research methods*, vol. 45, pp. 1191–1207, 2013.

[87] S. Föll, M. Maritsch, F. Spinola, V. Mishra, F. Barata, T. Kowatsch, E. Fleisch, and F. Wortmann, "FLIRT: A Feature Generation Toolkit for Wearable Data," *Computer Methods and Programs in Biomedicine*, 2021.

[88] M. Malik, J. T. Bigger, A. J. Camm, R. E. Kleiger, A. Malliani, A. J. Moss, and P. J. Schwartz, "Heart rate variability: Standards of measurement, physiological interpretation, and clinical use," *European Heart Journal*, vol. 17, no. 3, pp. 354–381, 03 1996. [Online]. Available: https://doi.org/10.1093/oxfordjournals.eurheartj.a014868

[89] M. Kamath and E. Fallen, "Correction of the heart rate variability signal for ectopics and missing beats, in: Malik m., camm aj (eds.): Heart rate variability," *Armonk, NY Futura Pub. Co. Inc*, 1995.

[90] B. Acar, I. Savelieva, H. Hemingway, and M. Malik, "Automatic ectopic beat elimination in short-term heart rate variability measurement," *Computer Methods and Programs in Biomedicine*, vol. 63, no. 2, pp. 123–131, 2000. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S016926070000081X

[91] G. Billman, "The LF/HF ratio does not accurately measure cardiac sympatho-vagal balance," *Frontiers in Physiology*, vol. 4, 2013.

[92] S. Estupiñán, F. Rebelo, P. Noriega, C. Ferreira, and E. Duarte, "Can virtual reality increase emotional responses (arousal and valence)? a pilot study," in *International conference of design, user experience, and usability*. Springer, 2014, pp. 541–549.

[93] M. D. Wiederhold, M. Crisci, V. Patel, M. Nonaka, and B. K. Wiederhold, "Physiological monitoring during augmented reality exercise confirms advantages to health and well-being," *Cyberpsychology, Behavior, and Social Networking*, vol. 22, no. 2, pp. 122–126, 2019.

[94] G. M. Reger, K. M. Holloway, C. Candy, B. O. Rothbaum, J. Difede, A. A. Rizzo, and G. A. Gahm, "Effectiveness of virtual reality exposure therapy for active duty soldiers in a military mental health clinic," *Journal of traumatic stress*, vol. 24, no. 1, pp. 93–96, 2011.

[95] G. N. Yannakakis, "Enhancing health care via affective computing," 2018.

[96] G. Chanel, C. Rebetez, M. Bétrancourt, and T. Pun, "Emotion assessment from physiological signals for adaptation of game difficulty," *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 41, no. 6, pp. 1052–1063, 2011.

[97] R. Lavoie, K. Main, C. King, and D. King, "Virtual experience, real consequences: the potential negative emotional consequences of virtual reality gameplay," *Virtual Reality*, vol. 25, no. 1, pp. 69–81, 2021.

[98] A. Drolet, P. Williams, and L. Lau-Gesk, "Age-related differences in responses to affective vs. rational ads for hedonic vs. utilitarian products," *Marketing Letters*, vol. 18, no. 4, pp. 211–221, 2007.

[99] R. R. McCrae and O. P. John, "An introduction to the five-factor model and its applications," *Journal of personality*, vol. 60, no. 2, pp. 175–215, 1992.

[100] A. Elor, M. Powell, E. Mahmoodi, M. Teodorescu, and S. Kurniawan, "Gaming beyond the novelty effect of immersive virtual reality for physical rehabilitation," *IEEE Transactions on Games*, vol. 14, no. 1, pp. 107–115, 2021.