Comparative Evaluation of the EEG Performance Metrics and Player Ratings on the Virtual Reality Games

Pratheep Kumar Paranthaman*, Nikesh Bajaj†, Nicholas Solovey*, and David Jennings*
*Department of Computer Science, Elon University, Elon, USA
Email: {pparanthaman, nsolovey, djennings4}@elon.edu
†School of Electronic Engineering and Computer Science, Queen Mary University of London, London, UK
Email: n.bajaj@qmul.ac.uk

***Abstract*—The low-cost electroencephalogram (EEG) devices are widely used by researchers in human-computer interaction, video games, and software systems to evaluate the impact of interaction design on user emotions. However, the performance metrics of emotion states provided by a low-cost EEG device suffer several reliability and accuracy issues, which can mislead the design decisions of the developers. In this research, we combined the EEG device with three virtual reality games to investigate the reliability of performance metrics extracted from the EEG data. We conducted the experiment with 14 players using virtual reality games with ranging levels of in-game actions. Our analysis shows that there is a significant difference between performance metrics provided by the EEG device and the actual players' experience. Finally, we used ad-hoc linear models to estimate the level of players' emotion states directly from the raw EEG. We also show the different brain activity maps for individual emotions, which reveal the commonly known relation between brain activity and specific emotions.**

***Index Terms*—virtual reality, EEG, player emotions, brain activity map**

I. Introduction

Virtual Reality (VR) games offer higher levels of immersion by combining audiovisual elements, haptic feedback, and interactive gameplay mechanics. Extra care needs to be taken when it comes to designing games for VR, as some in-game activities (frequent body movements and hand actions) in VR can lead to exertion and poor user experience. Specifically, the action and first-person VR games can cause discomfort in users [1]. So it is crucial to understand the players' experience in VR games, as it could be an informative aspect for designing user-centered VR games that are comfortable, engaging and safe for VR consumers. Therefore, in this paper, we decided to examine the players' emotion states in virtual reality games.

A. Low-Cost EEG Devices

Our next step was to identify an objective method to capture players' emotion states from VR games. One standard approach to measure the objective experience of users is through understanding the brain activity using an Electroencephalogram (EEG) device. Low-cost portable EEG devices are non-invasive and can capture the electrical activity in the brain through the sensors that can be attached to the user's head [2], [3]. Through the recorded brain activity from an EEG device, several aspects of mental activities can be recorded such as activities on brain areas (frontal, parietal, occipital), muscle movements (eye blink, jaw clench), and frequency bands [3]. The rise of low-cost portable EEG makers like Muse, Emotiv, and Neurosky [4] has opened up potential research space for investigating user experience (based on cognitive states) in several domains. However, the accuracy of low-cost EEG devices vary upon several aspects such as headset connectivity, sensor hydration, and device specifications. One study has reported the maximum accuracy in the low-cost EEG device as 60.57% [5]. Another research study conducted by [6] compared the accuracy in drowsiness detection from available low-cost EEG devices like Emotiv Epoc, the Neurosky MindWave, the OpenBCI, and the InteraXon Muse and this study reported 79.4% accuracy from OpenBCI [6].

B. Research Aim

In this research work: 1) we combined VR and EEG to evaluate the players' emotion states during the gameplay activity, and 2) we investigated the reliability of the performance metrics offered by Emotiv EPOC X EEG headset. The performance metrics (PM) refers to the level of the six emotion states (engagement, excitement, stress, relaxation, focus, and interest) provided by the Emotiv EPOC X headset as a score between 0 to 1. The process involved in computing these six PM is unknown to the developers due to the internal abstraction involved in Emotiv's performance metrics algorithm. Since, there is an increased use of EEG-based analysis in number of studies in games research [7], [8], it is crucial to check the reliability of these performance metrics (PM) values generated by the Emotiv 's EEG device. Further it is also important to explore the algorithmic approach to compute these emotion states with increased reliability and transparency. We believe, that this study of reliability check and exploration of algorithmic approach can benefit the research community involving brain research and VR games.

To achieve this goal of the PM reliability check, we plan to track the players' six emotion states in three VR game types with ranging in-game activities from high-action to low-action.

978-1-6654-3886-5/21/$31.00 ©2021 IEEE

For this we selected adventure - exploration/puzzle-solving, action shooter, and relaxation/meditation game types in VR.

C. Research Question

For determining the reliability of the EEG PM, we planned to compare the PM values computed by Emotiv with the player ratings (PR) on gameplay experience. The player ratings (PR) refer to the experience questionnaire, which we designed for the players to complete at the end of each VR game. The PR questionnaire corresponds to the six emotional experience questions related to the PM values in the three VR games. To investigate this association between the EEG performance metrics (PM) and player ratings (PR), we framed the following research question (RQ):

RQ: Are there any differences between player ratings (PR) and the EEG performance metrics (PM) scores provided by the Emotiv EPOC X?

D. Analysis and Evaluation

We conducted a research study with 14 participants. Each participant played the three VR games and during the gameplay we recorded their raw EEG, performance metrics (PM) scores, and frequency bands using the Emotiv EPOC X headset. Later, we compared the player ratings (PR) and performance metrics (PM). From the comparative analysis of PM and PR, we found that PM and PR were different. This gave us a scope to build an Ad-hoc Linear Model to explore the relationship between the brain activity, emotions, and PR. The Ad-hoc Linear model was developed based on associating the PR values with the raw EEG data from the device and then estimating the PM values (emotion states) from it. Through our Ad-hoc Linear Model Algorithm, we can estimate the PM values directly from the raw EEG data.

II. Related Work

A. Gameplay Experience Evaluation and Game Design through EEG

Ranging from practical applications in medicine and therapy to acting as a method of bolstering the entertainment value of games, Brain-Computer Interfaces (BCI) technology exists in a multitude of fields to provide information on brain activity [9], [10]. For analyzing gameplay experiences, low-cost non-invasive electroencephalogram (EEG) is a current focus in research. EEG devices are widely used to evaluate the in-game experiences, comfort levels and emotions in video games [11], [12].

In the research activity carried out by [12], the authors investigated the psychological aspects such as cognition, emotion and player experience on three level design conditions and the research findings highlight the fact on how gameplay experiences can be assessed using EEG. EEG systems can also be used in the dimension to incorporate Dynamic Difficulty Adjustment in games. Notably in the research activity [11], Dynamic Difficulty Adjustment (DDA) and Rhythm-Group theory aspects were implemented by combining the player's performance and EEG data in a 2D platformer game and the system was able to adapt accordingly towards the players' status. Similarly, in [13], DDA triggering algorithm was created in 3D person shooter game by incorporating an EEG device and in this system, the player excitement level was monitored using the Emotiv Epoc headset and based on the excitement level the DDA was triggered. In [14] , the authors used OpenBCI low-cost EEG device to evaluate the players' relaxation states based on presence and absence of sound in Candy Crush and Geometry Dash.

B. Virtual Reality Applications and EEG Data Analysis

EEG devices are also being used along with the virtual reality applications to evaluate the cognitive load, player emotions, and physical discomforts. However, combining VR and low-cost EEG devices can be challenging as the VR headset can affect the EEG sensor arrangement and can disturb the signal quality of EEG devices [15]. Specifically, the experiment conducted by [15] found that EEG signals beyond 50Hz were mainly affected by the intrusion of Head Mounted Displays (HMDs).

One study examined the ability to detect excitement in VR applications through an EEG headset. Participants used both VR goggles and an EEG headset while being exposed to a VR roller-coaster application intended to induce excitement. The researchers found that the deep learning approach with the EEG provided a 96% accuracy rate of detecting excitement [16]. Another study investigated possible differences between learning with VR and learning with a 2D display. The EEG data showed that learning in the VR environment resulted in a higher cognitive workload. The higher workload is believed to have had an influence over the test scores, which were lower for the VR users than the 2D display learners [17].

The study conducted by [18] compared data results from an EEG headset and a self-reporting questionnaire in a study regarding the effect of perspective in VR games on a player's level of engagement.The study found that there were conflicting results between the data reported by the EEG and data reported by the questionnaire. The researchers suggested further study into comparing EEG and subjective data collection in VR games. In [19], the authors analyzed the correlations between questionnaire and EEG data on physical discomforts induced by VR gaming. The researchers found strong correlations between the occurrence of physical discomforts and beta and theta waves, suggesting that EEG technology can accurately detect physical discomfort. The researchers recommended possible uses of these correlations, such as establishing an EEG warning system that can recommend when players should take breaks.

The related work highlights on the challenges in combining VR and EEG devices in experiments and the need for further analysis in comparing EEG outcomes and subjective data collection. These two points were of prime importance to our research study and we designed our research question and experiment to reflect on these aspects.

III. Experiment Setup

For this research study, we recruited 14 participants and all the participants were graduate/undergraduate students. For each participant, the research study lasted approximately between 60 - 100 minutes.

A. Device Description

In this experiment, we used the Emotiv EPOC X - 14 channel wireless EEG Headset for recording the brain activity and Oculus Quest VR headset (All-in-one VR gaming). We used the Emotiv Pro Suite to access the raw EEG, low resolution performance metrics data (six emotion states: excitement, engagement, stress, interest, focus, and relaxation), and frequency bands from the EEG headset. Emotiv Pro was installed on Dell Alienware M15 R3 and the Emotiv EPOC X headset was paired to the Emotiv Pro via Bluetooth and the realtime EEG data was monitored and recorded using Emotiv pro.

B. VR Games Used in the experiment

We chose three VR games that would facilitate in triggering the six PM values and the VR games used in this experiment can be categorized into three types based on the in-game actions involved and they are as follows:

Space Pirate Trainer: Space Pirate Trainer is an arcade action shooter game, where the game mechanics involve shooting, defending, and dodging. In the game, the player takes up a role of space shooter and each VR hand controller is equipped with a sci-fi gun and the player needs to fire at the enemies and dodge the enemy bullets either by blocking them by a shield or by tilting the body left or right. The game involves *high action* with more scope for continuous hand and body actions (such as aiming, shooting, and tilting the body left or right).

The Room VR: A Dark Matter: This is an adventure game involving the exploration and puzzle solving aspects. The player is thrown into a series of quests like solving puzzles and interacting with game elements to uncover a mystery. The game mechanics involve grabbing, teleporting, reading clues/texts, interacting with lockers, opening drawers, and using keys. This game involves *moderate actions*, where there is only medium scope for continuous hand actions.

Tripp: Tripp is a relaxation and meditation game targeted for providing emotional well-being through immersive VR experience. The game involves simple breathing exercises and gaze-based game mechanics (like selecting, moving, and controlling the flight of the figures in game). Primary interaction mechanics do not involve any hand interactions or body movements compared to the other two games. So the in-game actions in this game falls under *low action* category.

The three VR games used in this experiment were rated to be as "comfortable" on the Oculus Store Comfort ratings and these comfort ratings are measured based on camera movements, player motions, disorienting contents and effects [20].

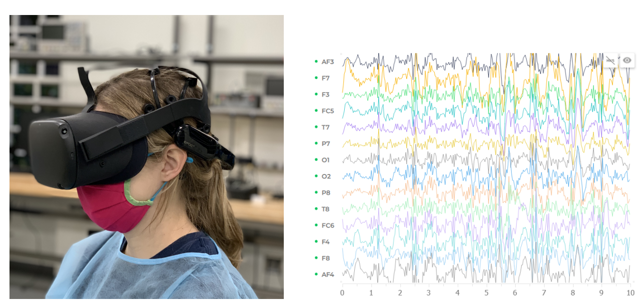
C. Demographics Information

Before the start of the experiment, we asked the participants to complete the research consent form and then to fill up a pre-test survey. In this pre-test survey, we collected data on age ($Mean = 21.57$; $SD = 2.17$), gender (male = 9 and female = 5), favorite video game genre, expertise level in using a VR device, VR games played from the given list, and whether they are right/left handed. Out of 14 participants, 5 of them have never used a VR device, 2 of them have moderately used VR devices and 7 of them have rarely used VR devices. In the three games used in this experiment, 2 participants had already played the Space Pirate Trainer before, and 1 participant had played The Room VR: Dark Matter and Tripp before. The remaining 11 users were completely new to all the three games used in the experiment.

D. Process

After the pre-test survey the participants were asked to wear the Emotiv EPOC X EEG headset. The research coordinator helped the participants to adjust sensors of the EEG device to ensure better connectivity. We monitored the sensor connection quality on the Emotiv Pro application. Once the better connectivity (above 98% of sensor connection quality) of the EEG device was established then the participants were given the Oculus Quest VR headset and this needs to be worn on top of the EEG device (see Fig 1.(a)).

The participants were given three VR games to play. Each VR game was played approximately for 5 minutes with a 10 minute break in between the games. During the 10 minute break, the participants removed the VR headset and relaxed for a while. Also, in this 10 minute break, the participants were asked to complete the player ratings (PR) questionnaire. In total, the participants filled one pre-test survey and three PR questionnaires (one for each VR game that they played). In addition to the surveys related to the experiment, we also followed the COVID safety screening procedure and in which, the participants filled pre-screening (approximately 24 hours before the experiment), on-site screening (before the start of the experiment) and a follow-up screening (approximately 24 - 48 hours after the experiment) surveys. Each participant was given a compensation of $10 Amazon gift card for participating in this research study.



(a) (b)

Fig. 1. (a) A participant wearing VR and EEG headsets during the experiment (b) Raw EEG plot (14-channel data) displayed on Emotiv Pro application

E. Data Collection

For each player, we randomized the sequence in which they played the three VR games and for this we used a randomizer program. The EEG data in each game was recorded for 5 minutes and the recorded data was stored in the Emotiv Pro Suite and later it was exported to a csv file for additional data processing. We recorded the following EEG data points using the Emotiv EPOC X [21], [22]:

Raw EEG: Raw EEG was captured at the rate of 128Hz from the 14 channels (AF3, F7, F3, FC5, T7, P7, O1, 02, P8, T8, FC6, F4, F8, AF4) by the EEG headset (See Fig 1.(b)).

Performance Metrics (PM): The performance metrics (PM) in the Emotiv EPOC X provides access to six emotion states (engagement, excitement, stress, relaxation, focus and interest) of users. Based on the Emotiv's documentation [21], all these six parameters are computed based on the mental activity acquired by the Emotiv headset in real time, but the details/process of deriving the PM values from the mental activity was not disclosed. The PM data was captured at the rate of 0.1Hz.

Each performance metric aspect holds a specific description from the Emotiv's documentation [21]. The aspect of *stress* is defined as overwhelming feeling and the fear of failure to satisfy the task needs. *Engagement* refers to immersion and the mixture of attention and concentration and the *excitement* is the sense of awareness and is characterized as positive physiological arousal. *Interest* is associated with the level of attraction/aversion within a task and it is referred as valence. Low interest corresponds to aversion and high interest corresponds to affinity. Whereas the mid-ranged score in interest denotes a neutral state. *Focus* is defined as the fixed attention on the given task and higher level of task switching induces distraction. *Relaxation* is referred to as being in the state of recovery from intense concentration [21].

Player Ratings (PR): The PR values were extracted through a questionnaire from the players after each VR game. The questionnaire contained six questions and each question was related to a particular emotion from PM. For example, one of the questions in the questionnaire was - "I felt *excitement* when playing this game" and the players were asked to mark their answer on a 5-point Likert scale. Likewise we had a question for each emotion state mentioned in PM (*excitement, engagement, stress, interest, focus*, and *relaxation*).

F. Problems Encountered during the Data Recording Phase

During the data recording process, our primary goal was to make sure that the EEG signal connectivity is 100% for all participants. There are several reasons in which the EEG headset's connectivity might drop less than 100% and in our case we noticed three reasons for poor EEG signal quality: 1) hair density - more hair could disturb/ break the connectivity of sensor felts on the scalp, 2) hydration level of the sensor felts - the sensor felts need to have good hydration at all times for better contact quality, and 3) motion/muscle artifacts. In our case, for the 5 participants, the EEG signal quality dropped less than 100% during the experiment. In the cases where the signal connectivity dropped , the PM values were also missed.

G. Data Processing

The scaled PM values from Emotiv range from 0 to 1 and this is produced every 10 seconds (0.1 Hz). The PR values were scaled to match the range of PM (i.e) from 0 to 1.

For comparative analysis, first we computed the average of PM values for the entire game duration and considered as final PM value for the game for a given emotional state. Secondly, we we mapped the PR to numerical values from 0 (Strongly Disagree) to 1 (Strongly Agree). Finally, we had six PM and six PR values for each player for a game. While computing statistics (Mean, SD, Correlation), the PR values corresponding to PM values that were not available from the Emotiv were ignored. On some instances, the PM values were not available from Emotiv due to the internal mechanisms and motion/muscle artifacts.

Although the Emotiv provides the power of five frequency bands namely; Theta, θ (4 - 8 Hz), Alpha, α (8 - 12 Hz), Low-beta β_L (12 - 16 Hz), High-beta β_H (16 - 25 Hz) and Gamma γ (25 - 45 Hz), for each sensor, the pre-processing techniques applied on the EEG data are unknown to users/developers. Further, it is unknown if any effective artifact removal algorithm is applied to avoid misleading computations. Therefore, to use the power of frequency bands for further analysis, we used the raw EEG data recorded from Emotiv. The EEG data was first, processed with an IIR high-pass filter of order 5 and cut-off frequency of 0.5 Hz, followed by an artifact removal algorithm (ATAR) [23]. The power of each frequency band was computed using Welch method.

IV. Results and Discussion

In this section, we firstly analyzed the EEG data between three VR games to understand whether the brain activity produced by the games used in this experiment are stimulating different emotions. Secondly, we conducted an independent analysis on PM and PR to analyze the trends and variations of emotion states between the games. Thirdly, we performed a comparative analysis of PM and PR, to investigate the reliability of the PM extracted by the EEG device. Finally, we used an Ad-hoc Linear Model to map the PR values with raw EEG data.

A. The VR Game types and their associated brain activity

We analyzed if the three VR games had produced different brain activities based on the recorded EEG data from the 14 participants. So we compared the power of five frequency bands (namely: θ, α , β_L, β_H and γ) of each sensor and the computation was performed as described in Section III-G. To analyze aggregated brain activity in three games, the power of each frequency band was averaged across participants. For each frequency band, the spatial density of power is obtained by interpolating and extrapolating the values computed from

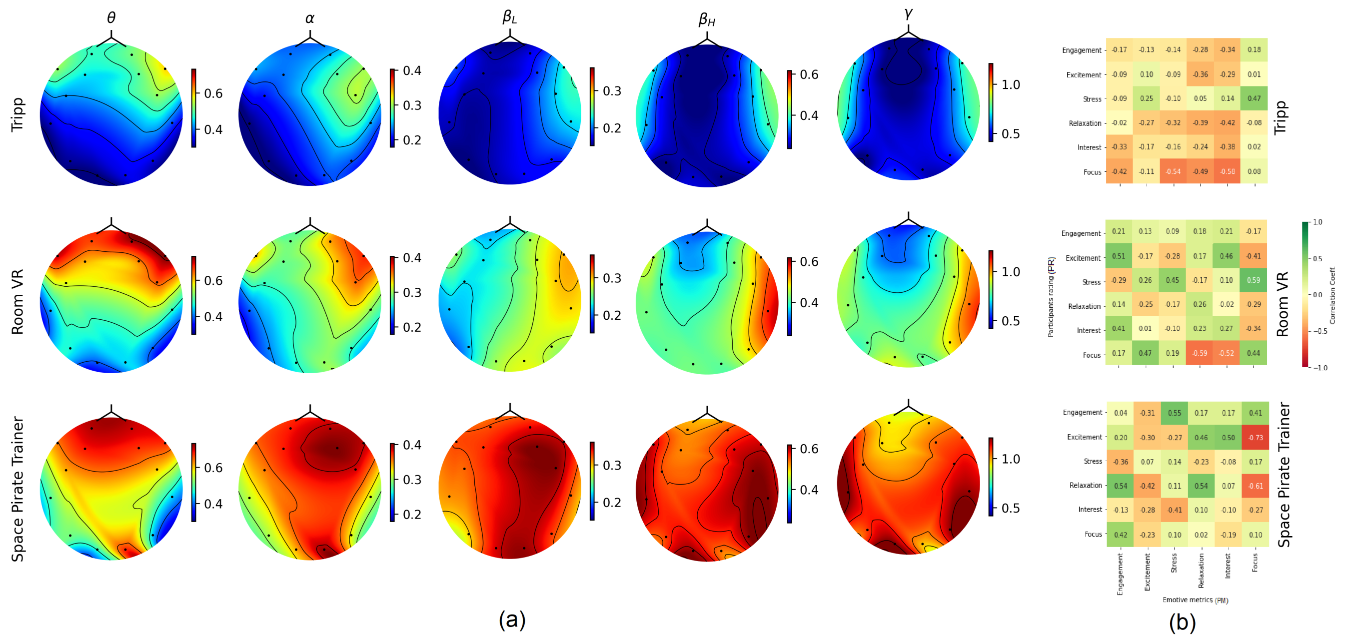


Fig. 2. (a) Topographic brain activity heatmap of average power band values of EEG data, across 14 participants, in five frequency bands (θ, α , β_L, β_H and γ). For comparison, the heatmap is scaled across the frequency bands. (b) Correlation matrix(Spearman's Rank Correlation) of six emotion states between PM and PR

the sensors using a Bicubic method adapted from the MNE library [24]. For the three game types, the spatial density of power is plotted as a topographical map reflecting the average brain activity in five frequency bands as shown in Fig 2 (a).

Discussion: From Fig. 2 (a), it can be observed the three games have different brain activities across the participants. In Fig. 2 (a), the higher brain activity is shown by red color and lower activity is shown by blue. Specifically, the game - Tripp, involves less brain activity in all five frequency bands (θ, α , β_L, β_H and γ) and this indicates that Tripp game consists of low in-game activities compared to the Space Pirate Trainer and the Room VR. Also, this low in-game activity in Tripp relates to the use of gaze-based interaction with slight head movement and without VR hand controllers/interactions. Looking into the game Space Pirate Trainer, the frequency band θ shows higher activity in pre-frontal cortex (the front part on the brain activity heatmap) compared to the Tripp and Room VR. This indicates the fact that Space Pirate Trainer game comprises higher executive brain functions (logical/reasoning) compared to others. This higher brain activity can be associated with the fast paced in-game hand actions involved in the Space Pirate Trainer and in addition to this the players were also required to tilt their body frequently in the game. In the game Room VR, the frequency band θ shows activity in pre-frontal cortex and this activity is higher compared to the Tripp. This indicates the fact that Room VR comprises moderate actions out of the three games. This moderate pre-frontal cortex activity in the Room VR can be related with low to medium paced hand movements such as grabbing items, opening lockers, and interacting with scene objects. The frequency and the pace of hand movements in Room VR were low compared to the Space Pirate Trainer.

So our initial analysis on understanding the differences between the brain activity and VR game types showed that these three VR games stimulate different brain regions with varying level of brain activity. Therefore, the VR game types used in this experiment have ranging in-game actions that correlate with our assumption on using the different game types in this experiment.

B. Independent Analysis of PM and PR

The EEG Performance Metrics (PM) Analysis: Table I shows the Mean and SD for the PM values extracted from the three VR games. It can be observed that there is no major difference in mean values of *engagement, excitement* and *focus* across all the games. On the other hand, there is a minor increment in *stress, relaxation*, and *interest* from Tripp to Room VR, followed by Space Pirate Trainer. To compare the emotion states between games, we conducted t-test and the results are shown in Table II. Firstly, as expected, there is no statistical significant difference observed between the games in three performance metrics features - *engagement, excitement, and focus*. There is a statistical significant difference between Tripp and Space Pirate Trainer in the *stress* ($p < 0.05$). The *relaxation* shows a statistical significant difference between all the three games. In the *interest*, a statistical significant difference is noted between Room VR and Space Pirate Trainer ($p < 0.05$), and Tripp and Space Pirate Trainer ($p < 0.01$).

The Player Ratings (PR) Analysis: Table III displays the Mean and SD for the PR values on six emotion states for the

TABLE I
MEAN ± SD OF THE EEG PM DATA FOR THE THREE VR GAMES

	Engagement	Excitement	Stress
Tripp	0.67±0.1	0.4±0.18	0.3±0.12
Room VR	0.65±0.1	0.38±0.13	0.39±0.18
Space Pirate Trainer	0.67±0.1	0.4±0.2	0.49±0.2
	Relaxation	Interest	Focus
Tripp	0.25±0.1	0.53±0.07	0.42±0.09
Room VR	0.34±0.11	0.6±0.08	0.36±0.19
Space Pirate Trainer	0.44±0.11	0.67±0.09	0.36±0.08

TABLE II
P-VALUES OF T-TEST (INDEPENDENT SAMPLES) BETWEEN GAMES FOR EACH CATEGORY OF SCORE FOR THE EEG PM DATA. $*p < 0.05$, $**p < 0.01$, $^{\dagger}\ p < 0.001$.

	Engagement	Excitement	Stress
Tripp - Room VR	0.7011	0.7241	0.211
Room VR - Space Pirate Trainer	0.8079	0.7003	0.2495
Tripp - Space Pirate Trainer	0.8893	0.9573	0.0198*
	Relaxation	Interest	Focus
Tripp - Room VR	0.0264*	0.0632	0.3655
Room VR - Space Pirate Trainer	0.0374*	0.045*	0.9226
Tripp - Space Pirate Trainer	0.0001†	0.0004†	0.2158

three VR games. There is a noticeable difference in mean values of all the emotion states across three games. *stress*, is lower for Tripp and higher for Space Pirate Trainer game, on the other hand, *relaxation* is higher for Tripp and lower for Space Pirate Trainer game. Another noticeable aspect is that the mean scores between the Tripp and Space Pirate Trainer are of major difference on the six emotional states. The levels of *engagement, excitement, interest*, and *stress* were incremental from Tripp to Room VR to the Space Pirate Trainer. Similar to the PM analysis, we conducted t-test for the three games for the player ratings. A statistical significant difference ($p < 0.05$) is noted in all the six emotion states between Tripp and Space Pirate Trainer. Similarly, there is a statistical significant difference in the majority of the emotion states (*engagement, excitement, interest, and focus*), when comparing the Room VR and Space Pirate Trainer. Also, a statistical significant difference was noted for the *stress* ($p < 0.01$) and *relaxation* ($p < 0.001$) emotions between Tripp and Room VR.

Discussion: The PM analysis shows that three emotions (*engagement, excitement*, and *focus*) do not change across three games, which is counter intuitive, because from the perspective of in-game activities involved in the three games, the levels of *focus, engagement* and *excitement* should vary. In Tripp, there is no scope for fast paced actions, but Space Pirate Trainer involves continuous hands and body movements. So this was indicated in the significant difference between the PM values in *stress* between Tripp and Space Pirate Trainer. Only the relaxation shows difference in all three games and again this highlights the ranging hand and body movements involved in the three games.

The PR analysis shows that all the emotions vary across games and the increment in level of four emotions (*engagement, excitement, interest*, and *stress*) reflect the differences in the gameplay involved in three games. One important aspect that we noticed was the stress and relaxation hold an inverse relationship, which is a known aspect and correctly reflects in the PR analysis.

In this independent analysis of PM and PR, we found that PM values are not reflective of the activities involved within the game types. The PR values reflect the expected differences in game types and this highlights the association of players' subjective emotions experienced within the game types are different.

TABLE III
MEAN ± SD OF PR FOR THE THREE VR GAMES (NORMALIZED)

	Engagement	Excitement	Stress
Tripp	0.75±0.23	0.46±0.21	0.18±0.17
Room VR	0.86±0.16	0.61±0.21	0.46±0.23
Space Pirate Trainer	0.98±0.06	0.91±0.12	0.52±0.29
	Relaxation	Interest	Focus
Tripp	0.79±0.19	0.71±0.19	0.73±0.22
Room VR	0.41±0.18	0.82±0.2	0.73±0.15
Space Pirate Trainer	0.3±0.22	0.96±0.09	0.95±0.1

TABLE IV
P-VALUES OF T-TEST (INDEPENDENT SAMPLES) BETWEEN GAMES FOR EACH CATEGORY OF SCORE FOR THE PR DATA. $*p < 0.05$, $**p < 0.01$, $^{\dagger}\ p < 0.001$.

	Engagement	Excitement	Stress
Tripp - Room VR	0.1778	0.0898	0.0014**
Room VR -Space Pirate Trainer	0.0127*	0.0001†	0.6059
Tripp - Space Pirate Trainer	0.0018**	0.0†	0.0013**
	Relaxation	Interest	Focus
Tripp - Room VR	0.0†	0.1674	1.0
Room VR -Space Pirate Trainer	0.1796	0.0254*	0.0002†
Tripp - Space Pirate Trainer	0.0†	0.0002†	0.0039**

C. Comparative Analysis between PM and PR

The goal of this analysis is to address the research question (RQ) by comparing the PM and PR. In this analysis, we had six PM and six PR values for each player for each game.

First, we performed the t-test between PM and PR for respective emotion state (for example, we compared PM-*excitement* with PR *excitement*). The results of t-test showed a statistically significant difference ($p < 0.05$) between 4 out of 6 PM values of the EEG headset with the respective PR values for each game. In Tripp, *stress, relaxation, interest*, and *focus* showed statistical significant difference ($p < 0.05$), while for other two games, *engagement, excitement, interest*, and *focus* showed statistical significant difference ($p < 0.05$).

For further analysis, we conducted the Spearman's Rank correlation test between PM and PR (see Fig 2 (b)). In the correlation analysis, the six emotion states from PM and PR were compared against each other for the three VR games. For Tripp, a negative correlation between PM and PR values was found for *engagement, stress, relaxation* and *interest*, and a low positive correlation between *focus* and *excitement*. However, for Room VR, a positive correlation was observed with all the emotion states between PM and PR, except for *excitement*. Lastly, the comparison between PM and PR could be observed from the trends found in independent analysis of PM and PR in previous section. We noticed that Mean values of PM and PR (see Table I and III) are different from each other for respective emotions for each game. Based on the t-test analysis from Table II and IV, we observed that compared to PM values, the PR values for most of the emotion states significantly differ between the games.

Discussion: The comparative analysis using t-test reflects that the PM and PR are different from each other and this answers our research question (RQ). The correlation analysis also shows that PM values does not capture the players' emotion states accurately. The reason for such major difference could be associated with the fixed internal configurations of Emotiv for PM computations. This indicates that players experienced difference in gameplay between the three game types. However, the PM values from the EEG device did not capture this difference.

D. Ad-hoc Linear Model for Emotion State Estimation

From comparative analysis, it can be seen the that estimation of PM values from Emotiv is not consistent with the PR (players' real experience), hence they are misleading. We used ad-hoc linear models to investigate the reliable computation of PM values, which reflects the actual players' experience. We show that a level of an emotion state (for example, *excitement*) can be reliably computed from the raw EEG data itself. To develop a linear model for an emotion state, we used respective PR values as a ground truth, reflecting the level of emotion state and power of each sensor for a given frequency band.

Let us denote, the level of an emotion state E as L_E, and the power of kth sensor as P_k^σ in a given frequency band $\sigma \in [\theta, \alpha, \beta_L, \beta_H, \gamma, All]$, where All is full frequency band by computing the average power for each sensor in $\theta, \alpha, \beta_L, \beta_H$, and γ. Then a linear model for an emotion state E can be defined as:

$$L_E = w_0 + \sum_{k=1}^{14} w_k P_k^\sigma \tag{1}$$

For each emotion state (E), a linear model was fit on the average power of each sensor in a frequency band, using linear regression by minimizing the Mean Square Error (MSE). The coefficient w_k reflects the relation of the level of emotion state L_E with kth sensor. In total, we developed six linear models for an emotion state with respect to five frequency bands and All frequency band.

To compare the linear models of six frequency bands and to show the closeness of linear approximation of the level of emotion states L_E, the MSE of all the linear models across θ, α, β_L, β_H, and γ are shown in Fig. 3.

The MSE for all linear models is below 0.07, where the PR values range from 0 to 1. This indicates that the linear approximation is quite close to the actual PR. It can further be observed that the MSE of linear model using All frequency bands is minimum for each emotion state, which indicates that combining all the frequency bands leads to a better approximation of the level of emotion state L_E with PR. Another point to note is that the MSE for *engagement* and *interest* is lower, while MSE for *stress* and *relaxation* is higher, indicating that *engagement* and *interest* can be better approximated when compared to *stress* and *relaxation*.

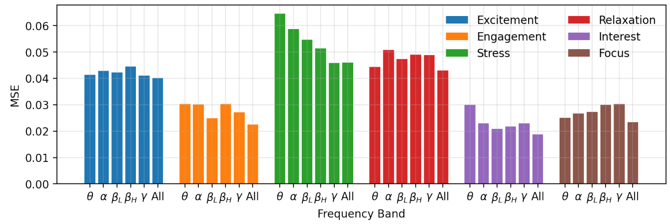


Fig. 3. MSE of linear models fit with different frequency bands.

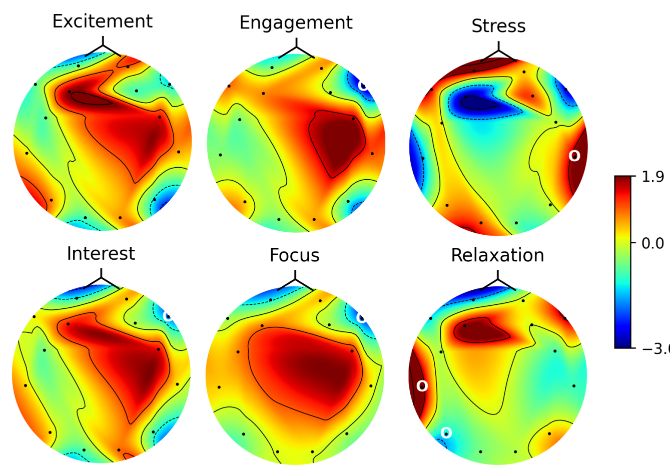


Fig. 4. Topographic brain activity heatmap: Linear relation between PR for each emotion state and overall power of different sensors, computed through Linear Model. Sensors with significant ($p < 0.05$) relations are highlighted as white circle.

Finally, the coefficients of the linear model with All frequency band were used to demonstrate the relation between the emotion states (in reference to PR) and different regions of the brain (sensors), as shown in Fig 4. Also, Fig. 4 can be seen as an average brain activity map for each emotion state.

Discussion: In Fig. 4, the *stress* and *relaxation* (which are quite opposite experiences) values show an inverse relationship. Specifically, the prefrontal lobe is positively associated with *stress* and negatively associated with *relaxation*. This means, high activity in prefrontal cortex shows the higher *stress* level, whereas, low activity shows the higher level of relaxation. This observation is aligned with the literature of Neuroscience, where prefrontal cortex has been associated with executive decisions, including logical and reasoning skills [25]. It is also aligned with our earlier observations of independent analysis of PR. Fig. 4 also shows the sensors that have significant ($p < 0.05$) relation with the emotion state, highlighted by white circles.

Thus, using this ad-hoc linear model we can estimate the level of six emotion states reliably compared to the in-built PM algorithm within the EEG device.

V. Conclusion

In this paper, we combined VR games and an EEG device to evaluate the players' emotion states in three VR games. From the results, we found that the players' subjective emotions recorded from the questionnaire was different from the emotion states determined by the EEG device. So we conducted a further analysis using ad-hoc linear models to map the relationship between emotion and raw EEG with the players' subjective experience data as a ground truth.

In addition to this, the emotions mapped using the ad-hoc linear models show better association with players' subjective experience data, specifically for the two emotions (stress and relaxation). The future work could expand on our foundational ad-hoc linear model analysis by improving the core of the linear model by combining multiple EEG parameters (such as facial expressions, in-game reactions and responses, body and head motions data) along with ground truth (i.e.) players' subjective responses to determine the emotions data from raw EEG.

In this research, combining EEG and VR together was challenging due to the fact that the sensors in the EEG device can induce issues in data capture even if there is a slight disturbance like moving the sensors to adjust the VR headset straps. So for the future research in combining these two devices, a custom headset arrangement/model can be created to reduce any discrepancies in the data collection process.

In conclusion, one of the crucial lessons learned from this experiment is to verify the essential parameters extracted from the EEG device as it can greatly affect the design decisions and user experience evaluation. Thus, in this research activity, we investigated the discrepancy between players' experience and data captured using the EEG device and we have provided an ad-hoc model to extract the player emotions data from the raw EEG signals.

References

[1] Dramamine, "Is vr gaming making you sick? use these 5 tips to keep playing," https://www.dramamine.com/blog/vr-gaming-making-you-sick-use-these-5-tips-to-keep-playing, 2021.

[2] R. Chang, C. Wu, K. He, and B. Lin, "A flexible and wearable EEG device," in *2016 Third International Conference on Computing Measurement Control and Sensor Network (CMCSN)*, 2016, pp. 48–51.

[3] EMOTIV, "What is an EEG machine? definition & faqs," https://www.emotiv.com/glossary/eeg-machine/, 2021.

[4] Imotions Publish, "EEG headset prices – an overview of 15+ eeg devices - imotions," https://imotions.com/blog/eeg-headset-prices/, 2021.

[5] S. Pathirana, D. Asirvatham, and G. Johar, "A critical evaluation on low-cost consumer-grade electroencephalographic devices," in *2018 2nd International Conference on BioSignal Analysis, Processing and Systems (ICBAPS)*, 2018, pp. 160–165.

[6] J. Larocco, M. Le, and D.-G. Paeng, "A systemic review of available low-cost EEG headsets used for drowsiness detection," *Frontiers in Neuroinformatics*, vol. 14, 10 2020.

[7] J. Kosiński, K. Szklanny, A. Wieczorkowska, and M. Wichrowski, "An analysis of game-related emotions using emotiv epoc," in *2018 Federated Conference on Computer Science and Information Systems (FedCSIS)*, 2018, pp. 913–917.

[8] T. N. Malete, K. Moruti, T. S. Thapelo, and R. S. Jamisola, "Eeg-based control of a 3d game using 14-channel emotiv epoc+," in *2019 IEEE International Conference on Cybernetics and Intelligent Systems (CIS) and IEEE Conference on Robotics, Automation and Mechatronics (RAM)*, 2019, pp. 463–468.

[9] G. Schalk, D. Mcfarland, T. Hinterberger, N. Birbaumer, and J. Wolpaw, "Bci2000: a general-purpose brain-computer interface (bci) system," *IEEE Trans. Biomed. Eng.*, vol. 51, pp. 1034–, 07 2004.

[10] H. R. Khairuddin, A. S. Malik, W. Mumtaz, N. Kamel, and L. Xia, "Analysis of EEG signals regularity in adults during video game play in 2d and 3d," in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2013, pp. 2064–2067.

[11] H. B., K. Mikami, and K. Kondo, "Adaptable game experience based on player's performance and EEG," 06 2017, pp. 1–8.

[12] L. E. Nacke, S. Stellmach, and C. A. Lindley, "Electroencephalographic assessment of player experience: A pilot study in affective ludology," *Simulation & Gaming*, vol. 42, no. 5, pp. 632–655, 2011. [Online]. Available: https://doi.org/10.1177/1046878110378140

[13] A. Stein, Y. Yotam, R. Puzis, G. Shani, and M. Taieb-Maimon, "EEG-triggered dynamic difficulty adjustment for multiplayer games," *Entertainment Computing*, vol. 25, pp. 14–25, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1875952117301325

[14] S. Cano, N. Araujo, C. Guzman, C. Rusu, and S. Albiol-Pérez, "Low-cost assessment of user experience through EEG signals," *IEEE Access*, vol. 8, pp. 158 475–158 487, 2020.

[15] S. Hertweck, D. Weber, H. Alwanni, F. Unruh, M. Fischbach, M. E. Latoschik, and T. Ball, "Brain activity in virtual reality: Assessing signal quality of high-resolution EEG while using head-mounted displays," in *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 2019, pp. 970–971.

[16] J. Teo and J. Chia, "EEG-based excitement detection in immersive environments: An improved deep learning approach," vol. 2016, 09 2018, p. 020145.

[17] G. Makransky, T. S. Terkildsen, and R. E. Mayer, "Adding immersive virtual reality to a science lab simulation causes more presence but less learning," *Learning and Instruction*, vol. 60, pp. 225–236, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0959475217303274

[18] D. Monteiro, H. Liang, A. Abel, N. Bahaei, and R. d. C. Monteiro, "Evaluating engagement of virtual reality games based on first and third person perspective using eeg and subjective metrics," in *2018 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*, 2018, pp. 53–60.

[19] J. Heo and G. Yoon, "EEG studies on physical discomforts induced by virtual reality gaming," *Journal of Electrical Engineering & Technology*, vol. 15, 02 2020.

[20] Oculus Quest, "Oculus homepage," https://support.oculus.com/comfort/, 2021.

[21] EMOTIV, "Performance metrics," https://www.emotiv.com/knowledge-base/performance-metrics/, 2021.

[22] EMOTIV-Frequency Bands, "What are they and how do i access them?" https://www.emotiv.com/knowledge-base/frequency-bands-what-are-they-and-how-do-i-access-them/, 2021.

[23] N. Bajaj, J. R. Carrión, F. Bellotti, R. Berta, and A. De Gloria, "Automatic and tunable algorithm for eeg artifact removal using wavelet decomposition with applications in predictive modeling during auditory tasks," *Biomedical Signal Processing and Control*, vol. 55, p. 101624, 2020.

[24] A. Gramfort, M. Luessi, E. Larson, D. A. Engemann, D. Strohmeier, C. Brodbeck, R. Goj, M. Jas, T. Brooks, L. Parkkonen *et al.*, "Meg and eeg data analysis with mne-python," *Frontiers in neuroscience*, vol. 7, p. 267, 2013.

[25] S. Sanei and J. A. Chambers, *EEG signal processing*. John Wiley & Sons, 2013.