Int. J. Human–Computer Studies 160 (2022) 102762

Contents lists available at ScienceDirect

International Journal of Human - Computer Studies

journal homepage: www.elsevier.com/locate/ijhcs

# Effects of interacting with facial expressions and controllers in different virtual environments on presence, usability, affect, and neurophysiological signals

Arindam Dey [*,a], Amit Barde [b], Bowen Yuan [a], Ekansh Sareen [c], Chelsea Dobbins [a], Aaron Goh [a], Gaurav Gupta [a], Anubha Gupta [c], Mark Billinghurst [d]

[a] University of Queensland, Australia
[b] University of Auckland, Australia
[c] IIIT Delhi, India
[d] University of South Australia, Australia

ARTICLE INFO

*Keywords:*
Virtual reality
Facial expression
Interaction
User study
Neurophysiological signals

ABSTRACT

Virtual Reality (VR) interfaces provide an immersive medium to interact with the digital world. Most VR interfaces require physical interactions using handheld controllers, but there are other alternative interaction methods that can support different use cases and users. Interaction methods in VR are primarily evaluated based on their usability, however, their differences in neurological and physiological effects remains less investigated. In this paper—along with other traditional qualitative matrices such as presence, affect, and system usability—we explore the neurophysiological effects—brain signals and electrodermal activity—of using an alternative facial expression interaction method to interact with VR interfaces. This form of interaction was also compared with traditional handheld controllers. Three different environments, with different experiences to interact with were used—happy (butterfly catching), neutral (object picking), and scary (zombie shooting). Overall, we noticed an effect of interaction methods on the gamma activities in the brain and on skin conductance. For some aspects of presence, facial expression outperformed controllers but controllers were found to be better than facial expressions in terms of usability.

## 1. Introduction

Virtual Reality (VR) is a medium that immerses users in a fully simulated graphical world supported by physical interactions. There have been different interaction methods and devices developed to support interactions in this medium, with the most common being handheld controllers. However, alternative methods such as body movement (Feng et al., 2016), touch (Benzina et al., 2012; Yan et al., 2016), and other embodied interactions (Galvan Debarba et al., 2017) have also been implemented. However, all of these interaction methods require users to be at least partially physically able in order to interact in VR.

There has been a considerable amount of research carried out in the VR domain to enable hands free interactions. Researchers have used eye-gaze (Piumsomboon et al., 2017), walking-in-place (Tregillus and Folmer, 2016), head movement (Lu et al., 2019), and speech (Manuri et al., 2016) to facilitate interactions in VR environments. An emerging area of research in VR is the ability to implement alternate methods of interaction, such as using a Brain-Computer Interface (BCI) (Coogan and He, 2018; Lotte et al., 2012). The BCI captures and interprets neural activity which is then used to drive interactions with and in the VR environment. This enables hands-free interactions, and does not require the user to exercise any of the muscles in the body for interactions. Unfortunately, BCI systems require careful calibration to work well, and users must invest a significant amount of time and effort in order to learn how to operate such a device.

In order to address these shortcomings, we propose a novel interaction technique for use in VR – using facial expressions (FEs). Recently, there has been commercial interest in using facial expression in VR, as evidenced by HTC's release of the Vive facial tracker[1]. The main motivations for using facial expressions as an interaction method is two fold.

* Corresponding author.:.
*E-mail address:* a.dey@uq.edu.au (A. Dey).

[1] https://www.vive.com/eu/accessory/facial-tracker/

https://doi.org/10.1016/j.ijhcs.2021.102762
Received 10 September 2020; Received in revised form 7 July 2021; Accepted 17 December 2021
Available online 30 December 2021
1071-5819/

First, facial expressions can offer an additional interaction modality to complement and/or extend current interaction paradigms available in VR. They can also be beneficial for certain tasks such as blowing a candle and/or bubbles or kissing a loved one. Second, it can increase the accessibility of VR systems to those users who cannot use other modes of interaction due to physical limitations. Additionally, given that facial expressions are an intrinsic part of the human condition, and are a natural way to interact and convey information; implementing them as a means to interact with a virtual environment (VE) seems logical. Common facial expressions such as anger, happiness, and surprise are well known and easy to reproduce, and be easily implemented to effect actions and movements in a VE. Associations between facial expressions and movements can be learned by users, and allow users to interact in VR without the use of traditional methods such as controllers or treadmills (Virtuix, 2013; Warren and Bowman, 2017). However, it is important to systematically evaluate and understand what effects interacting with facial expressions has on overall experience in the VE including presence, usability, emotional states, and other neurophysiological signals. This is the primary objective of this study.

The capture of facial expressions is facilitated via an electromyography (EMG) sensor. These signals are then processed in order to classify them into a pre-selected set of facial expressions, each associated with its own interaction type in the VR environment. Facial expressions have been the subject of research in VR for over a decade. However, the focus has primarily been on using facial expressions as a means to increase realism and presence (Sanchez-Vives and Slater, 2005) in social interactions (Riva et al., 2007; Roth et al., 2016) in VEs. They have also been used to assess emotions or affect (Magnenat-Thalmann et al., 2005) and sharing emotion (Hart et al., 2018). The use of facial expressions as a means of interaction—navigation and manipulation of objects (actions)—in VR, to the best of our knowledge, has not been explored before. It offers the opportunity to facilitate hands-free interaction in VR. Such an interaction methodology could help increase the reach of VR applications by making them available to a wider base of users, both differently-abled and able bodied. The proposed interaction method can also be integrated with existing interaction methods to augment the user experience by providing additional interactions complementing hand-held controllers where facial expressions are more natural form of interaction than handheld controllers.

To facilitate this exploration, we have designed three virtual environments (VEs) that provide three different experiences—happy, neutral, and scary. All of these environments provided two main interactions—navigation and action (touch, pick, and shoot). Using an Emotiv Epoc+ 14-channel wireless electroencephalogram (EEG) device[2] we enabled facial expressions to be used in VR to accomplish both navigation and actions by training our system to recognise facial movement via the EEG headset. The integration of the EEG device and other electrical sensors with the headset is feasible as the external noise produced by the headsets is minimal (Si-Mohammed et al., 2018). We used hand-held controllers as a control condition, to compare the facial expression-based interaction with, as this is the most widely used interaction method in VR.

As presence is one of the most important factors in VR, the effect of these interaction methods on presence was measured, as well as emotional arousal and overall system usability ratings. Besides these qualitative measures, we also measured physiological data—electrodermal activity (EDA)—which was used as a quantitative measure of physiological arousal. The sympathetic nervous system controls sweat gland activity, which is measured using EDA sensors, and thus increases physiological arousal (Sugenoya et al., 1990). Additionally, neurological data was collected to quantitatively measure cognitive states of the users.

[2] https://www.emotiv.com/epoc/

### 1.1. Novelty and contribution

Despite the advances in VR, the technology remains inaccessible to a portion of the population not having the physical ability to use the interaction techniques in and/or with such devices. While our primary motivation is to make VR more accessible with hands-free interaction methods, these methods are also beneficial to complement the currently used methods by providing additional interaction opportunities. The research covered in this paper evaluates the use of facial expressions as a means of interaction in VR. However, this study did not involve any differently-abled participants to explore aspects of accessibility. However, it is the first study to implement and evaluate facial expression for interaction in VR. While the primary use of facial expressions in VR has been as a means to enhance realism and/or increase presence, our research uses it to interact in VR, which is the key novelty.

The research detailed in this paper makes two vital contributions. Firstly, we identified system usability and emotional effects of facial expressions- and controller-based interactions. Standard system usability measures, such as the System Usability Scale (Brooke et al., 1996), and emotional/affect measures, such as the Positive and Negative Affect Schedule (PANAS) (Crawford and Henry, 2004), were used to evaluate the seemingly *intangible* aspects of this form of interaction. These tell us how participants *felt* when using the two different forms of interaction in VR evaluated for this study. Secondly, we identified neurophysiological effects of the two interaction methods. These measures include neural activity and EDA. These measures provide quantitative data that indicates the physical state of participants during the interactions that cannot otherwise be collected using questionnaires.

Together, both these measures are able to provide us with a comprehensive subjective and objective outlook on the two interaction methods that were tested. This has also helped us identify the design challenges and opportunities to facilitate a better interaction experience using facial expressions. Overall, our results demonstrate that using facial expressions can be a viable technique to facilitate interactions in VEs.

The rest of the paper is organized as follows. In the next section, we discuss some of the earlier work undertaken in the area. This is followed by a description of our experimental system and the virtual environments, including the facial expressions used for the interaction. We then describe our user evaluation in detail. The next section provides the results of the data analysis. Then we discuss the results with respect to the proposed hypotheses. Finally, we conclude by pointing towards the future research directions.

## 2. Related work

In this section we discuss some of the earlier work in the field of interaction in VR.

From the early days of VR, researchers aimed to make interactions natural, mimicking to a great degree, those that people use in the real-world. For example, if one were to reach out and push a door in VR, one would expect it to react in a manner similar to the real-world. In doing so the VR environment would, in effect, replicate the entire chain of events and objects that enable the interaction. Carrying over real-world affordances into VR has been the focus of research in both academia and industry alike. To that end, the remainder of this section covers some of the interaction techniques that have been implemented in VR to enable natural and/or seamless interaction with the environment.

**Gloves:** Our hands allow us to interact and manipulate objects with incredible precision and dexterity. To have similar abilities in VR would enable us to enjoy these same benefits in the virtual world. Previously, researchers have developed a variety of different gloves that enable us to interact intuitively in a virtual environment (Bowman et al., 2002). This type of physically-based interaction with force-feedback has been found to be more realistic (Borst and Indugula, 2005) and subjectively preferred (Prachyabrued and Borst, 2012). However, despite their intuitiveness and the natural interaction abilities they afford, gloves have not gained ground as a means for interaction in VR. A recent study found that gloves are less usable than controllers in VR, although users preferred the haptic feedback that gloves provided (Fahmi et al., 2020).

**Hand-held Controllers:** The use of hand-held controllers in VR seems to be a viable alternative to the use of hands themselves. While hand-held controllers do not offer the dexterity and flexibility afforded by the human hand, they make up for this by enabling other interactions that would not have been possible if one were to only use their hands. For example, the touch-pad on the upper half of a controller used with the HTC Vive VR headset allows users to interact within the VR environment by simply moving their thumb along its surface. Once the desired manipulation has been achieved using the touch-pad, it can be clicked in order to select it. This dual functionality of the touch-pad allows designers to exploit these features for range of interaction types in VR. Controllers also provide multiple buttons and/or joysticks (Oculus Rift, Oculus Quest) for selection and manipulation tasks. Such choice provides a distinct advantage over the use of hands where it would be greater challenge to integrate touch-pad like interaction without engaging the entire hand for the process. Despite these advantages, it is well known that controllers tend to be less realistic when it comes to interactions in VR. Actions like grabbing and pointing are not easily achieved and require the user to learn how these can be achieved using controllers.

**Hands-free interaction:** Hands-free techniques as means to interact within a VR environment have existed for several years. These range from different kinds of walking techniques to the use of eye gaze and head movements. For the purposes of this paper, we also consider locomotion as a form of interaction within VR.

*Eye Gaze:* The use of eye gaze as means to interact with a computing environment has been around for close to four decades (Bolt, 1982). The eyes provide us with the most direct information in terms of areas and objects of interest (Jacob, 1990). This property of eye movement can be exploited in order to enable efficient hands-free interactions in VR. Eye gaze has been used to enable locomotion and selection tasks, and has demonstrated good usability in both cases. Both these features can be implemented by using the "dwell time" as described by Sibert and Jacob (Sibert and Jacob, 2000). This refers to the time a user needs to fix his/her gaze on a target in order to trigger an interaction. However eye gaze has been shown to suffer from some drawbacks. The nature of eye movements means that users find it relatively hard to concentrate their gaze on a particular area or object for extended periods. The natural tendency of the eyes to 'scan' the environment - saccades - makes it hard to interpret eye gaze information accurately. This technique can also suffer from what has been termed the "Midas Touch Problem" (Jacob, 1990), where everything a user looks at within the environment can potentially lead to an interaction being triggered. Keeping these drawbacks in mind, it is appears that eye gaze is most suited to being a secondary method of interaction that supplements a more robust primary form of interaction in the VE.

*Locomotion:* Locomotion gives users the ability to traverse through a VE, much like they would in the real-world. This allows the user to be more immersed in an environment. The quality of this form of interaction can determine the user experience in the VE. There are several ways in which locomotion has been implemented in VEs. This involves the use of eye-gaze, teleportation (Bozgeyikli et al., 2016), walk-in place (Slater et al., 1995; Usoh et al., 1999) and other locomotion techniques such as treadmills (Ruddle et al., 2013) and trigger walking (Sarupuri et al., 2017). All these techniques aim to make the user feel like they are interacting with the environment in manner that does not break the sense of immersion and presence. Studies have demonstrated that the effectiveness of the type of locomotion implemented varies greatly depending on the environment in which the user is placed, the perceived or possible interactions that a user can perform within the environment and the constraints that the environment imposes on the user (Boletsis, 2017). For an overview of locomotion types in VR please read (Boletsis, 2017).

Other popular hands-free interaction techniques that have been adopted for interaction include the use of cameras and infra-red (IR) sensors like the LeapMotion[3], and even ultra-sound emitters such as the ones used by Ultrahaptics[4]

**Brain-Computer Interface (BCI):** A relatively recent addition to the technology that can mediate the interaction between a user and a VE has been the BCI. A BCI records neural activity and translates this into real-time action. For example, if a user desires to move forward in the VE, they only need think of such an action. The BCI detects the neural activity that represents this thought and converts it into motion in the VE. While this form of interaction has shown to be extremely useful in several cases (Amores et al., 2016; Dey et al., 2019; Friedman, 2015; Johnson et al., 2018; Martišius and Damaševičius, 2016; Salisbury et al., 2016; Škola et al., 2019), the immense work that goes into processing neurological data to extract meaningful information means that such a system requires a significant amount of computing resources. For a BCI to work well it also must be calibrated to understand a user's neural activity, and a user needs to spend a considerable amount of time learning how to use a BCI in order to be able to fully benefit from it. The use of BCIs in VR has been an ongoing area of research. A overview of some this research can be found in (Lotte et al., 2012).

This section has covered some of the most popular interaction techniques in VR that have been explored over the years. We have also covered some of the interaction techniques that are visible in the commercial space such as hand-held controllers[5] and omnidirectional treadmills[6]. However, a majority of these interaction techniques, exclude a sub-set of the population - the differently-abled. An inability to use limbs means most of the interaction techniques reviewed in this section cannot be implemented. While the BCI can be used to fill this void, another alternative is to use facial expressions that can also be used by users without or limited access to their hands.

**Facial Expressions (FE):** Given these shortcomings in the other interaction methods, this paper puts forth a novel interaction technique that we believe can provide an alternate solution to the differently-abled community using facial expressions. However, it must be noted that facial expressions can also be used to complement currently available interaction methods for all users. Tasks, such as blowing bubbles, kissing, and putting out a candle are more natural using facial expressions than a hand-held controller. It has been suggested that despite the varying nature of facial musculature, all humans possess a small set of common muscles that allow for reproduction of "universal facial expressions (Waller et al., 2008). We believe this generalisable nature of facial expressions makes them a unique interaction tool which, to the best of our knowledge, have not been explored in this context.

Several researchers have explored how to capture facial expressions in VR HMDs. Previous methods have typically used a camera-based approach, a contact-based approach, or optical sensors. For example, Thies et al. used a camera-based approach by placing an infrared camera inside the HMD (Thies et al., 2018), while others detected face expressions using a camera to observe mouth and lower face movement (Burgos-Artizzu et al., 2015). Contact sensors, such as electromyographic sensors, have been used to measure facial muscle movement in a way that is compatible with HMDs (Gruebler and Suzuki, 2014), or have been embedded in the HMD faceplate (Bernal et al., 2018; Cha et al., 2020). For optical sensors, Susuki et al. (Suzuki et al., 2017) showed how photo reflective sensors embedded in the HMD could track face muscle motion. Sometimes these approaches are combined together such as in the work of Li et al. (Li et al., 2015). They placed strain gauges in the

3 https://www.ultraleap.com/product/leap-motion-controller/
4 https://www.ultraleap.com/haptics/
5 https://www.vive.com/, https://www.oculus.com/
6 https://www.virtuix.com/

foam face padding of the HMD to measure upper face expressions, combined with an RGBD camera for mouth tracking. These methods have almost entirely been used in VR to create real-time expressions on virtual avatars or character animation.

Besides enabling the detection of facial expressions, other researchers have used facial expressions to convey emotions in a VR game (Hart et al., 2018) and in social VR platforms (Tanenbaum et al., 2020). Another work has evaluated the effect of displayed facial expression on peripersonal and interpersonal spaces in a social VR setting (Ruggiero et al., 2017). Despite having some interest in facial expression in VR, none of the earlier work has used facial expressions as a mode of interaction paradigm to enable basic operations, such as navigation and action in VR. As such, our current experiment is novel because we have used the detected facial expression to implement interactions in VR.

## 3. Experimental system design

In order to use facial expressions as a means of interaction, three environments were designed using the Unity3D game engine[7] (version 2019.1.0f2), utilising various C# scripts and some prefabs from the asset store. SteamVR was utilised to enable the use of the HTC VIVE Head Mounted Display (HMD) and for the controllers to issue commands directly to Unity. The Emotiv EPOC+ EEG headset was used as means to gauge the facial expressions and record neural activity. It was decided that an EEG headset would be a good tool to record facial activity since it is well known that movement artefacts tend to contaminate EEG signals. In the case of this study, this contamination of EEG signals was used as a marker for facial expressions in order to facilitate interactions within VR. Additionally, an EEG headset was thought to be significantly comfortable in comparison to affixing EMG sensors to participants' faces.

The system interfaced the Emotiv EPOC+ via Node-RED, a flow-based programming tool, which allowed facial expressions and mental command information to trigger keystrokes. These keystrokes were directed at the application in focus (Unity), where the keys were bound to actions in the system exercisable by the player in the virtual environment. These actions included moving along a fixed path, stopping, and an action that would interact with the environment (picking up an object, shooting, or waving a net). See Fig. 1 for a detailed schematic diagram of the experimental system.

### 3.1. Interacting with facial expressions (FE)

There are two ways to interact with the system using facial expressions. The first is to build a framework directly in Unity. The other is using Node-RED to map each Facial Expression to the keys on the keyboard. These two methods have their own advantages and disadvantages. The Emotiv framework can respond faster because it sends instructions directly to the player. However, the disadvantage of the framework is if the command is not triggered, it is difficult to determine whether it is a failure due to bugs or because an FE has not been identified. Node-RED was more generalized and easier to test. It mapped commands to the keyboard and the results were displayed more intuitively in the console.

The schematic of the Node-RED code is illustrated in Fig. 1. It connected the Emotiv app to accept FE commands and simulated keyboard input. A threshold was set within the app to control the level of the signal strength that could trigger keyboard inputs. When the strength of the command signal met or exceeded the set threshold, the simulated key would be pressed. Three different facial expressions were used in this experiment—smile, frown, and clench (see Fig. 2). A smile was used to trigger the "move" command, the frown for the "stop" command and the clench for the "action" command. It should be noted that an action command refers to waving a bat in the happy environment, picking up objects in the neutral environment, and shooting in the scary environment. These actions were chosen to match, in some sense, the most logical actions that would not surprise participants and these were easier to perform for interaction than other facial expressions commonly recognized by the Emotiv system.

## 4. Virtual environments

For this experiment, we developed three different virtual environments - happy, neutral and scary - with different experiences. The motivation for developing these three different virtual environments was to explore if changes in the environment affected the use of FEs based on the emotional and physiological responses these environments were developed to evoke. A recent study identified that easier interaction and positive emotion caused higher presence (Pallavicini et al., 2020). As we are proposing a novel interaction method, it is important to investigate its performance in different VEs inducing different emotions. In addition to monitoring FEs, the use of the Epoch headset allowed us to monitor the neural activity as well. The experience in each of the environments lasted for four minutes. Within each environment, participants were initially stationary, and could issue a move command to start moving along a predefined path. The stop command could be used to stop whenever required in order to interact with the environment. We fixed the movement speed to be equal in all environments to avoid any confounding effects. After four minutes, the experience automatically ended. The following subsections explain what participants encountered and the actions they were expected to perform on each of the three VEs.

### 4.1. Happy

In the happy environment (see Fig. 3a), participants were tasked with navigating a park to catch butterflies. Participants were required to use a net to catch the butterflies. For those participants that used the HTC VIVE, the behaviour of the net was governed by the controller i.e. the net moved depending on how the controller was moved by the participant. In case of the participants using FEs to interact with the environment (Emotiv Epoc+), two facial expressions - smile and frown - were used to start and stop movement respectively. The clench' FE was used to initiate the action command. This resulted in the movement of the net, enabling the participants to capture butterflies.

### 4.2. Neutral

In the Neutral environment (Fig. 3b), participants were tasked with navigating a workshop to pick up items strewn throughout. When using the HTC VIVE controllers, hovering them in close proximity to an object highlighted it in green. Pulling the trigger resulted in the object being picked up. Those who used FEs as a means of control were required to position the item in the centre of their vision, at which point the item would be highlighted. The interaction command (clench) could then be used to initiate the action of picking up the item. Start and stop commands were initiated with the smile and frown FEs as described earlier.

### 4.3. Scary

In the Scary environment (Fig. 3c), participants were tasked with navigating an underground base to shoot zombies. Participants could shoot while at rest or in motion. The direction in which the participants shot was dependent on the position of their heads. A red dot in the centre of their field of view served to inform participants of the direction of their gaze. This helped them accurately focus shots during the virtual experience. The zombies were programmed to walk towards the participants and attack the participants where they were less than or equal to two meters away. Participants were able to shoot at the zombies from any distance and eliminate a zombie after scoring five hits on it.

[7] https://unity.com/

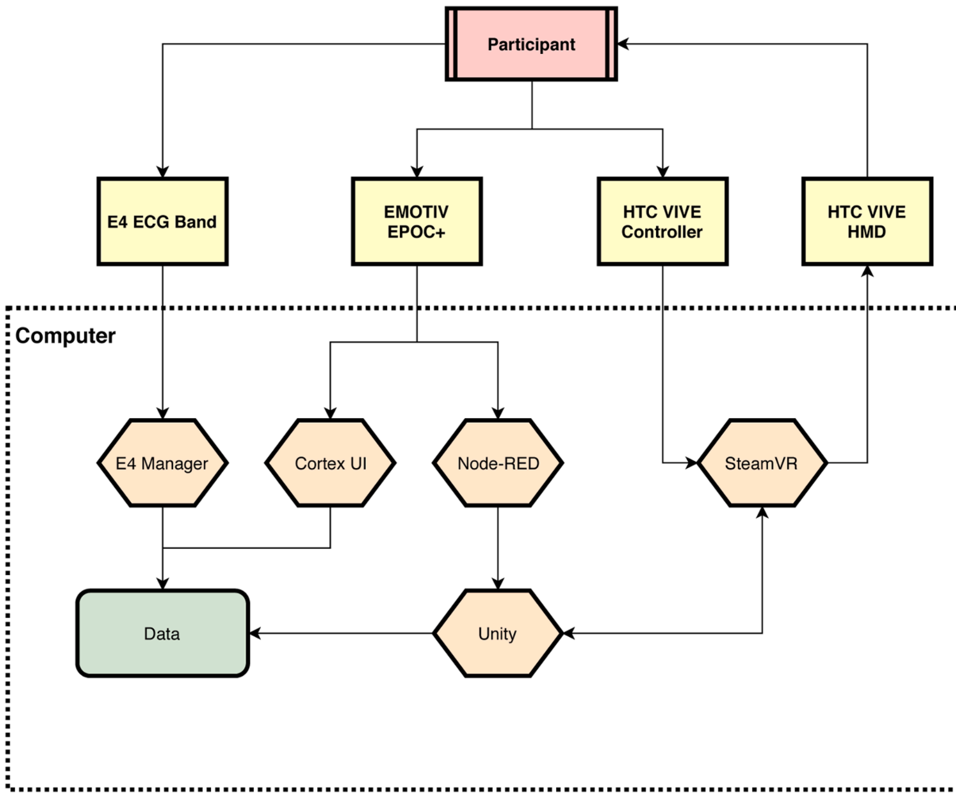



**Fig. 1.** A schematic diagram of the experimental system showing each of the individual components.

**Fig. 2.** Facial expressions used to interact with the virtual environments. Left - smile; middle- frown, and right - clench.

Participants encountered up to three zombies at a time in the VE. For participants using controllers to interact in the VE, shooting was accomplished by depressing the trigger button on the controller. Those using FEs to interact with the environment used the same sets of FEs as described earlier.

## 5. User evaluation

To evaluate the performance of the interaction methods we performed a *mixed-factorial* experiment with two independent variables.

### 5.1. Independent variables

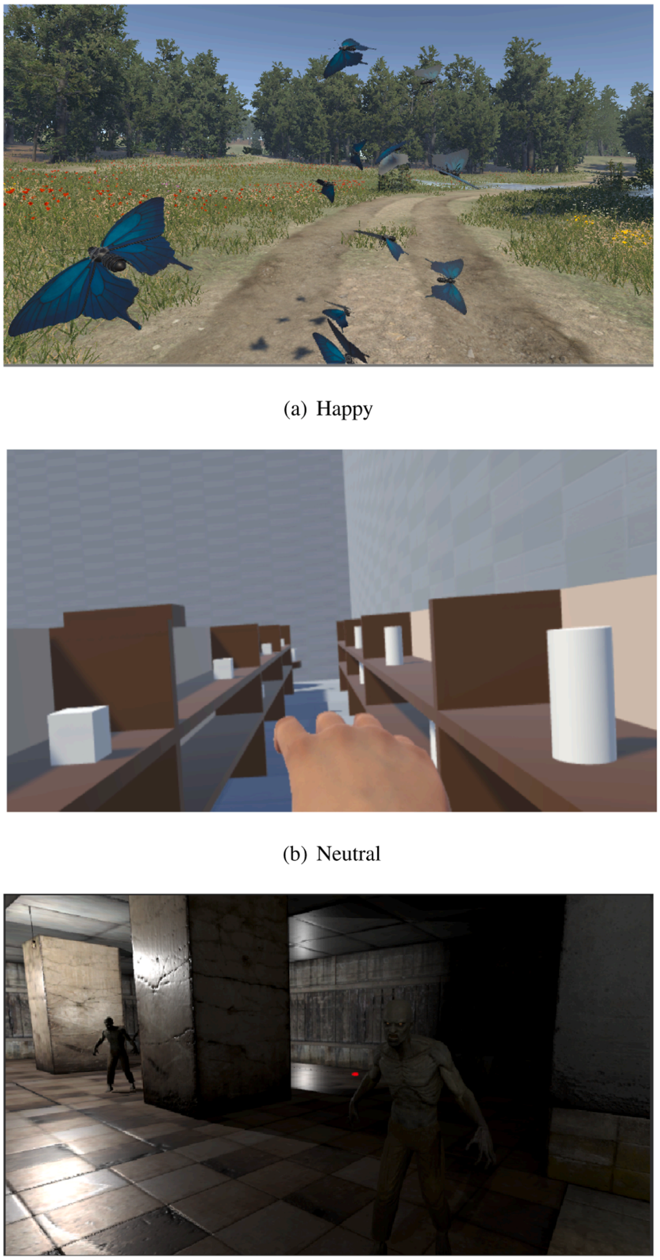
- Interaction Methods (Facial Expressions, Controller) – *between-subjects*
  We compared the performance of interactions in virtual environments using facial expressions and handheld controllers. The hand-held controllers were used as a baseline condition. The details about these interaction methods are provided in the section Experimental System Design. Being a between-subject variable, each participant used only one method of interaction.
- Virtual Environments (Happy, Neutral, Scary) – *within-subjects*
  We created three virtual environments as described in the section Virtual Environments. Each of these environments required participants to initiate movement, stop movement (optional), and select actions. The actions participants were required to undertake were; catching butterflies with a net in the happy environment, picking up objects in the neutral environment, and shooting zombies in the scary environment. Each of the three environments provided differing experiences and cognitive load. All participants experienced all of these environments using their assigned interaction method.

### 5.2. Dependent variables

We also used a mix of subjective and objective variables. Three subjective variables that were measured were presence, using the Slater-Usoh-Steed Presence questionnaire (Slater et al., 1994), emotions, using the Self Assessment Manikin (Bradley and Lang, 1994), and usability, using the System Usability Scale (SUS) (Brooke et al., 1996). Objective data collected during the study, included EEG recordings and and EDA. Finally, participants were also asked to provide an overall impression of their experience of using the designated interaction method in an open ended and informal interview.

### 5.3. Hypotheses

Overall, we expected the controllers to perform better, as they are a more intuitive interaction method than facial expressions. Our goal was to find out how FEs compared with the most commonly used interaction method in VR - hand-held controllers. As stated earlier, FEs were considered in order to study their feasibility as an interaction type for users who are differently-abled. Our study was informed by the following hypotheses:



(a) Happy

(b) Neutral

(c) Scary

**Fig. 3.** Experimental Environments. Each of them lasted for four minutes.

- **H1:** A higher cognitive load is observed with higher task difficulty (Held et al., 2017) and learning novel information (Sweller, 2011). As interacting with FE is a novel skill that requires more concentration and physical effort, we expected that a higher cognitive load will be observed while using FE, which will be reflected as increased Gamma activity.
- **H2:** Beta activity increases during emotionally negative experiences, and tasks that require large cognitive resources (Ray and Cole, 1985). Using facial expressions as a means to interact in a VE is likely to increase cognitive load and may also induce negative emotions. This is primarily due to the difficulty involved in learning and using FEs.
- **H3:** Using FEs is a novel method to interact in a VE. Given that this method is novel, we expected that participants would find it relatively hard to use. Accordingly, we expected FEs to receive a lower SUS rating.
- **H4:** Due to the difficulty in mastering the FE based interaction participants are likely to feel less dominant in this condition than the controllers. At the same time, they will be less aroused while using facial expressions than the controllers.
- **H5:** Using controllers requires more physical movement and we expected that this will increase skin conductance (EDA) more than the facial expressions.

### 5.4. Procedure

The experiment was of a mixed-factorial design. Hence, every participant used only one interaction method, but experienced all three environments. In the beginning, we welcomed participants, asked them to sit on a comfortable chair, read the information sheet, and an experimenter explained the overall study and answered all questions that the participants had. Participants then signed the consent form, and filled out a demographic questionnaire.

They were then equipped with all the sensors. For the purpose of this study Electrodermal Activity (EDA) have been collected using the Empatica E4 wristband (Garbarino et al., 2014), whilst EEG was captured using an Emotiv Epoc+. After the sensors were attached to the body and a reliable connection was established with the computer, we collected data for two minutes that served as a baseline measurement. These two minutes were split equally into data collected in an eyes closed condition for a minute and another minute with their eyes open. During this phase of data collection participants were not made to wear the VR headset. Next, the tasks that participants were required to perform - start, stop, and action - in the environments were explained to them. For participants using FE, this was followed by a training session that lasted approximately fifteen minutes. Participants were required to use FEs during this session in order to help train the system to effectively recognise the FEs during the study. This training phase was required for each individual participant as the FE recognition system we used is sensitive to individual differences in performing each of the facial expressions. Following the training session, the first environment was displayed and participants started the task. After a fixed duration of four minutes, the experiment was automatically halted. Participants then removed the VR HMD, answered the questionnaires and rested for as long as they needed to feel comfortable before beginning the experiment in the next environment. The entire experiment took about 80 minutes per participant on average. Participants were seated for the entire duration of the experiment on a chair that could be rotated. Irrespective of the interaction methods participants used, they were all required to hold the controllers. This was done to curtail potential confounding variables that could have caused problems during the analysis phase (See Fig. 4). However, they were instructed to avoid physical movement beyond necessary to avoid inducing noise in the data. For the FE condition, participants rested their hands (holding the controllers) on their thighs, as they were not required to use the controllers. The entire experiment was performed in a noise free air-conditioned room at the university.The study was approved by the relevant Human Participants Ethics Committee prior to recruitment and data collection.

### 5.5. Participants

We recruited a total of 18 adult participants[8] (nine in each group) without any physical or psychological disability. Among the participants, two were female and one participant preferred to abstain from disclosing their gender. The average age of the participants was 23.6 years (SD = 4.2). A self-rated score (out of 5) of familiarity and/or experience with VR demonstrated that participants did not have much experience with VR (Mean = 2.7, SD = 1.5). The 18 participants were

[8] Recruitment of more participants was not possible due the COVID restrictions.

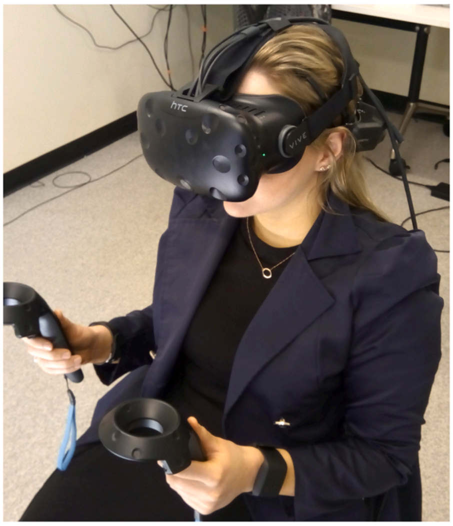



**Fig. 4.** A participant (reenacted) in the experimental sessions wearing Emotiv Epoc+ EEG, Empatica E4 wristband, and HTC Vive.

divided equally between the two interaction types - FE and hand-held controllers.

## 6. Results

In this section we detail the results of the data analysis that was performed using SPSS v25 statistical package. We used a mixed-factorial ANOVA for all of the dependent variables, except for the neurological data. Where an ANOVA showed a statistically significant difference ($p < .05$) for virtual environments, we performed further pair-wise comparisons using Bonferroni corrections. While there is not a consensus on the use of ANOVA for Likert-scale based data, we found it is an appropriate test to use following (Carifio and Perla, 2007; Glass et al., 1972). Given the between subjects nature of the interface conditions, it must be noted that the 18 participants were divided into two sets of nine each for each of the interactions - controller and FE.

Overall, we noticed an effect of interaction methods on the gamma activities in the brain. An interaction effect was found on EDA. In some aspects of presence facial expression outperformed controllers but controllers were found to be better than facial expressions in terms of usability.

### *6.1. Electroencephalogram (EEG)*

#### *6.1.1. Data acquisition and pre-processing*

Raw EEG data was collected using the 14 channel Emotiv Epoc+ EEG device at a sampling rate of 256 Hz. This device uses the standard 10–20 electrode system (Jasper, 1958) and has AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, and AF4 as the channels of the device.

The pre-processing for the controller group was done using the pre-processing pipeline proposed in (Sareen et al., 2020b), was executed in EEGLAB (Delorme and Makeig, 2004), and previously implemented in (Sareen et al., 2020a). The pre-processing steps involved a bandpass filter of 1–60 Hz over the raw EEG data to filter out unnecessary signal information and the DC-offset of the device. The filtering was followed by the removal of the 50 Hz line-noise. The filtered and noise-free data was decomposed into its constituent components using Independent Component Analysis (ICA) to remove the muscular and ocular artifacts from the data.

For the FE group, similar steps were followed up to the artifact rejection stage. It was pivotal in the case of the FEs to account for the nature of the interaction design. In this group, the data was heavily corrupted with the artifacts of facial Electromyograph (fEMG), which were of relatively higher amplitude than the neural signals. This lead to a relatively suppressed response of the neural signals. Moreover, this form of interference corrupted the EEG data with generic discontinuities as well. To address this experiment-specific artifact, the power spectrum of each component was analyzed, and their contribution to the data spectrum estimated in EEGLAB. The fMEG artifacts have been reported to have relatively higher spectral amplitudes than the neural signals across the full spectrum (Yong et al., 2008) and thus have a higher contribution percentage than the latter.

In our analysis, we also observed a characteristically similar nature of components contributing to fMEG artifacts, as mentioned in (Yong et al., 2008). Two of the highest contributing components were identified and rejected from the data. Finally, the clean and artifact-free data of the groups was filtered into five frequency bands, namely delta (1–4 Hz), theta (4–8 Hz), alpha (8–14 Hz), beta (13–25 Hz), and gamma (25–60 Hz) for further analysis.

#### *6.1.2. Analysis*

Considering the difference between the two interactions methods, we hypothesized that an increased cognitive load or a stressful affective brain state would be observed in FE interaction than the controller-based interaction. The effect of cognitive load is known to be reflected in the frontal and parietal regions of the brain (Chai et al., 2018). Thus, to evaluate our hypothesis, we selected a subset of channels linked to these brain regions, namely F7, F3, FC5, P7, P8, FC6, F4, and F8. All further analysis was done on this subset of channels only.

To compare the differences between the two interactions methods in the three environments across the participants, the Power Spectral Density (PSD) was computed using Welchs spectral power estimate for all the selected channels. Fig. 5 depicts the PSD plots of the channels where significant differences were observed. The figure only depicts the frequency resolution of 1–40 Hz were characteristic differences were visible. It was observed from this figure that, in the mean sense, the facial expression-based interaction group depicted characteristically higher peaks between 1–3 Hz (lower delta band), 16–22 Hz (Beta band), and 26–30 Hz (lower gamma band) than the controller based interaction group.

To further validate these observable differences, we compared the two groups statistically, especially for delta, beta, and gamma bands, where these differences were observed in PSD plots. We applied the Mann-Whitney $U$ test for statistical comparison, considering the small sample size of nine participants in two independent groups with an unknown distribution. This test is a non-parametric equivalent of the two-sample $t$-test. The total spectral power for each of the eight channels of interest, for each participant and each of the three frequency bands of interest, was computed and statistically compared between the two interaction methods. A right-tailed Mann-Whitney $U$ test was used where the rejection of null-hypothesis would indicate higher spectral power in the facial expression-based interaction group than the controller-based interaction group. Due to the small sample size of participants, a less strict significance level of $p < 0.10$ was kept as the threshold for significance. In the delta and beta band, no significant differences were observed at $p < 0.10$. Whereas in the gamma band, significant differences were observed in all three environments at $p < 0.10$ significance level. Specifically for the happy and neutral environment types, for most of the channels of interest, a much stricter significance level of $p < 0.05$ was also satisfied. Further, due to the small sample size, we also estimated the effect size using Cohens d. Cohens effect size value ($0.5 < d < 0.8$), which suggested a moderate to high practical significance at F7, FC5, and FC6 positions in all the three environments; moderate to high ($0.5 < d < 0.8$) practical significance at F3 and high ($d > 0.7$) practical significance at F8 in Neutral and Scary environments and high ($d > 0.8$) practical significance at P8 position in Happy environment. However, at F4 and P7 positions small effect size was observed ($d < 0.5$). Table 1 depicts the results of significant statistical comparisons observed in the gamma band for all three

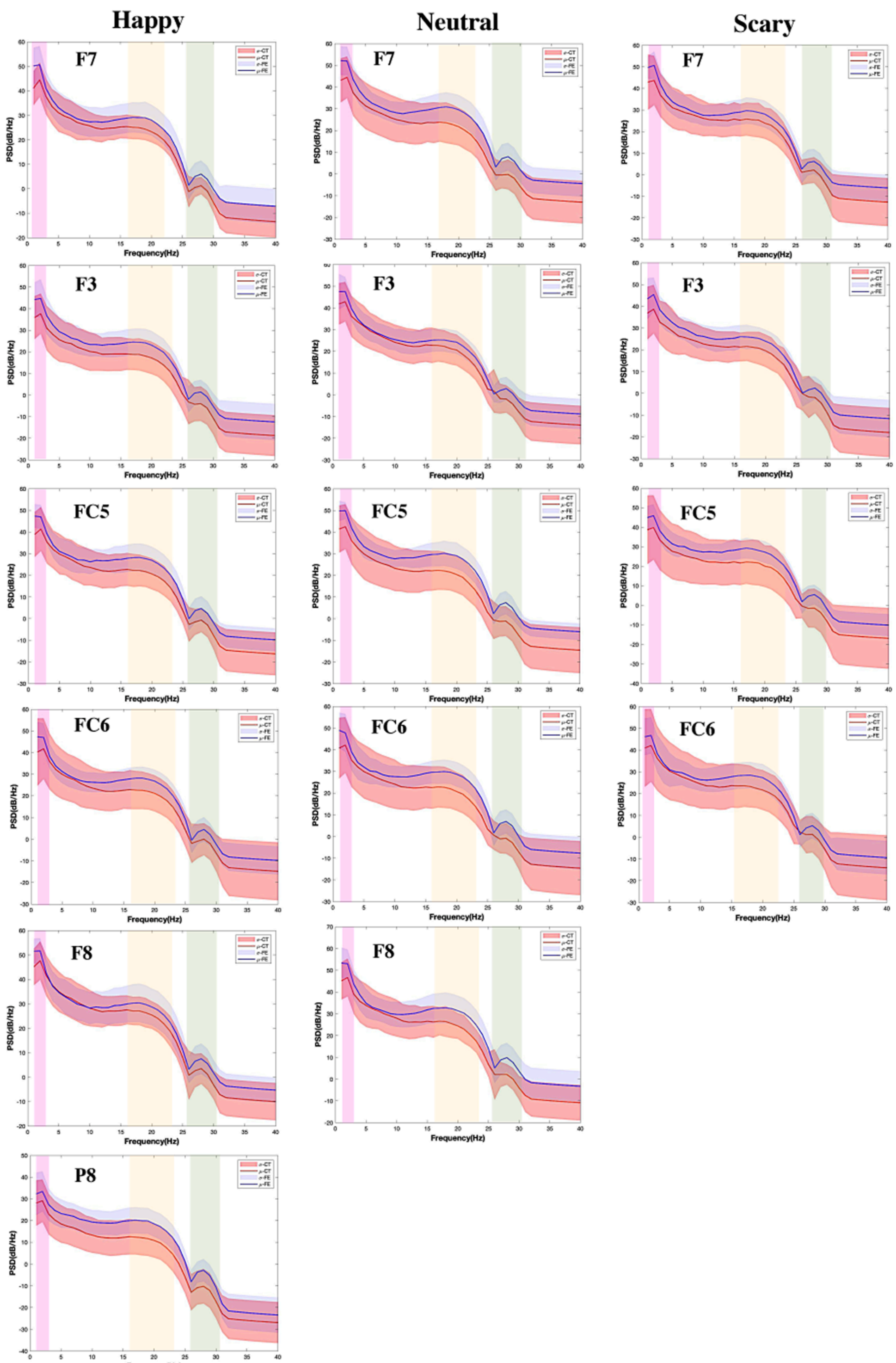



**Fig. 5.** PSD plots of channels of interest for comparison between facial expression-based interaction (blue) and controller-based interaction (red). In each plot the red/blue colored line and shaded area depicts the mean and the standard deviation respectively across the participants in that group. The magenta, yellow and green shaded area across the frequency spread depicts the region where characteristic peaks were observed in 1–3 Hz (lower delta band), 16–22 Hz (Beta band), and 26–30 Hz (lower gamma band) respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

environments.

Observations from the PSD plots and statistical analysis suggest that a characteristic differences exist in the gamma band with relatively higher activity in the facial expression-based interaction group as compared to the controller-based interaction group. Research studies have suggested that the gamma band is known to be engaged in cognitive functioning, and the prominence of activity in this band reflects a state of anxiety and stress (Abhang et al., 2016). Thus, observing significantly higher activity in the gamma band of facial expression-based interaction group might suggest an increased cognitive load experienced by the participants. This finding aligns with our initial hypothesis of observing cognitive load while using an unconventional interaction method like FEs as the participants will require more training and concentration compared to conventional interaction methods like controllers.

Further, we wanted to explore any characteristic differences between the three environments, namely happy, neutral, and scary in the facial expression-based interaction group and the controller-based interaction group. For this, we compared the boxplots of the averaged spectral power of all the eight channels for all the participants in a group while

**Table 1**
Total spectral power statistical comparison between facial expression-based interaction group and controller-based interaction group in gamma band for three environment types. Due to the limited number of participants we included results with $p < .1$ level and effect size measured by Cohen's d.

| Channels | Happy | Neutral | Scary |
|---|---|---|---|
| F7 | U = 107, z = 1.85, p = 0.03, d = 0.73 (large) | U = 106, z = 1.77,p = 0.04, d = 0.70 (large) | U = 101, z = 1.30, p = 0.09, d = 0.50 (medium) |
| F3 | U = 106, z = 1.77, p = 0.04, d = 0.48 (small) | U = 101, z = 1.29, p = 0.09, d = 0.50 (medium) | U = 101, z = 1.32, p = 0.09, d = 0.63 (large) |
| FC5 | U = 103, z = 1.50, p = 0.07, d = 1.00 (large) | U = 107, z = 1.85, p = 0.03, d = 0.86 (large) | U = 101, z = 1.29, p = 0.09, d = 0.72 (large) |
| P7 | – | – | – |
| P8 | U = 111, z = 2.21, p = 0.01, d = 0.85 (large) | – | – |
| FC6 | U = 104, z = 1.59, p = 0.06, d = 0.92 (large) | U = 108, z = 1.94, p = 0.03, d = 0.88 (large) | U = 101, z = 1.29, p = 0.09, d = 0.64 (large) |
| F4 | – | U = 93, z = 0.57, p = 0.28, d = 0.25 (small) | – |
| F8 | U = 101, z = 1.33, p = 0.09, d = 0.47 (small) | U = 110, z = 2.12, p = 0.02, d = 0.79 (large) | – |

experiencing happy, neutral, and scary environments for each of the five frequency bands. We again used the Mann-Whitney *U* test to statistically evaluate if any significant differences existed between the three environments for each of the five frequency bands for both the groups in comparison. The boxplots are depicted in Fig. 6 and Fig. 7. It was observed from the statistical comparison that there was no statistically significant difference between the three environments for each of the five frequency bands at $p < 0.10$ significance level. This might suggest that the different interaction modes do not affect the participants response in three different environments. This is an encouraging finding from the VR environment development point of view as a new interaction medium tested here i.e., facial expression, does not behave differently in different environment types; similar to standard interaction mediums like controllers. Hence, just like hand-held controllers, facial expression can possibly be used as an interaction medium in different VR environments without worrying about different environment types. We also observed a marked increase of spectral power in both beta and gamma bands for all the three environments in both the interaction mediums. Studies have suggested that increased beta and gamma-band power indicates a highly attentive state (gam, 2018; Abhang et al., 2016). Thus, increased beta and gamma band activity in both the interaction mediums might suggest that the participants were actively engaged in the task in all three environments.

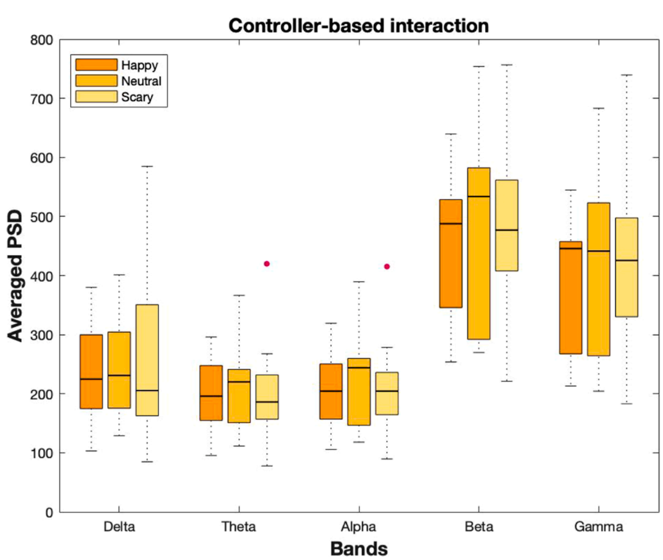



**Fig. 6.** Boxplots for comparison between the three environments in controller-based interaction group.

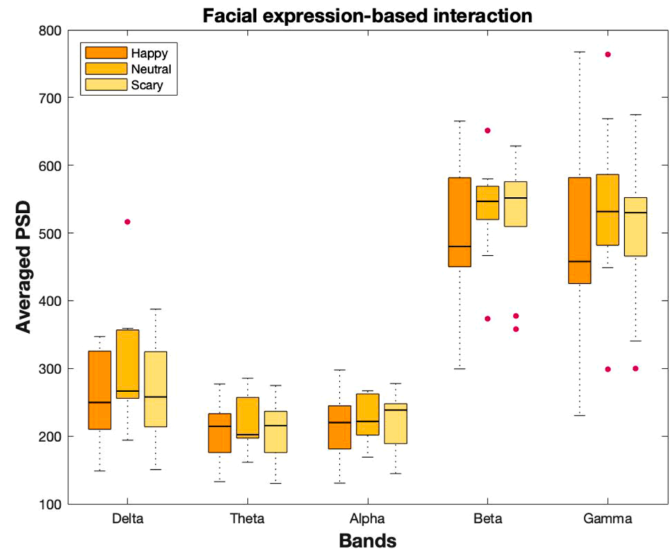



**Fig. 7.** Boxplots for comparison between the three environments in Facial expression-based interaction group.

### 6.2. *Electrodermal activity (EDA)*

Physiological data enables us to capture spontaneous and subconscious aspects of the users' state as they interact with different environments and controllers (Fairclough, 2009). EDA refers to the "variation of the electrical properties of the skin in response to sweat secretion" (Benedek and Kaernbach, 2010). A raw EDA signal is composed of two components—Skin Conductance Level (SCL) [tonic component] and Skin Conductance Response (SCR) [phasic component]. SCL is a general measure of psychophysiological activation that is slow moving (Setz et al., 2009), whilst SCRs depict higher-frequency changes that are directly related to an external stimulus (Greco et al., 2016). Typically, SCR and heart rate are the best discriminators for arousal detection (Can et al., 2019). These signals have been used in this study to evaluate the short-term effects of each environment/controller by determining increased sympathetic activity (sympathetic arousal), which elevates heart rate, blood pressure, and sweating (Poh et al., 2010).

The raw EDA signal has been pre-processed using the cvxEDA algorithm (Greco et al., 2015), which decomposes the signal into the SCL and SCR components. Statistical features were then extracted from the SCR signals, including mean, standard deviation, minimum, maximum and percentiles. This analysis was undertaken using MATLAB R2019b.

The results in Fig. 8 illustrates that EDA differed the most during the neutral controller environment. The overall results illustrate that the controller seemed to elicit a greater reaction in the majority of environments. When analysing using a mixed-factorial ANOVA we noticed a strong but not significant effect of environments—F(2, 32)=5.01, $p = .08$, $\eta_p^2$=.14, OP=.5—where neutral environment caused highest EDA. However, we noticed a significant interaction effect—F(2, 32)=7.6, $p = .03$, $\eta_p^2$=.2, OP=.7. For Happy environment, facial expression caused higher EDA than controllers but it was the opposite in other two environments as shown in Fig. 8b.

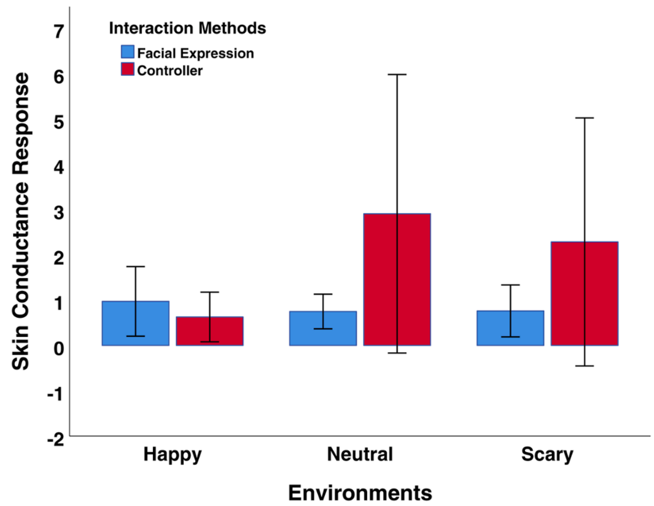



**Fig. 8.** Electrodermal activity. Whiskers represent ±95% confidence intervals.

### 6.3. Self-Assessment manikin (SAM)

SAM is a pictorial scale that measures overall emotional valance, arousal, and dominance of an experience. We used a nine-point scale.

**Valance:** For valance, we noticed a significant main effect of the environments—F(2, 30)=6.5, $p = .005$, $\eta_p^2$=.3, OP=.9. Post-hoc test showed that the Happy (M=6.7, SD=1.6) environment had significantly (p=.01) higher valance than the Scary (M=4.4, SD=2.0) environment. There was no significant effect of the interaction methods (Fig. 9).

We did not notice any significant main or interaction effect of either of the variables on arousal and dominance.

### 6.4. System usability scale (SUS)

We calculated the SUS scale score from the questionnaire as described by Brooke (Brooke et al., 1996). We noticed a significant main effect of environments—F(2, 30)=3.4, $p = .048$, $\eta_p^2$=.18, OP=.6. However, a pair-wise comparison after Bonferroni adjustments did not yield a significant difference but the highest difference (p=.1) was noted between the Happy (M=61.6, SD=25.7) and Scary (M=50.7, SD=21.8) environments.

We also noticed a significant main effect of the interaction methods—F(1, 15)=12, $p = .003$, $\eta_p^2$=.4, OP=.9—where facial expression (M=40.8, SD=19.3) received significantly lower SUS score than the controllers (M=69.4, SD=18.9) as shown in Fig. 10.

### 6.5. Presence questionnaire

There were six different questions in the presence questionnaire. The first question relates to the *overall presence or the sense of being there.* We did not notice any significant main effects of the interaction methods but we noticed a significant effect of VEs on overall presence—F(2, 30)=3.8, $p = .03$, $\eta_p^2$=.2, OP=.65. Happy environment provided a higher sense of being there than the Scary one. We also noticed a significant interaction effect—F(2, 46)=7.3, $p = .003$, $\eta_p^2$=.3, OP=.91. As shown in Fig. 11a, for Neutral and Scary environments, facial expressions had higher presence than controllers. However, for the Happy environment the effect was reversed.

The second question asked about the *feeling of realism.* We did not find any significant difference. However, there was a trend towards a significant interaction effect (p=.07), as Fig. 11b shows that facial expression provided higher realism than controllers in the Scary environment whereas it was the opposite in other environments.

The fourth question referred to the participants *feeling of being the environment or elsewhere during the experience.* Here we noticed a significant main effect of the interaction methods—F(1, 15)=4.9, $p = .04$, $\eta_p^2$=.25, OP=.6. Facial expressions made participants feel more in the environment than the controllers (Fig. 11c).

For the third, fifth, and sixth questions we did not notice any significant effect. When analyzing the overall presence by summing all of the six questions together, we did not notice a significant difference. However, there was a strong trend towards a significant main effect of the environments—F(2, 30)=3.2, $p = .055$, $\eta_p^2$=.18, OP=.6. User's felt that they had higher presence in the happy environment than the scary environment.

## 7. Discussion

We presented the first ever study that compared interacting with facial expressions in VR with a traditional handheld controller using multiple quantitative and qualitative measures, such as presence, usability, affect, neurological, and physiological effects. Previously, FE has never been used for performing interactions, such as navigation and action, in VR environments. At the outset we had five hypotheses and in this section we discuss the results in relation to those hypotheses.

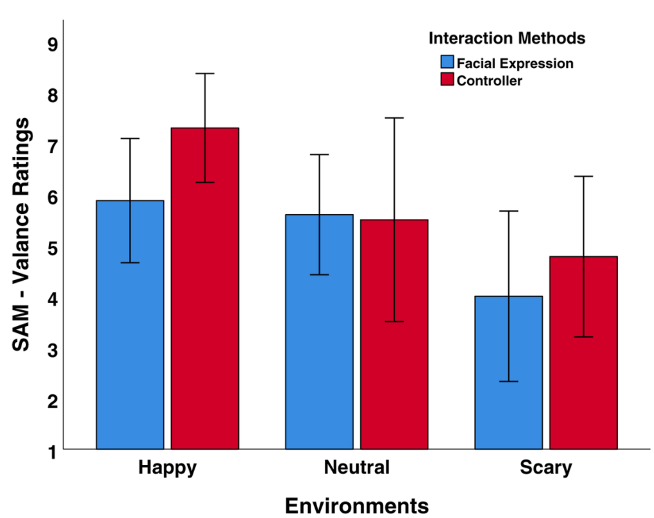



**Fig. 9.** Valance ratings from the SAM questionnaire. Whiskers represent ±95% confidence intervals.

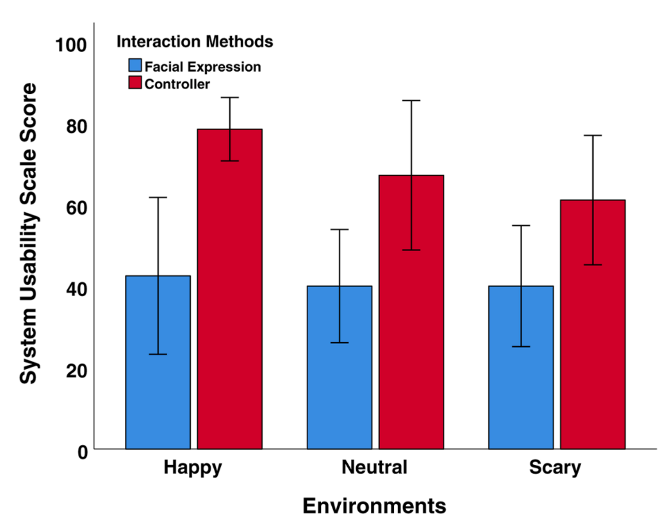
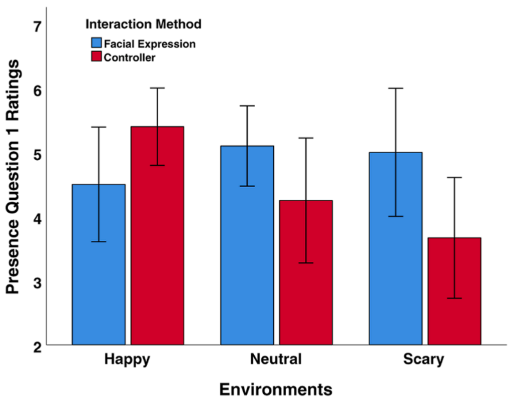
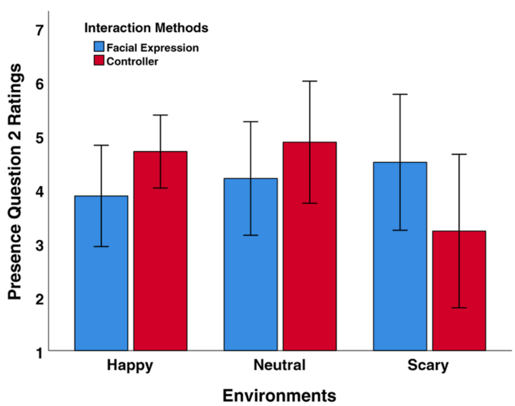


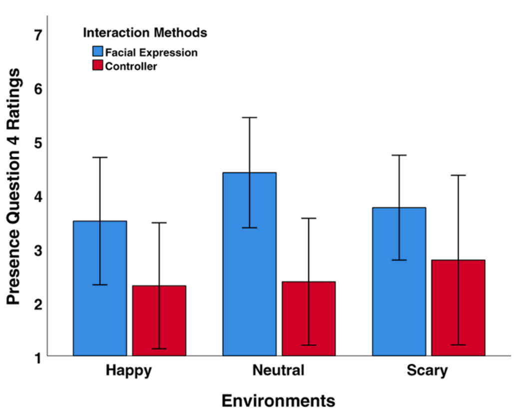

**Fig. 10.** System Usability Scale score. Whiskers represent ±95% confidence intervals.




(a) Presence Question 1




(b) Presence Question 2




(c) Presence Question 4

**Fig. 11.** Slater-Usoh-Steed Presence Questionnaire [43] ratings (only significant effects). Whiskers represent ±95% confidence intervals.

Our first hypothesis (H1) expected higher Gama activity in FE than the controllers and this hypothesis has been accepted. As Gama activity is associated with higher cognitive load (Abhang et al., 2016), we believe learning a new method for interaction and constantly working with facial muscles increased the cognitive load for FE interactions when compared with the hand-held controllers. The analysis has demonstrated that there exists a statistically significant difference between the FE and hand-held controllers. Furthermore, in the gamma band, the moderate to large effect size on most of the electrode locations (except for P8, P7, and F4) in the three environments, supplements the practical significance of the statistical results estimated even at relaxed significance criterion. The higher spectral power in the gamma band is significantly more pronounced ($p < 0.05$) in the happy and neutral environments. We partially attribute this finding to the need of using clench gesture significantly more than one would in the scary environment. This is because, while only five shots were required to neutralise a zombie, something that was fairly easily accomplished, there were no such restrictions on how many time one had to wave a net before being able to catch a butterfly or the number of objects one could collect with the period for which the simulation ran i.e. four minutes. Additionally, we observed that participants tended to lean forward and backwards when using the FE interaction technique - a possible consequence of not fully understanding and/or being used to the technique. Although, one participant (P3) mentioned that it felt natural to move forward and back when navigating using a smile and frown while she understood that body movement is not required for performing the experimental tasks.

H2 stated that an increase in Beta activity would be observed during an emotionally negative experience and/or one requiring high cognitive resources. While our analysis does show this to be the case, the increase in Beta activity between both interaction conditions across three environment types is not statistically significant. This result is a likely indication that while participants do find it harder to use the FE based interaction method (H3), there is sufficient evidence to warrant a deeper investigation of this phenomena. That no statistically significant difference exists points to the fact that the FE based interaction technique could potentially be implemented with the understanding that it would have a higher initial learning curve; following which there would be little or no difference in the cognitive load caused by it on the user in comparison to using the hand-held controllers.

Our third hypothesis (H3) expected FEs to be scored lower on the SUS. Our results demonstrate that this is the case in our study. Two major contributing factors for such a result could be; the lack of intuitiveness of such an interaction methodology and the fact that participants had to invest close to fifteen minutes prior to the study in order to learn the interactions and train a system to recognise them. However, it must also be noted that the use of FEs may be appropriate in cases where hand-held controllers cannot be used due to physical constraints or in cases where facial expressions are more natural way of interaction than the controllers. A case in point, P9 stated "I may not use it (facial expression) when I can use my hands. But I'd definitely use facial expressions for interactions if I cannot use my hands and in this case I think it provides higher opportunities than the challenges." There is a possibility that with more training and practice the usability of facial expression-based interaction will improve. Another participant (P17) said "I initially struggled to interact with it but I got a hold of it later."

H4 predicted lower dominance and arousal in the case of facial expression-based interaction than controller-based. This hypothesis was not supported as we did not find any significant difference between the interaction methods when dominance and arousal were measured using the SAM scale. This finding is encouraging as it demonstrates, similar to H1, that facial expression-based interaction can be used without any additional emotional burden.

Our final hypothesis (H5) states that the use of hand-held controllers would result in a higher skin conductance (EDA) due to greater levels of movement required for the interactions using this technique. Our results demonstrate that there is a statistically significant difference between the hand-controllers and the FE interaction methods in the neutral and scary environments, with the controllers demonstrating higher EDA levels in both cases. No such differences were observed between the two interaction types in the happy environment. We have been unable to fully understand the nature of these results, other than to speculate that the bodily motions induced by the neutral and scary conditions were significantly greater for the two environments than in the happy environment. These movements could have resulted in the recorded data being saturated by extraneous artifacts which were not properly removed in the processing stage.

It is interesting to note that we have found FE to provide a significantly higher sense of presence in the virtual environments than controllers. We believe the explicit interaction with the environment afforded by the FE technique may have lead the participants to feel connected to the environments in a manner that the hand-held controllers could not. It shows that FEs present an opportunity to improve the experiences in VR while increasing its accessibility.

Overall, from this first ever comparative exploration of FE against commonly used controllers have shown that FE is indeed a viable options to navigate and perform actions in virtual environments. While it initially requires higher cognitive resources and results in lower usability compared to hand-held controllers as it is a new method of interaction but participants can get used to it after awhile and it can increase presence in VR. Facial expressions do not induce any additional emotional burden. Hence, for users who cannot use hand-held controllers or in situations where hand-held controllers do not provide a natural interaction, facial expressions can be used.

### 7.1. Limitations & future work

While we have obtained some interesting insights from this study and showed the viability of FEs as an interaction method for VR application, the findings of the current study has a few limitations.

First, this study made use of only three FEs. These enabled the participants to execute only the most basic interactions in VR. For an interaction method to be seamless and intuitive, it must be able to leverage actions that are natural to a user. Forcing the user to learn a set of complicated and disjointed actions can make an interaction method cumbersome to use. Future studies can explore how to exploit the natural movements of human facial musculature to provide more natural and intuitive interaction. Additionally, it will also be of significant importance if this method of interaction can be combined with existing methods in order to provide a greater level of control over the interactions being performed in a VR environment.

Second, the EEG headset used in this study was an inexpensive, consumer-grade hardware that provided us with 14 channels. While this shows that the proposed interaction technique can be implemented with relative ease using such an inexpensive system, more neurological data needs to be acquired during such studies. This will allow us to gain deeper insights into how the brain works when confronted with interaction methods that may not, at first glace, seem intuitive to use. To enable this, future studies will make use of high quality gel based electrode systems with 32 or more channels. This higher density of electrodes will be able to provide a better picture of neural activity and how it related to different interaction methods.

Third, the current study did not identify the neural activity based on the individual events or actions that the participants performed such as start, stop, and action. Future studies can use the stratification of the different actions to identify the neural effects of these actions to help create more efficient and usable interaction strategies.

Fourth, our statistical analyses and the stated cognitive associations should be interpreted with caution, given the relaxed significance criterion followed in this work and largely varying literature on frequency band and cognitive associations. Future studies should expand our presented framework and aim for a larger dataset, stricter significance criteria, and depth investigation of frequency band-cognitive associations to further validate our findings.

In addition to the limitations and future work listed above, an important aspect to consider in the future is the social aspect of interaction methodologies in collaborative virtual environments. Careful consideration must be given to the use of FEs as an interaction method in collaborative virtual spaces in order to avoid socially embarrassing or seemingly offensive interactions stemming from an incorrect implementation of the FE interaction technique. A similar caution needs to be used in FE interactions to avoid unintended interactions by expressions caused by accidental or natural reactions to stimuli in the environment.

## 8. Conclusions

In this paper, we have presented one of the first studies that has explored facial expressions as an interaction method in VR and compared this with traditional handheld controllers. We have used neurophysiological signals, emotional effects, presence, and usability to evaluate the differences between these two modes of interaction. We used an off-the-shelf EEG headset to obtain both neural activity and facial expressions. EDA was obtained using the Empatica E4. We expected the hand-held controllers to outperform FE based interaction. However, we have found that, both of the interaction methods are comparable in most measures. Task engagement was equally high for both of the interaction methods. Although, we noticed an increased cognitive load when FEs were used, participants were able to use them successfully to perform specific tasks. In terms of presence, facial expressions provided higher realism and feeling of being in the environment than the controllers. Overall, it shows that FE can be used as a viable alternative interaction method to hand-held controllers where the users cannot use a hand-held controller or they do not provide a natural interaction. However, the fact that there was such a clear difference in the usability ratings between the two methods, there is a need for further research to make FEs easier and more intuitive to use.

## CRediT authorship contribution statement

**Arindam Dey:** Conceptualization, Methodology, Formal analysis, Resources, Data curation, Supervision, Writing – original draft, Writing – review & editing, Project administration. **Amit Barde:** Writing – original draft, Writing – review & editing, Formal analysis. **Bowen Yuan:** Software, Investigation. **Ekansh Sareen:** Formal analysis. **Chelsea Dobbins:** Formal analysis. **Aaron Goh:** Software, Investigation. **Gaurav Gupta:** Software, Investigation. **Anubha Gupta:** Formal analysis. **Mark Billinghurst:** Supervision, Writing – review & editing.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References

2018. Gamma brain waves: 40 hz to 100 hz. http://mentalhealthdaily.com/2014/03/12/gamma-brain-waves-40-hz-to-100-hz/.
Abhang, P.A., Gawali, B.W., Mehrotra, S.C., 2016. Introduction to EEG-and speech-based emotion recognition. Academic Press.
Amores, J., Benavides, X., Maes, P., 2016. Psychicvr: Increasing mindfulness by using virtual reality and brain computer interfaces. Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems. ACM.2–2
Benedek, M., Kaernbach, C., 2010. A continuous measure of phasic electrodermal activity. J. Neurosci. Methods 190 (1), 80–91.
Benzina, A., Dey, A., Toennis, M., Klinker, G., 2012. Empirical evaluation of mapping functions for navigation in virtual reality using phones with integrated sensors. Proceedings of the 10th asia pacific conference on Computer human interaction, pp. 149–158.
Bernal, G., Yang, T., Jain, A., Maes, P., 2018. Physiohmd: a conformable, modular toolkit for collecting physiological data from head-mounted displays. Proceedings of the 2018 ACM International Symposium on Wearable Computers, pp. 160–167.
Boletsis, C., 2017. The new era of virtual reality locomotion: a systematic literature review of techniques and a proposed typology. Multimodal Technologies and Interaction 1 (4), 24.
Bolt, R.A., 1982. Eyes at the interface. Proceedings of the 1982 conference on Human factors in computing systems, pp. 360–362.
Borst, C., Indugula, A., 2005. Realistic virtual grasping. IEEE Proceedings. VR 2005. Virtual Reality, 2005., pp. 91–98. https://doi.org/10.1109/VR.2005.1492758.
Bowman, D.A., Wingrave, C.A., Campbell, J.M., Ly, V.Q., Rhoton, C., 2002. Novel uses of pinch glovesfor virtual environment interaction techniques. Virtual Real 6 (3), 122–129.
Bozgeyikli, E., Raij, A., Katkoori, S., Dubey, R., 2016. Point & teleport locomotion technique for virtual reality. Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play, pp. 205–216.
Bradley, M.M., Lang, P.J., 1994. Measuring emotion: the self-assessment manikin and the semantic differential. J Behav Ther Exp Psychiatry 25 (1), 49–59.

Brooke, J., et al., 1996. Sus-a quick and dirty usability scale. Usability evaluation in industry 189 (194), 4–7.

Burgos-Artizzu, X.P., Fleureau, J., Dumas, O., Tapie, T., LeClerc, F., Mollet, N., 2015. Real-time Expression-sensitive Hmd Face Reconstruction. SIGGRAPH Asia 2015 Technical Briefs, pp. 1–4.

Can, Y.S., Arnrich, B., Ersoy, C., 2019. Stress detection in daily life scenarios using smart phones and wearable sensors: a survey. J Biomed Inform 103139.

Carifio, J., Perla, R.J., 2007. Ten common misunderstandings, misconceptions, persistent myths and urban legends about likert scales and likert response formats and their antidotes. Journal of Social Sciences 3 (3), 106–116.

Cha, H.-S., Choi, S.-J., Im, C.-H., 2020. Real-time recognition of facial expressions using facial electromyograms recorded around the eyes for social virtual reality applications. IEEE Access 8, 62065–62075. https://doi.org/10.1109/ACCESS.2020.2983608.

Chai, W.J., Abd Hamid, A.I., Abdullah, J.M., 2018. Working memory from the psychological and neurosciences perspectives: a review. Front Psychol 9, 401.

Coogan, C.G., He, B., 2018. Brain-computer interface control in a virtual reality environment and applications for the internet of things. IEEE Access 6, 10840–10849.

Crawford, J.R., Henry, J.D., 2004. The positive and negative affect schedule (panas): construct validity, measurement properties and normative data in a large non-clinical sample. British journal of clinical psychology 43 (3), 245–265.

Delorme, A., Makeig, S., 2004. Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis. J. Neurosci. Methods 134 (1), 9–21.

Dey, A., Chatburn, A., Billinghurst, M., 2019. Exploration of an eeg-based cognitively adaptive training system in virtual reality. 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), pp. 220–226. https://doi.org/10.1109/VR.2019.8797840.

Fahmi, F., Tanjung, K., Nainggolan, F., Siregar, B., Mubarakah, N., Zarlis, M., 2020. Comparison study of user experience between virtual reality controllers, leap motion controllers, and senso glove for anatomy learning systems in a virtual reality environment. IOP Conference Series: Materials Science and Engineering, Vol. 851. IOP Publishing, p. 012024.

Fairclough, S.H., 2009. Fundamentals of physiological computing. Interact Comput 21 (1–2), 133–145.

Feng, M., Dey, A., Lindeman, R.W., 2016. An initial exploration of a multi-sensory design space: Tactile support for walking in immersive virtual environments. 2016 IEEE Symposium on 3D User Interfaces (3DUI), pp. 95–104. https://doi.org/10.1109/3DUI.2016.7460037.

Friedman, D., 2015. Brain-computer interfacing and virtual reality. Handbook of Digital Games and Entertainment Technologies 1–22.

Galvan Debarba, H., Bovet, S., Salomon, R., Blanke, O., Herbelin, B., Boulic, R., 2017. Characterizing first and third person viewpoints and their alternation for embodied interaction in virtual reality. PLoS ONE 12 (12), 1–19. https://doi.org/10.1371/journal.pone.0190109.

Garbarino, M., Lai, M., Bender, D., Picard, R.W., Tognetti, S., 2014. Empatica e3a wearable wireless multi-sensor device for real-time computerized biofeedback and data acquisition. 2014 4th International Conference on Wireless Mobile Communication and Healthcare-Transforming Healthcare Through Innovations in Mobile and Wireless Technologies (MOBIHEALTH). IEEE, pp. 39–42.

Glass, G.V., Peckham, P.D., Sanders, J.R., 1972. Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance. Rev Educ Res 42 (3), 237–288.

Greco, A., Lanata, A., Citi, L., Vanello, N., Valenza, G., Scilingo, E., 2016. Skin admittance measurement for emotion recognition: a study over frequency sweep. Electronics (Basel) 5 (3), 46.

Greco, A., Valenza, G., Lanata, A., Scilingo, E.P., Citi, L., 2015. Cvxeda: a convex optimization approach to electrodermal activity processing. IEEE Trans. Biomed. Eng. 63 (4), 797–804.

Gruebler, A., Suzuki, K., 2014. Design of a wearable device for reading positive expressions from facial emg signals. IEEE Trans Affect Comput 5 (3), 227–237.

Hart, J.D., Piumsomboon, T., Lawrence, L., Lee, G.A., Smith, R.T., Billinghurst, M., 2018. Emotion sharing and augmentation in cooperative virtual reality games. Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts, pp. 453–460.

Held, D., Meudt, S., Schwenker, F., 2017. Bimodal recognition of cognitive load based on speech and physiological changes. In: Schwenker, F., Scherer, S. (Eds.), Multimodal Pattern Recognition of Social Signals in Human-Computer-Interaction. Springer International Publishing, Cham, pp. 12–23.

Jacob, R.J., 1990. What you look at is what you get: eye movement-based interaction techniques. Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 11–18.

Jasper, H.H., 1958. The ten-twenty electrode system of the international federation. Electroencephalogr. Clin. Neurophysiol. 10, 370–375.

Johnson, N., Carey, J., Edelman, B., Doud, A., Grande, A., Lakshminarayan, K., He, B., 2018. Combined rtms and virtual reality brain–computer interface training for motor recovery after stroke. J Neural Eng 15 (1), 016009.

Li, H., Trutoiu, L., Olszewski, K., Wei, L., Trutna, T., Hsieh, P.-L., Nicholls, A., Ma, C., 2015. Facial performance sensing head-mounted display. ACM Transactions on Graphics (ToG) 34 (4), 1–9.

Lotte, F., Faller, J., Guger, C., Renard, Y., Pfurtscheller, G., Lécuyer, A., Leeb, R., 2012. Combining Bci with Virtual Reality: Towards New Applications and Improved Bci. Towards Practical Brain-Computer Interfaces. Springer, pp. 197–220.

Lu, X., Yu, D., Liang, H.-N., Feng, X., Xu, W., 2019. Depthtext: Leveraging head movements towards the depth dimension for hands-free text entry in mobile virtual reality systems. 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR). IEEE, pp. 1060–1061.

Magnenat-Thalmann, N., Kim, H., Egges, A., Garchery, S., 2005. Believability and interaction in virtual worlds. 11th International Multimedia Modelling Conference. IEEE, pp. 2–9.

Manuri, F., Sanna, A., Lamberti, F., Paravati, G., 2016. Vocal one switch (vos) selection interfaces for virtual and augmented reality hands-free tasks. Proceedings of the Conference on Smart Tools and Applications in Computer Graphics. Eurographics Association, Goslar Germany, Germany, pp. 79–87. https://doi.org/10.2312/stag.20161367.

Martišius, I., Damaševičius, R., 2016. A prototype ssvep based real time bci gaming system. Comput Intell Neurosci 2016, 18.

Pallavicini, F., Pepe, A., Ferrari, A., Garcea, G., Zanacchi, A., Mantovani, F., 2020. What is the relationship among positive emotions, sense of presence, and ease of interaction in virtual reality systems? an on-site evaluation of a commercial virtual experience. PRESENCE: Virtual and Augmented Reality 27 (2), 183–201.

Piumsomboon, T., Lee, G., Lindeman, R.W., Billinghurst, M., 2017. Exploring natural eye-gaze-based interaction for immersive virtual reality. 2017 IEEE Symposium on 3D User Interfaces (3DUI), pp. 36–39. https://doi.org/10.1109/3DUI.2017.7893315.

Poh, M.-Z., Swenson, N.C., Picard, R.W., 2010. A wearable sensor for unobtrusive, long-term assessment of electrodermal activity. IEEE Trans. Biomed. Eng. 57 (5), 1243–1252.

Prachyabrued, M., Borst, C.W., 2012. Virtual grasp release method and evaluation. Int J Hum Comput Stud 70 (11), 828–848. https://doi.org/10.1016/j.ijhcs.2012.06.002.

Ray, W.J., Cole, H.W., 1985. Eeg alpha activity reflects attentional demands, and beta activity reflects emotional and cognitive processes. Science 228 (4700), 750–752.

Riva, G., Mantovani, F., Capideville, C.S., Preziosa, A., Morganti, F., Villani, D., Gaggioli, A., Botella, C., Alcañiz, M., 2007. Affective interactions using virtual reality: the link between presence and emotions. CyberPsychology & Behavior 10 (1), 45–56.

Roth, D., Lugrin, J.-L., Galakhov, D., Hofmann, A., Bente, G., Latoschik, M.E., Fuhrmann, A., 2016. Avatar realism and social interaction quality in virtual reality. 2016 IEEE Virtual Reality (VR). IEEE, pp. 277–278.

Ruddle, R.A., Volkova, E., Bülthoff, H.H., 2013. Learning to walk in virtual reality. ACM Transactions on Applied Perception (TAP) 10 (2), 1–17.

Ruggiero, G., Frassinetti, F., Coello, Y., Rapuano, M., Di Cola, A.S., Iachini, T., 2017. The effect of facial expressions on peripersonal and interpersonal spaces. Psychol Res 81 (6), 1232–1240.

Salisbury, D.B., Dahdah, M., Driver, S., Parsons, T.D., Richter, K.M., 2016. Virtual reality and brain computer interface in neurorehabilitation. Baylor University Medical Center Proceedings, Vol. 29. Taylor & Francis, pp. 124–127.

Sanchez-Vives, M.V., Slater, M., 2005. From presence to consciousness through virtual reality. Nat. Rev. Neurosci. 6 (4), 332.

Sareen, E., Singh, L., Gupta, A., Verma, R., Achary, G.K., Varkey, B., 2020. Functional brain connectivity analysis in intellectual developmental disorder during music perception. IEEE Trans. Neural Syst. Rehabil. Eng. 28 (11), 2420–2430.

Sareen, E., Singh, L., Varkey, B., Achary, K., Gupta, A., 2020. Eeg dataset of individuals with intellectual and developmental disorder and healthy controls under rest and music stimuli. Data Brief 105488.

Sarupuri, B., Hoermann, S., Steinicke, F., Lindeman, R.W., 2017. Triggerwalking: a biomechanically-inspired locomotion user interface for efficient realistic virtual walking. Proceedings of the 5th Symposium on Spatial User Interaction, pp. 138–147.

Setz, C., Arnrich, B., Schumm, J., La Marca, R., Tröster, G., Ehlert, U., 2009. Discriminating stress from cognitive load using a wearable eda device. IEEE Trans. Inf. Technol. Biomed. 14 (2), 410–417.

Si-Mohammed, H., Petit, J., Jeunet, C., Argelaguet, F., Spindler, F., Evain, A., Roussel, N., Casiez, G., Lécuyer, A., 2018. Towards bci-based interfaces for augmented reality: feasibility, design and evaluation. IEEE Trans Vis Comput Graph.

Sibert, L.E., Jacob, R.J., 2000. Evaluation of eye gaze interaction. Proceedings of the SIGCHI conference on Human Factors in Computing Systems, pp. 281–288.

Škola, F., Tinková, S., Liarokapis, F., 2019. Progressive training for motor imagery brain-computer interfaces using gamification and virtual reality embodiment. Front Hum Neurosci 13, 329.

Slater, M., Usoh, M., Steed, A., 1994. Depth of presence in virtual environments. Presence: Teleoperators & Virtual Environments 3 (2), 130–144.

Slater, M., Usoh, M., Steed, A., 1995. Taking steps: the influence of a walking technique on presence in virtual reality. ACM Transactions on Computer-Human Interaction (TOCHI) 2 (3), 201–219.

Sugenoya, J., Iwase, S., Mano, T., Ogawa, T., 1990. Identification of sudomotor activity in cutaneous sympathetic nerves using sweat expulsion as the effector response. Eur J Appl Physiol Occup Physiol 61 (3), 302–308.

Suzuki, K., Nakamura, F., Otsuka, J., Masai, K., Itoh, Y., Sugiura, Y., Sugimoto, M., 2017. Recognition and mapping of facial expressions to avatar by embedded photo reflective sensors in head mounted display. 2017 IEEE Virtual Reality (VR). IEEE, pp. 177–185.

Sweller, J., 2011. Cognitive Load Theory. Psychology of learning and motivation, Vol. 55. Elsevier, pp. 37–76.

Tanenbaum, T.J., Hartoonian, N., Bryan, J., 2020. "how do i make this thing smile?": An inventory of expressive nonverbal communication in commercial social virtual reality platforms. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. Association for Computing Machinery, New York, NY, USA, pp. 1–13. https://doi.org/10.1145/3313831.3376606.

Thies, J., Zollhöfer, M., Stamminger, M., Theobalt, C., Nießner, M., 2018. Facevr: real-time gaze-aware facial reenactment in virtual reality. ACM Transactions on Graphics (TOG) 37 (2), 1–15.
Tregillus, S., Folmer, E., 2016. Vr-step: Walking-in-place using inertial sensing for hands free navigation in mobile vr environments. Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. ACM, pp. 1250–1255.
Usoh, M., Arthur, K., Whitton, M.C., Bastos, R., Steed, A., Slater, M., Brooks Jr, F.P., 1999. Walking> walking-in-place> flying, in virtual environments. Proceedings of the 26th annual conference on Computer graphics and interactive techniques, pp. 359–364.
Virtuix, O., 2013. Move naturally in your favorite game, kickstarter.
Waller, B.M., Cray Jr, J.J., Burrows, A.M., 2008. Selection for universal facial emotion. Emotion 8 (3), 435.
Warren, L.E., Bowman, D.A., 2017. User experience with semi-natural locomotion techniques in virtual reality: the case of the virtuix omni. Proceedings of the 5th Symposium on Spatial User Interaction.163–163
Yan, Z., Lindeman, R.W., Dey, A., 2016. Let your fingers do the walking: A unified approach for efficient short-, medium-, and long-distance travel in vr. 2016 IEEE Symposium on 3D User Interfaces (3DUI), pp. 27–30. https://doi.org/10.1109/3DUI.2016.7460027.
Yong, X., Ward, R.K., Birch, G.E., 2008. Facial emg contamination of eeg signals: Characteristics and effects of spatial filtering. 2008 3rd International Symposium on Communications, Control and Signal Processing. IEEE, pp. 729–734.