PhysioHMD: a conformable, modular toolkit for collecting physiological data from head-mounted displays

Leave Authors Anonymous
for Submission
City, Country
e-mail address

Leave Authors Anonymous
for Submission
City, Country
e-mail address

Leave Authors Anonymous
for Submission
City, Country
e-mail address

ABSTRACT

Virtual and augmented reality headsets are unique as they have access to our facial area: an area that presents an excellent opportunity for always-available input and insight into the user's state. Their position on the face makes it possible to capture bio-signals as well as facial expressions. This paper introduces the PhysioHMD, a software and hardware modular interface built for collecting affect and physiological data from users wearing a head-mounted display. The platform enables researchers and developers to aggregate and interprets signals in real-time, and use those to develop novel, personalized interactions and evaluate virtual experiences. Our design offers seamless integration with standard HMDs, requiring minimal setup effort for developers and those with less experience using game engines. The PhysioHMD platform is a flexible architecture that offers an interface that is not only easy to extend but also is complemented by a suite of tools for testing and analysis. We hope that PhysioHMD can become a universal, publicly available testbed for VR and AR researchers.

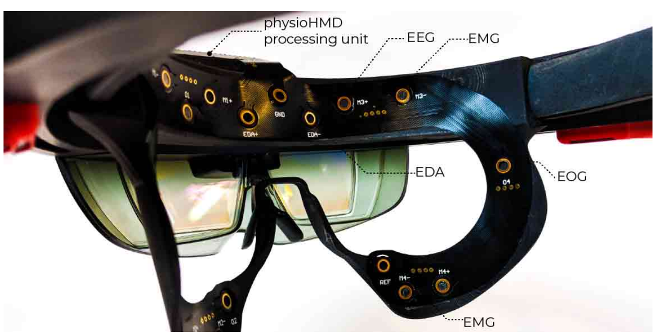

Figure 1: View of PhysioHMD hardware setup for AR experience. Gold plated electrodes and the flexible PCB record data through the contact with the skin.

CHI'16, May 07–12, 2016, San Jose, CA, USA

ISBN 123-4567-24-567/08/06. . . $15.00

DOI: http://dx.doi.org/10.475/123_4

ACM Classification Keywords

H.5.1. [Information interfaces and presentation]: Multimedia Information Systems - Artificial, augmented, and virtual realities. H.5.2. [User Interface]: Input devices and strategies

Author Keywords

Affect recognition; virtual reality; augmented reality; physiological signals; BCI; behavioural measures

INTRODUCTION

Augmented Reality (AR) and Virtual Reality (VR) technologies, hereafter referred jointly in the context related to our system as MR technologies, have enjoyed increased popularity in the last few of years thanks to the emergence of inexpensive and easy to deploy headsets. While MR technologies primarily support applications in the entertainment and gaming industry, they are also increasingly used in healthcare and human behaviors research to treat anxiety, phobias, psychosis, and post-traumatic stress disorder (PTSD). In both entertainment and healthcare applications, it is essential to understand the behavior, performance, and engagement of the user [22, 13, 25, 5].

The main contribution for this project is the development of an open-source platform for affect and attention sensing in virtual environments (VEs) to enable novel applications in the areas of mental and physical therapy, social VR, and VR games and entertainment.

In clinical settings, VR technology has recently gained much interest because it enables novel, promising methods for treating anxiety and other mental disorders [17]. VR-based therapy is also opening up new and exciting opportunities for pain management and personalized physical and sports therapy [18, 7].

Typically VR therapy requires that the user not just wear the VR technology, but also an array of other sensors and devices that enable real-time monitoring of the user's physiological and cognitive state (EEG, EDA, EOG, eye gaze, heart rate, facial expression and more). This sensing technology is not just cumbersome to set up and wear, but it is also very costly, not standardized and often errors prone. To overcome these problems, we are developing a new platform, called PhysioHMD, that consists of both hardware and software, and that will make collecting and using information about the internal state of the user cheap and easy. Offering a standard for collecting

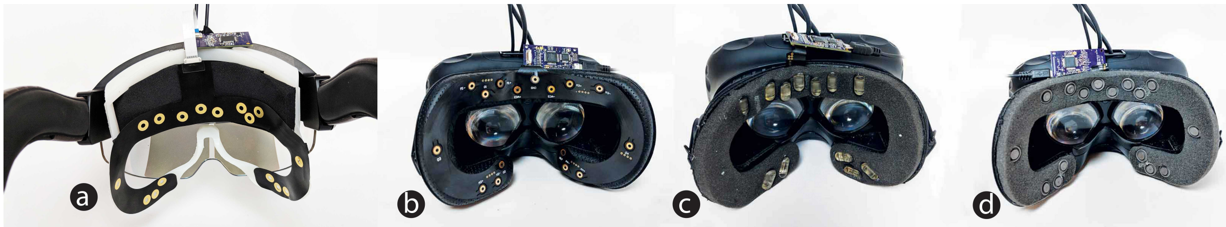

Figure 2: The image depicts every headset variation explored during this research. a) AR headset with flexible pcb & gold plated electrodes. b) VR headset with flexible PCB & gold plated electrodes. c) VR headset with hydrogel electrodes. d) VR headset with Ag/AgCI electrodes,

and comparing data across different users, sessions and study settings across multiple disciplines.

RELATED WORK

VR offers the potential to develop human testing and training environments that allow for the precise control of elaborate stimulus presentations in which human cognitive and functional performance can carefully be evaluated and rehabilitated. It is also important to mention that this research transfers into other areas where it is essential to understand the user's response to the content presented, thus allowing the system to integrate the user even further.

In the case of AR, measuring focus and attention in learning/training scenarios is an essential component of the evaluation of this novel application which is gaining traction in industry. However, fundamental issues need to be addressed for VR technology to be reasonably and efficiently applicable to cognitive rehabilitation. Specifically, it is essential to be able to measure the affect and physiological response of the user in real-time.

Emotions are multifaceted events with corresponding physiological signs as well as human expressions [27]. Even though the majority of existing methods for automatic emotion recognition are based on audio-visual analysis [9], there is an increasing body of research on emotion recognition from peripheral and central nervous system physiological responses [23, 20].

There are advantages to using physiological signals for emotion recognition as opposed to using audio-visual signals. They cannot easily be faked, they do not require a front-facing camera, and they can be used in any degree of illumination/noise. Moreover, they can be combined with audio-visual modalities to construct a more robust and accurate multi-modal emotion recognizer [9]. A system like BIOPAC [26] is now being employed in a variety of applications [30, 8, 32] for measuring facial movements using standardized tools such as Ekman and Friesen's Facial Action Coding System (FACS) and facial electromyography (sEMG). Systems like BIOPAC, SHIMMER, and GTECH are end to end solutions that are robust and produce reliable data through signals that have been thoroughly tested, but they also feature proprietary data that cannot be shared amongst users and many features must be bought in order to gain access.

The facilities that have access to physiological sensing instrumentation and software for data aggregation and syncing are extremely expensive, and few research labs have the resources to conduct such studies. Home grown solutions may not produce refined data, and systems like iMotions [3] are prohibitively expensive and require a high degree of knowledge and time to set up and use properly. With PhysioHMD, we present a high-fidelity platform that is state-of-the-art but can be made available at orders of magnitude lower cost than what is currently on the market so users are not limited by financial or geographical restrictions. There is no other product available that collects all the individual signals that PhysioHMD tests in such a cost-effective, clean, and compact way.

PhysioHMD uses a Convolution Neural Networks (CNN) [16] a model shown to have had great success in solving classification problems like those mentioned above. Some researchers use CNN to recognize gestures from sEMG signals [28]. The state-of-the-art deep auto-encoder architecture is also designed to extract discriminant features in the multi-modal data [24].

Filmmakers, entertainers and other storytellers are trying to figure out what MR as a medium might mean for their respective fields. Some interesting experiments that make use of physiological or affect data include PsychicVR [4], a VR system that integrates a brain-computer interface device with a VR headset to improve mindfulness while enjoying a playful, immersive experience. The interactive storytelling platform PINTER [11] uses physiological data to drive the unfolding of a plot. PINTER features an underlying narrative that consists of a medical drama which combines aspects of medical practice with the evolution of personal relationships between lead characters.

PhysioHMD could also help create more realistic experiences for art installations like The Enemy, where the audience gets to be face-to-face with combatants from three conflict zones: the Maras in El Salvador, the Democratic Republic of the Congo, and Israel and Palestine. Their testimonies and confessions about their lives, experiences, and perspectives on war will allow the spectator to better understand their motivations and their humanity [6]. This installation monitors the body movement of the person experiencing the installation to gauge how the spectator's gaze and body motion are reacting to the enemy base. These inputs then lead the spectator to the end of the experience to meet with the "enemy" that the system infers the spectator is feeling more empathetic towards. Art installations like The Enemy could benefit from a system like PhysioHMD to drive the immersive experience with the audience, and to explore new ways of telling a story.

PHYSIOHMD

PhysioHMD is a sensor and computing platform developed to support the analysis of multi-modal data related to the behavior and responses of a user, with the goal of enabling evaluation and customization of virtual experiences. There are two main components to the PhysioHMD hardware. First there is the main PCB, an analog front end that collects bio-potential signals from muscle movements, eye movements, electrodermal activity and brain signals. Second, there is an ergo-electronics face-pad, a flexible PCB with gold-plated pickup electrodes that can connect to electrodes like Ag/AgCl (silver/silver chloride), hydro-gels or can be used by themselves as depicted in Figure 2.

The software side is similarly composed of two main components. First, a game engine package that can be dropped in any virtual scene, and second, a signal processing component with normative data for signal pre-processing, feature extraction and a multivariate visualization method for data interpretation by end-users.

Although hardware costs have come down for AR and VR devices, researchers still do not have access to simple interfaces for deploying Virtual Environments (VEs), interfaces that require little knowledge of game engine content creation, sensor data, data logging and data visualization. Given these constraints, we compiled a list of requirements for the PhysioHMD platform:

1. A plug and play pipeline that can be deployed with minimal development effort.
2. Physical form factor must be comfortable to the user and easy to use.
3. The system should be as modular as possible for data collection.
4. System supports standard implementations of existing algorithms, feature extraction, and classification.
5. Offers a publicly available open-source code base for use and further improvement by the community interested in this body of work.
6. Has a flexible architecture that allows for convenient extensions.
7. Includes a game engine interface with sample scenes and relevant tools.

Hardware

The PhysioHMD system collects sEMG, EEG, EDA, ECG and eye-tracking data. EDA data reflects emotional arousal and is used to identify the magnitude of the emotional response. It is used to complement other bio-signals that gather valence data. sEMG data gathers the user's facial expression which helps estimate the valence. EEG data provides relevant information about attention as well as valence. ECG with Heart Rate Variability also provides significant information about both arousal and valence. We also use the aGlass [2] eye tracker to gauge the user's attention to a stimulus as well as the point of regard (POR).

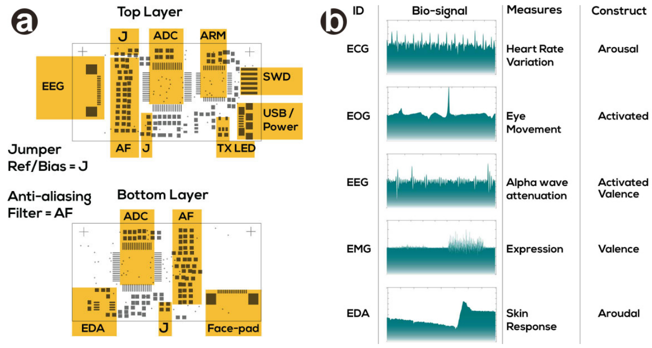

Figure 3: a) PCB configuration is depicting locations of main components on top and bottom planes. b) A Sample of signals gathered from PhysioHMD and their relevance in gathering affective state data.

Electronics

We chose to build our own hardware to capture the bio-signals coherently and to avoid the inconvenience caused during the setup of multiple signal acquisition devices. The configuration of the PCB is shown in Figure 3a. We used a pair of TI ADS1299 as the front-end for the ADC operations: one ADS is used for sEMG and the possibility of EOG data acquisition while the other acquires the frontal EEG data. Both the ADS are controlled by ARM Cortex M0+ using SPI communication standard. We used this configuration instead of a daisy-chained configuration to allow for flexibility in the configuration of individual channels of each ADC. In daisy chain mode (OpenBCI), the second ADC mimics the register configuration of the first ADC and doesn't allow for individual channel configuration. The PCB has multiple jumpers for the configuration of both ADC's Reference and Bias settings as desirable. The flexibility of configuration for individual channels also allows us to integrate ECG measurements from the same hardware. The ARM communicates with the PC using the inbuilt USB interface. We use anti-aliasing filters before both ADC's and sample at 500Hz. We also integrated EDA measurement by using a voltage divider and a bandpass filter of 1.5 Hz to 15 Hz to remove artifacts. Then, we buffer the signal with an amplifier of gain 2 V/V and use ARM's ADC to sample the data.

Electrodes

We built and tested the sensing face pads that integrate the bio-signal sensors for detecting affect of users into two HMD platforms: one into the face cushion of an HTC VIVE VR headset and the other on the Meta 2 AR headset. In order to fit into a HMD facepad, we determined it best to develop a flexible PCB that connects directly into the PhysioHMD's PCB. The flexible PCB has gold plated pads which can either be used as standalone electrodes or are compatible with external electrodes. We chose to place the EDA electrodes on the forehead region because it is one of the regions most dense with sweat glands. We placed sEMG electrodes above the eyebrows

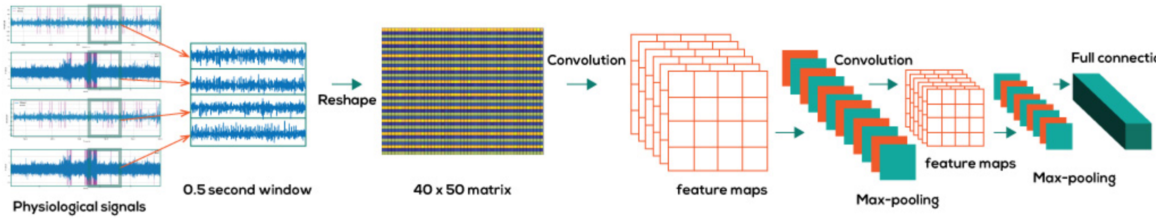

Figure 4: The process of our classification systerm and the CNN architecture used.

on the *frontalis* muscle and on cheeks on the *zygomaticus* muscle. We also set EOG Vertical (EOGV) and EOG Horizontal (EOGH) electrodes in a standard placement. Further, we set EEG electrodes according to the 10-20 international electrode system on the user's frontal lobe.

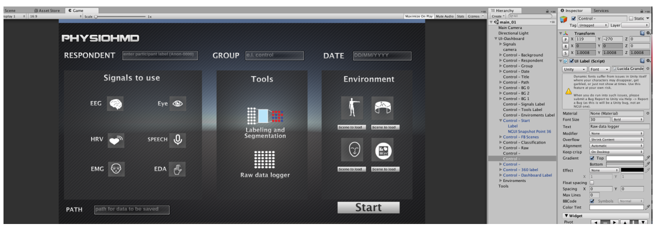

Figure 5: Unity package main dashboard, where users can select signals to measure, methods in which to segment data, and environments to test within.

Software

Game Engine Integration

The data from sensors is streamed to a Python script which processes the data and estimates the emotion label. The data and the emotion label is communicated to Unity using a UDP port. To facilitate the integration, we have encapsulated it into a Unity3D package. By encapsulating the platform, less experienced users can drop the package into an empty or already built environment and access our tools. The sample scenes included within the package are set with default configurations that can easily be customized with the exposed parameters in the editor. The main scene is a dashboard (Figure 5) for the person running the study, and here the user can select the signals of interest, choose our data segmentation tool, or merely record raw data. Once those parameters are selected, the user can choose to use one of the demo scenes. Lastly, the user can also choose to take information from external API or SDKs. We show those utilities by integrating aGlass SDK [2] to provide a point of regard (POR) data and Beyond Verbal affective speech recognition. These demo scenes were created with the intention to meet most user cases in MR behavioral research.

1. A full body IK scene, where the user can have a room scale experience while embodying a full-size avatar.
2. A mimicry scene where a 3D avatar replicates the facial expressions made by the user wearing the HMD.
3. A 360° scene where 360° video is played.
4. A scene with a particle system that can instantiate animals or objects that the user might have a phobia.

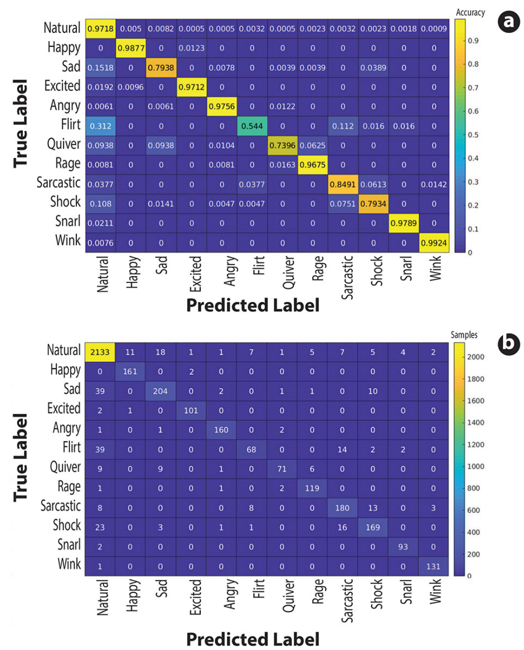

Figure 6: Confusion matrix. a) Confusion matrix of prediction accuracy. b) Confusion matrix of prediction amount.

Machine Learning for Pattern Recognition

We use a typical LeNet-5 [12] five-layers CNN architecture to classify the collected data as shown in Figure 4. There are two convolution layers and two pooling layers, with one full-connection layer. Convolution layers are used to extract the main feature while pooling layers will subsample the feature maps. We reshape the multichannel data into a matrix and use a 3*3 size convolution kernel in the first layer and third layer to improve feature sensing capabilities of convolution layers. In the second and fourth layers, the subsample window size is 2*2, and max-pooling is used. In the full-connection layer, rectified linear units [21] are used to improve the nonlinear performance of the network.

To train the network, we collected data from 6 users (three females and three males ages 19-30) for 12 different expressions. As part of the data collection procedure, we asked the users to put on our physioHMD prototype and repeat the expression for 12 times for 5 seconds with intervals of 3 seconds. The

Table 1: A comparison of facial expression recognition accuracy between Katsuhiro's method and our method

Method	Natural	Happy	Sad	Excited	Angry	Flirt	Quiver	Rage	Sarcastic	Shock	Snarl	Wink
Kats.	95.7%	98.9%	76.2%	—	80.0%	—	—	—	—	92.1%	—	—
Ours	97.2%	98.8%	79.4%	97.1%	97.6%	54.4%	74.0%	96.8%	84.9%	79.3%	97.9%	99.2%

captured data is then pre-processed before feeding into the network. Notch filters are used to remove the power line interference at 60Hz and 120Hz, and a high-pass filter is applied to cut off the low frequencies below 30Hz. We then label the time sequence sEMG data for each expression.

A 0.5 second time window is used to segment the multichannel time series data into subsequences. Every subsequence is a training or testing sample. The label for a sample is decided by choosing the principal type, which takes the maximum percentage in marker vector. Then follow a one-hot encoding [19] style to re-encode the labels.

Following Karpathy's paper [14], we take advantage of data augmentation to reduce the effects of over-fitting. The augmentation consists of generating sEMG signal translations and horizontal reflections increasing the size of our dataset. After the data segmentation and augmentation, we get 12,838 samples. Whenever we collect different expression samples we get neutral expression samples in addition to the specific expression being measured, so we have most neutral samples in our dataset.

To avoid over-fitting, we use 0.5 dropout rate [29] in the training process. From this, we get a learning rate of 0.001 this parameter tells the optimizer how far to move the weights in the direction of the gradient for a mini-batch. In 30 minutes and 800 iterations, our optimized network sees a classification accuracy in the training set of 99.8% and 92.3% in the testing dataset.

PLATFORM EVALUATION

We conducted a set of tests to evaluate the accuracy and of our system and the signal quality vs. ergonomic comfort levels once worn by the user. We were mainly interested in testing the robustness and usability of the prototype for long periods of time in multiple scenarios. 8 participants (four females and four males), 18-32 years old for 12 different expressions As part of the data collection procedure, we asked the users to put on our physioHMD prototype and repeat the expression for 12 times for 5 seconds with intervals of 3 seconds. We also tested each electrode on the following parameters: level of comfort, signal quality, and shelf life. After wearing the headset for 15 minutes for each face-pad, we asked the user to self-report on the level of comfort.

Qualitative analysis

The training and identification processes were done for each individual. The recognition accuracy values of the facial expressions are shown in Figure 6. This figure shows the testing confusion matrix, which shows the different performance of each expression. Due to obvious signal patterns and high-intensity signal amplitude, Happy, Excited, Angry, Rage, Snarl, and Wink have the highest recognition accuracy. For Sad, Flirt, Quiver, Sarcastic and Shock, because of the similar signal pattern with other expressions and weak signal amplitude, we obtained lower accuracy relatively. Compared with the state-of-the-art results, our platform can deal with more complex expressions and shows better performance in recognizing the basic ones which Katsuhiro et al. also showed in their paper [31]. Table 2 compares the facial recognition model's accuracy of Katsuhiro's et al. research in comparison to our facial recognition model. We also use our trained model in real time expressions recognition. The 3D animation, shown in Figure 8 are driven by real time expressions information coming from user's physiological data recognition by using our model.

Faceplate	*Comfort*	*Signal Gain(dB)*	*Shelf Life(months)*
Facepad	4.5	NA	NA
Gold Plated	3.1	0	>12
Ag/AgCl	4.1	-1	<6
Hydrogel	4.4	+7	<1

Table 2: Comparison of Comfort, Signal Quality and Shelf life of different electrodes

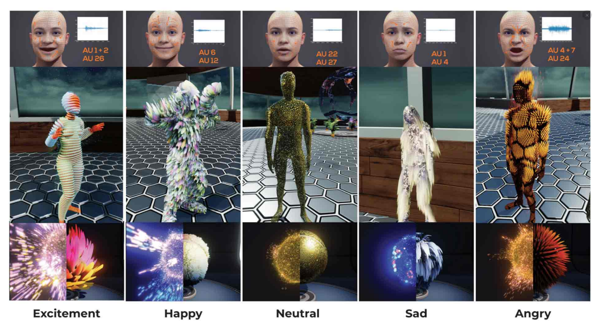

Figure 7: The depiction of the range of possible transformations and qualities for an avatar's emotional expressions with a particle system. The figure shows the particle system's variations in particle size, density, brightness, and color which can all adjust to express the emotions of the user visually.

Ergo-Electronics Evaluation

To meet the design goals that we had imposed on ourselves, we tested our prototypes and compared the three different face-pad electrodes: the gold-plated pads, standard Ag/AgCl electrodes, and hydrogel-based electrodes. We tested each electrode on the following parameters: level of comfort, signal quality, and shelf life. For the experiment, we asked the participants to evaluate comfort levels, we asked the participants how comfortable each type of electrode on the face pad was while creating different facial expressions in the HMD during a trial. We used a scale of 1-5, where 1 is uncomfortable and

Figure 8: Facial expressions: Confused, Surprised , Fear, Happy, Sad, Angry

5 is maximum comfort. We also used a standard face pad as a the neutral reference. Hydrogels match closely to the standard facepad in terms of comfort whereas gold plated electrodes were the least comfortable for the participants. Table 2 gives a summary of the average comfort felt by the participants, the average signal gain and the shelf life of each electrode.

Further, we compared the signal-to-noise ratio of the signals acquired by each different electrode and evaluated the signal gain with reference to the gold-plated electrode. We found the Ag/AgCl electrodes had -1dB signal gain and hydrogels had +7dB signal gain on average with the same expression. The shelf life of gold plated electrodes is estimated to be >1year whereas the shelf life of Ag/AgCl and hydrogels is <6months and <1month respectively. The hydrogels also require frequent treatment with saline solution for keeping the signal quality high. Also, based on observations during signal analysis from all three different electrodes, we found that hydrogels and Ag/AgCl had better mechanical contact compared to gold-plated electrodes because they protrude from the facepad. Hence, we concluded that hydrogels will be suitable for physiological data acquisition where high signal quality and comfort is desirable. Ag/AgCl electrodes will be desirable where both contact requirement and cost are constraints. Gold plated electrodes will be desirable where longevity and minimum cost are required.

APPLICATIONS

In this section, we present three applications to demonstrate the capabilities of the PhysioHMD system. A first application uses the system to create more expressive avatars in a social VR setting. The user's real-time expression and emotion are mapped into the user's VR avatar. A second application uses the data to customize the virtual experience to the user's affective state. It illustrates the use of the system for Cognitive Behavior Therapy and for the detection and treatment of disorders such as anxieties and fears. For the third application, we show a scene that can be used in VR exposure therapy settings that offers gradual hierarchies of fearful stimuli; in the study presented here, the scene spawns insects in varying quantities according to the arousal response from the user.

Expression and Emotion Mapping

Affective avatars

The second demo scene demonstrates how the system can allow users to express emotions in abstract ways using full-size avatars in a virtual environment (Figure 7). An affect event triggers a change in the avatar's texture and color, enabling enhanced interpersonal communications at a distance within social VR settings.

The setup begins by the user putting on the physioHMD on. On our only intervention with the user during the experiment is when we asked to make sure the headset has to be tightly secured to their head. We do this to make sure we have proper contact between the user and the electrode.

Once the scene begins to play unity and python begin to send data bi-linearly. The facial expression signals are then passed to tensorflow [1] and the labels send from tensorflow back to Unity are then used to trigger a blend shaper deformation on the avatars. This is done so the face so that the cheek and forehead muscles of the avatar can be modulated by the user to show his or her expression in real time.

The affect of the user is represented visually in two different ways: (1) the fur of the avatar can grow when arousal is high and (2) the color of the avatar can intensify in brightness or change color to highlight when the user is in a high arousal situation. The creation of this application allowed us to experiment with the manipulation of a user's self-expression in VR space as well as the perception of others in it, providing valuable tools to evoke a desired affect reaction.

Within this application area, we are looking to provide agency and express affect in VR through avatars to produce the compelling human-to-human connection. We made two sample scenes that highlight multiple possibilities when exploring the implication and responses for the user in VR. The first demo scene, as shown in Figure 8 maps the facial expression of the user wearing PhysioHMD into a 3d rigged model avatar. We investigated mimicry due to is relevance in areas, such as autism spectrum disorder research [10]. Participants played an imitation game with both a socially engaged avatar and socially disengaged avatar. This application presents a direct mapping of the user's facial expressions and affect state onto the VR avatar.

Adaptable Exposure

As part of our unity package, we present a case scenario scene of our system in a phobia treatment scenario where we present a subject with the stimulus of a feared object using a particle system. The images (sprites) spawned through the particle system can be modified (speed, size, the rate of spawn, movement) in the unity inspector to increase or decrease the arousal level of the user. Figure 11 shows a participant with entomophobia and her response to the stimulus of the spawning of more insects. The images (sprites) spawned by the particle system can be modified in the unity inspector, allowing the organizer of the study to control the virtual animal by choosing different functions (e.g., increase/decrease the number of animals; increase/decrease the size of animals; make the animal move continuously or randomly; make the animal stay still).

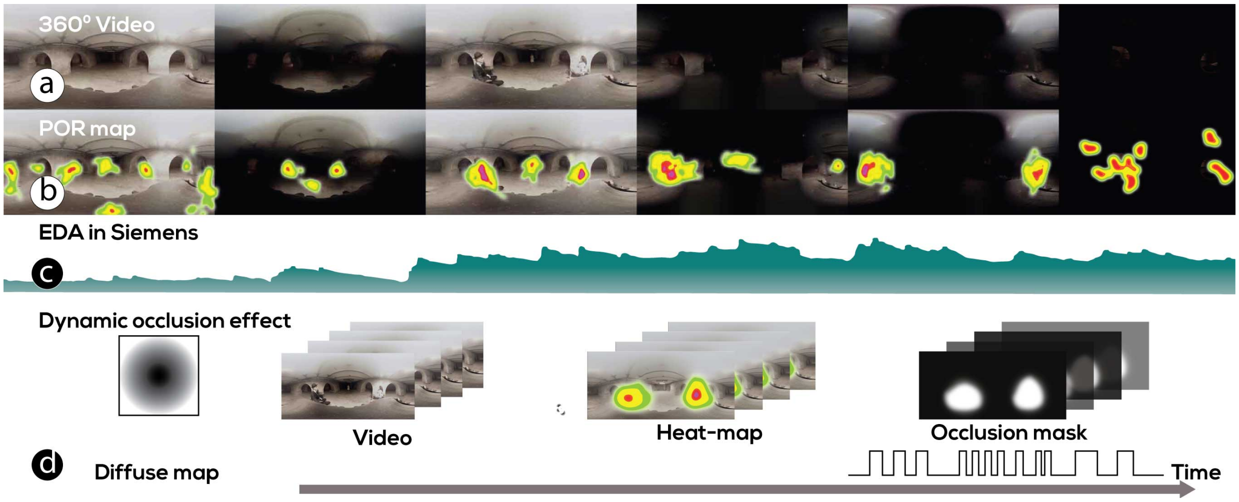

Figure 9: a) Frames from the 360 experience. b) Heat-map from point-of-regard (POR) from user's gaze. c) Electrodermal activity in Siemens. d) Diagram showing how the occlusion shader works.

This allows both the therapist and the respondent to observe and interact with the AR content. Thus, although the therapist fully controls the environment, the patient can interact freely with the virtual critter that react to his/her behavior.

he behavior of the stimuli in our application can be controlled either by parameters in the inspector or by the physiological response from the user. Taking data input from the EDA signal, control changes were sent to the particle system and synchronously relayed to trigger occurrences in the behavior of the critters in the scene meant to represent and/or provoke arousal in the participants. Three levels of participant arousal were determined to range from low to high. Such levels were established based on simple rules regarding how the data from the sensors changed in the short, medium and long-term. Since EDA readings can vary significantly from one participant to another, where possible, the control system was designed to change the criteria on which these rules were based in order to more accurately reflect the arousal levels of the user group throughout the course of the installation.

Skin conductance response (SCR) and is considered to be useful as it signifies a response to internal/external stimuli. We follow Kim's et al. [15] method for SCR extraction from EDA signals by reducing the sampling rate to 20 samples per second, differentiation and subsequent convolution with a 20-point Bartlett window. This method correctly detects the occurrence of SCR, as shown in Figure 10a. This procedure yielded the output waveform shown in Figure 10b for the input signal shown in Figure 10. The occurrence of the SCR was detected by finding two consecutive zero-crossings, from negative to positive and positive to negative. The amplitude of the SCR was obtained by finding the maximum value between these two zero-crossings. One significant benefit to this method is that it does not require explicit determination of the threshold level. With conventional SCR detection by visual inspection by a human supervisor, where the threshold level is determined arbitrarily, and thus objective analysis can hardly be achieved. From this, we get the SCR and the peak which does not show exponential decay, depending on the context (e.g., if two SCRs occur close together in time, the first response may not decay before the second begins, yet this is not considered an artifact)

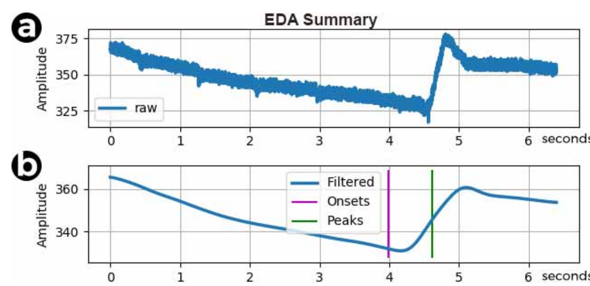

Figure 10: a) Typical waveform of EDA under emotional stimulation. b)Output signal from detection module from signal in a)

Dynamic Occlusion and POR Tool

Monitoring the user's reactions towards the VE content has been a hot topic, as it enables the generation of personalized VR experiences. Our system's use of arousal levels to provides real-time, reliable information about the user's reception of the content and can help the system adapt the content seamlessly. In our 360 video demo player scene, we use the gaze data and the skin response data to increase the levels of arousal in the users. Our demo flashes a 360 captured video of people with masks in a basement, similar to those seen in horror movies.

$$D = \frac{kA - Ymin}{Ymax - Ymin} \tag{1}$$

1: D is Duty Ratio of the PWM signal. A is the amplitude of EDA signal. k is the constant coefficient, Ymin and Ymax are the minimal and maximum value about the PWM signal.

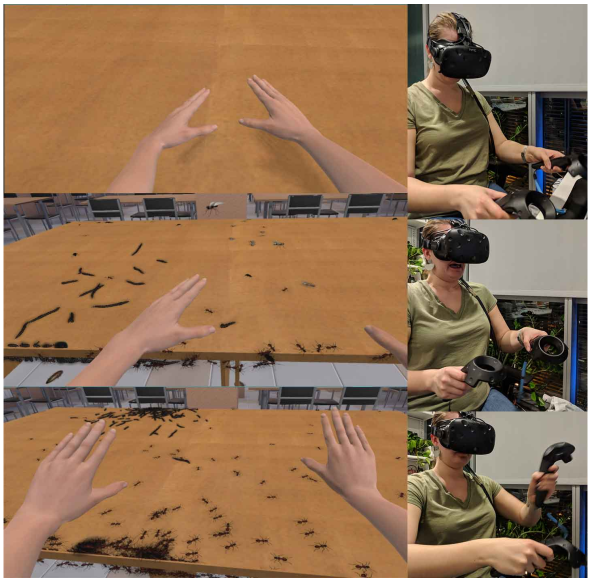

Figure 11: Visualization of how users can be exposed to phobias in a therapeutic VR setting. Here, a user with entomophobia is exposed to a virtual setting containing insects. The physiological response recorded by the PhysioHMD modulates the quantity of insects the user is exposed to within the environment.

In order to direct the focus of the user to the people within the video, a surfaces shader dynamically occludes locations informed by the point of regard (POR) data from the gaze tracking system as areas that are not of interest to the user. POR is the gaze point mapped onto a plane at a specific location within the eye tracker coordinate system.

The screen of HTC Vive HMD is divided into two halves and each half screen is used for one eye. So, for monocular eye aGlass module, the coordinate is mapping to the half of the screen which size is half of the HTC HMD screen (1080 pixel * 1200 pixel). The coordinate system is normalized, the coordinate of top left is (0, 0) and the coordinate of right bottom corner is (1, 1). For example, the pixel coordinate of HTC Vive screen where aGlass coordinate (0.5, 0.5) map to is (540, 600).

LIMITATIONS AND FUTURE WORK

The PhysioHMD platform enables researchers and developers interpret signals in real-time, and use those signals to develop novel, personalized interactions to help evaluate VR experiences. Our design offers seamless integration with standard HMDs, requiring minimal setup effort for developers and those with less experience using game engines compared to baseline systems. The applications presented in this research aimed to expose advantages in using physiological computing tools that involved multi-modal sensing for VR psychotherapy, behavioral studies or expression explorations, but there are many features we plan to add in upcoming versions.

- More studies need be done to thoroughly understand the role of our tool with the rest of our target community. Like-wise, usability issues regarding the navigation and data visualization and interpretation by health professionals should be carefully studied to improve the likelihood of extended use of the tool. We plan to host workshops with psychology, cognitive science department as well with HCI groups to build a community.
- Current unity3d sample content still requires the user to have a basic understanding of the game engine, which could lead to frustration to those unfamiliar with Unity3d. We plan to improve our UI and the system as whole further so the platform can be run as an executable system without losing access to parameters in the editor.
- We are working to incorporate signals like skin temperature and respiration used by some the literature referenced in the related work section.
- Our current CNN model takes upwards of 30 minutes to train to recognize gestures and desired patterns, but we hope to integrate our system with deep learning cloud clusters to provide an additional solution for retraining that would be less time-intensive. We also plan to provide a trained model to get developers testing their systems faster.

For future work, more channels will be used in experiments. The performance of our CNN on the raw data from more than four channels in the dataset should be investigated. Since the dimensionality of the data is high, an effective channel selection algorithm is necessary. Secondly, the relationship between our CNN structure and its performance is of interest for finding appropriate tradeoffs between solution quality and training time. Thirdly, in the fine-tuning (supervised learning) stage, there exists an active learning problem for selecting appropriate examples for labeling in some applications. For example, human experts are needed to evaluate the true conditions of subjects as the labels of training data, which is time-consuming and expensive. When only a limited number of examples can be labeled, an intruiging question presents itself in which examples should be labeled. For these reasons we would like to expand the PhysioHMD platform to a larger community in order to help grow the quality and quantity of the data.

CONCLUSION

In this paper, we introduced PhysioHMD: a sensor and computing platform developed to support the analysis of multi-modal data related to the behavior and responses of a user utilizing MR technology to enable evaluation and customization of virtual experiences. The toolkit is intended to assist both researchers and non-experts in the arduous task of collecting and processing physiological signals and creating experiences in a game engine. The software provides signal processing methods and data logging for EMG, ECG, EEG and EDA signals in order to provide researchers with accurate, real-time information regarding a user's response to content in a virtual environment. Our intention is to grow a community that contributes to HCI and XR technology research through the pluggable open source platform.

ACKNOWLEDGMENTS

We thank all the volunteers, and all publications support and staff, who wrote and provided helpful comments on previous versions of this document. We thank http://www.brainco.tech for offering us the hydrogel electrodes. And http://www.root-motion.com/ for giving us a free access to their IK system.

REFERENCES

1. MartÃ■n Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, Manjunath Kudlur, Josh Levenberg, Rajat Monga, Sherry Moore, Derek G. Murray, Benoit Steiner, Paul Tucker, Vijay Vasudevan, Pete Warden, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. 2016. TensorFlow: A System for Large-scale Machine Learning. In *Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation (OSDI'16)*. USENIX Association, Berkeley, CA, USA, 265–283. http://dl.acm.org/citation.cfm?id=3026877.3026899
2. System aGlass. 2017. Eye tracking. (2017). http://www.aglass.com
3. alex. iMotions: Biometric Research, Simplified. (????). https://imotions.com/
4. Judith Amores, Xavier Benavides, and Pattie Maes. 2016. PsychicVR: Increasing Mindfulness by Using Virtual Reality and Brain Computer Interfaces. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '16)*. ACM, New York, NY, USA, 2–2. DOI: http://dx.doi.org/10.1145/2851581.2889442
5. Jakki O. Bailey, Jeremy N. Bailenson, and Daniel Casasanto. 2016. When does virtual embodiment change our minds? *PRESENCE: Teleoperators and Virtual Environments* 25, 3 (2016), 222–233.
6. Karim Ben Khelifa. 2017. Meet The Enemy. (2017). http://theenemyishere.org
7. C. Bryanton, J. Bosse, Marie Brien, Jennifer Mclean, Anna McCormick, and Heidi Sveistrup. 2006. Feasibility, motivation, and selective motor control: virtual reality compared to conventional home exercise in children with cerebral palsy. *Cyberpsychology & behavior* 9, 2 (2006), 123–128.
8. Huan Deng and Ping Hu. 2018. Matching Your Face or Appraising the Situation: Two Paths to Emotional Contagion. *Frontiers in Psychology* 8 (2018). DOI: http://dx.doi.org/10.3389/fpsyg.2017.02278
9. Sidney K. D'mello and Jacqueline Kory. 2015. A Review and Meta-Analysis of Multimodal Affect Detection Systems. *ACM Comput. Surv.* 47, 3 (Feb. 2015), 43:1–43:36. DOI:http://dx.doi.org/10.1145/2682899
10. Paul A. G. Forbes, Xueni Pan, and Antonia F. de C. Hamilton. 2016. Reduced Mimicry to Virtual Reality Avatars in Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders* 46, 12 (Dec. 2016), 3788–3797. DOI: http://dx.doi.org/10.1007/s10803-016-2930-2
11. Stephen Gilroy, Julie Porteous, Fred Charles, and Marc Cavazza. 2012. PINTER: interactive storytelling with physiological input. ACM Press, 333. DOI: http://dx.doi.org/10.1145/2166966.2167039
12. Song Han, Jeff Pool, John Tran, and William Dally. 2015. Learning both weights and connections for efficient neural network. In *Advances in neural information processing systems*. 1135–1143.
13. Rorik Henrikson, Bruno Araujo, Fanny Chevalier, Karan Singh, and Ravin Balakrishnan. 2016. Multi-Device Storyboards for Cinematic Narratives in VR. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology (UIST '16)*. ACM, New York, NY, USA, 787–796. DOI: http://dx.doi.org/10.1145/2984511.2984539
14. Andrej Karpathy, George Toderici, Sanketh Shetty, Thomas Leung, Rahul Sukthankar, and Li Fei-Fei. 2014. Large-scale Video Classification with Convolutional Neural Networks. (2014).
15. K. H. Kim, S. W. Bang, and S. R. Kim. 2004. Emotion recognition system using short-term monitoring of physiological signals. *Medical and Biological Engineering and Computing* 42, 3 (May 2004), 419–427. DOI:http://dx.doi.org/10.1007/BF02344719
16. Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. 2012. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*. 1097–1105.
17. Kate Laver, Stacey George, Susie Thomas, Judith E. Deutsch, and Maria Crotty. 2012. Virtual reality for stroke rehabilitation. *Stroke* 43, 2 (2012), e20–e21.
18. Mindy F. Levin, Patrice L. Weiss, and Emily A. Keshner. 2015. Emergence of virtual reality as a tool for upper limb rehabilitation: incorporation of motor control and motor learning principles. *Physical therapy* 95, 3 (2015), 415–425.
19. Alireza Makhzani, Jonathon Shlens, Navdeep Jaitly, Ian Goodfellow, and Brendan Frey. 2015. Adversarial autoencoders. *arXiv preprint arXiv:1511.05644* (2015).
20. Christian MÃijhl, Brendan Allison, Anton Nijholt, and Guillaume Chanel. 2014. A survey of affective brain computer interfaces: principles, state-of-the-art, and challenges. *Brain-Computer Interfaces* 1, 2 (April 2014), 66–84. DOI: http://dx.doi.org/10.1080/2326263X.2014.912881
21. Vinod Nair and Geoffrey E Hinton. 2010. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)*. 807–814.

22. Cuong Nguyen, Stephen DiVerdi, Aaron Hertzmann, and Feng Liu. 2017. CollaVR: Collaborative In-Headset Review for VR Video. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology (UIST '17)*. ACM, New York, NY, USA, 267–277. DOI: http://dx.doi.org/10.1145/3126594.3126659

23. Rosalind W. Picard, Elias Vyzas, and Jennifer Healey. 2001. Toward machine emotional intelligence: Analysis of affective physiological state. *IEEE transactions on pattern analysis and machine intelligence* 23, 10 (2001), 1175–1191.

24. Ahmed Ben Said, Amr Mohamed, Tarek Elfouly, Khaled Harras, and Z. Jane Wang. 2017. Multimodal deep learning approach for joint EEG-EMG data compression and classification. In *Wireless Communications and Networking Conference (WCNC), 2017 IEEE*. IEEE, 1–6.

25. Pilar Salgado-Pineda, RamÃşn Landin-Romero, Francisco Portillo, Clara Bosque, Ausias Pomes, Bernhard Spanlang, Jose Carlos Franquelo, Cristina Teixido, Salvador SarrÃş, Raymond Salvador, Mel Slater, Edith Pomarol-Clotet, and Peter J. McKenna. 2016. Examining hippocampal function in schizophrenia using a virtual reality spatial navigation task. *Schizophrenia Research* 172, 1-3 (April 2016), 86–93. DOI: http://dx.doi.org/10.1016/j.schres.2016.02.033

26. Wataru Sato, Tomomi Fujimura, and Naoto Suzuki. 2008. Enhanced facial EMG activity in response to dynamic facial expressions. *International Journal of Psychophysiology* 70, 1 (2008), 70–74.

27. Klaus R. Scherer. 2005. What are emotions? And how can they be measured? *Social Science Information* 44, 4 (Dec. 2005), 695–729. DOI: http://dx.doi.org/10.1177/0539018405058216

28. Natalie Segura Velandia, Robinson Moreno, and Ruben Hernandez BeleÃśo. 2017. CNN architecture for robotic arm control in a 3D virtual environment by means of by means of EMG signals. *Contemporary Engineering Sciences* 10 (Jan. 2017), 1377–1390. DOI: http://dx.doi.org/10.12988/ces.2017.711162

29. Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. 2014. Dropout: A simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research* 15, 1 (2014), 1929–1958.

30. Freya Sukalla, Helena Bilandzic, Paul D. Bolls, and Rick W. Busselle. 2015. Embodiment of Narrative Engagement. *Journal of Media Psychology* 28, 4 (Jan. 2015), 175–186. DOI: http://dx.doi.org/10.1027/1864-1105/a000153

31. K. Suzuki, F. Nakamura, J. Otsuka, K. Masai, Y. Itoh, Y. Sugiura, and M. Sugimoto. 2017. Recognition and mapping of facial expressions to avatar by embedded photo reflective sensors in head mounted display. In *2017 IEEE Virtual Reality (VR)*. 177–185. DOI: http://dx.doi.org/10.1109/VR.2017.7892245

32. Elizabeth A Velkoff, Lauren N Forrest, Dorian R Dodd, and April R Smith. 2016. I Can Stomach That! Fearlessness About Death Predicts Attenuated Facial Electromyography Activity in Response to Death-Related Images. *Suicide & Life-Threatening Behavior* 46, 3 (June 2016), 313–322. DOI: http://dx.doi.org/10.1111/sltb.12194