VREED: Virtual Reality Emotion Recognition Dataset Using Eye Tracking & Physiological Measures

LUMA TABBAA, University of Kent, United Kingdom
RYAN SEARLE, University of Kent, United Kingdom
SABER MIRZAEE BAFTI, University of Kent, United Kingdom
MD MOINUL HOSSAIN, University of Kent, United Kingdom
JITTRAPOL INTARASISRISAWAT, Prince of Songkla University, Thailand
MAXINE GLANCY, BBC Research & Development, United Kingdom
CHEE SIANG ANG, University of Kent, United Kingdom

The paper introduces a multimodal affective dataset named VREED (VR Eyes: Emotions Dataset) in which emotions were triggered using immersive 360° Video-Based Virtual Environments (360-VEs) delivered via Virtual Reality (VR) headset. Behavioural (eye tracking) and physiological signals (Electrocardiogram (ECG) and Galvanic Skin Response (GSR)) were captured, together with self-reported responses, from healthy participants (n=34) experiencing 360-VEs (n=12, 1-3 min each) selected through focus groups and a pilot trial. Statistical analysis confirmed the validity of the selected 360-VEs in eliciting the desired emotions. Preliminary machine learning analysis was carried out, demonstrating state-of-the-art performance reported in affective computing literature using non-immersive modalities. VREED is among the first multimodal VR datasets in emotion recognition using behavioural and physiological signals. VREED is made publicly available on Kaggle [1]. We hope that this contribution encourages other researchers to utilise VREED further to understand emotional responses in VR and ultimately enhance VR experiences design in applications where emotional elicitation plays a key role, i.e. healthcare, gaming, education, etc.

CCS Concepts: • **Human-centered computing** → **Virtual reality**.

Additional Key Words and Phrases: Dataset, Virtual Reality, ECG, GSR, Affective Computing

ACM Reference Format:
Luma Tabbaa, Ryan Searle, Saber Mirzaee Bafti, Md Moinul Hossain, Jittrapol Intarasisrisawat, Maxine Glancy, and Chee Siang Ang. 2021. VREED: Virtual Reality Emotion Recognition Dataset Using Eye Tracking & Physiological Measures. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 4, Article 178 (December 2021), 20 pages. https://doi.org/10.1145/3495002

[1]https://www.kaggle.com/dataset/e9e93acd547401db81cd9988c96760d823142fa894a4a79c6af971949d4bdf85

Authors' addresses: Luma Tabbaa, University of Kent, Canterbury, United Kingdom, L.A.Tabbaa@kent.ac.uk; Ryan Searle, University of Kent, Canterbury, United Kingdom, RLS47@kent.ac.uk; Saber Mirzaee Bafti, University of Kent, Canterbury, United Kingdom, SM2121@kent.ac.uk; Md Moinul Hossain, University of Kent, Canterbury, United Kingdom, M.Hossain@kent.ac.uk; Jittrapol Intarasisrisawat, Prince of Songkla University, Songkhla, Thailand, Jittrapol.psu@gmail.com; Maxine Glancy, BBC Research & Development, London, United Kingdom, Maxine.Glancy@bbc.co.uk; Chee Siang Ang, University of Kent, Canterbury, United Kingdom, C.S.Ang@kent.ac.uk.

2474-9567/2021/12-ART178 $15.00
https://doi.org/10.1145/3495002

1 INTRODUCTION

Emotions can be described as subjective experiences that involve psychological and physiological reactions and responses [31]. They are a crucial aspect in our lives, constituting an essential part in decision-making, social interaction, perception, memory, learning and creativity [81, 88]. Emotions can be expressed through verbal or non-verbal cues (i.e. facial expressions, gestures, voice, etc.) [10]. A plethora of literature has explored the use of technology-based emotion assessment and recognition modalities to further understand human experiences in many domains such as psychology [34, 38, 57], healthcare [39], education [41], advertisement [28] and tourism [53], to mention a few. Assessing and evaluating emotional states are of specific importance in the HCI community [89]; understanding emotional responses can help researchers design better user experiences, hence, improving usability, acceptability and accessibility of technologies. For instance, HCI researchers have used physiological responses to enhance game design and experience [47], evaluate website design and usability [66], and examine the acceptability of assisted interactions using smart conversational agents (such as Apple Siri and Amazon Alexa) [44]. Furthermore, emotion recognition is paramount in Affective Computing research; enabling intelligent systems to recognise, infer and interpret human reactions to improve user experiences as well as enhance their outcomes [38].

To this end, researchers have utilised an array of sensors to evaluate and monitor human's behavioural and physiological biomarkers such as Electrocardiogram (ECG) [54, 88], Galvanic Skin Response (GSR) [48, 83], Electroencephalogram (EEG) [35, 38, 81] and eye tracking [75]. In such studies, emotional elicitation is normally induced by exposing subjects to emotional experiences, situations, or stimuli. To boost the advances in HCI in general and affective computing in particular, many researchers have generated and shared annotated multimodal affective datasets in which other researchers can easily access data that are time-consuming to generate otherwise and directly develop their methodologies and test their hypotheses or algorithms. There are currently several popular affective datasets that combine psychological, behavioural, and physiological data, such as MAHNOB-HCI [80], DECAF [5], DEAP [37] , DREAMER [36] and WESAD[77]. In all these datasets, emotional responses were triggered using non-immersive modalities such as video clips [5, 36, 37, 80] or audio [58].

Virtual Reality (VR), an immersive platform where users "step into" a Virtual Environment (VE), has become a popular modality to trigger or modulate emotional responses. Unlike other non-immersive modalities, users become completely isolated from the real world when using VR, in which they become fully immersed and emotionally present in the VE [11, 12, 71]. A growing body of research has shown that VR can induce varied emotions such as fear [51], happiness [11], sadness [11], relaxation [9] and anxiety [56] using both computer-generated Three-Dimensional Virtual Environments (3D-VEs) [48] and 360° Video-Based Virtual Environments (360-VEs) captured with specialised cameras [85]. Despite the popular use of VR as an immersive stimuli modality in many research disciplines [39, 42, 53], to our knowledge, there are no publicly available affective VR datasets combining psychological, behavioural and physiological data.

Herein, we present a novel multimodal affective dataset where users' emotions were elicited using 360-VEs delivered via a VR headset. The "VR Eyes: Emotions Dataset" (VREED) [**?**] include self-reported questionnaires, eye tracking data, ECG data, and GSR data. The dataset consists of data collected from healthy participants (n=34) who engaged in 360-VEs (n=12, 1-3 minutes each), yielding a total of 408 trials. The 360-VEs were selected through focus groups and a pilot trial (with 12 additional volunteers). The dataset also includes additional documentation of the verbal instructions protocol and questionnaire samples were used in data collection. In this paper, the stimuli selection process and experimental method are described in depth. Furthermore, we present the statistical analysis which validates the selection of the emotionally eliciting 360-VEs. Finally, we carried out a baseline Machine Learning (ML) analysis to demonstrate the application of the dataset in emotion recognition research. We hope that this dataset will serve the research community and provide a much-needed resource that could be used to further understand emotional elicitation and recognition in VR.

2 RELATED WORK

2.1 Measuring Emotion Responses

In order to study emotions and emotional elicitation, it is crucial to understand the underlying structure of emotions and how they are intercorrelated. There is an extensive body of research that proposed several affect models through which researchers can describe, quantify and measure emotions, such as The Tree of Emotions [59], Plutchik's wheel of emotions [64] and the Circumplex Model of Affects (CMA) [73]. In particular, the CMA is a widely used model that interprets emotional affects as a continuum of highly interrelated states. The CMA is a bi-dimensional model, where the valence dimension ranges from positive (i.e. happy, relaxed) to negative (i.e. nervous, sad), and the arousal dimension ranges from low arousal (i.e. calm, depressed) to high arousal (i.e. tense, excited).

In addition to measuring emotional elicitation through subjective self-reports, a plethora of literature has examined the use of physiological and behavioural markers to understand and assess emotional responses during exposure to affective stimuli. Examples of such work explored the use of ECG for heart rate, respiratory rate and blood volume pressure [23, 54, 83, 88], GSR for skin sweat activity [48, 83, 88], EEG for brain activity [35, 81], eye tracking [75], and head movement [45].

Furthermore, eye tracking has gained popularity in many domains as a modality of emotion assessment, as it gives researchers a window into the user's visual, emotional and cognitive processes [65, 75]. Eye-gaze metrics include blinking, fixation (when the eye temporarily remains still over a period of time, typically occur during visual and cognitive processing over informative regions of interest), saccade (the rapid motion of the eye from one fixation to another) and micro-saccade (an intra-fixational eye movement feature where the eye jitters during a fixation) [32]. Studies showed that eye tracking could be used as a robust assessment modality for emotional responses when exposed to stimuli. For instance, pupillary and eye gaze behaviour measures have been used in emotion recognition on positive, negative and neutral video clips, demonstrating 77.80% accuracy using Support Vector Machine (SVM) with linear kernel [2]. Another study found that eye-gaze measures can predict arousal and valence with 71.4% and 58.57% accuracy, respectively, using the Feature Level Fusion (FLF) strategy [8].

Other physiological measures such as ECG; graphic recordings of the changes occurring in electrical potentials as a result of cardiac activity have been previously examined in eliciting emotional responses [6]. In particular, ECG features such as R-R intervals can be used to recognise certain states such as panic, fear and depression [19]. Other ECG features such as heart rate, blood volume pressure, rhythm and electrical activity have also been explored in the literature as a modality to measure emotional states [7, 36]. For instance, using the Least Squares Support Vector Machine (LS-SVM), ECG signals were classified into arousal, valence and the four quadrants of the CMA with a reported accuracy of 82.78%, 72.91%, and 61.52%, respectively [33]. Furthermore, ensemble methods have been shown to be one of the most successful methods for emotion detection using ECG signals [17], achieving an accuracy of 78.12% when classifying the four quadrants of the CMA.

Finally, GSR has been identified as a robust measure of emotional intensity. GSR is an involuntary (autonomic) response of eccrine (sweat glands) activity that is influenced by sympathetic nervous system activity as a result of environmental stimuli such as touch, sight, sound, odour or taste, where GSR increases with excitement and nervousness. GSR data relates to the changes of the sweat gland activity, which usually responds to the intensity of the user's emotional responses. Measuring GSR involves measuring skin conductance; as skin conductance increases, skin resistance drops, which is defined as the improvement in the skin's ability to conduct electric current [21]. Research shows a direct link between the GSR and the emotional arousal the user experiences when exposed to visual or auditory stimuli [74, 86]. Dominguez et al. (2020) compared different ML techniques for emotion recognition using GSR signals. SVM with a linear kernel was found to perform best when evaluated on the test dataset, detecting amusement, sadness, and neutral states. The need to identify effective stimuli to elicit emotions was also highlighted in this study.

In order to elicit psychological, behavioural and physiological reactions, researchers normally induce emotional responses through designing experiences, environments, or stimuli that elicit the desired set of emotions. The body of literature has explored various resources of stimuli to elicit emotional responses such as images [83], sounds [54], film [23], television video commercials [50], and most recently, VR [46]. Using VR, researchers found that 3D-VEs can trigger or elicit a range of emotions, including fear (see a review in [16]), anxiety (see reviews in [16, 60]), as well as relaxation [38, 71]. 360-VEs in VR were also found to be emotionally engaging by eliciting a range of emotions such as anger [48], relaxation [9, 45], sadness [9, 45], anxiety and fear [45].

2.2 Emotional Elicitation & Assessment in Virtual Reality

An emerging body of research has utilised VR technology to create an emotional space where users "step into" and emotionally engage with the VR experience. Such research has explored how emotional responses can be triggered and assessed in a variety of applications such as healthcare [51, 56], forensics [69], education[41], training [38], gaming [25] and advertisement [28], to mention a few. There are many reasons why VR has received such significant interest in the research community. Firstly, the immersive nature of VR experiences allows the user to suspend their awareness of the real world and feel like "being there" in the virtual space [78]. Secondly, the sense of presence in VR enables the user to engage with affective stimuli more deeply and profoundly than non-immersive or semi-immersive mediums [11, 71]. Finally, VR Head Mounted Displays (HMDs) are becoming more readily available in the consumer market with a variety of interaction modalities, price ranges and working mechanisms, including system-dependent, all-in-one and mobile VR, making VR more accessible and easy to deploy to end-users.

Considering that VR can simulate a real response, emotions elicited during VR are equally real [78]. In addition to the psychological influence of emotional elicitation in VR, researchers found that emotional stimuli in VR also influence users physiologically, such as blood pressure and heart rate [27, 48], skin conductance response [27], brain activity [38], eye-gaze behaviour [62, 68, 82], head movement [45], and physical sexual arousal [69, 82].

Despite the plethora of literature exploring the use of VR and supporting its efficacy in eliciting emotional responses, only a little research has employed the use of eye tracking using VR in this context. One reason could be that the eye tracking hardware and software have only been recently embedded in mainstream VR HMDs, available in the consumer market. For instance, one study reproduced a VR version of a well-known cognitive task used for cognitive ability evaluation using eye tracking VR [62]. Another study used eye tracking VR is in the assessment of sexual deviancy [82]. In both cases, 3D-VEs were used where eye tracking data was captured by projecting the data onto 3D objects. As for eye tracking in VR using 360-VEs, one study examined gaze-guided adaptive narratives to enhance the user's experience in virtual tourism, where the relevant text-based and audio-based information appeared depending on the user's gaze at pre-defined points of interest [40]. Another study explored emotional elicitation in 360-VEs and explored the correlation between head movement in VR and emotional elicitation, finding that head yaw positively correlated with valence and head pitch with arousal [45].

2.3 Affective Datasets

Although data-gathering is essential for many studies, it can often be a lengthy, difficult or expensive process that requires significant resources. One alternative to streamline this process is the use of publicly available datasets or open data. Access to high-quality data is vital for improving upon existing research and technologies in many domains. Specifically, in HCI and affective computing, datasets available for research use are important for testing research questions, hypotheses, algorithms, or the design of new ML models that better recognise emotions and understand how humans interact with computers. There are a number of well-established affective datasets which are widely used in current literature. Two of the most popular datasets which categorise emotional stimuli based on subjective ratings are the International Affective Picutres System (IAPS) [43] and the Affective Digital

Sounds System (IADS) [42], in which, they provide hundreds of digital picture-based or sound-based emotional stimuli. In both datasets, the stimuli content and participants' ratings of each stimuli item are included in the dataset.

In addition to subjective ratings, many affective datasets provide researchers with physiological and behavioural data. There are datasets that incorporate psychological, behavioural and physiological signals captured while subjects engage in affective stimuli. For instance, MAHNOB-HCI is an emotion recognition dataset in which face videos, audio signals, eye-gaze data, and peripheral/central nervous system physiological signals were recorded while participants watched 20 different emotional videos [80]. Furthermore, the Database for Emotion Analysis using Physiological signals (DEAP) is a publicly available dataset that contains affective music videos, the subjective ratings of these videos and the recorded EEG data for participants when engaging in these videos [37]. Similarly, The "Database for Emotion Recognition through EEG and ECG Signals from Wireless Low-cost Off-the-Shelf Devices" (DREAMER) utilises EEG signals as well as ECG signals recorded from participants watching audio-visual stimuli [36]. Finally, the "Database for Decoding Affective Physiological Responses" (DECAF) recorded brain signals similar to those found in DEAP and MAHNOB-HCI but was recorded using a Magnetoencephalogram (MEG) which facilitates a naturalistic affective response [5].

It is important to note that all of the aforementioned datasets have utilised non-immersive modalities to elicit and record emotional responses. In the context of affective stimulation using VR, publicly available resources are scarce. Only one affective dataset was found, in which a list of affective 360-VEs is shared with the research community [46]. Despite the wide use of VR as an immersive medium to elicit emotional responses, multimodal VR datasets that would allow researchers to test their hypotheses, algorithms, and theories to further understand emotional elicitation in VR are lacking. To our knowledge, there is no publicly available dataset of VR-based affective stimuli that combines psychological, physiological and behavioural data.

To this end, we present a publicly available dataset called "VR Eyes: Emotions Dataset" (VREED)[2] which can be found on Kaggle. The dataset consists of self-reported questionnaires, pre-processed eye tracking data, ECG data and GSR data, as well as the extracted feature set (eye tracking, ECG and GSR) used in our baseline machine learning analysis. This data is not synchronised with the Youtube videos due to the pre-processing steps. In addition, VREED includes the initial list of 360-VEs that were used as part of the stimuli selection process through focus groups as well as the results of a pilot trial where additional volunteers rated a selection of 360-VEs. Due to licensing restrictions, the actual 360-VEs could not be included in the dataset; however, full description and Youtube links are provided in VREED. Finally, the dataset provides documentation including a verbal instructions protocol and questionnaire samples used during the data collection process. Table 1 presents with a summary of the dataset contents.

Throughout this paper, we introduce the VREED dataset, validate the stimuli used for emotional elicitation, and demonstrate its potential through ML analysis. We hope that VREED will provide a much-needed source to the growing community of HCI researchers in emotional elicitation and recognition in VR.

3 STIMULI SELECTION PROCESS

The selection of effective stimuli is essential for eliciting appropriate affective responses; therefore, the selection process underwent several stages. The study aimed to select three 360-VEs within each quadrant of the CMA. The final 360-VEs (n=12) used in the study were a result of focus groups and a pilot trial; each explained in the following subsections.

[2]https://www.kaggle.com/dataset/e9e93acd547401db81cd9988c96760d823142fa894a4a79c6af971949d4bdf85

Table 1. Dataset Content Overview

Stimuli Selection Process Data	
Focus Groups	Researchers (n=6) met over three 1-hour sessions to extensively discuss and experience 360-VEs (n=126) in the aims of finding suitable VEs for the pilot trial
Pilot Trial	In a span of one week, volunteers (n=12) spent around 1 hour each engaging in and rating the selected 360-VEs (n=21) using SAM (arousal, valence) and VAS (joy, anger, calmness, sadness, disgust, relaxation, happiness, fear, anxiousness and dizziness)
Experiment Data	
Stimuli Selection Method	Based on the pilot trial results (n=12)
Participant Profile & Pre-Exposure Data	Participants (n=34) provided demographic information (i.e. sex, sexual orientation, age, ethnicity), answered a health inclusion questionnaire, and provided baseline SAM (arousal, valence) and VAS (joy, anger, calmness, sadness, disgust, relaxation, happiness, fear, anxiousness and dizziness) ratings
Post-Exposure Data (repeated measure)	SAM (arousal, valence), VAS (joy, anger, calmness, sadness, disgust, relaxation, happiness, fear, anxiousness and dizziness), and PQ ratings
Recorded Physiological Signals	Pre-processed and features extracted of eye tracking, ECG, and GSR data
Additional Materials	Questionnaire samples, a verbal instructions protocol, randomisation table of 360-VEs per-participant

3.1 Focus Groups

Using the Youtube online platform, the research team attempted to manually find potential 360-VEs using the filter "360" (which refers to 360-VEs) available on the platform (n=126). Six researchers with HCI and psychology expertise engaged in three focus group sessions to identify suitable VEs. As a result, the following exclusion criteria emerged to the selection process:

- VEs which are purely or heavily computer-generated; i.e. for the consistency of the selection.
- Monochrome 360-VEs; i.e. maintain the colour scheme across all selected VEs consistent.
- Blurring, moving, shaking, unstable cameras; i.e. to avoid inducing adverse effects of VR such as motion sickness.
- Bad stitching techniques; as it may cause unwanted distraction or annoyance.
- Resolution less than 2K (2048×1080); i.e. to avoid compromising content resolution quality.
- Audio content with low recording quality such as scratching, unclear, or white/ambient noise to avoid annoyance.
- 360-VEs that are less than one minute long are to be excluded; i.e. to keep minimal engagement time consistent. As for 360-VEs that are more than three minutes long were to be capped at the three-minute

mark; since the participants are expected to engage in twelve 360-VEs, prolonged sessions may cause exhaustion.

After applying the exclusion criteria, researchers in the focus groups identified the initial selection list (n=81). Then, researchers excluded 360-VEs (n=38) which were perceived as neutral or received highly conflicting ratings where the focus groups could not agree on the emotion categorisation. Finally, for the remaining 360-VEs (n=43), to ensure a diverse selection of the 360-VEs, the research team voted for the most emotionally intense 360-VEs among 360-VEs that shared highly similar content. For example, only one 360-VE was selected among 360-VEs containing baby animals such as puppies and kittens. In this process, 22 360-VEs were excluded. As a result, the pilot trial included twenty-one 360-VEs.

3.2 Pilot Trial

Twelve volunteers (six females and six males) aged between 19 and 33 (M=24.17, SD=4.19) engaged in and rated the selected 360-VEs (n=21). The volunteers watched the 360-VEs in a randomised order. Volunteers rated each 360-VE using the following tools:

- Self-Assessment Manikin (SAM) [14] is a well-established affective state measurement using picture-based of a cartoon-like manikin shapes (SAM) to plot basic the CMA dimensions. The valence scale ranges from 1="sad" to 9="happy" pictures of SAM, while the arousal scale ranges from 1="calm" to 9="excited" pictures of SAM.
- The Visual Analog Scale (VAS) [30] is a horizontal scale ranging across a continuum from 0 to 100, anchored by two verbal descriptors at each end. Using VAS, volunteers rated how they felt whilst engaging in 360-VEs using a scale of: joy, happiness, calmness, relaxation, anger, disgust, fear, anxiousness and sadness.

The full results of the pilot trial are included in the VREED dataset. Table 2 shows the final selection of 360-VEs for the study (n=12) and the volunteer's valence and arousal ratings.

Table 2. Pilot Trial Volunteers' Ratings Results

CMA	ID	Title	Rated Valence		Rated Arousal	
			M	SD	M	SD
High Arousal Positive	125	Rope Walking	6.42	1.38	7.17	1.34
	012	Brazilian Dance	6.38	1.60	6.00	2.45
	024	Dancing with the Stars	6.67	1.51	6.00	1.26
Low Arousal Positive	117	Beautiful Resorts	7.83	1.59	3.83	2.55
	051	Pond in a Forest	7.08	1.44	3.58	2.68
	013	Cute Bunnies	7.42	1.68	3.50	2.58
High Arousal Negative	108	The Exorcist	3.75	2.18	6.75	1.86
	109	Alone in a Tent	3.83	2.21	6.50	2.39
	116	Zombies Eating Flesh	3.33	2.39	6.33	2.39
Low Arousal Negative	076	Post Terror Attacks	3.25	1.96	3.42	2.19
	075	Refugee Stories	2.75	1.76	3.50	2.02
	080	Refugee Boats	2.17	1.80	3.83	2.44

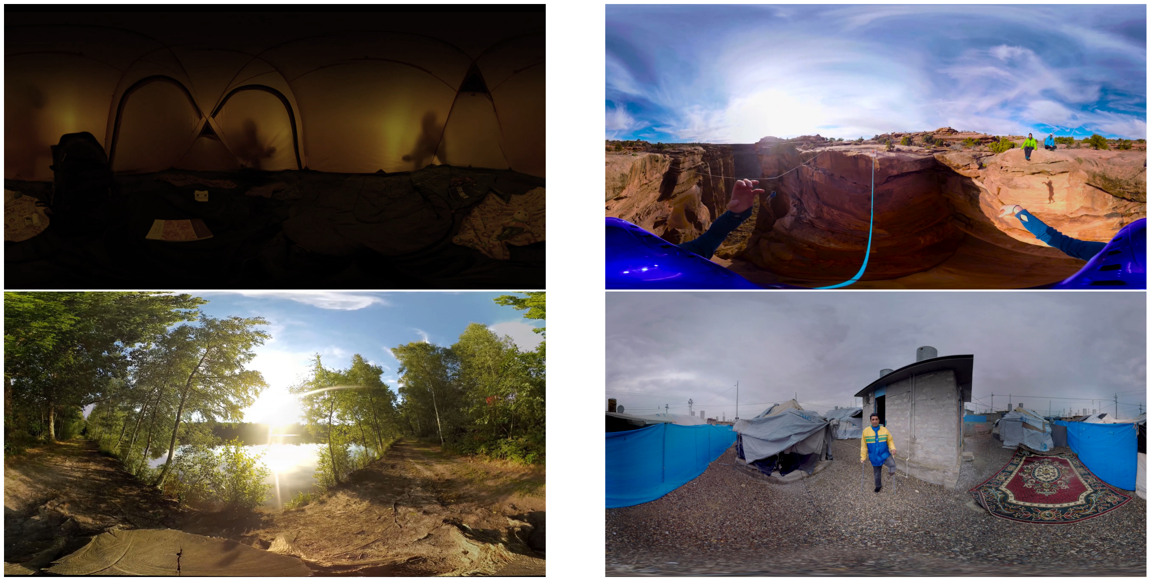

Fig. 1. Examples of 360-VEs Used in the Study

4 EXPERIMENTAL SETUP AND DATA COLLECTION

4.1 Participant Screening Criteria

An invite was sent to various mailing lists within the University of Kent. The email contained an overview of the study and a link to a survey which described participation information and an eligibility checker.

Specifically, individuals who reported they have or have had a seizure(s), seizure disorder, epilepsy, heart condition(s), heart arrhythmia or hypertension were excluded from participation. Furthermore, individuals who reported to have or have had a vestibular disorder, any medical condition(s) affecting balance, frequent headaches, light-headedness or dizziness, visual or hearing impairment, head injury, neurological disease(s), learning disability, psychological disorders or clinical depression were also excluded. In addition, for individuals who have a perfect or close-to-perfect vision with the assistance of glasses were excluded from participation due to the rigidity of the VR headset. Furthermore, individuals who rated six or higher on a Likert scale on "how easily do you get motion or carsick?" where 1="never been motion sick" and 7="get motion sick very easily" were excluded considering that participants were expected to engage in VR for an extended period. Finally, due to the nature of materials used for ECG and GSR experimental setup, individuals who reported to have skin rash from non-precious metal and rubbing alcohol were excluded from participation.

4.2 Ethics

All participants signed a consent form prior to the study. All procedures conformed to the Declaration of Helsinki. The study was approved by the Research Ethics Review Group at the University of Kent.

Privacy and security is a significant concern with physiological data collection, especially eye tracking data, where biometric content may be retrieved from eye movements [13]. As such, to preserve the anonymity of our participants, raw data is not provided as part of VREED. Instead, pre-processed data (eye tracking, ECG and GSR) and extracted features are included.

4.3 Participants

Thirty-four individuals (17 female and 17 male) aged between 18 and 61 years (M=25.0, SD=7.65) volunteered to take part in this study. 55.9% of participants (n=19) reported having used VR before; of which, none have reported feeling motion sick amid or post exposure to VR. On a Likert scale from one to seven on "how easily do you get motion or carsick?" where 1="never been motion sick" and 7="get motion sick very easily" participants reported they do not easily get motion sick (M=1.35, SD=1.12).

4.4 Apparatus and Setup

All computers and machines were set up in a "control room", whilst the "participant room" only contained a table and chair, a VR headset and headphones, a laptop as well as ECG and GSR cables.

Headset & Headphones: The FOVE-0 [24] headset and a set of headphones were wire-connected to a dedicated computer to stream the visual and auditory content. The FOVE-0 is a hands-free headset secured with its 3-point harness adjustable Velcro head straps. The headset has a WQHD OLED display (2560x1440 pixels) and renders at a frame rate of 60 fps with a field of view up to 100 degrees. The head orientation tracking system uses Inertial Measurement Units (IMU), and the eye tracking system uses infrared-based technology on each eye with tracking accuracy less than 1 degree at 120 fps and running at a sampling frequency of 60Hz. Headphones were used to stream auditory content.

Biopac System: The Biopac MP 150 system [4] was used to continuously acquire ECG (using the ECG100C hardware module) and GSR (using the EDA100C hardware module) signals and each were set to a unique channel. For both ECG and GSR, Biopac data acquisition software AcqKnowledge 4.1 [3] was used. The Biopac system was wire-connected to a dedicated computer. The participant's right arm was secured using Velcro tape to ensure minimal noise to ECG and GSR signals.

ECG: Continuous signals were acquired using a Lead II configuration (LEAD110-Series cables and 3-meter MEC110C extension cables) where electrodes were placed on the participant's right arm (Vin-) wrist and left calf (Vin+). Pre-gelled disposable electrodes (EL500-Series) were used, enhanced with electrode gel (GEL100) to increase the conductivity between the skin and the electrode and secured with an adhesive tape to ensure minimal noise.

GSR: Continuous signals were acquired using two disposable and adhesive electrodes (BIOPAC EL507) placed on the participant's right hands' index and middle fingers and wire-connected to the hardware module (using LEAD110A cables).

Lastly, an 11" Macbook Pro laptop and mouse were provided for the participant to fill all self-reported questionnaires digitally.

4.5 Experimental Procedure

Each participant visited the laboratory once, where each session lasted approximately two hours. A verbal instructions protocol was used to ensure that instructions are held constant for all participants. Prior to the start of the session, participants were informed about the study but were not informed about the purpose or the hypotheses of the study. Participants then signed the consent form and filled the "participant profile & pre-exposure" questionnaire. Then, ECG and GSR electrodes were placed onto the participant's body and secured with adhesive and Velcro tapes. Afterwards, the equipment was tested to ensure that the electrodes were picking up signals correctly, followed by a 3 minute rest that was recorded as baseline. During rests (baseline or between 360-VEs) participants were asked to close their eyes, breathe normally and try to not to think of anything too exciting or stressful. Once completed, participants were introduced to the use of the VR; they were explained how the headset can be fitted using the adjustment straps and how to navigate VEs by rotating their head and upper body whilst seated. Eye-gaze calibration was required using the standard FOVE-0 calibration program where they were asked to look and follow a green dot. Once the calibration has been completed, participants were asked to engage in the 360-VE from beginning to end. At the end of each VE, participants filled the "Post-Exposure" questionnaire then had a two-minute cool-down period (where they were asked to relax and sit quietly) before the next one. This procedure was repeated until participants engaged in all 360-VEs. The order of the quadrants and the 360-VEs within the quadrants were randomised using the Latin Square Design to avoid order effects. At the end of the session, participants were fully briefed about the study aims and received a £10 Amazon voucher as a token of appreciation for their participation.

4.6 Psychological Measures

Samples of the "Participant Profile & Pre-Exposure" and "Post-Exposure" questionnaires are included in the VREED as supplementary material. Participants completed the following self-reported measures:

- As part of the "Participant Profile & Pre-Exposure" questionnaire, participants were asked to report demographic information such as age, sex, sexual orientation, ethnicity, dominant hand, and English proficiency.
- As part of the "Participant Profile & Pre-Exposure" questionnaire, the SAM and VAS (described in 3.2) were used as a baseline measure. The questions were adapted to ask participants how they felt "right now, at this moment".
- As part of the "Post-Exposure" questionnaire, the SAM and VAS were used immediately after engaging in each 360-VE. The questions were adapted to ask participants how they felt "whilst watching the video".
- As part of the "Post-Exposure" questionnaire, participants were asked to complete the Presence Questionnaire (PQ) [84] immediately after each 360-VE experience. The PQ is composed of eight questions related to feelings of presence rated on a 7-point Likert scale.
- Using VAS, participants were asked to rate how dizzy they felt, once at the beginning of the session (as part of the "Participant Profile & Pre-Exposure" questionnaire) and once after exposure to each 360-VE (as part of the "Post-Exposure" questionnaire).

4.7 Eye Tracking, ECG & GSR Measures

Eye Tracking: The raw data (per participant per 360-VE, sampled at 60 Hz rate) included vector data of each eye independently, binary eye-closed/open for each eye independently, head orientation in degrees and timestamp.

ECG: ECG signals were sampled at 2000 Hz with a 35 Hz filter and 1 Hz high pass filter using the Biopac Acqknowledge software according to the manufacturer guidelines.

GSR: Tonic GSR signals were recorded at a sample rate of 2000 Hz with a low pass filter at 10 Hz and a high pass filter set to DC. The signal was recorded at 15 Hz with the Biopac Acqknowledge software.

4.8 Feature Extraction

The study included 34 participants engaging in twelve 360-VEs, yielding a total of 408 trials. Data from eight participants were excluded from further analysis due to the poor quality in one or more of the data modalities recorded caused by technical problems. Therefore, the final dataset included 26 participants and 312 trials.

4.8.1 *Eye Tracking Feature Extraction.*

Eye tracking raw data per participant were generated for each 360-VE from beginning to end, including data for the left eye, right eye, and head rotation. The GazeParser library [79] was used to extract eye-gaze features for analysis. In preparation for feature extraction, eye-gaze and head-movement data were used in conjunction to produce horizontal and vertical viewing angle (X, Y) data per eye. The feature extraction algorithm is based on velocity threshold method or what is also known as velocity-based identification; where the algorithm distinguishes fixations (low velocity) from saccades (high velocity) based on gaze point-to-point velocities (see full taxonomy of identifying eye-gaze metrics and eye tracking protocols including velocity-based algorithms [76]). Finally, statistical calculations were carried out for each feature. Considering that the 360-VEs had different time lengths, the count (number of instances) were normalised by length (number of instances/time). Table 3 describes the eye tracking features that were extracted, brief description, threshold and the statistical calculations carried out for each feature, including Normalised Count (NormCount), Mean (M), Maximum (Max), Standard Deviation (SD) and Skewness (Skew); a measure of asymmetry in a distribution. All eye tracking features can be found in Table 3.

Table 3. Eye Tracking Features

Main Feature	Brief Description	Threshold	Statistical Metrics
Fixation	Temporal stillness in the eye movement over time	Fixation minimum duration=300ms	Number of fixations (NormCount) per second First fixation duration Duration (M, Max, SD, Skew)
Micro-Saccade	Intra-fixational movement where the eye jitters during a fixation	Micro-saccade minimum duration during a fixation=400ms	Number of micro-saccades (NormCount) per second Peak velocity (M, Max, SD, Skew) Direction (M, Max, SD, Skew) Horizontal amplitude (M, Max, SD, Skew) Vertical amplitude (M, Max, SD, Skew)
Saccade	Rapid motion of the eye from one fixation to another	Saccade velocity threshold=35ms Saccade acceleration threshold=400ms Saccade minimum duration=30ms Saccade minimum amplitude=5ms	Number of saccades (NormCount) per second Duration (M, Max, SD, Skew) Direction (M, Max, SD, Skew)
Blink	Eye closed	Blink minimum duration =50ms	Number of blinks (NormCount) per second Duration (M, Max, SD, Skew)

4.8.2 ECG Feature Extraction. Prior to feature extraction, the signals were filtered using a high pass filter with a threshold of 10 to correct noise and baseline problems. Features shown in literature [15] to be reliable predictors of emotional state are used. Features were extracted using an implementation by heartpy [26]. Considering that the 360-VEs had different time lengths, the count (number of instances) were normalised by length (number of instances/time). A total of 18 features were extracted, as described in table 4.

Table 4. ECG Features

Feature	Brief Description
Signal Amplitude (M, Max, Min)	The voltage of electrical activity recorded
R-R Intervals (M, Med, Max, Min, SD, SDSD, RMSSD)	Time elapsed between two successive R-waves
ibi	Inter beat interval
bpm	Beats per minute
pnn50	Percentage of successive RR intervals that differ by more than 50ms
pnn20	Percentage of successive RR intervals that differ by more than 20ms
pnn50pnn20	Ratio of pnn50 and pnn20
VLF	Absolute power of the very-low-frequency band (0.0033–0.04 Hz)
LF	Absolute power of the low-frequency band (0.04–0.15 Hz)
HF	Absolute power of the high-frequency band (0.15–0.4 Hz)

4.8.3 *GSR Feature Extraction.*

A low pass filter that removes muscle noise which allows the detection of sweating peaks to be more accurate, was first applied. Features shown in literature [15] to be reliable predictors of emotional state are used. Considering that the 360-VEs had different time lengths, the count (number of instances) were normalised by length (number of instances/time). A total of eight features were extracted, as described in table 5.

Table 5. GSR Features

Feature	Brief Description
Skin Conductance level (M, Max, Min, SD, Var)	Tonic level of electrical conductivity of skin
Number of local minima per second	Number of valleys per second
Number of local maxima per second	Number of peaks per second
Peaks/Time	Ratio of peaks and time

5 RESULTS

5.1 Validation of 360-VEs

This analysis aimed to understand whether 360-VEs triggered the desired arousal and valence effects among the participants. Table 1 describes the valence and arousal ratings using SAM in each quadrant. For the valence dimension, negative VEs were perceived as negative (low arousal negative; M=2.46, high arousal negative; M=4.29) and positive VEs were perceived as positive (low arousal positive; M=6.48, high arousal positive; M=6.46). For the arousal dimension, the intended low arousal VEs were perceived as low arousing (low arousal negative; M=3.52, low arousal positive; M=2.51). However, for the intended high arousal VEs, only the intended high arousal negative VEs were perceived as high arousing (high arousal negative; M=6.00, high arousal positive; M=3.80).

Table 6. Ratings of Valence and Arousal Per CMA Quadrant

Intended CMA Quadrant	Rated Arousal		Rated Valence	
	M	SD	M	SD
High Arousal Positive	3.80	1.97	6.46	0.77
Low Arousal Positive	2.51	1.12	6.48	1.00
High Arousal Negative	6.00	1.60	4.29	1.59
Low Arousal Negative	3.52	1.69	2.46	1.06

Rated Valence ANOVA: Two-way repeated-measures ANOVA, followed by Tukey's HSD test were carried out to determine the significance of engaging in 360-VEs in the four CMA quadrants over the valence dimension. The rated valence significantly differed in the four quadrants of the CMA, F(132, 2)=21.62, p<.001. Tukey's HSD test indicated that the mean of rated valence in negative VEs (M=3.38, SD=1.63) was significantly lower than positive VEs (M=6.47, SD=0.89), t(132, 2)=-15.59, p<.001. The mean of rated valence in low arousal VEs (M=4.44, SD=2.71) was significantly lower than high arousal VEs (M=5.36, SD=1.66), t(132, 2)=-4.55, p<.001; meaning that participants rated high arousal VEs significantly more positively than low arousal VEs.

Rated Arousal ANOVA: Two-way repeated-measures ANOVA, followed by Tukey's HSD test were carried out to determine the significance of engaging in 360-VEs in the four CMA quadrants over the arousal dimension.

The rated arousal significantly differed in the four quadrants of the CMA, F(132, 2)=4.55, p=.035. Tukey's HSD test indicated that the mean of rated arousal in low arousal VEs (M=3.02, SD=1.51) was significantly lower than high arousal VEs (M=4.92, SD=2.09), t(132, 2)=-6.71, p<.001. The mean value of rated arousal in negative VEs (M=4.76, SD=2.05) was significantly higher than positive VEs (M=3.16, SD=1.72), t(132, 2)=5.71, p<.001; meaning that negative VEs were perceived as significantly more arousing than positive VEs.

In summary, the participants experienced four distinct emotional states over the arousal dimension (high, low) and valence dimension (negative, positive). Even though participants perceived the high arousal positive 360-VEs as low arousing, the ratings of arousal in this quadrant were still significantly higher than low arousing positive 360-VEs.

5.2 Baseline Machine Learning Analysis

Three classification tasks, two binary and a four-class problem, were carried out: high/low arousal (binary), positive/negative valence (binary) and the four classes in accordance with the four quadrants of the CMA. Based on initial tests with a range of classifiers and drawing from literature [18, 36], SVM was selected and was used for the baseline analysis. Using an SVM classifier with a radial basis function kernel, 10-fold cross validation was performed on 90% of the data leaving 10% for unseen testing. In particular, stratified K-fold validation was used to ensure balanced classes. SciKit learns [61] implementation of grid search was used to select the hyper-parameters of the SVM. The classification accuracy's obtained from testing the model on unseen data and the precision (P), recall (R) and f1 scores are presented in Table 8. These results showed the benefits of using combined modalities especially in the four-class classification task. The eye tracking features substantially outperforms the ECG features for arousal and four-class classification but performs similarly in recognising valence. Eye Tracking features outperform the GSR features in valence and four-class classification tasks and they perform equally in the arousal classification task. The combination of features of all three modalities achieves the highest accuracy than all of the individual modalities for all three classification tasks.

Tables 7, 8 and 9 present a summary of these results, specifically highlighting the results obtained using similar modalities as VREED as well as the modalities which achieved the best accuracy for each dataset. Direct comparisons with the literature work must be proceeded with caution due to the differences in experimental setups, apparatus, conditions and feature extraction and analysis methodologies. Furthermore, research in affective computing examining the use of VR is scarce. However, some similarities with studies which examined affect in non-immersive mediums can be observed. The baseline results achieved with the VREED dataset are consistent with other baseline results reported in other affective datasets that used non-immersive mediums. Our results are comparable with DEAP [37] where EEG and peripheral physiological responses of which included GSR, blood volume pressure, respiration pattern, skin temperature and Electromyography (EMG) were recorded whilst engaging in video clips using a PC monitor. 104 features were extracted from these modalities. Using peripheral features (excluding EEG), the authors obtained accuracy of 57% when classifying arousal and 62.7% accuracy when classifying valence which compare to our 87.5% and 56.25% respectively using only GSR features. When these features were fused with EEG and multimedia content analysis features, higher accuracy was obtained (61.6% for arousal & 64.7% for valence). Dreamer [36] contains ECG and EEG data, the classification accuracy using the ECG data with SVM reported on the baseline analysis was 62.37% for both arousal valence compared to our 75% and 78% accuracy for arousal and valence. DECAF [5] includes various peripheral physiological responses including ECG. Using late fusion, arousal was classified with 73.5% and valence with 77.5% accuracy. The MAHNOB-HCI dataset [80] includes four peripheral nervous system signals: GSR, respiration amplitude, skin temperature, and ECG along with EEG and eye-gaze data. A fusion of modalities performed best followed by the eye-gaze data, achieving 68.8% and 63.5% for binary classification of arousal and valence. Our eye tracking results compare well achieving 87.5% and 78.13% for arousal and valence. In conclusion, our baseline results compare similarly to other well-established emotion recognition datasets, proving VREED to be a robust dataset for further research ML analysis.

Table 7. Accuracy's of Different Modalities in the Recognition of Emotions

Modality	Arousal	Valence	4-Class
ECG	75%	78%	50%
GSR	87.5%	56.25%	53.13%
Eye Tracking	87.5%	78.13%	62.5%
ECG + GSR + Eye Tracking	90.63%	84.38%	71.88%

Table 8. Precision, Recall and F1 Scores

Modality	Arousal			Valence			4-Class		
	P	R	f1	P	R	f1	P	R	f1
ECG	0.75	0.75	0.75	0.79	0.78	0.78	0.52	0.5	0.5
GSR	0.88	0.88	0.87	0.62	0.56	0.53	0.53	0.53	0.52
Eye Tracking	0.88	0.88	0.87	0.74	0.74	0.72	0.63	0.62	0.62
ECG + GSR + Eye Tracking	0.91	0.91	0.91	0.86	0.84	0.84	0.76	0.72	0.72

Table 9. Comparison with Other Datasets

Dataset	Modality	Arousal	Valence
VREED	ECG	75%	78%
	GSR	87.5%	56.25%
	Eye Tracking	87.5%	78.13%
	ECG + GSR + Eye Tracking	90.63%	84.38%
DEAP	Peripheral	57%	62.7%
	MCA	65.1%	61.8%
DREAMER	ECG	62.37%	62.37%
	ECG + EEG	62.32%	61.84%
DECAF	Peripheral	55%	60%
	Late Fusion	73.5%	77.5%
MAHNOB-HCI	Peripheral	46.2%	45.5%
	Eye Gaze	63.5%	68.8%
	Fusion	67.7%	76.1%

6 DISCUSSIONS & CONCLUSIONS

In this paper, we present VREED, a publicly available affective dataset that utilises VR as a stimulus which enables users to engage with deeper emotional engagement compared to non-immersive or semi-immersive mediums [1, 70]. This dataset contains ECG, GSR and eye tracking data from 26 participants. The data were recorded while participants engaged in twelve 360-VEs specially selected to elicit certain emotions following the CMA emotion model. Limitations in the dataset relate to the size and sample of participants, which included University volunteers. As such, a larger study is needed to produce a bigger dataset that generalise to the general population.

The participants experienced the correct emotion whilst engaging in the 360-VE's. Two-way repeated-measures ANOVA, followed by Tukey's HSD tests were used to validate this. The arousal and valence ratings significantly differed in the four quadrants, meaning that the participants experienced four distinct emotional states over the arousal dimension (high, low) and valence dimension (negative, positive).

There are many publicly available datasets [5, 36, 37, 77, 80] available. Although these datasets are not directly comparable (due to different sensor modalities and the use of non-immersive emotion triggers), the machine learning results obtained using VREED are promising and show an improvement in accuracy in similar classification problems. We achieved 90.63%, 84.38% and 71.88% accuracy classifying arousal, valence and the four sections of the CMA quadrant, respectively.

The dataset we have produced can be used to develop more advanced emotion recognition models. These models could lead to more effective communication and improve applications that rely on human-computer interaction in many domains [29, 41, 53], including a variety of healthcare applications. For instance, an important use of affective systems is to assist parents, teachers and carers of children with autism [63]. A reliable affective model, in which VREED could aid in the development of, could be used to detect the childrens' emotional states when it might not be otherwise visible, and communicate this to the children themselves or others. This type of system could enhance communication and help assist social interaction. Affective systems have also been used as diagnostic and treatment tools in various stress and anxiety disorders. Specifically, post-traumatic stress disorder is one of these disorders in which the diagnostics and treatment have benefited greatly from affective systems where VEs have been used to elicit stress in a controlled environment [67, 72]. The ability to detect and record affective states within these VEs may improve treatments by allowing clinicians to personalise the environment to the level that helps the patient feel the correct emotion or intensity of emotion [87].

Future work on this topic could include the use of EEG signals which many established datasets have used and shown to be a good predictor of emotion [36, 37, 80]. EEG signals were not included into our dataset due to the difficulties in collecting such data whilst wearing a VR HMD without significant redesign of hardware. To overcome this, a reliable and compact EEG sensor that can be used in conjunction with a VR HMD needs to be identified and/or developed. For instance, there has been significant development efforts in designing flexible sensors that captures physiological signals, including EEG for brain-machine interactions [49], and there has been pioneering work integrating such sensors with VR headsets [52].

In addition, more in-depth investigation using more advanced ML techniques could be carried out to improve the current emotion recognition performance further. One technique that has produced very good results in many domains within healthcare, including robotic surgery and diagnostic tasks, is deep learning [20] and has been successful whilst using physiological datasets. However larger datasets are needed to develop robust deep learning models [22]. These can be time-consuming and expensive to create for physiological and affective datasets. This problem can be overcome using transfer learning; this is where one can develop a model on a similar dataset and then refine it on a smaller dataset. This technique has already successfully been used for emotion recognition [55]. Further research would involve producing larger datasets or identifying existing datasets where it would be appropriate to develop transfer models.

Finally, future work could focus on developing end-to-end therapeutic systems for use, especially in a home environment, combining the sensors, emotion recognition models and VR therapy application. To achieve this, assessing the viability of using off-the-shelf heart rate and GSR sensors will be vital. The ability to easily use this system outside of a more controlled laboratory/hospital environment is key in allowing this technology to reach and aid more people. VREED is made publicly available on Kaggle, and it is hoped that other researchers would utilise VREED to test their methods, hypotheses and algorithms.

REFERENCES

[1] 2005. Commentary on Riva, G., Virtual Reality in Psychotherapy: Review. *CyberPsychology Behavior* 8, 3 (2005), 231–240. https://doi.org/10.1089/cpb.2005.8.231

[2] 2015. *Combining Eye Movements and EEG to Enhance Emotion Recognition.* AAAI Press.

[3] 2021. AcqKnowledge Data Acquisition and Analysis Software - Win: ACK100W, ACK100M: Research: BIOPAC. https://www.biopac.com/product/acqknowledge-software/

[4] 2021. Data Acquisition, Loggers, Amplifiers, Transducers, Electrodes: BIOPAC. https://www.biopac.com/

[5] M. K. Abadi, R. Subramanian, S. M. Kia, P. Avesani, I. Patras, and N. Sebe. 2015. DECAF: MEG-Based Multimodal Database for Decoding Affective Physiological Responses. *IEEE Transactions on Affective Computing* 6, 3 (2015), 209–222. https://doi.org/10.1109/TAFFC.2015.2392932

[6] Foteini Agrafioti, Dimitris Hatzinakos, and Adam K. Anderson. 2012. ECG Pattern Analysis for Emotion Detection. *IEEE Transactions on Affective Computing* 3, 1 (2012), 102–115. https://doi.org/10.1109/T-AFFC.2011.28

[7] F. Agrafioti, D. Hatzinakos, and A. K. Anderson. 2012. ECG Pattern Analysis for Emotion Detection. *IEEE Transactions on Affective Computing* 3, 1 (2012), 102–115. https://doi.org/10.1109/T-AFFC.2011.28

[8] Ashwaq Alhargan, Neil Cooke, and Tareq Binjammaz. 2017. Multimodal affect recognition in an interactive gaming environment using eye tracking and speech signals. *Proceedings of the 19th ACM International Conference on Multimodal Interaction - ICMI 2017* (2017). https://doi.org/10.1145/3136755.3137016

[9] Allison P. Anderson, Michael D. Mayer, Abigail M. Fellows, Devin R. Cowan, Mark T. Hegel, and Jay C. Buckey. 2017. Relaxation with Immersive Natural Scenes Presented Using Virtual Reality. *Aerospace Medicine and Human Performance* 88, 6 (2017), 520–526. https://doi.org/10.3357/amhp.4747.2017

[10] Michael Argyle, Veronica Salter, Hilary Nicholson, Marylin Williams, and Philip Burgess. 1970. The Communication of Inferior and Superior Attitudes by Verbal and Non-verbal Signals*. *British Journal of Social and Clinical Psychology* 9, 3 (1970), 222–231. https://doi.org/10.1111/j.2044-8260.1970.tb00668.x

[11] R.m. Baños, C. Botella, M. Alcañiz, V. Liaño, B. Guerrero, and B. Rey. 2004. Immersion and Emotion: Their Impact on the Sense of Presence. *CyberPsychology & Behavior* 7, 6 (2004), 734–741. https://doi.org/10.1089/cpb.2004.7.734

[12] Stéphane Bouchard, Geneviève Robillard, Patrice Renaud, and François Bernier. 2011. EXPLORING NEW DIMENSIONS IN THE ASSESSMENT OF VIRTUAL REALITY INDUCED SIDE EFFECTS. *Journal of Computer and Information Technology* (2011), 20–22.

[13] Efe Bozkir, Onur Günlü, Wolfgang Fuhl, Rafael F. Schaefer, and Enkelejda Kasneci. 2021. Differential privacy for eye tracking with temporal correlations. *Plos One* 16, 8 (2021). https://doi.org/10.1371/journal.pone.0255979

[14] M. M. Bradley and P. J. Lang. 1994. Measuring Emotion: The Self-Assessment Manikin and the Semantic Differential. *Journal of Behavior Therapy and Experimental Psychiatry* 25, 1 (1994), 49–59. https://doi.org/10.1016/0005-7916(94)90063-9

[15] Adrián Colomer Granero, Félix Fuentes-Hurtado, Valery Naranjo Ornedo, Jaime Guixeres Provinciale, Jose M. Ausín, and Mariano Alcañiz Raya. 2016. A Comparison of Physiological Signal Analysis Techniques and Classifiers for Automatic Emotional Evaluation of Audiovisual Contents. *Frontiers in Computational Neuroscience* 10 (2016), 74. https://doi.org/10.3389/fncom.2016.00074

[16] Julia Diemer, Georg W. Alpers, Henrik M. Peperkorn, Youssef Shiban, and Andreas Mã¼Hlberger. 2015. The impact of perception and presence on emotional reactions: a review of research in virtual reality. *Frontiers in Psychology* 6 (2015). https://doi.org/10.3389/fpsyg.2015.00026

[17] Theekshana Dissanayake, Yasitha Rajapaksha, Roshan Ragel, and Isuru Nawinne. 2019. An Ensemble Learning Approach for Electrocardiogram Sensor Based Human Emotion Recognition. *Sensors* 19, 20 (2019), 4495. https://doi.org/10.3390/s19204495

[18] J.a. Domínguez-Jiménez, K.c. Campo-Landines, J.c. Martínez-Santos, E.j. Delahoz, and S.h. Contreras-Ortiz. 2020. A machine learning model for emotion recognition from physiological signals. *Biomedical Signal Processing and Control* 55 (2020), 101646. https://doi.org/10.1016/j.bspc.2019.101646

[19] Maria Egger, Matthias Ley, and Sten Hanke. 2019. Emotion Recognition from Physiological Signal Analysis: A Review. *Electronic Notes in Theoretical Computer Science* 343 (2019), 35–55. https://doi.org/10.1016/j.entcs.2019.04.009

[20] Andre Esteva, Alexandre Robicquet, Bharath Ramsundar, Volodymyr Kuleshov, Mark Depristo, Katherine Chou, Claire Cui, Greg Corrado, Sebastian Thrun, Jeff Dean, and et al. 2019. A guide to deep learning in healthcare. *Nature Medicine* 25, 1 (2019), 24–29.

https://doi.org/10.1038/s41591-018-0316-z
[21] B. Farnsworth. [n.d.]. What is GSR (galvanic skin response) and how does it work? https://imotions.com/blog/gsr/
[22] Oliver Faust, Yuki Hagiwara, Tan Jen Hong, Oh Shu Lih, and U Rajendra Acharya. 2018. Deep learning for healthcare applications based on physiological signals: A review. *Computer Methods and Programs in Biomedicine* 161 (2018), 1–13. https://doi.org/10.1016/j.cmpb.2018.04.005
[23] Cristina Fernández, Juan C. Pascual, Joaquim Soler, Matilde Elices, Maria J. Portella, and Enrique Fernández-Abascal. 2012. Physiological Responses Induced by Emotion-Eliciting Films. *Applied Psychophysiology and Biofeedback* 37, 2 (2012), 73–79. https://doi.org/10.1007/s10484-012-9180-7
[24] Inc. FOVE. [n.d.]. FOVE0 Headset Specification. url: https://fove-inc.com/product/;. https://fove-inc.com/product/; Accessed: 4-May-2020.
[25] Sarah-Christin Freytag and Carolin Wienrich. 2017. Evaluation of a virtual gaming environment designed to access emotional reactions while playing. In *2017 9th International Conference on Virtual Worlds and Games for Serious Applications (VS-Games)*. 145–148. https://doi.org/10.1109/VS-GAMES.2017.8056585
[26] Paul Van Gent, Haneen Farah, Nicole Van Nes, and Bart Van Arem. 2019. HeartPy: A novel heart rate algorithm for the analysis of noisy signals. *Transportation Research Part F: Traffic Psychology and Behaviour* 66 (2019), 368–378. https://doi.org/10.1016/j.trf.2019.09.015
[27] Alessandra Gorini, Eric Griez, Anna Petrova, and Giuseppe Riva. 2010. Assessment of the emotional responses produced by exposure to real food, virtual food and photographs of food in patients affected by eating disorders. *Annals of General Psychiatry* 9, 1 (2010). https://doi.org/10.1186/1744-859x-9-30
[28] Dan M. Grigorovici and Corina D. Constantin. 2004. Experiencing Interactive Advertising beyond Rich Media. *Journal of Interactive Advertising* 5, 1 (2004), 22–36. https://doi.org/10.1080/15252019.2004.10722091
[29] Aya Hassouneh, A.m. Mutawa, and M. Murugappan. 2020. Development of a Real-Time Emotion Recognition System Using Facial Expressions and EEG based on machine learning and deep neural network methods. *Informatics in Medicine Unlocked* 20 (2020), 100372. https://doi.org/10.1016/j.imu.2020.100372
[30] G. A. Hawker, S. Mian, T. Kendzerska, and M. French. 2011. Measures of Adult Pain: Visual Analog Scale for Pain (VAS Pain), Numeric Rating Scale for Pain (NRS Pain), McGill Pain Questionnaire (MPQ), Short-Form McGill Pain Questionnaire (SF-MPQ), Chronic Pain Grade Scale (CPGS), Short Form-36 Bodily Pain Scale (SF). *Arthritis Care & Research* 63, s11 (2011), 240–252. https://doi.org/10.1002/acr.20543
[31] Don H. Hockenbury and Sandra E. Hockenbury. 2010. *Discovering psychology and study guide.* Macmillan.
[32] K. Holmqvist, M. Nyström, R. Andersson, R. Dewhurst, H. Jarodzka, and J. Van de Weijer. 2011. *Eye tracking: a comprehensive guide to methods and measures.* Oxford University Press.
[33] Y. Hsu, J. Wang, W. Chiang, and C. Hung. 2020. Automatic ECG-Based Emotion Recognition in Music Listening. *IEEE Transactions on Affective Computing* 11, 1 (2020), 85–99. https://doi.org/10.1109/TAFFC.2017.2781732
[34] Benjamin Iffland, Fabian Klein, Rita Rosner, Babette Renneberg, Regina Steil, and Frank Neuner. 2020. Cardiac reactions to emotional words in adolescents and young adults with PTSD after child abuse. *Psychophysiology* 57, 1 (2020), e13470. https://doi.org/10.1111/psyp.13470
[35] Amir Jalilifard, Ednaldo Brigante Pizzolato, and Md Kafiul Islam. 2016. Emotion classification using single-channel scalp-EEG recording. *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (2016). https://doi.org/10.1109/embc.2016.7590833
[36] S. Katsigiannis and N. Ramzan. 2018. DREAMER: A Database for Emotion Recognition Through EEG and ECG Signals From Wireless Low-cost Off-the-Shelf Devices. *IEEE Journal of Biomedical and Health Informatics* 22, 1 (2018), 98–107. https://doi.org/10.1109/JBHI.2017.2688239
[37] S. Koelstra, C. Muhl, M. Soleymani, J. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras. 2012. DEAP: A Database for Emotion Analysis ;Using Physiological Signals. *IEEE Transactions on Affective Computing* 3, 1 (2012), 18–31. https://doi.org/10.1109/T-AFFC.2011.15
[38] Ilkka Kosunen, Mikko Salminen, Simo Järvelä, Antti Ruonala, Niklas Ravaja, and Giulio Jacucci. 2016. RelaWorld. *Proceedings of the 21st International Conference on Intelligent User Interfaces - IUI 16* (2016). https://doi.org/10.1145/2856767.2856796
[39] M. Arun Kumar, R. Vimala, and K.r. Aravind Britto. 2019. A cognitive technology based healthcare monitoring system and medical data transmission. *Measurement* 146 (2019), 322–332. https://doi.org/10.1016/j.measurement.2019.03.017
[40] Tiffany C.k. Kwok, Peter Kiefer, Victor R. Schinazi, Benjamin Adams, and Martin Raubal. 2019. Gaze-Guided Narratives. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (2019). https://doi.org/10.1145/3290605.3300721
[41] H. Chad Lane and Sidney K. D'Mello. 2018. Uses of Physiological Monitoring in Intelligent Learning Environments: A Review of Research, Evidence, and Technologies. *Mind, Brain and Technology Educational Communications and Technology: Issues and Innovations* (2018), 67–86. https://doi.org/10.1007/978-3-030-02631-8_5
[42] Peter J. Lang and Margaret M. Bradley. 2007. *The International Affective Digitized Sounds (2nd Edition; IADS-2): Affective Ratings of Sounds and Instruction Manual.* Technical Report. University of Florida. Technical report B-3.
[43] Peter J. Lang, Margaret M. Bradley, and B. N. Cuthbert. 2008. *International affective picture system (IAPS): affective ratings of pictures and instruction manual.* Technical Report. University of Florida. Technical Report A-8.

[44] Seungyeon Lee, Hyejin Ryu, Byungho Park, and Myung Hwan Yun. 2020. Using Physiological Recordings for Studying User Experience: Case of Conversational Agent-Equipped TV. *International Journal of Human–Computer Interaction* 36, 9 (2020), 815–827. https://doi.org/10.1080/10447318.2019.1693166

[45] Benjamin J. Li, Jeremy N. Bailenson, Adam Pines, Walter J. Greenleaf, and Leanne M. Williams. 2017. A Public Database of Immersive VR Videos with Corresponding Ratings of Arousal, Valence, and Correlations between Head Movements and Self Report Measures. *Frontiers in Psychology* 8 (2017). https://doi.org/10.3389/fpsyg.2017.02116

[46] Benjamin J. Li, Jeremy N. Bailenson, Adam Pines, Walter J. Greenleaf, and Leanne M. Williams. 2017. A Public Database of Immersive VR Videos with Corresponding Ratings of Arousal, Valence, and Correlations between Head Movements and Self Report Measures. *Frontiers in Psychology* 8 (2017), 2116. https://doi.org/10.3389/fpsyg.2017.02116

[47] Tao Lin and Atsumi Imamiya. 2006. Evaluating usability based on multimodal information. *Proceedings of the 8th international conference on Multimodal interfaces - ICMI 06* (2006). https://doi.org/10.1145/1180995.1181063

[48] Mary F. Macedonio, Thomas D. Parsons, Raymond A. Digiuseppe, Brenda A. Weiderhold, and Albert A. Rizzo. 2007. Immersiveness and Physiological Arousal within Panoramic Video-Based Virtual Reality. *CyberPsychology & Behavior* 10, 4 (2007), 508–515. https://doi.org/10.1089/cpb.2007.9997

[49] Musa Mahmood, Deogratias Mzurikwao, Yun-Soung Kim, Yongkuk Lee, Saswat Mishra, Robert Herbert, Audrey Duarte, Chee Siang Ang, and Woon-Hong Yeo. 2019. Fully portable and wireless universal brain–machine interfaces enabled by flexible scalp electronics and deep learning algorithm. *Nature Machine Intelligence* 1, 9 (2019), 412–422. https://doi.org/10.1038/s42256-019-0091-7

[50] Anca Cristina Micu and Joseph T. Plummer. 2010. Measurable Emotions: How Television Ads Really Work. *Journal of Advertising Research* 50, 2 (2010), 137–153. https://doi.org/10.2501/s0021849910091300

[51] Matthew C. Mishkind, Aaron M. Norr, Andrea C. Katz, and Greg M. Reger. 2017. Review of Virtual Reality Treatment in Psychiatry: Evidence Versus Current Diffusion and Use. *Current Psychiatry Reports* 19, 11 (2017). https://doi.org/10.1007/s11920-017-0836-0

[52] Saswat Mishra, Yun-Soung Kim, Jittrapol Intarasirisawat, Young-Tae Kwon, Yongkuk Lee, Musa Mahmood, Hyo-Ryoung Lim, Robert Herbert, Ki Jun Yu, Chee Siang Ang, and et al. 2020. Soft, wireless periocular wearable electronics for real-time detection of eye vergence in a virtual reality toward mobile eye therapies. *Science Advances* 6, 11 (2020). https://doi.org/10.1126/sciadv.aay1729

[53] Nur Shuhadah Mohd, Hairul Nizam Ismail, Norliza Isa, and Syed Muhammad Rafy Syed Jaafar. 2019. Millennial Tourist Emotional Experience In Technological Engagement At Destination. *International Journal of Built Environment and Sustainability* 6, 1-2 (2019), 129–135. https://doi.org/10.11113/ijbes.v6.n1-2.396

[54] M. Nardelli, G. Valenza, A. Greco, A. Lanata, and E. P. Scilingo. 2015. Recognizing Emotions Induced by Affective Sounds through Heart Rate Variability. *IEEE Transactions on Affective Computing* 6, 4 (2015), 385–394. https://doi.org/10.1109/TAFFC.2015.2432810

[55] Hong-Wei Ng, Viet Dung Nguyen, Vassilios Vonikakis, and Stefan Winkler. 2015. Deep Learning for Emotion Recognition on Small Datasets using Transfer Learning. *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction* (2015). https://doi.org/10.1145/2818346.2830593

[56] Kazuyuki Niki, Yoshiaki Okamoto, Isseki Maeda, Ichiro Mori, Ryouhei Ishii, Yoshinobu Matsuda, Tatsuya Takagi, and Etsuko Uejima. 2019. A Novel Palliative Care Approach Using Virtual Reality for Improving Various Symptoms of Terminal Cancer Patients: A Preliminary Prospective, Multicenter Study. *Journal of Palliative Medicine* 22, 6 (2019), 702–707. https://doi.org/10.1089/jpm.2018.0527

[57] Heather J. Nuske, Giacomo Vivanti, Kristelle Hudry, and Cheryl Dissanayake. 2014. Pupillometry reveals reduced unconscious emotional reactivity in autism. *Biological Psychology* 101 (2014), 24–35. https://doi.org/10.1016/j.biopsycho.2014.07.003

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

[59] W. Gerrod. Parrott. 2001. *Emotions in social psychology: essential readings.* Psychology Press.

[60] Thomas D. Parsons and Albert A. Rizzo. 2008. Affective outcomes of virtual reality exposure therapy for anxiety and specific phobias: A meta-analysis. *Journal of Behavior Therapy and Experimental Psychiatry* 39, 3 (2008), 250–261. https://doi.org/10.1016/j.jbtep.2007.07.007

[61] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12 (2011), 2825–2830.

[62] J. Pettersson, A. Albo, J. Eriksson, P. Larsson, K. W. Falkman, and P. Falkman. 2018. Cognitive Ability Evaluation using Virtual Reality and Eye Tracking. *2018 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA)* (2018). https://doi.org/10.1109/civemsa.2018.8439999

[63] Rosalind W. Picard. 2009. Future affective technology for autism and emotion communication. *Philosophical Transactions of the Royal Society B: Biological Sciences* 364, 1535 (2009), 3575–3584. https://doi.org/10.1098/rstb.2009.0143

[64] Robert Plutchik. 2001. The Nature of Emotions. *American Scientist* 89, 4 (2001), 344. https://doi.org/10.1511/2001.4.344

[65] Luca Puviani, Sidita Rama, and Giorgio Matteo Vitetta. 2016. Computational Psychiatry and Psychometrics Based on Non-Conscious Stimuli Input and Pupil Response Output. *Frontiers in Psychiatry* 7 (2016). https://doi.org/10.3389/fpsyt.2016.00190

[66] Qing-Xing Qu, Fu Guo, and Vincent G. Duffy. 2017. Effective use of human physiological metrics to evaluate website usability. *Aslib Journal of Information Management* 69, 4 (2017), 370–388. https://doi.org/10.1108/ajim-09-2016-0155

[67] Greg M. Reger, Kevin M. Holloway, Colette Candy, Barbara O. Rothbaum, JoAnn Difede, Albert A. Rizzo, and Gregory A. Gahm. 2011. Effectiveness of virtual reality exposure therapy for active duty soldiers in a military mental health clinic. *Journal of Traumatic Stress* 24, 1 (2011), 93–96. https://doi.org/10.1002/jts.20574

[68] Patrice Renaud, Joanne L. Rouleau, Luc Granger, Ian Barsetti, and Stéphane Bouchard. 2002. Measuring Sexual Preferences in Virtual Reality: A Pilot Study. *CyberPsychology & Behavior* 5, 1 (2002), 1–9. https://doi.org/10.1089/109493102753685836

[69] Patrice Renaud, Dominique Trottier, Joanne-Lucine Rouleau, Mathieu Goyette, Chantal Saumur, Tarik Boukhalfi, and Stéphane Bouchard. 2013. Using immersive virtual reality and anatomically correct computer-generated characters in the forensic assessment of deviant sexual preferences. *Virtual Reality* 18, 1 (2013), 37–47. https://doi.org/10.1007/s10055-013-0235-8

[70] Bacchetta M. Cesa G. Conti S. Riva, G. and E. Molinari. 2004. The Use of VR in the Treatment of Eating Disorders. *Studies in Health Technology and Informatics* 99 (2004), 121–163. https://doi.org/10.3233/978-1-60750-943-1-121

[71] Giuseppe Riva, Fabrizia Mantovani, Claret Samantha Capideville, Alessandra Preziosa, Francesca Morganti, Daniela Villani, Andrea Gaggioli, Cristina Botella, and Mariano Alcañiz. 2007. Affective Interactions Using Virtual Reality: The Link between Presence and Emotions. *CyberPsychology & Behavior* 10, 1 (2007), 45–56. https://doi.org/10.1089/cpb.2006.9993

[72] Albert Rizzo, Greg Reger, Greg Gahm, JoAnn Difede, and Barbara O. Rothbaum. 2009. Virtual Reality Exposure Therapy for Combat-Related PTSD. *Post-Traumatic Stress Disorder* (2009), 375–399. https://doi.org/10.1007/978-1-60327-329-9_18

[73] James A. Russell. 1980. A circumplex model of affect. *Journal of Personality and Social Psychology* 39, 6 (1980), 1161–1178. https://doi.org/10.1037/h0077714

[74] Valorie N. Salimpoor, Mitchel Benovoy, Gregory Longo, Jeremy R. Cooperstock, and Robert J. Zatorre. 2009. The Rewarding Aspects of Music Listening Are Related to Degree of Emotional Arousal. *PLOS ONE* 4, 10 (10 2009), 1–14. https://doi.org/10.1371/journal.pone.0007487

[75] Dario D. Salvucci and Joseph H. Goldberg. 2000. Identifying fixations and saccades in eye-tracking protocols. *Proceedings of the symposium on Eye tracking research & applications - ETRA 00* (2000). https://doi.org/10.1145/355017.355028

[76] Dario D. Salvucci and Joseph H. Goldberg. 2000. Identifying Fixations and Saccades in Eye-Tracking Protocols. In *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications (ETRA '00)*. Association for Computing Machinery, New York, NY, USA, 71–78. https://doi.org/10.1145/355017.355028

[77] Philip Schmidt, Attila Reiss, Robert Duerichen, Claus Marberger, and Kristof Van Laerhoven. 2018. Introducing WESAD, a Multimodal Dataset for Wearable Stress and Affect Detection. *Proceedings of the 20th ACM International Conference on Multimodal Interaction* (2018). https://doi.org/10.1145/3242969.3242985

[78] Mel Slater and Maria V. Sanchez-Vives. 2016. Enhancing Our Lives with Immersive Virtual Reality. *Frontiers in Robotics and AI* 3 (2016). https://doi.org/10.3389/frobt.2016.00074

[79] Hiroyuki Sogo. 2012. GazeParser: an open-source and multiplatform library for low-cost eye tracking and analysis. *Behavior Research Methods* 45, 3 (2012), 684–695. https://doi.org/10.3758/s13428-012-0286-x

[80] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic. 2012. A Multimodal Database for Affect Recognition and Implicit Tagging. *IEEE Transactions on Affective Computing* 3, 1 (2012), 42–55.

[81] Samarth Tripathi, Shrinivas Acharya, Ranti Dev Sharma, Sudhanshi Mittal, and Samit Bhattacharya. 2017. Using Deep and Convolutional Neural Networks for Accurate Emotion Classification on DEAP Dataset. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI'17)*. AAAI Press, 4746–4752.

[82] Dominique Trottier, Joanne-Lucine Rouleau, Patrice Renaud, and Mathieu Goyette. 2013. Using Eye Tracking to Identify Faking Attempts During Penile Plethysmography Assessment. *The Journal of Sex Research* 51, 8 (2013), 946–955. https://doi.org/10.1080/00224499.2013.832133

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

[84] Bob G. Witmer and Michael J. Singer. 1998. Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence: Teleoperators and Virtual Environments* 7, 3 (1998), 225–240. https://doi.org/10.1162/105474698565686

[85] Kate Wolitzky, Robyn Fivush, Elana Zimand, Larry Hodges, and Barbara Olasov Rothbaum. 2005. Effectiveness of virtual reality distraction during a painful medical procedure in pediatric oncology patients. *Psychology & Health* 20, 6 (2005), 817–824. https://doi.org/10.1080/14768320500143339

[86] G. Wu, G. Liu, and M. Hao. 2010. The Analysis of Emotion Recognition from GSR Based on PSO. In *2010 International Symposium on Intelligence Information Processing and Trusted Computing*. 360–363. https://doi.org/10.1109/IPTC.2010.60

[87] Georgios N Yannakakis. 2018. Enhancing health care via affective computing. *Malta Journal of Health Sciences* 5, 1 (2018), 38–42. https://doi.org/10.14614/HEALTHCOMP/9/18

[88] Shanbin Zhang, Guangyuan Liu, and Xiangwei Lai. 2015. Classification of Evoked Emotions Using an Artificial Neural Network Based on Single, Short-Term Physiological Signals. *Journal of Advanced Computational Intelligence and Intelligent Informatics* 19, 1 (2015), 118–126. https://doi.org/10.20965/jaciii.2015.p0118

[89] Chiara Zucco, Barbara Calabrese, and Mario Cannataro. 2017. Sentiment analysis and affective computing for depression monitoring. *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (2017). https://doi.org/10.1109/bibm.2017.8217966