IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 11 July 2025, accepted 5 August 2025, date of publication 1 September 2025, date of current version 10 September 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3597793

RESEARCH ARTICLE

Affect Recognition in Immersive Room-Scale Environments: A Large-Scale VR Study With Custom Facial Sensing at the Science Museum in London

IFIGENEIA MAVRIDOU[1], (Member, IEEE), EMILI BALAGUER-BALLESTER[2], THEODOROS KOSTOULAS[3], (Member, IEEE), AND ELLEN SEISS[4]

[1]Department of Cognitive Science and Artificial Intelligence, Tilburg University, 5037 AB Tilburg, The Netherlands
[2]Department of Computing and Informatics, Bournemouth University, BH12 5BB Poole, U.K.
[3]Department of Information and Communication Systems Engineering, University of the Aegean, 832 00 Samos, Greece
[4]Department of Psychology, Faculty of Science and Technology, Bournemouth University, BH12 5BB Poole, U.K.

Corresponding author: Ifigeneia Mavridou (i.mavridou@tilburguniversity.edu)

This work was supported by U.K. Engineering and Physical Sciences Research Council (EPSRC) via the Centre for Digital Entertainment, Bournemouth University, under Grant EP/L016540/1.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Bournemouth University Research Governance and Ethics Committee on 13/02/2018 under Application No. ID: 18848.

ABSTRACT Recent technological advances have provided the chance to conduct Virtual Reality (VR) experiments with increased ecological validity, which in turn can elicit more naturalistic responses in immersed users. However, many studies still prefer highly controlled setups and passive stimulation, often due to the practical complexities in effectively associating cause (stimulus) to response in highly interactive and dynamic VR experiences. Many of these studies also rely on subjective ratings from participants recorded either after the experience (relying on memory) or during the experience (interrupting immersion). In this paper, we advance this experimental protocol in a large-scale feasibility study by 1) investigating affective changes in terms of valence and arousal ratings in various interactive 3D room-scale VR environments with 2) continuous valence and arousal self-ratings from a controller and 3) a novel wireless physiological facial EMG and PPG sensor setup specifically designed to record affect, without relying on memory or interrupting immersion. In this study, n=291 participants experienced neutral, positive and negative virtual environments in 'passive' and 'active' conditions. Continuous self-ratings and physiological measures confirmed the feasibility of detecting affective states in room-scale VR conditions. To our knowledge, this is the highest *n* in a feasibility study in affect detection to date. Our study generated the most populated physiological data library collected in VR, which also compares passive and active VR settings. This setup can provide a solid experimental foundation for VR affective computing studies in more unconstrained, ecologically valid environments.

INDEX TERMS Affective computing, facial expression, database, multimodal analysis, physiological signals, technology and devices for affective computing, guidelines, three-dimensional graphics, virtual reality.

I. INTRODUCTION

Commercially available Virtual Reality (VR) technologies can simulate far more immersive real-world scenarios and interactions than traditional experimental affect

The associate editor coordinating the review of this manuscript and approving it for publication was Alessandro Floris.

manipulations. These simulated real-world scenarios can elicit more intense emotional reactions in an experimentally controlled setting [1]. Growing evidence has shown that participants who experience higher levels of presence in VR tend to interact in more naturalistic ways [2], [3], [4] and have natural affective reactions to lifelike scenarios.

To achieve these affective responses in naturalistic VR setups, three design factors must effectively integrate with three essential aspects: the depth of presence, the immersive impact, and the user's holistic engagement with the mediated experience [5]. These aspects are contingent upon the carefully designed VR technology's inherent attributes alongside the meticulous structuring of sensory-rich environments.

Recent VR technologies offer advanced, lightweight user tracking methods combined with high-resolution audio and graphical capabilities. This amalgamation holds the potential to generate experiences with heightened ecological validity (a term linked to the content's richness/vividness [6]). These advances encourage the shifting support a shift away from traditional, seated laboratory setups—often reliant on distracting and cumbersome sensors—toward unobtrusive, wearable technologies that enable more naturalistic exploration and interaction. As such, sensor-enabled VR has the capacity to become the ideal experimental tool for behavioural and affective sciences.

Virtual environments (VE), were utilised in conjunction with physiological response acquisition as a psychological research tool in manifold applications, typically related to stress/arousal induction or stress reduction/meditation (see review by [7]). Recent examples have shown that VR can be used in exposure therapy to induce relaxation and to reduce anxiety [8], in pain research to distract users from painful experiences (see review by [9]), and for stress reduction in clinical contexts [10], [11]. VR scenarios were also used to induce negative emotions and stress (negative arousal), for example in public speaking scenarios [12] and by introducing phobic elements into the VE [13]. More broadly, reported presence levels are higher in stressful than neutral VE scenarios in the literatures [14] and [15].

These VR interventions greatly benefit from an effective affect recognition system. Such a system could foster the effectiveness of VR applications in phobia therapy [16], meditation and relaxation [3], [17], [18], training and exergames ([19], [20], [21], [22]), spectrum disorders [23], [24], mental health therapies [25], and aesthetic experiential research [26].

Affect typically consists of the instinctual fluctuations of two main dimensions: valence and arousal (as conceptualised in the Circumflex Model [27]). Arousal refers to the level of physiological activation accompanying the emotional state, ranging from calmness to intense excitement. On the other hand, valence describes the quality of that activation, ranging from negative to positive. Sensing approaches, such as heartrate changes and skin conductance, can indicate affective variations and thus are common in psychophysiological research. Recently, we have witnessed the integration of such sensing methods within headset-induced VR studies [7]. However, physiological activation is only one way of inferring affect:

Intense affective experiences often trigger facial muscle activations [28], and movement patterns, like approach and withdrawal, can be effectively tracked via integrated sensors [29]. Recent works have demonstrated the potential of camera-based facial tracking and voice cues for affect recognition in multimedia experiences (e.g., [30]). This growing interest in facial expression analysis is further reflected in the announcements of recent commercial HMDs which include built-in face tracking capable of estimating Facial Action Units through computer vision and machine learning techniques [31], [32]. Nonetheless, the integration of EMG in XR research remains a valuable approach, particularly given its widespread use as a ground truth in validating these more indirect methods [33].

Notwithstanding VR's transformative potential to study naturalistic affective responses in room-scale, (inter)active, immersive experiences, the technology is not fully developed for experimental usage. For example, the VEs used to elicit responses are often passive, two-dimensional, pre-recorded, and in seated settings, restraining natural movements [34].

Additionally, arousal and valence levels explored in the literature often skew towards negative and extreme values, typically comparing high vs low arousing and negative versus neutral [35], where the level of variation is limited. Ideally, such exploration would benefit from standardising the experimental paradigm using validated stimuli materials for both affective dimensions in a balanced manner, for example, by including negative and positive validated VEs and stimuli.

This study is designed to address these points by providing further insights into the effects of room-scale VE-induced affective modulation on physiological signals, self-reported affect scores and memory. The concept of a dynamic interactive experience urges us to explore innovative methods for examining affective responses to content. The new approach involves (a) a practical sensor setup for freely walking settings and (b) a re-examination of the conventional approach of defining an experience as a solitary stimulus unit or breaking it down solely along linear stimulus segments. Therefore, we developed novel custom-designed sensing technology designed to detect valence and arousal responses, interaction tracking, and prototypic event-tracking solutions. The following sections provide an overview of the study, its objectives and methods.

II. STUDY OVERVIEW

The study had three key aims. Firstly, it compared the effects of positive, neutral and negative VEs on affect elicitation using subjective and electrophysiological data. Secondly, the study investigated the feasibility of a custom wearable sensor setup for affect detection on both dimensions (valence and arousal) in room-scale VR settings during active exploration. Lastly, it compared passive prerecorded walk-through VR

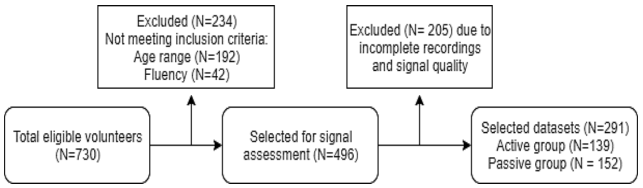

FIGURE 1. **Datasets selection flowchart.**

settings and (inter)active dynamic walk-through VR settings to shed light on the inherent differences of the two induction modes and their effects on subjective aspects of the experience.

Four VEs were used from the AVEL library [34] comprising pre-validated 3D room-scale affective VEs, each with individual event-stimuli. These included one baseline VE and three affective VEs (a neutral, a positive, and a negative) consisted of replicas of a physical office room. Each affective VE was populated with 14 audiovisual stimuli (referred to as 'events') designed to evoke different ranges of valence and arousal.

The active mode supported dynamic, room-scale exploration, allowing users to freely walk, look around, and interact naturally (users can look and move freely in various directions). In contrast, the passive mode limited interactivity and movement (seated, linear experience). The passive mode user was vicariously experiencing the same VEs, only this time via recorded videos from the viewpoint of the user in the active mode. In both modes, the Continuous Affect Self-Rating (CASR) [36] tool was employed using the controller's trackpad, to register valence and arousal self-reported ratings throughout the experience, by moving the thumb atop the VR handheld-controller's trackpad. This allowed continuous VR experience and event-based analysis. Traditional single post-VE affect ratings were also collected for comparison. To effectively link stimuli to response, we developed a novel user-gaze based interaction and event tracking system, along with an event-marker system. These systems enabled us to track the virtual objects that the user was looking at, thus allowing for the analysis of physiological changes in relation to context.

Data were recorded at the Science Museum in London from a diverse population during a six-week period, where museum visitors were invited to participate. The present study is presumed as the largest-scale VR study with physiological sensors conducted up to date. For the continuous physiological tracking, we used four identical sensor face masks of the emteqVR prototype which were mounted onto an HTC VIVE headset [37]. Each mask comprised of seven facial electromyographic (EMG) and two photoplethysmographic (PPG) sensors. The interaction modes, active and passive, were compared using subjective affective ratings, facial EMG, physiological, and movement measures. Potential confounding variables as alexithymia and expressivity were recorded with standardised questionnaires to control for potential group differences. Note, affect intensity is also known to influence cognition, including memory recognition and recall, with notably emotional stimuli being observed as more memorable [38], [39].

A. HYPOTHESES

H1: VE Comparison across groups using post-VE and continuous affective ratings, presence, and memory recognition accuracy. It is predicted that:

H1a: VEs evoke the targeted valence and arousal ratings in VR. These should be similar, albeit more intense, compared to the values recorded for the validated virtual environments in the online survey reported in [34] (manipulation check).

H1b: The expected valence and arousal ratings in VR are affected by expressivity group (stronger ratings for affective VEs for the more expressive group) and alexithymia group (less strong ratings for affective VEs for the high alexithymia group).

H1c: Positive and negative VEs result in interactive VR experiences (active group) inducing enhanced presence scores and memory recognition accuracy.

H2: Comparison between the active and passive group using continuous affective ratings, presence scores and memory recognition accuracy. It is predicted that:

H2a: The active group elicits more extreme affective subjective ratings compared to the passive group.

H2b: The active group elicits higher levels of presence and memory recognition accuracy compared to the passive group.

H3: Feasibility of valence and arousal detection with physiological sensors in active and passive VR. It is predicted that:

H3a: EMG sensors can reliably detect spontaneous affective changes in VR settings.

H3b: PPG sensor can reliably detect arousal changes in VR settings.

H3c: Enhanced affective EMG and PPG responses are expected in the active compared to the passive group, de-spite the reduced signal-to-noise ratio.

H4: Feasibility of valence detection from movement. It is predicted that:

H4a: People show approach behaviour for positive events and withdrawal behaviour for negative events in the active group.

III. METHODS

A. PARTICIPANTS

From an initial pool of 730 volunteers, participants were screened for cardiovascular, medical, and psychological conditions. Other exclusion criteria are described in the Supplementary Material. After exclusions, 291 participants with an age range of 18-35 years and good English fluency were selected for data analysis (Fig.1). Most participants had little to no experience with VR (81.44%). From those, NA=139 participants were randomly allocated in the 'Active' group (76 females (54.7%); 63 males (45.3%), mean age = 25.22 ± 4.69 years), and NP=152 in the 'Passive' group (87 females (46.1%), 64 males (42.1%), 1 transgender/non-

FIGURE 2. **360-degree screenshots of the VEs: (1) Training CASR VE, (2) VR adaptation, nd (7) Negative VE.**

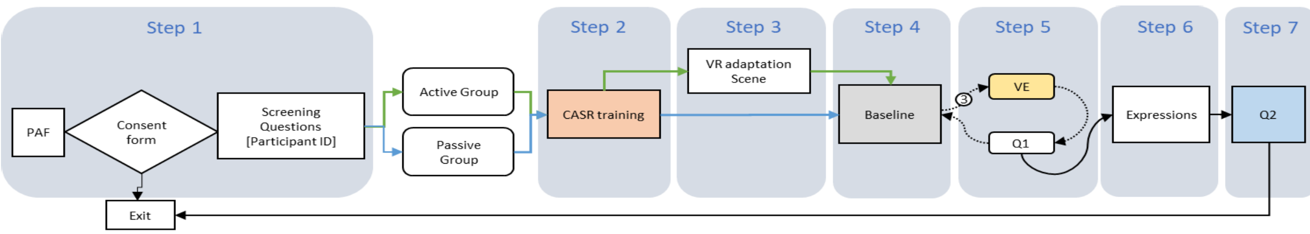

FIGURE 3. **Outline of the experimental protocol used for the VR study divided in steps (1-7). Step 1 included the introduction to the study, the Participant Information Form (PAF), the consent form, a demographic questionnaire with screening questions and the allocations of participant IDs. Participants were then divided into an active and a passive group. They were trained on how to use the CASR interface using the VIVE Controller (Step2). Active group users were introduced to VR adaptation scene (Step 3). Both groups experienced the affective scenarios preceded by a baseline recording session and followed by a short experience questionnaire (Step 4 & 5). Step 6 included a short recording of facial expressions of three emotions. All participants were asked to complete questionnaire (Q2; Step 7) which included questions about their personality, as well as alexithymia and expressivity scales.**

binary individual (0.7%), mean age = 24.35 ± 4.51 years). None of the participants were compensated for their time. The study was approved by the Bournemouth University Research Ethics Committee (ID: 18848).

B. MATERIAL

Apparatus and instrumentation. For this study, four HTC Vive HMDs and six EmteqVR interfaces were used in this study (for further reference see progression of the sensor setups, also known as 'Faceteq' [40], 'emteqVR' [37], and 'emteqPRO' [41] implemented for different headsets). The HTC Vive HMD was selected for its state-of-the-art consumer-grade room-scale tracking, powerful PC-based performance, and compatibility with the emteqVR EMG system at the time of the study. The EmteqVR devices collected multimodal electro-physiological data from seven facial EMG sensors and PPG. They were connected via Bluetooth Low Energy (BLE) to custom applications provided by Emteq Labs. Each participant's data were recorded locally on dedicated desktop computers. In addition, two external webcams were positioned in the corners of the physical room to monitor the participants' movement, and two web-cameras (1080p, 60 fps) were positioned on the wall to provide the virtual camera feed (virtual mirror for enhanced presence [42]; for complete stimuli overview see [34]). Furthermore, six desktop computers (OS: Windows 10) and four tablets (OS: Android) were required for the completion of forms and questionnaires. The OBS software was used to record the VR and camera feed which were used for synchronisation and detection of movement artefacts. This architecture allowed for independent and synchronized data collection across multiple participants in parallel, all within the same semi-controlled environment.

Questionnaires. wo questionnaires were used in this study. The first questionnaire (Q1) was administered after each VE presentation, included single-item ratings (ranged 1–9) for valence, arousal, presence [43], motion sickness, enjoyment, comfort (Post-VE ratings). These Post-VE ratings were collected once per VE to capture overall impressions. In contrast, continuous affect ratings (CASR) were gathered throughout the experience, allowing us to examine in-experience events and linking them to the continuous electrophysiological responses recorded. Using both methods enabled us to compare overall impressions with moment-to-moment experiences, providing deeper insights into the effects of specific events. The questionnaire also included a short memory recognition task which included all events for each affective VE (Neutral, Positive and Negative) (as in the AVEL library [34]). The second questionnaire (Q2) was completed only once at the end of the study. It consisted of a short demographic survey with questions about age, gender, immersive technology experience, the Toronto Alexithymia Scale (TAS, 20 items) [44], and the Berkeley Expressivity Questionnaire [45]. All questionnaires were designed using the Qualtrics Software [46].

Virtual Environments (VE). Seven virtual environments were used in total (Fig. 2). Four VEs were taken from the AVEL library. These are: the 'Baseline VE' (4), the 'Neutral VE' (5), the 'Positive VE' (6), and the 'Negative VE' (7) (see description in [34]). In addition, three other VE scenes were developed: the 'training on CASR' (1), a 'VR adaptation VE' (2), and a 'home cinema' VE (3). The 'training CASR' was a simple VE with a screen where various short videos were played, and instructions were given on how to self-rate using the CASR tool with a wireless hand controller (Fig. 2.1). The VR adaptation VE was designed to familiarize users with the room-scale VR technologies and the movement boundaries. It was used as practice space for training on CASR while exploring the 3D space. The VE featured a 3D path that has the exact same size as the VEs office-replicas (Fig. 2.2). The home-cinema VE, for the passive mode, included a screen for video presentation visible to the participant, and a control user interface only visible to the experimenter (Fig. 2.3). In all the virtual office-room versions, the dimensions of the room and the point of entrance of the user were kept identical. Before running the main VR study, all affective virtual environments were validated online using screen-recorded videos [34]. An independent participant group rated each scenario's valence and arousal using standard scales. This validation ensured the intended emotional tone of each VE.

Physical Space. The study was conducted at the 'Who am I?' gallery space within the Science Museum in London (Fig. S2). For the active condition, the dimension of the VR walkable area was 2.5m x 3m which included the desk where the experimenter was sitting. This desk was successfully masked in VR (as performed in previous studies [47]).

IV. EXPERIMENTAL PROCEDURE

The experiment consisted of seven steps in Fig. 3. All participants were introduced to the study, donned the VR headset with the EmteqVR mask, and completed the CASR training. Participants were allocated randomly into groups ('Active' and 'Passive'). The Active group started with a VR familiarisation exercise, in which instructions for continuous self-rating using the VR controller were given (x axis for valence ratings and y axis for arousal ratings; see Fig S2.1). Participants of the active group navigated the room-scale environments using SteamVR's walk-in-place functionality, which simulates virtual walking by detecting subtle in-place steps without requiring full physical movement across the room. Both groups started with a baseline recording of 2 minutes, followed by the randomised presentation of the three affective VEs (neutral, positive and negative VE). After each VE experience, participants answered the "Q1" questionnaire. Afterwards, participants performed three facial expressions with three repetitions each, smiling, frowning (squeezing the eyebrows) and surprised (raising the eyebrows) in high intensity (referred to as 'expression recording'). Finally, they were asked to complete the "Q2" questionnaire. The entire study had a duration of approximately 40 minutes.

V. SIGNAL PRE-PROCESSING FOR DATA ANALYSIS

EMG. Data were processed with the Signal Processing toolbox in MATLAB using a similar approach as in [36]. A notch filter on 50 Hz and harmonics (from 100 to 450 Hz) was applied on all signals (sampling rate: 1000 Hz) prior to other pre-processing steps. Next, a Butterworth bandpass filter at 50–450 Hz (6th order) was applied in order to envelop the most essential spectrum of f-EMG signal [48]. Baseline correction was applied by subtracting the mean value of the signal per channel. Extreme outliers were removed using a Hampel filter [49]. Post-filtering, the data were visually inspected for malfunctions and low signal-to-noise ratio, which in our case could be caused by interference with movement artefacts or faulty fitting of the sensors. The first 1000 samples and last 1000 samples from each recording were excluded from the processed signal, before segmenting data into epochs. Next, the signals were normalised using the min-max normalisation method (1), where x1–5 are obtained from the four separate VE recordings and the expression recording.

$$x_{norm} = \frac{x - \min(x_1, x_2, x_3, x_4, x_5)}{(\max(x_1, x_2, x_3, x_4, x_5) - \min(x_1, x_2, x_3, x_4, x_5))} \tag{1}$$

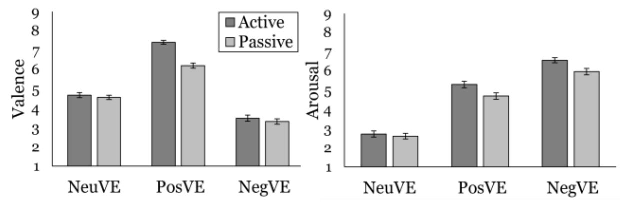

FIGURE 4. **Mean event-based CASR valence (left) and arousal ratings(right) per group. SE Error bars.**

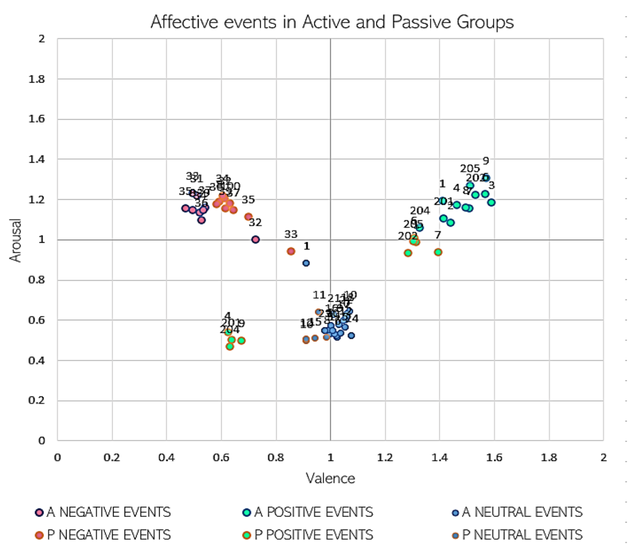

FIGURE 5. **Valence-Arousal coordinates for each event marker, grouped by colour for each VE (negative:pink, positive:green, neutral:blue). The event markers are divided in those derived from the Active ("A") denoted by a blue outline and the Passive group ("P") denoted by orange outline.**

TABLE 1. **3 X2 mixed anova on valence and arousal scores.**

	Main effect VE	Main effect Group	Interaction VE*Group
Valence scores	F(1.88, 542.53) = 340.27, $p < .001$*	F(1, 29) = 22.04, $p < .001$*	F(1.88, 542.53) = 10.708, $p < .001$*
Arousal Scores	F(1.99, 574.19) = 337.60, $p < .001$*	F(1, 289) = 7.99, $p = .005$*	F(1.99, 574.19) = 1.95, p= .143 (ns)

Next, an event-based epoching approach was employed. Each participant encountered a sequence of audiovisual stimuli which were spatially distributed across the room-scale virtual environment. Some were visible throughout the exploration (static) while others activated based on the participants spatial location and duration within the VE (in active scenario). The interaction and activation of those events (N=14) were dynamically annotated in time, based on the behaviour of the user within the VE. The event stimuli are described in the AVEL dataset documentation [34]. Specifically, the onset of the events taking place in each VE was tracked using the custom-build interaction tracking solution with millisecond accuracy. From this, an epoch was calculated from an event's onset (minus 250ms). The duration of each epoch was variable and relative to the user's interaction,

TABLE 2. **3 × 2 Mixed ANOVA on arousal and valence CASR scores.**

	Main effect VE	Main effect Group	Interaction VE*Group
Valence scores	*F(1.91, 550.52) = 511.98, p< .001**	F(1, 289) = 204.83, $p < .001$*	F(1.91 550.52) = 16.72, $p < .001$*
Arousal Scores	F(1.98, 571.39) = 246.70, $p < .001$*	F(1, 289) = 368.09, $p = .005$*	F(1.98, 571.39) = 9.47, $p < .001$*

with a minimum length of 10 seconds. On average the participant interacted with 10 to 14 different events in each VE, resulting into varied number of epochs per user. From each epoch, the root mean square (RMS) value of the epoch was computed per EMG channel. Then, VE averages were calculated as the average across all epochs extracted from each VE.

PPG. The PPG data from the EmteqVR interface (sampling rate: 1000Hz) were filtered with a Butterworth bandpass filter (0.5 – 4Hz). Outliers were removed via Hampel filter. The data were epoched in the same way as the EMG data and inspected for artifacts manually. A strict dataset inclusion protocol was applied based on visual inspection of each user's recordings (7 segments per participant). While baseline segments were generally clean, noise or artifacts were often present during the VE exploration. For inclusion in heart rate and variability analysis, recordings needed to contain at least 25–30 seconds of clean, continuous data. If such segments could not be identified, or if the clean periods did not overlapped with event annotations that dataset was excluded from analysis. This process substantially reduced the usable dataset (N=291 from 496; Fig. 1). Afterwards, R-peaks were detected from which the features [50] of pulse rate (beats-per-minute; BPM) were calculated. In addition, the pulse-rate variability (PRV) was measured by calculating root mean square of the successive differences (RMSSD), and the standard deviation of the NN (R-R) intervals (SDNN) using the "HRV tool" [50].-1

CASR. Data recorded from the CASR tool (from all sessions and groups), were synchronised with the physiological signals using the system timestamps. The data were epoched along with the physiological signals while considering an average human rating response-delay, at 200ms [51], [52], [53]. Median valence and arousal scores were calculated for each epoch from each participant, and the mean scores across epochs were calculated for each VE.

User Movement. Data related to the user's movement were recorded within the Active VR group. The data recorded within the Unity environment, corresponded to the user's normalised vector distance from active virtual events. The custom-build event annotation via gaze tracking system, indicated the interaction onset and off-set, during which the vector distance ('Dis') was recorded. During analysis, the mean Dis per event was calculated and averaged across events of the same VE, giving us three Dis scores per user. I.e. for the positive, neutral and negative VE.

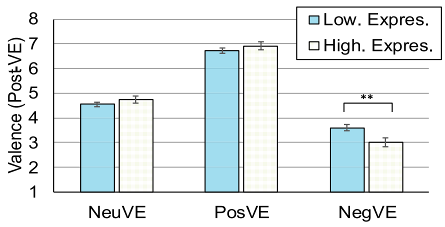

FIGURE 6. Mean valence ratings per expressivity group for the *Neutral* (NeuVE), *Positive* (PosVE), and *Negative* (NegVE) VEs. Error bars display Standard Error (SE).

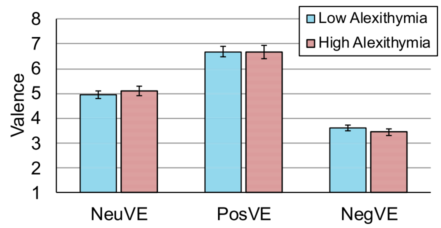

FIGURE 7. Mean valence ratings per alexithymia group for the *Neutral* (NeuVE), *Positive* (PosVE), and *Negative* (NegVE) VEs. Error bars display Standard Error (SE).

VI. RESULTS

A. VALENCE AND AROUSAL RATINGS

Valence and arousal ratings were analysed in several ways. Firstly, we analysed it the traditional way by using post-VE ratings (Q1) where participants were asked how they felt about the overall experience. Secondly, we compared these with the mean valence and arousal ratings from the CASR method where participants rated continuously during the experience. Similar ratings would validate CASR which can be better suited to investigate changes over time during dynamic environments and avoids reliance on memory. Thirdly, we analysed valence and arousal levels of the specific events displayed during the VEs which is only possible when using CASR method. Finally, we investigated whether factors such as expressivity and alexithymia do influence CASR mean valence and arousal ratings.

B. POST-VE VALENCE AND AROUSAL RATINGS (Q1)

Figure 6 displays the mean valence and arousal ratings reported after each VE experience and group (see for mean scores and SD). The figure shows that all VE ratings followed the expected pattern which was confirmed with mixed 3 × 2 ANOVAs with the within-participant factor VE (positive, neutral, negative) and the between-participant factor Group (passive vs. active group) conducted for the valence and arousal Post-VE ratings, separately (Table 1).

Valence Ratings. The mixed ANOVA results revealed a main effect of VE (see Table 1). Post-hoc paired t-tests showed that valence ratings were significantly higher (more positive) for positive than for neutral VEs (t(290)=18.285, $p<.001$), for neutral than for negative VEs (t(290)=8.674, $p<.001$) and for positive than for negative VEs (t(290)=23.694, $p<.001$). Secondly, the main effect of group was also significant, with the active group having overall higher valence ratings across all VEs (5.19±0.92 vs 4.70±0.90). Importantly, the interaction between the factors VE and group was also significant. This interaction was caused by significantly higher (more positive) valence ratings of the active group in the Positive VE compared to the passive group (t(280.321)=6.943, $p<.001$). There were no group differences for the neutral (t(289)=0.650, $p=.516$) and negative VEs (t(270.001)=0.847, $p=.398$).

Arousal Ratings. The mixed ANOVA results revealed a main effect of VE. Posthoc paired t-tests showed that arousal ratings were significantly higher for positive than for neutral VEs (t(290)=17.057, $p<.001$), and for negative than for neutral VEs (t(290)=24.584, $p<.001$) and negative than for positive VEs (t(290)=9.071, $p<.001$). Secondly, the main effect of group was significant, with the active group having overall higher arousal ratings across all VEs compared to the passive group. The interaction between the factors VE and group was not significant.

C. VALENCE AND AROUSAL CASR RATINGS PER VE

The CASR ratings were first averaged across events of each VE, and they are displayed in Fig 4. The results from the mixed 3 × 2 ANOVA are displayed in Table 2.

CASR Valence Ratings. The mixed ANOVA results revealed a main effect of VE, meaning that valence ratings were significantly higher (more positive) for positive then for neutral VEs (t(289)=18.671, $p<.001$), for neutral than for negative VEs (t(289)=14.255, $p<.001$) and for positive than for negative VEs (Z=13.560, 28.127, $p<.001$). Secondly, the main effect of group was significant, with the active group having higher valence ratings across all VEs (5.26±1.85 vs 4.86±1.27). The interaction between the factors VE and group was also significant, with post-hoc paired t-tests indicating significant valence rating difference between all three Vs (all t≥8.585, all p<.001). The groups only differed in their valence ratings for the positive VEs (t(256.460)=7.423, p<.001), not for neutral (t(272.263)=1.492, p =.131) and negative VE (t(289)=.426, p =.670). These findings show that the post-VE and the CASR valence ratings showed a similar pattern of findings.

CASR Arousal Ratings. The mixed ANOVA results revealed significant main effect of the VE category. Post-hoc t-tests showed that arousal ratings were significantly higher for positive and negative than for neutral VEs (Both t(289)>=18.670, $p<.001$), but there was no difference between the positive and negative VEs (t(289)=2.163, $p=.093$, not significant after Bonferroni correction), which is different from the post-VE arousal ratings. Secondly, the main effect of group was significant, meaning that the active group had higher arousal ratings across all VEs compared to the passive group (4.86±1.30 vs 4.39±1.41). Finally, the interaction between the factors VE and group was significant which is again different from the post-VE arousal ratings. For the passive group, post-hoc t-tests revealed

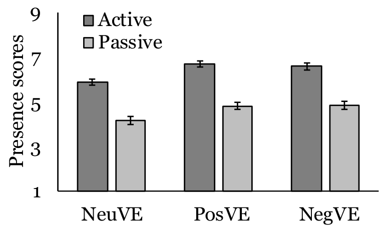

FIGURE 8. **Presence scores for active and passive group. Errors bars are representing standard errors (SE).**

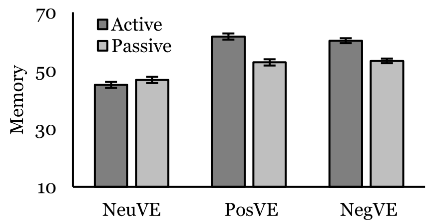

FIGURE 9. **Memory Recognition Scores. The mean scores expressed in percentages (%) (calculated across all events for each VE) showed overall average score across VEs and Memory recognition (accuracy in percentage) per group for each VE. Errors bars are representing standard errors (SE).**

significant arousal rating differences between all VEs (all t's(138)≥4.790, *p's<.001).* In the active group, however, arousal ratings were only higher for the positive compared to the neutral VE (t(138) = 13.673, *p<.001*) and for the negative compared to the neutral VE (t(138)=11.613, *p<.001*). Furthermore, independent post-hoc t-tests comparing both groups for each VE separately. These showed that arousal ratings were higher in the active compared to the passive group for the positive VE (t(289)=5.261, *p<.001*) but not for the neutral VE (t(289)=2.394, *p=.050* and negative VE (t(289)=0.391, *p=.554*). In summary, arousal ratings differed between the post-VE and CASR ratings which will be addressed in the discussion.

CASR ratings per event. Figure 5 shows the mean valence and arousal ratings for all events. The events per group are clustered by outline colours: blue for the Active group ('A') and orange for the passive group ('P'). Figure 7 shows that the event ratings follow the expected distribution within the AV space as in AVEL library [34], with the exception of the passive group ratings for the positive events (reduced valence and arousal), which explains the positive valence rating differences between the groups. As you can see in the figure the affective ratings have generally higher arousal for affective events than for neutral events resulting in the typical where the so-called 'V-shape curve' within the cartesian space [55].

D. EXPRESSIVITY AND ALEXITHYMIA

Further, the effect of subjective factors of expressivity and alexithymia level on self-reported scores on valence and arousal ratings were analysed. The CASR ratings (instead of the post-VE ratings) were used for this analysis.

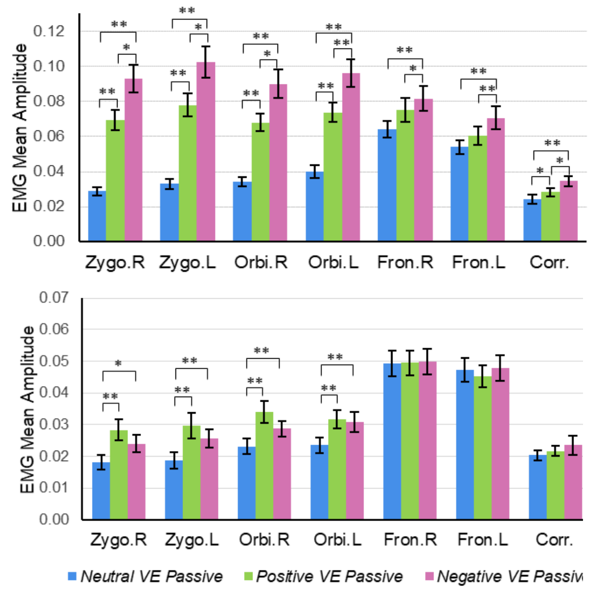

FIGURE 10. **EMG results (mean RMS) for the active group (top figure) and passive group (bottom figure). VE: neutral (blue), positive (green) and negative (pink). Error bars represent SE. Channels showed: Zygomaticus Right/Left (Zygo.R/L), Frontalis (Fron.), Orbicularis (Orbi.), Corrugator (Corr.). Significant pairwise comparisons signified with ∗ for p<.05, and ∗∗ for p<.01.**

TABLE 3. **3 × 2 Mixed ANOVA on EMG channels.**

EMG Channel	Main Effect Group	Main Effect VE	Interaction VE x Group
Combined	(***p***<.001**)	(***p***<.001**)	Sign (***p***<.001**)
Channel 1	(***p***<.001**)	(***p***<.001**)	Sign (***p***<.001**)
Channel 2	(***p***<.001**)	(***p***<.001**)	Sign (***p***<.001**)
Channel 3	(***p***<.001**)	(***p***<.003**)	Sign (***p***<.006**)
Channel 4	(***p***<.006**)	(***p***<.001**)	Sign (***p***<.001**)
Channel 5	Sign(***p***<.001**)	(***p***<.001**)	Sign (***p***<.001**)
Channel 6	(***p***<.001**)	(***p***<.001**)	Sign (***p***<.001**)
Channel 7	(***p***<.006**)	(***p***<.001**)	(*p*=.065)

* significance level *p* <.05, ** significance level *p* <. 01

Expressivity. Mixed ANOVAs with factors VE (positive, vs negative vs neutral) and expressivity group (scores calculated based on [45] and divided in low vs high groups; cut off = 80) were performed on valence and arousal CASR ratings, separately. For the valence ratings, the ANOVA showed significant main effect of VE, as expected ($F(2{,}578)$=.459.61, *p<.001*). No significant main effect on expressivity group was found (F(1,289)=.089, *p=.765*). However, there was a significant interaction between both factors ($F(2{,}578)$=3.27, *p=.039*). A subsequent planned independent t-test indicated that valence ratings were less different between the positive and negative VE conditions for the low compared to the high expressivity group ($t(289)$=2.21, *p=.028*), meaning that the low expressivity group has less spread in valence ratings (Fig 6). For the arousal ratings, again the ANOVA showed significant main effects on VE (F(2,578)=.311, *p<.001*). Also, no significant effects on expressivity group were found (F(1,289)=.103, *p=.103*), and nor on the interaction both factors (F(2,578)=.311, *p=.732*). This analysis

TABLE 4. 3 × 2 Mixed ANOVAs findings for the PPG features.

PPG Feature	Main Effect Group	Main Effect VE	Interaction VE x Group
PR (rBPM)	**(*p*<.001**)**	**(*p*<.001**)**	**(*p*=.024*)**
PRV (RMSSD)	**(*p*<.001**)**	n.s. (*p*=.417)	n.s. (*p*=.087)
RV (SDNN)	**(*p*<.001**)**	n.s. (*p*=.245)	n.s. (*p*=.094)

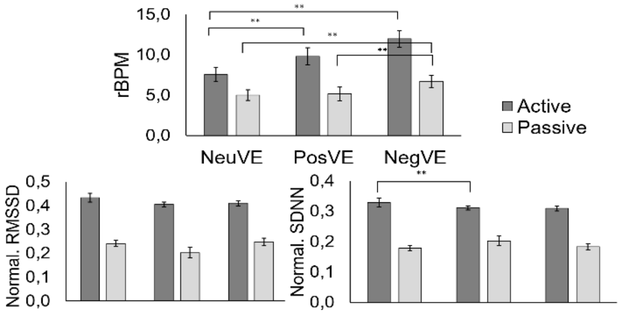

FIGURE 11. Mean rBPM, RMSSD, SDNN for each VE for the active and passive groups. Significant effects at <.05. SE Error bas.

shows that low expressivity reduces the spread of affective valence ratings. However, arousal ratings are not affected by expressivity. Please note, the mean expressivity scores for the 'Active' group (71.90±16.79) and 'Passive' group (74.71±13.43) were not significantly different (t(289)=1.59, *p=.115*).

Alexithymia. Mixed ANOVAs with factors VE (positive, vs negative vs neutral) and alexithymia group (low vs high; cut off = 51) were performed on valence and arousal CASR ratings, separately. For the valence ratings, the ANOVA showed the previously reported main effect of VE (F(2, 578)=460.42, *p<.001*)., More importantly, there was no significant main effect of alexithymia group (F(1, 289)=.637, p=.426)) and no significant interaction both factors (F(2, 578)=.587, *p=.556*).For arousal ratings, a similar picture emerged. Again, the ANOVA revealed the previously reported main effect of VE (F(2, 578)=235.19, *p<.001*), More importantly, there was no significant main effect of alexithymia group (F(1,289)=.067, *p=.796*) and no significant interaction both factors (F(2, 578)=.908, *p=.401*). This means that alexithymia scores did not influence CASR valence and arousal ratings (Fig. 7).

PRESENCE AND MEMORY RECOGNITION (Q1)

Presence.Mean presence self-rated scores were calculated for each VE across participants of the two group (see Fig. 8). A mixed 3 × 2 ANOVA with the factors VE category and Group (active vs passive) showed that mean presence ratings were significantly higher in the active compared to the passive group (6.47±1.42 vs 4.54±2.00, F(1, 289) = 88.863, *p*<.001). There was also a significant effect of VE (F(2,578) = 31.626, *p<.001*), showing that presence scores were higher for the negative and positive VEs compared to the neutral VE (*Both Z(pos/neg vs neu) >=*4.634, *p′s*<.001). groups (45.93±12.99 for the neutral VE, 57.01±13.47 for the positive VE, and 56.60 ±10.48 for the negative VE; See

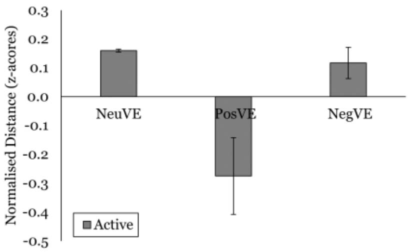

FIGURE 12. Normalised distance averaged across stimuli for each VE: Neutral, Positive, Negative. SE Error bars.

Fig. 9). The three VEs were found to have significantly effects on memory across all users(*F(2,578)=115.8, p<.001*). Post-hoc comparisons revealed significant differences between the positive and neutral VE, and between negative and neutral VE (*All t(290)>=11.21, p<.001*). The related ANOVA revealed a significant main effect of group on memory scores (F(1, 289) = 21.46, *p*<.001), meaning that the active group remembered the events better (55.48% on average across VES) than the passive group (50.93%). The interaction between VE category and group was also significant (F(1.88, 543) = 22.363, *p*<.001). Post-hoc t-tests showed higher memory scores in the active group for the negative and positive VEs (all t(289)'s ≥5.90, all p's <.001) compared to the passive group. Interestingly, significant correlations between memory recognition and presence scores were also found for the negative and positive VE (all r's≥.156, all p's ≤.008), thus confirming the effect of affective intensity on memory recognition and presence. Notably, our results indicate emotional intensity in VR enhances recognition memory, regardless of valence, and is further enhanced in active immersive VR.

E. FACIAL EMG ANALYSIS

We analysed the mean RMS amplitudes for all f-EMG channels separately by using mixed 3 × 2 ANOVAs with the factors VE category (neutral, positive, negative) and Group (active vs. passive group). Findings are shown in Table 3.

In short, for all channels there was a significant main effect of Group (all F's ≥ 7.614; all p's ≤.006) showing that EMG activity was always higher in the active compared to the passive group (see Fig. 10). In addition, there was always a significant main effect of the VE category (all F's ≥ 5.940; all p's ≤.003). Finally, there was also a significant interaction between the factors VE category and group (all F's ≥ 5.443; all p's ≤.006), with exception of the findings for EMG channel 7. Here, the interaction was only marginally significant (F(1.693, 484.124)= 2.901, p=.065) but it showed a similar pattern as seen for all other channels. Generally, EMG RMS amplitudes were more strongly modulated by the VE category in the active compared to the passive group.

See detailed description of the group post-hoc tests in the Supplementary Materials.

F. FACIAL PPG ANALYSIS

ANOVA analyses were conducted to determine whether PPG features reflecting pulse rate (BPM minus baseline BPM (rBPM)) and pulse rate variability measures (RMSSD, SDNN), can dissociate between VE categories for the active and passive group (see Fig. 11 and Table 4). For all features, the main effect of Group was significant (all F's $\geq$ 12.895; all ***p's*** **<.001**) meaning that higher values for the active compared to the passive group.

The main effect of VE category was only significant for the heart rate (rBPM; $F(2, 572) = 19.706$, ***p*<.001**). The interaction between VE category and group was only significant for the rBPM feature ($F(2, 572) = 3.774$, ***p*=.024**), meaning that the effect of VE category on rBPM was stronger for the active compared to the passive groups. This was further evaluated when conducting Friedman and post-hoc tests for the active and passive groups separately (see S2).

G. DISTANCE FROM EVENT STIMULUS IN ACTIVE CONDITION

In the active group we computed an additional measure, the distance (Dis) because participants of this group were able to walk towards or away from an event in the virtual space. The mean vector distance '*Dis*' was calculated between the normalised user's position and the event stimulus' position during an event interaction. From those, the average Dis across all events per VE was computed (see Fig. 12; neutral VE: 0.16±0.06, positive VE:-0.28±1.59, negative VE: 0.12±0.64). Negative value denotes that the user moved closer to an event (approach) and positive moving away from the event (avoidance). ANOVA's analysis showed that VE category was significant (*F(2,137)=8.61, p=.002*). Post-hoc tests revealed that the distance observed in the Positive VE was significantly decreased compared to both the Neutral VE ($t(139)=3.26$, *p<.001*) and Negative VE ($t(137) = -2.83$, *p=.003*). The difference between the negative and neutral VEs were not significant ($t(137)=.78$, $p=.22$). The results suggest that participants moved towards the stimuli presented in the Positive VE. The avoidance behaviour expected for the Negative VE was not observed to a significant level compared to the neutral VE, an observation that supports further the use of control stimuli VEs (as the NeutralVE) for such comparisons.

VII. DISCUSSION

Recent progress in VR technologies offers advanced immersive capabilities for room-scale 3D interactive environments while tracking the user's behaviour. These environments have the potential to evoke naturalistic behavioural and emotional responses, which can be recorded in well-controlled scenarios. Unfortunately, pre-validated affective VR environments and measurements for experimental usage are scarce and guidelines for robust experimental design are still underdeveloped [51]. The current study aims to further this development by comparing the usage of different affective self-ratings in conjunction with memory recognition tests, as well as behavioural and physiological measures in an interactive VR set up. This study developed several methodological improvements, and it is an example for comprehensive and robust feasibility testing of experimental VR designs, which advances methods usage in the VR field, and contributes towards the design of guidelines for experimental VR designs.

Examples for our design innovations are, firstly, we used an online validation of the scenarios before conduction the VR study [34]. Secondly, throughout the VE experience a continuous self-rating tool (CASR [36]) was used and compared to the frequently used single-point post VE affective rating scales. Thirdly, a custom gaze-based event detection system was implemented to link user responses to contextual information (events or objects) within the VR environment during the interactive condition and trigger subsequent audio-visual properties. Thus, when users looked at specific objects, the system logged event annotations containing timestamps, metadata about the object in focus, gaze coordinates, and the duration of the viewing interaction. The event-stimuli were activated by the user's interaction and gaze, and whose onset timestamps were recorded to allow for granular, event-based analysis (as seen in [34], [56]). Fourthly, a multimodal continuous electrophysiological wearable sensor setup was used in traditional seated VR settings but also in the room-scale configuration, offering insights into their applicability in mobile and interactive applications. Finally, we recorded the effect of affect and presentation mode on not only presence scores but also on cognition, i.e. memory accuracy.

To this end a large-scale VR study was conducted at the Science Museum in London with over 291 participants, which, to the authors' best knowledge, is one of the largest interactive VR experiments with a broad diverse sample to date. The semi-controlled museum setting introduced natural variability that enhanced ecological validity, aligning with recent efforts to move affective and physiological research beyond traditional laboratory environments [57].

The purpose of this study was to explore the effects of room-scale free-walking HMD-VR settings on users' affective responses. Notably, this is the first study to explore the affective impact of 3D fully immersive content in room-scale and seated conditions using an amalgamation of subjective and sensing data. For this, the experiment compared two distinct settings and modes of interaction: the 'active ', naturalistic room-scale VR configuration, and the 'passive', seated, vicarious experience VR mode. The 'active' VR experience design was exploration-based (for the three VE scenarios neutral, positive and negative) and, hence, it was dynamic and personalized for each user, resulting in a highly variable user journey within the virtual space. Such 'active' experiences can more closely simulate everyday conditions [8], [58] and thus allow the study of naturalistic behaviours. The usage of 'active' and 'passive' conditions allowed the investigation of the effect of interactive VR on self-ratings and physiological measures, e.g. the quality and sensitivity of physiological measures in interactive settings.

The affective VEs elicited affective responses which were recorded and validated via different data sources: affect self-ratings presence ratings, memory scores, as well as physiological and movement measures (fig. 3). For the physiological EMG and PPG measures, a novel, custom, unobtrusive, wireless sensor-configuration was used to ascertain the feasibility of detecting changes in both valence and arousal from distinct signal features.

Using this very large dataset, we confirmed the feasibility of the experimental design in both seated passive and interactive room-scale VR conditions. We demonstrated that clear differences between affective conditions can be measured in by the passive and interactive conditions, and that the enhanced level movements in the active condition did not compromise the sensitivity of the signals despite slightly lower signal quality. In fact, the active condition showed clearly stronger affective ratings across multiple measures than the passive condition. These findings were also reflected in presence scores and memory accuracy, showing a clear influence of affect on cognition.

We showed that continuous affective self-ratings can be successfully implemented during the experiment (and not afterwards, as often done in other studies ([5])), and that they are a valid alternative to post-VR ratings. These also allow event–based analysis with CASR self-ratings as ground truth. This can be better used in conjunction with continuous physiological measures.

Physiological PPG and EMG measures were recorded with wireless wearable sensors and showed clear group (active vs passive group) and affective state effects, again showing that they can be feasibly used in interactive conditions with even stronger effects. These findings are discussed in more detail below.

Affect induction differences between interactive room-scale vs passive seated VR-mode on affective self-ratings, expressivity and presence. As a reminder, affective valence and arousal ratings were recorded in three ways to compare their validity in the study. These were single-point post-VE self-ratings, CASR VE-ratings and event-based ratings. This comparison was conducted because CASR is better suited for high-temporal-resolution analysis in conjunction with continuous physiological signals than post-VE. Key findings were as follows:

Firstly, findings from the analysed post-VE and CASR self-ratings confirmed the ability of the three VEs and their events to induce the predefined variations of valence and arousal in VR, confirming our first hypothesis (**H1a**). Importantly, a clear dissociation between VEs, neutral, positive and negative was reflected in the self-reported data from both affective self-ratings (post-VE, and CASR). These effects were also observed in the presence scores, indicating that affect intensity is linked to presence (as [39], [59]) consistent with **H1c**.

Secondly, the mode of interaction showed differences between the two groups, in line with our second hypothesis (**H2b**). The self-ratings reflected that the passive group was less susceptible to the affective manipulations, reporting overall lower arousal ratings than the active group (fig. 4,5) reduced memory recognition [39] (fig. 8) and lower presence scores (fig. 7). Overall, the interaction modes showed significant effects on a person's experience with VR content, further supporting the use of active-VR interaction mode for VR mood induction.

Thirdly, we have shown that participants with lower expressivity (as measured with the expressivity questionnaire) showed less variation of affective self-ratings between positive and negative VEs (supporting **H1d**). Arousal ratings, however, were unaffected by expressivity. Alexithymia, measured separately, was also had no observable effect on either self-rating type.

Importantly, the continuous affect self-rating (CASR) method proved more sensitive than the single post-VE ratings, revealing additional difference between the two groups. Post-VE ratings across individuals showed that the negative VE was rated as the more arousing than the positive VE for both groups. This asymmetry was somewhat expected as suggested by the evaluative space model [33] as negative stimuli generate higher arousal responses than the equivalent positive ones [60], [61] due to the so-called *negative bias* [62].

Looking at the event-ratings, this asymmetry was not as prominent for the active group; the positive events were rated of similar intensity to the negative events. By comparison, the passive group's event ratings were characterised by overall lower arousal, reduced positive valence but also increased negative stimulation, regardless of the VE. This was surprising because all these events were pre-validated [6] in screen-based settings like the passive group's condition [34]. We however witnessed a stronger modulation of interactivity on self-reported affect scores, and disproportionate negative affective ratings in the passive mode in-situ study. This negative bias may be tentatively reflecting potential lack of engagement, habituation and boredom [61], [63] although no engagement measures were collected to confirm this. It is also possible that the passive group did not meet their expectations of the VR experience thus generating negative affective reactions (in line with the predictive coding theory [65, 7]).

Our active-VR CASR event-ratings showcase a clear V-shape 'boomerang' curve [54] considerably evident on ratings of the IAPS picture library, e.g. [65], showing that arousal reflects the intensity of valence, towards higher pleasure (positive) or displeasure (negative). The further supports **H2c**, that active VR facilitates naturalistic appraisal patterns. These results show that interactivity meaningfully shapes affective processing in immersive environments and underscore the need to explore additional moderators like user expectations and engagement.

In summary, the study demonstrated that the affective impact of VEs is closely linked to presence, with higher intensity of affect—both positive and negative—correlating with stronger feelings of immersion in the VR experience. It was shown that CASR and event-based ratings during user-driven interactive VR experiences are a valid and highly sensitive, context-aware alternative to post-VR assessments. In the future, we recommend using CASR and event-based ratings for more accurate and continuous affect measurement in VR studies, especially in conjunction with continuous physiological measures.

The effect of affect induction on cognition as measured by memory accuracy. Memory accuracy scores were significantly higher for the positive and negative VEs compared to the Neutral (similar to the presence scores), with the active group outperforming the passive group. Overall, emotional intensity in VR, regardless of valence direction (negative or positive), was found to enhance memory recognition, particularly in the active VR environments, with significant correlations between memory and presence for both positive and negative VEs. These findings support **H1b** and **H2d**, that both affective intensity and interaction mode enhance memory performance.

EMG and Heart Rate Measures Discriminate Affective Conditions. A key benefit of the CASR method is its compatibility with continuous physiological data, allowing moment-by-moment comparisons between subjective ratings and bodily signals. Our findings show similar affective effects mirrored in the physiological data recorded within the three VEs supporting the feasibility of recognising valence and arousal from physiological sensors (**H3a and H3b).**

The EMG activation from the seven facial sensors not only discriminated between the positive and negative conditions (esp. for the *active* group) but also between the affective conditions and the neutral one for both groups (table 3, fig.9). Similarly, PPG measures, which have been traditionally used for the detection of arousal responses, were sensitive to the affective VE and the group type. More specifically, the pulse rate was higher in the active group compared to the passive group. This was expected, because physical movement can directly affect the physiological signals (e.g., raise heart rate). More importantly, in the active group the pulse rate was enhanced in affective conditions compared to the neutral condition. This shows that PPG had good discriminatory power between high arousing VEs against the low arousing neutral VE, esp. in the active group, despite degraded signal quality by introducing undesired variations and artefacts in the signal [66], especially in upright positions [67] and due to motion-induced artifacts [68], [69]. Note, pulse rate variability was enhanced in the active compared to the passive condition but there were hardly any differences between affective VEs. These reduced effects for the PRV can be explained by the higher sensitivity to artifacts because the computation of PVR measures in VR is more sensitive to motion changes compared to rBPM and EMG features.

In summary, our study investigated the feasibility of the usage of physiological measures in an 'active' compared to a 'passive' condition. Most studies detecting affect outside VR record data from users in seated positions facing one direction (e.g. [70], [71], [72], [73]). This approach is sensibly chosen to enhance to signal quality albeit at the cost of reducing the ecological validity, limiting users, and contriving the study setting. In our study, despite physical movement and the inherent effects on the extracted EMG pulse-rate measures, it was clearly demonstrated that with appropriate filtering and dynamic *epoching* techniques, it is possible to extract EMG and PPG measures of wearable sensors in interactive immersive settings. These measures show clear differences between affective VEs, esp. in the interactive conditions, supporting **H3c** that physiological sensing can detect affective responses in active VR.

Distance from event-stimulus; a promising valence measure for room-scale VR. In the active group, distance analysis revealed that participants moved closer (approach) to events in the positive VE compared to the neutral and negative VE. Interestingly, avoidance behaviour in the negative VE was not significantly different from the neutral VE, suggesting that the neutral stimuli served as a useful control. This finding partly supports **H3d**, that approach/avoidance behaviour can indicate valence—but suggests avoidance may be more context-dependent in immersive VR. This may indicate that avoidance in VR is expressed differently, potentially influenced by the virtual environment's design or the nature of the negative stimuli, warranting further exploration into how avoidance manifests in immersive settings.

VIII. LIMITATIONS AND FUTURE DIRECTIONS

This study demonstrated the feasibility of using interactive, room-scale VR environments to induce and measure affective responses, leveraging continuous self-ratings and physiological data. However, there are some limitations and areas for future exploration.

First, the scarcity of pre-validated affective VR environments and comprehensive experimental guidelines remains a challenge, highlighting the need for further development in this field. Although the study introduced several innovative methods, including online scenario validation and continuous self-ratings (CASR), it also faced constraints related to the varying sensitivity and susceptibility of physiological measures and potential motion artifacts in different VR conditions. Motion-related artefacts were a known source of noise. Our study design intentionally relied on native room-scale locomotion via SteamVR's walk-in-place functionality to preserve naturalistic movement and evaluate affective responses under minimal technological mediation. Future work could explore alternative movement mechanisms and platform-based locomotion approaches which were beyond the scope of this work.

Second, physiological signals such as PPG and EMG can be highly sensitive to motion artifacts, particularly in active VR conditions. In our study, a strict exclusion protocol for motion contaminated datasets was implemented, which ensured signal quality but substantially reduced the available sample size. This underscores the need for more sophisticated motion correction strategies. Future studies could integrate inertial motion sensor data and/or apply offline motion artifact correction (e.g. with ICA methods) or use real-time correction algorithms to improve the robustness of physiological analyses, for wearable PPG sensors (see [66], [74], [75]).

Future research should address these limitations by refining the experimental designs and signal processing pipeline to improve overall signal quality. Studies should also expand the range of affective and contextual VR scenarios and incorporate state-of-the-art interaction technologies to increase ecological validity. Additionally, incorporating machine learning techniques could provide deeper insights into physiological data mapping and subjective traits affecting emotional responses. Moreover, exploring how avoidance behaviour manifests in VR, given its potential to be expressed differently in immersive environments compared to traditional settings, would be valuable.

Future research should build upon our study's findings, which not only provide a valuable library of experimental data and analysis but also outline steps towards a systematic approach for affective experiment analysis in VR. Our guidelines, specifically tailored for room-scale interactive settings, offer a foundation for developing robust experimental principles and advancing the field's understanding of affect detection in immersive environments.

IX. CONCLUSION

This study successfully demonstrated the potential of user-centred affect detection approaches within immersive VR environments, showcasing the advantages of integrating unobtrusive, miniaturized wearable sensors. By using these technologies, we established the feasibility of capturing continuous electrophysiological data and detecting affective responses in VR settings. Our findings reveal that interactive, room-scale VR environments are highly effective in inducing and measuring variations in emotional valence and arousal, with continuous self-ratings and physiological data providing valuable insights into user experiences.

The results underscore the superiority of active interaction modes over passive settings for affect elicitation. The active VR experience led to stronger affective ratings, higher memory scores, and enhanced presence compared to passive VR. This is supported by our event-based analysis of physiological signals and affective ratings, indicating that free-walking interactive VR experiences offer more robust affective stimulation than traditional seated experiences. The study also highlights the benefits of incorporating pre-validated 3D interactive events in VR, which can better manipulate emotional responses through naturalistic exploration.

Overall, this study underscores the feasibility of developing context-aware systems for affect detection in VR, which could significantly benefit future applications in entertainment, simulation training, and healthcare by offering more objective assessments of user emotions.

ACKNOWLEDGMENT

The authors would like to thank Dr. Charles Nduka, Dr. Graeme Cox, Dr. Mohsen Fatoorechi, Dr. Mahyar Hamedi, Andy Cleal, Dr. James Archer, Milky Asefa, Simon Clarke, and Claire Baert of Emteq Laboratories for their support before and during data collection. They also thank all the participants, the Science Museum, London, for hosting the study through the Live Science programme, and of course the team of research assistants led by Maya Perry.

REFERENCES

[1] C. J. Wilson and A. Soranzo, "The use of virtual reality in psychology: A case study in visual perception," *Comput. Math. Methods Med.*, vol. 2015, pp. 1–7, Apr. 2015, doi: 10.1155/2015/151702.

[2] R. M. Baños, C. Botella, M. Alcañiz, V. Liaño, B. Guerrero, and B. Rey, "Immersion and emotion: Their impact on the sense of presence," *CyberPsychol. Behav.*, vol. 7, no. 6, pp. 734–741, Dec. 2004, doi: 10.1089/cpb.2004.7.734.

[3] A. Chirico, P. Cipresso, D. B. Yaden, F. Biassoni, G. Riva, and A. Gaggioli, "Effectiveness of immersive videos in inducing awe: An experimental study," *Sci. Rep.*, vol. 7, no. 1, pp. 1–16, Apr. 2017, doi: 10.1038/s41598-017-01242-0.

[4] A. Chirico and A. Gaggioli, "When virtual feels real: Comparing emotional responses and presence in virtual and natural environments," *Cyberpsychol., Behav., Social Netw.*, vol. 22, no. 3, pp. 220–226, Mar. 2019, doi: 10.1089/cyber.2018.0393.

[5] J. Diemer, G. W. Alpers, H. M. Peperkorn, Y. Shiban, and A. Mühlberger, "The impact of perception and presence on emotional reactions: A review of research in virtual reality," *Frontiers Psychol.*, vol. 6, pp. 1–9, Jan. 2015, doi: 10.3389/fpsyg.2015.00026.

[6] J. Freeman, J. Lessiter, K. Pugh, and E. Keogh, "When presence and emotion are related, and when they are not," in *Proc. 8th Annual Int. Workshop Presence*, 2005, pp. 1–16.

[7] A. Halbig and M. E. Latoschik, "A systematic review of physiological measurements, factors, methods, and applications in virtual reality," *Frontiers Virtual Reality*, vol. 2, Jul. 2021, Art. no. 694567, doi: 10.3389/frvir.2021.694567.

[8] A. Gorini and G. Riva, "Virtual reality in anxiety disorders: The past and the future," *Expert Rev. Neurotherapeutics*, vol. 8, no. 2, pp. 215–233, Feb. 2008, doi: 10.1586/14737175.8.2.215.

[9] J. I. Gold, K. A. Belmont, and D. A. Thomas, "The neurobiology of virtual reality pain attenuation," *CyberPsychol. Behav.*, vol. 10, no. 4, pp. 536–544, Aug. 2007, doi: 10.1089/cpb.2007.9993.

[10] G. Riva, "Virtual reality in psychotherapy: Review," *CyberPsychol. Behav.*, vol. 8, no. 3, pp. 220–230, Jun. 2005, doi: 10.1089/cpb.2005.8.220.

[11] K. M. Malloy and L. S. Milling, "The effectiveness of virtual reality distraction for pain reduction: A systematic review," *Clin. Psychol. Rev.*, vol. 30, no. 8, pp. 1011–1018, Dec. 2010, doi: 10.1016/j.cpr.2010.07.001.

[12] D.-P. Pertaub, M. Slater, and C. Barker, "An experiment on public speaking anxiety in response to three different types of virtual audience," *Presence*, vol. 11, no. 1, pp. 68–78, Feb. 2002, doi: 10.1162/105474602317343668.

[13] M. Price and P. Anderson, "The role of presence in virtual reality exposure therapy," *J. Anxiety Disorders*, vol. 21, no. 5, pp. 742–751, Jan. 2007, doi: 10.1016/j.janxdis.2006.11.002.

[14] I. Alsina-Jurnet and J. Gutiérrez-Maldonado, "Influence of personality and individual abilities on the sense of presence experienced in anxiety triggering virtual environments," *Int. J. Hum.-Comput. Stud.*, vol. 68, no. 10, pp. 788–801, Oct. 2010, doi: 10.1016/j.ijhcs.2010.07.001.

[15] S. Bouchard, J. St-Jacques, G. Robillard, and P. Renaud, "Anxiety increases the feeling of presence in virtual reality," *Presence*, vol. 17, no. 4, pp. 376–391, Aug. 2008, doi: 10.1162/pres.17.4.376.

[16] S. Notzon, S. Deppermann, A. Fallgatter, J. Diemer, A. Kroczek, K. Domschke, P. Zwanzger, and A.-C. Ehlis, "Psychophysiological effects of an iTBS modulated virtual reality challenge including participants with spider phobia," *Biol. Psychol.*, vol. 112, pp. 66–76, Dec. 2015, doi: 10.1016/j.biopsycho.2015.10.003.

[17] A. P. Anderson, M. D. Mayer, A. M. Fellows, D. R. Cowan, M. T. Hegel, and J. C. Buckey, "Relaxation with immersive natural scenes presented using virtual reality," *Aerosp. Med. Human Perform.*, vol. 88, no. 6, pp. 520–526, Jun. 2017, doi: 10.3357/amhp.4747.2017.

[18] C. Crescentini, L. Chittaro, V. Capurso, R. Sioni, and F. Fabbro, "Psychological and physiological responses to stressful situations in immersive virtual reality: Differences between users who practice mindfulness meditation and controls," *Comput. Hum. Behav.*, vol. 59, pp. 304–316, Jun. 2016, doi: 10.1016/j.chb.2016.02.031.

[19] N. Vaughan, V. N. Dubey, T. W. Wainwright, and R. G. Middleton, "A review of virtual reality based training simulators for orthopaedic surgery," *Med. Eng. Phys.*, vol. 38, no. 2, pp. 59–71, Feb. 2016, doi: 10.1016/j.medengphy.2015.11.021.

[20] A. Borrego, J. Latorre, M. Alcañiz, and R. Llorens, "Comparison of oculus rift and HTC vive: Feasibility for virtual reality-based exploration, navigation, exergaming, and rehabilitation," *Games for Health J.*, vol. 7, no. 3, pp. 151–156, Jun. 2018, doi: 10.1089/g4h.2017.0114.

[21] S. Bronner, R. Pinsker, R. Naik, and J. A. Noah, "Physiological and psychophysiological responses to an exer-game training protocol," *J. Sci. Med. Sport*, vol. 19, no. 3, pp. 267–271, Mar. 2016, doi: 10.1016/j.jsams.2015.03.003.

[22] S. V. Adamovich, G. G. Fluet, E. Tunik, and A. S. Merians, "Sensorimotor training in virtual reality: A review," *NeuroRehabilitation*, vol. 25, no. 1, pp. 29–44, Aug. 2009, doi: 10.3233/nre-2009-0497.

[23] M. Bellani, L. Fornasari, L. Chittaro, and P. Brambilla, "Virtual reality in autism: State of the art," *Epidemiol. Psychiatric Sci.*, vol. 20, no. 3, pp. 235–238, Sep. 2011, doi: 10.1017/s2045796011000448.

[24] S. Parsons and P. Mitchell, "The potential of virtual reality in social skills training for people with autistic spectrum disorders," *J. Intellectual Disability Res.*, vol. 46, no. 5, pp. 430–443, Jun. 2002, doi: 10.1046/j.1365-2788.2002.00425.x.

[25] L. R. Valmaggia, L. Latif, M. J. Kempton, and M. Rus-Calafell, "Virtual reality in the psychological treatment for mental health problems: An systematic review of recent evidence," *Psychiatry Res.*, vol. 236, pp. 189–195, Feb. 2016, doi: 10.1016/j.psychres.2016.01.015.

[26] D. Västfjäll, P. Larsson, and M. Kleiner, "Emotion and auditory virtual environments: Affect-based judgments of music reproduced with virtual reverberation times," *CyberPsychol. Behav.*, vol. 5, no. 1, pp. 19–32, Feb. 2002, doi: 10.1089/109493102753685854.

[27] J. A. Russell, "A circumplex model of affect," *J. Personality Social Psychol.*, vol. 39, no. 6, pp. 1161–1178, Dec. 1980, doi: 10.1037/h0077714.

[28] A. J. Fridlund and J. T. Cacioppo, "Guidelines for human electromyographic research," *Psychophysiology*, vol. 23, no. 5, pp. 567–589, Sep. 1986, doi: 10.1111/j.1469-8986.1986.tb00676.x.

[29] J. T. Cacioppo, G. G. Berntson, C. J. Norris, and J. K. Gollan, "The evaluative space model," in *Handbook Theories Social Psychology*. Thousand Oaks, CA, USA: SAGE Publications Inc., Aug. 2012, pp. 50–73, doi: 10.4135/9781446249215.n4.28].

[30] G. Bingöl, S. Porcu, A. Floris, and L. Atzori, "QoE estimation of WebRTC-based audio-visual conversations from facial and speech features," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 20, no. 5, pp. 1–23, Jan. 2024, doi: 10.1145/3638251.

[31] Meta. *Face Tracking–Movement SDK*. Accessed: Aug. 7, 2025. [Online]. Available: https://developers.meta.com/horizon/documentation/unity/move-face-tracking/

[32] PICO. *Face Tracking–Developer Documentation*. Accessed: Aug. 7, 2025. [Online]. Available: https://developer.picoxr.com/

[33] L. Kulke, D. Feyerabend, and A. Schacht, "A comparison of the affectiva iMotions facial expression analysis software with EMG for identifying facial expressions of emotion," *Frontiers Psychol.*, vol. 11, pp. 1–9, Feb. 2020, doi: 10.3389/fpsyg.2020.00329.

[34] I. Mavridou, E. Balaguer-Ballester, C. Nduka, and E. Seiss, "A reliable and robust online validation method for creating a novel 3D affective virtual environment and event library (AVEL)," *PLoS ONE*, vol. 18, no. 4, Apr. 2023, Art. no. e0278065, doi: 10.1371/journal.pone.0278065.

[35] D. R. Herring, M. H. Burleson, N. A. Roberts, and M. J. Devine, "Coherent with laughter: Subjective experience, behavior, and physiological responses during amusement and joy," *Int. J. Psychophysiol.*, vol. 79, no. 2, pp. 211–218, Feb. 2011, doi: 10.1016/j.ijpsycho.2010.10.007.

[36] I. Mavridou, E. Seiss, M. Hamedi, E. Balaguer-Ballester, and C. Nduka, "Towards valence detection from EMG for virtual reality applications," in *Proc. 12th Int. Conf. Disability*, 2018, pp. 4–6. [Online]. Available: http://eprints.bournemouth.ac.uk/31022/

[37] I. Mavridou, E. Seiss, T. Kostoulas, M. Hamedi, E. Balaguer-Ballester, and C. Nduka, "Introducing the EmteqVR interface for affect detection in virtual reality," in *Proc. 8th Int. Conf. Affect. Comput. Intell. Interact. Workshops Demos (ACIIW)*, Sep. 2019, pp. 83–84, doi: 10.1109/ACIIW.2019.8925297.

[38] J. Bailey, J. Bailenson, A. Won, J. Flora, and K. Armel, "Presence and memory: Immersive virtual reality effects on cued recall," in *Proc. Int. Soc. Presence Res. Annual Conf.*, 2011, pp. 24–26.

[39] T. Sharot and E. A. Phelps, "How arousal modulates memory: Disentangling the effects of attention and retention," *Cognit., Affect., Behav. Neurosci.*, vol. 4, no. 3, pp. 294–306, Sep. 2004, doi: 10.3758/cabn.4.3.294.

[40] I. Mavridou, J. T. McGhee, M. Hamedi, M. Fatoorechi, A. Cleal, E. Ballaguer-Balester, E. Seiss, G. Cox, and C. Nduka, "FACETEQ interface demo for emotion expression in VR," in *Proc. IEEE Virtual Reality (VR)*, Mar. 2017, pp. 441–442, doi: 10.1109/VR.2017.7892369.

[41] M. Gnacek, J. Broulidakis, I. Mavridou, M. Fatoorechi, E. Seiss, T. Kostoulas, E. Balaguer-Ballester, I. Kiprijanovska, C. Rosten, and C. Nduka, "EmteqPRO—Fully integrated biometric sensing array for non-invasive biomedical research in virtual reality," *Frontiers Virtual Reality*, vol. 3, pp. 1–17, Mar. 2022, doi: 10.3389/frvir.2022.781218.

[42] E. J. Kim and J. Kim, "Virtual reflections: Self-mirroring and its influence on self-presence in virtual reality environments," *Vis. Comput.*, vol. 41, no. 11, pp. 8271–8283, Mar. 2025, doi: 10.1007/s00371-025-03868-y.

[43] S. Bouchard, G. Robillard, J. St-Jacques, S. Dumoulin, M. J. Patry, and P. Renaud, "Reliability and validity of a single-item measure of presence in VR," in *Proc. 3rd IEEE Int. Workshop Haptic, Audio Vis. Environ. Appl.*, Jun. 2004, pp. 59–61, doi: 10.1109/HAVE.2004.1391882.

[44] D. Leising, T. Grande, and R. Faber, "The Toronto alexithymia scale (TAS-20): A measure of general psychological distress," *J. Res. Personality*, vol. 43, no. 4, pp. 707–710, Aug. 2009. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0092656609000828

[45] J. Gross, O. John, and J. Richards. *Berkeley Expressivity Questionnaire*. Accessed: Sep. 16, 2019. [Online]. Available: http://www.midss.org/content/berkeley-expressivity-questionnaire

[46] Qualtrics, Provo, UT, USA. (2019). *Qualtrics*. [Online]. Available: https://www.qualtrics.com

[47] A. L. Simeone, I. Mavridou, and W. Powell, "Altering user movement behaviour in virtual environments," *IEEE Trans. Vis. Comput. Graphics*, vol. 23, no. 4, pp. 1312–1321, Apr. 2017, doi: 10.1109/TVCG.2017.2657038.

[48] M. Hamedi, S.-H. Salleh, K. Ismail, A. M. Noor, M. Astaraki, and H. Aslian, "Time-frequency facial gestures EMG analysis using bilinear distribution," in *Proc. IEEE Int. Conf. Signal Image Process. Appl. (ICSIPA)*, Kuala Lumpur, Malaysia, Oct. 2015, pp. 169–173, doi: 10.1109/ICSIPA.2015.7412184.

[49] S. Bhowmik, B. Jelfs, S. P. Arjunan, and D. K. Kumar, "Outlier removal in facial surface electromyography through Hampel filtering technique," in *Proc. IEEE Life Sci. Conf. (LSC)*, Dec. 2017, pp. 258–261, doi: 10.1109/LSC.2017.8268192.

[50] M. Vollmer, "A robust, simple and reliable measure of heart rate variability using relative RR intervals," *Comput. Cardiol.*, vol. 2025, pp. 609–612, Feb. 2015, doi: 10.1109/CIC.2015.7410984.

[51] R. T. Wilkinson and S. Allison, "Age and simple reaction time: Decade differences for 5,325 subjects," *J. Gerontology*, vol. 44, no. 2, pp. P29–P35, Mar. 1989, doi: 10.1093/geronj/44.2.p29.

[52] A. S. Vincent, T. Roebuck-Spencer, K. Gilliland, and R. Schlegel, "Automated neuropsychological assessment metrics (v4) traumatic brain injury battery: Military normative data," *Mil. Med.*, vol. 177, no. 3, pp. 256–269, Mar. 2012, doi: 10.7205/milmed-d-11-00289.

[53] D. L. Woods, J. M. Wyma, E. W. Yund, T. J. Herron, and B. Reed, "Factors influencing the latency of simple reaction time," *Frontiers Human Neurosci.*, vol. 9, pp. 1–12, Mar. 2015, doi: 10.3389/fnhum.2015.00131.

[54] P. Kuppens, F. Tuerlinckx, J. A. Russell, and L. F. Barrett, "The relation between valence and arousal in subjective experience," *Psychol. Bull.*, vol. 139, no. 4, pp. 917–940, Jul. 2013, doi: 10.1037/a0030811.

[55] M. Vasser and J. Aru, "Guidelines for immersive virtual reality in psychological research," *Current Opinion Psychol.*, vol. 36, pp. 71–76, Dec. 2020, doi: 10.1016/j.copsyc.2020.04.010.

[56] R. Warp, M. Zhu, I. Kiprijanovska, J. Wiesler, S. Stafford, and I. Mavridou. *Moved By Sound: How Head-Tracked Spatial Audio Affects Autonomic Emotional State and Immersion-Driven Auditory Orienting Response in VR Environments*. Accessed: Jun. 28, 2025. [Online]. Available: www.aes.org/tmpFiles/elib/20250628/21703.pdf

[57] G. Vigliocco, L. Convertino, S. De Felice, L. Gregorians, V. Kewenig, M. A. E. Mueller, S. Veselic, M. Musolesi, A. Hudson-Smith, N. Tyler, E. Flouri, and H. J. Spiers, "Ecological brain: Reframing the study of human behaviour and cognition," *Roy. Soc. Open Sci.*, vol. 9, no. 11, pp. 1–12, Nov. 2024, doi: 10.1098/rsos.240762.

[58] E. F. Risko, D. C. Richardson, and A. Kingstone, "Breaking the fourth wall of cognitive science: Real-world social attention and the dual function of gaze," *Current Directions Psychol. Sci.*, vol. 25, no. 1, pp. 70–74, Feb. 2016, doi: 10.1177/0963721415617806.

[59] A. Sutcliffe, B. Gault, and J.-E. Shin, "Presence, memory and interaction in virtual environments," *Int. J. Hum.-Comput. Stud.*, vol. 62, no. 3, pp. 307–327, Mar. 2005, doi: 10.1016/j.ijhcs.2004.11.010.

[60] J. Heinström, "Negative affectivity—The emotional dimension," in *From Fear To Flow*. Amsterdam, The Netherlands: Elsevier, 2010, pp. 75–103, doi: 10.1016/b978-1-84334-513-8.50007-8.

[61] J. T. Cacioppo, J. T. Larsen, N. K. Smith, and G. G. Bertson, "The affect system: What lurks below the surface of feelings?" in *Proc. Feelings Emotions Amsterdam Conf.*, 2004, pp. 223–242. Accessed: Jan. 29, 2020. [Online]. Available: https://psycnet.apa.org/record/2004-95227-014

[62] C. J. Norris, "The negativity bias, revisited: Evidence from neuroscience measures and an individual differences approach," *Social Neurosci.*, vol. 16, no. 1, pp. 68–82, Dec. 2019, doi: 10.1080/17470919.2019.1696225.

[63] J. D. Eastwood, A. Frischen, M. J. Fenske, and D. Smilek, "The unengaged mind: Defining boredom in terms of attention," *Perspect. Psychol. Sci.*, vol. 7, no. 5, pp. 482–495, Sep. 2012, doi: 10.1177/1745691612456044.

[64] G. Stefanics, J. Heinzle, A. A. Horváth, and K. E. Stephan, "Visual mismatch and predictive coding: A computational single-trial ERP study," *J. Neurosci.*, vol. 38, no. 16, pp. 4020–4030, Apr. 2018, doi: 10.1523/jneurosci.3365-17.2018.

[65] N. Acharya and N. Mithulananthan, "Locating series FACTS devices for congestion management in deregulated electricity markets," *Electr. Power Syst. Res.*, vol. 77, nos. 3–4, pp. 352–360, Mar. 2007, doi: 10.1016/j.epsr.2006.03.016.

[66] K. T. Sweeney, T. E. Ward, and S. F. McLoone, "Artifact removal in physiological signals—Practices and possibilities," *IEEE Trans. Inf. Technol. Biomed.*, vol. 16, no. 3, pp. 488–500, May 2012, doi: 10.1109/TITB.2012.2188536.

[67] S. Lu, H. Zhao, K. Ju, K. Shin, M. Lee, K. Shelley, and K. H. Chon, "Can photoplethysmography variability serve as an alternative approach to obtain heart rate variability information?" *J. Clin. Monitor. Comput.*, vol. 22, no. 1, pp. 23–29, Jan. 2008, doi: 10.1007/s10877-007-9103-y.

[68] J. Harju, A. Tarniceriu, J. Parak, A. Vehkaoja, A. Yli-Hankala, and I. Korhonen, "Physiological measurement related content monitoring of heart rate and inter-beat intervals with wrist plethysmography in patients with atrial fibrillation," *Physiol. Meas.*, vol. 39, no. 6, p. 84001, 2018, doi: 10.1088/13616579/aad2c0.

[69] A. Alqaraawi, A. Alwosheel, and A. Alasaad, "Heart rate variability estimation in photoplethysmography signals using Bayesian learning approach," *Healthcare Technol. Lett.*, vol. 3, no. 2, pp. 136–142, Jun. 2016, doi: 10.1049/htl.2016.0006.

[70] D. Wu, C. G. Courtney, B. J. Lance, S. S. Narayanan, M. E. Dawson, K. S. Oie, and T. D. Parsons, "Optimal arousal identification and classification for affective computing using physiological signals: Virtual reality stroop task," *IEEE Trans. Affect. Comput.*, vol. 1, no. 2, pp. 109–118, Jul. 2010, doi: 10.1109/T-AFFC.2010.12.

[71] M. Soleymani, M. Pantic, and T. Pun, "Multimodal emotion recognition in response to videos," *IEEE Trans. Affect. Comput.*, vol. 3, no. 2, pp. 211–223, Feb. 2011, doi: 10.1109/TAC.2011.37.

[72] S. Poria, E. Cambria, R. Bajpai, and A. Hussain, "A review of affective computing: From unimodal analysis to multimodal fusion," *Inf. Fusion*, vol. 37, pp. 98–125, Sep. 2017, doi: 10.1016/j.inffus.2017.02.003.

[73] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "DEAP: A database for emotion analysis ;Using physiological signals," *IEEE Trans. Affect. Comput.*, vol. 3, no. 1, pp. 18–31, Jan. 2012, doi: 10.1109/T-AFFC.2011.15.

[74] W. Wu and J. Lee, "Improvement of HRV methodology for positive/negative emotion assessment," Institute Comput. Sci., Social Informatics Telecommunications Eng. (ICST), Washington, DC, USA, Tech. Rep., 2009, doi: 10.4108/ICST.COLLABORATECOM2009.8296. [Online]. Available: https://eudl.eu/pdf/10.4108/ICST.COLLABORATECOM2009.8296

[75] S. Stankoski, I. Kiprijanovska, I. Mavridou, C. Nduka, H. Gjoreski, and M. Gjoreski, "Breathing rate estimation from head-worn photoplethysmography sensor data using machine learning," *Sensors*, vol. 22, no. 6, p. 2079, Mar. 2022, doi: 10.3390/s22062079.

IFIGENEIA MAVRIDOU (Member, IEEE) is currently an Assistant Professor with Tilburg University, The Netherlands, and the Coordinator of the Virtual and Extended Reality Laboratory, MINDLABS. She specializes in extended reality (XR), affective computing, and wearable physiological sensing. After receiving her EPSRC-funded Ph.D. from the Centre for Digital Entertainment, Bournemouth University, U.K., she worked in the industry on wearable sensor development for virtual reality headsets and as a consultant on physiological sensing for international academic and commercial projects. To streamline physiological research in XR, she has designed and led numerous studies and funded projects bridging academia and industry in U.K. and Europe. Her research interests include electrophysiological state recognition in real and immersive environments, with applications in training, industrial settings, entertainment, AI ethics, and societal impact.

EMILI BALAGUER-BALLESTER received the Ph.D. degree. He has been an Associate Professor with the Department of Computing and Informatics, Bournemouth University, since 2017. He has also been the Co-Champion of the Interdisciplinary Neuroscience Research Centre at this institution. Previously, he was a Senior Lecturer/a Lecturer, from 2012 to 2017, a Project Leader with the Bernstein Centre for Computational Neuroscience Heidelberg–Mannheim, Durstewitz Laboratory, from 2008 to 2012, a Postdoctoral Researcher with Denham Laboratory, Plymouth, from 2005 to 2008, and a Research Fellow of the Spanish Science Ministry, from 2001 to 2005. He supervised 13 Ph.D. students to completion. His expertise stands at the interface between *Computational Neuroscience* and *Machine Learning* (neural dynamics reconstruction, ensemble modeling, data-driven model discovery, auditory system models, and affect decoding) through collaborations with human and animal neurophysiology laboratories. His Ph.D. dissertation from the University of Valencia, in 2001, and received the Culture Institute "Juan Gil-Albert" (Spain) Prize. He is an Editor in two journals *PLOS Computational Biology* and *Frontiers in Computational Neuroscience*.

THEODOROS KOSTOULAS (Member, IEEE) is currently an Associate Professor with the Department of Information and Communication Systems Engineering, University of the Aegean. Before that, he was a Senior Lecturer with the Faculty of Science and Technology, Bournemouth University, U.K. Even before that, he was a Postdoctoral Researcher and a Lecturer with the Computer Vision and Multimedia Laboratory and the Swiss Center for Affective Sciences, University of Geneva, Switzerland. He has 20 years of experience in affective computing, multimodal interaction, and machine learning, particularly being active in the realization of more than ten multidisciplinary national and international research and research and innovation projects. Over these years, he has been working towards bringing together expertise from a variety of disciplines, including art, psychology, and medical domains, for the corresponding heterogeneous groups of end-users.

ELLEN SEISS received the Diploma degree (M.Sc. equivalent) in psychology from Humboldt University, Berlin, in 2000, and the Ph.D. degree in ''sensorimotor processing in Parkinson's and Huntington's disease'' from the University of Birmingham, in 2004. She has expertise in cognitive, clinical, and affective neuroscience, including affective computing. She has worked in the area of affective computing for the last eight years, investigating affect and cognition using experimental VR designs and affective measures, such as self-ratings and physiological measures (such as PPG, facial EMG, GSR, and EEG/ERP measures). She is a Founding Member of the Multimodal Immersive NEuro-sensing (MINE) Research Cluster, Bournemouth University, which combines cognitive, affective, and computational neuroscience with virtual reality (VR) and uses physiological and electroencephalography (EEG) technology within interactive virtual environments.