Total VREcall: Using Biosignals to Recognize Emotional Autobiographical Memory in Virtual Reality

KUNAL GUPTA, University of Auckland, New Zealand
SAM W. T. CHAN, University of Auckland, New Zealand
YUN SUEN PAI, Keio University Graduate School of Media Design, Japan
NICHOLAS STRACHAN, University of Auckland, New Zealand
JOHN SU, University of Auckland, New Zealand
ALEXANDER SUMICH, Nottingham Trent University, United Kingdom
SURANGA NANAYAKKARA, National University of Singapore, Singapore
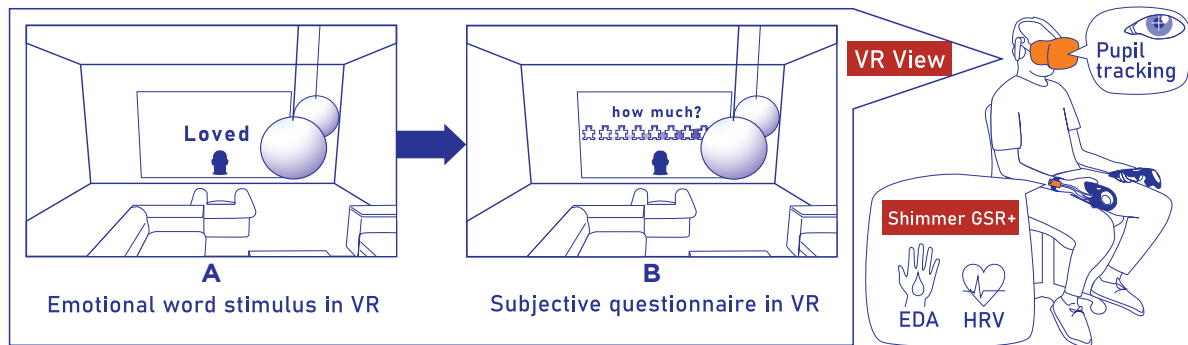MARK BILLINGHURST, University of Auckland, New Zealand

Fig. 1. A: Participant performing tasks in VR wearing Head Mounted Display and Physiological Sensors. B: Participant completing the Self-Assessment Manikin (SAM) subjective questionnaire in VR

Our memories and past experiences contribute to guiding our perception and action of future affective experiences. Virtual Reality (VR) experiences are more vividly memorized and recalled than non-VR ones, but there is little research on how to detect this recall in VR. We investigate the feasibility of recognizing autobiographical memory (AM) recall in VR using

Authors' addresses: Kunal Gupta, kgup421@aucklanduni.ac.nz, University of Auckland, 70, Symonds Street, Grafton, Auckland, Auckland, New Zealand, 1011; Sam W. T. Chan, sam@ahlab.org, University of Auckland, 70, Symonds Street, Grafton, Auckland, New Zealand; Yun Suen Pai, Keio University Graduate School of Media Design, 4 Chome-1-1 Hiyoshi, Kohoku Ward, Yokohama, Japan; Nicholas Strachan, nstr173@aucklanduni.ac.nz, University of Auckland, 70, Symonds Street, Grafton, Auckland, New Zealand; John Su, jsu678@aucklanduni.ac.nz, University of Auckland, 70, Symonds Street, Grafton, Auckland, New Zealand; Alexander Sumich, Nottingham Trent University, United Kingdom; Suranga Nanayakkara, National University of Singapore, 21 Lower Kent Ridge Rd, Singapore, Singapore; Mark Billinghurst, University of Auckland, 70, Symonds Street, Grafton, Auckland, New Zealand.

physiological cues: skin conductance, heart-rate variability, eye gaze, and pupillary response. We devised a methodology replicating an existing AM Test in VR. We conducted a user study with 20 participants recalling AM using three valence categories cue words: positive, negative, and neutral. We found a significant effect of AM recalls on EDA peak, and eye blink rate, with a generalized recognition accuracy of 77.1% and person dependent accuracy of up to 95.1%. This shows a promising approach for detecting AM recall in VR and we discuss the implications for VR experience design.

CCS Concepts: • **Human-centered computing** → **Empirical studies in HCI**; **Virtual reality**.

Additional Key Words and Phrases: Autobiographical Memory; Virtual Reality; Physiological Signals; Eye-tracking; Electro-dermal Activity; Biosignals; Emotion; Memory

1 INTRODUCTION

Autobiographical memories (AM) can be defined as the collection of past experiences remembered by people about their own life and are of self-relevance [15]. These memories are essential in forming individual's identity [6] and contribute as the directive function, helping in educating current and future activities and decisions based on past experiences [82]. For example, memories associated while remembering your first breakup, your visit to Taj Mahal when you were 8, or presenting your work at the Ubicomp conference.

AM recall lead to re-experiencing past events involving emotions, intention, thoughts, and actions [35]. In addition, AM recall advances mental health and well-being by supporting crucial elements of emotional and cognitive functioning like emotion regulation, positive mood, and feelings of the meaning of life [47, 55, 84]. Usually, the recall of more personal autobiographical memories results in higher empathic concern, the component of empathy which is focused on the comprehension of others' emotional states including feelings of sympathy, compassion, and warmth [80]. Previous research demonstrated that AM recall is an appropriate emotion elicitation method for the emotion recognition study [28, 58]. These research indicates that detecting AM recall and its associated emotions could potentially augment empathy, enable the design of better personalized human-computer interactions, and overall improve the user experience.

Emotional autobiographical memories recall has displayed reliable activation of the Autonomic Nervous System of our body [67], responsible to regulate involuntary physiological processes such as heart rate, respiration, pupil responses, blood pressure, and sweat responses. These physiological changes are measured using electrodermal activity (EDA) for sweat responses, heart rate variability (HRV) for heart rate, pupillometry sensing (eye response and gaze behavior) for pupil responses during recall. Having demonstrated a link between change in physiology and self-reports of emotional AM characteristics as compared to the neutral AM [20, 24, 67], the notion of using machine learning techniques to find physiological patterns of emotional AM seems promising.

With the advances in experiential technologies, immersive platforms such as virtual reality (VR) are suggested as a powerful tool to reproduce real, complex situations and environments, creating unique opportunities to explore human behavior and psychology in closely controlled conventional laboratory conditions [18]. A VR experiment enables verisimilitude (i.e. the extent to which experiment task replicates real-life) as well as its veridicality (i.e. the extent to which experiment results accurately indicate the psychological event) [32, 68]. Researchers have successfully investigated memory recall in VR using techniques involving various cues such as words [40], environmental lighting [44], objects [45], auditory cues [1], and visual cues [12]. Rivu et al. [63] demonstrated that stronger emotions can be induced in VR using AM, yet, there has been very little research on understanding how emotional AM recall responds in VR, its physiological implications and if it can be recognized. The knowledge of past experience and it's associated emotions could be helpful in designing more engaging and

personalized emotional VR experiences tailored to user's AM. This motivated us to ask the following research questions:

RQ1: What is the relationship between emotional autobiographical memory recall with elicited emotions in VR?

RQ2: How does the recalled emotional AM impacts the physiological signals such as EDA, HRV, and Pupil responses in VR?

RQ3: How can we detect if the experience being recalled is autobiographical memory using physiological signals?

RQ3a: How can we detect if the recalled autobiographical memory in VR elicits positive, negative or neutral valence emotional states?

In order to answer these RQs, we investigated the effects of emotional autobiographical memory recall on human physiology by conducting a traditional Autobiographical Memory Test (AMT) in a VR setting. As being one of the first studies focused on the attempt to detect AM recall and the emotions associated with it, we wanted to introduce the least number of independent factors such as images, videos, or other visual, and auditory cues that could affect the induced emotions. To get a better understanding through this early procedure, we chose to display an emotional word in the neutral valence virtual room since it is the least noisy and proven. Section 2 provides an overview of the related literature. In section 3, we discusses the experimental evaluation where we collected physiological signals such as electrodermal activity (EDA), photoplethysmography (PPG)/ heart rate variability (HRV), pupil responses, and eye movement from the participants while they were recalling their AM using positive, negative, and neutral valence words as stimuli cues from a database. The participants self-reported their subjective measures of emotion and recalled AM feedback using self-reported affective and memory experience questionnaires.

Section 4 explains the analysis methods for the AM recall experiences, collected physiological data, and machine learning workflow. When participants were recalling AM as compared to no recall, the analysis results show a significant main effect in physiological cues, such as EDA peak number (EDA-PN), EDA mean peak amplitude (EDA-PA), HRV Standard Deviation of NN intervals (SDNN), mean pupil diameter (mean PD), blink count, and mean blink duration. No significant difference was reported between positive, negative and neutral emotional words recall using physiological data. However, Personalized AM recall and emotions in AM recall (EAMR) were detected up to an accuracy of 95.1% and 84.1% respectively using machine learning techniques. Finally, these results are reported and further discussed in Section 5 and 6; and Section 7 proposes the potential applications of the system. Section 8 and 9 provides the limitations, future work, and concluding remarks.

To our knowledge, this work is among the first to model emotional AM recall in VR using physiological signals. The main contributions of this work are three-fold: (1) an empirical experimental paradigm to investigate emotional AM recall and its relationship with physiological information in VR using a simple text-based visual stimulus; (2) introduction of the analysis of physiological information such as EDA, HRV, and pupil responses as encouraging modalities to infer the emotional autobiographical memory recall status in VR; (3) a user study in VR to investigate the automatic recognition of emotional autobiographical memory recall status.

2 RELATED WORK

This section discusses how our work builds on previous research on emotional autobiographical memory with respect to physiological signals and virtual environments.

2.1 Autobiographical Memories and Emotions

In affective computing research, emotional experiences are generally examined using continuous dimensions primarily as valence and arousal responses [59, 64]. *Valence* dimension indicates the pleasantness or unpleasantness

of the experience and usually reported in terms of it's positive or negative emotional state. On the other hand, *arousal* dimension refers to the emotional intensity of the experience showing the level of arousal experienced by the person while going through the event. Several multi-disciplinary studies including psychology [38], computer science [61], and neuroscience [46, 66] have demonstrated that subjective experience of valence and arousal are linked to the physiological changes in the person while they experience the emotional events [37].

Autobiographical memories (AM) includes both emotionally intense events and marginal episodes as a recollection of self-relevant and meaningful experiences in one's own life [15, 16]. Notably, an influence on the memory encoding and subsequent phenomenological characteristics during recollection has been shown by the emotional characteristics associated with the event [48]. Previous AM research suggested that highly emotional experiences tends to be better remembered with more sense of reliving and re-experiencing by the individuals as compared to the events devoid of emotion or of neutral emotional states [13]. Numerous studies have found that retrieval of central details in an autobiographical memory was enhanced [9] and vivid [2] when the memories were highly arousing . Researchers also found a greater recall of peripheral and sensory details of positive memories [17, 78, 79] whereas negative memories correlate with an enhanced recall of better accuracy [77] and central features [51, 62].

Previous works in the field of mnemonics or memory research have investigated memory recall in a conventional laboratory setting using various paradigms such as inducing micro-events before the memory recognition tests [7]. However, this challenges the ecological validity of the test, especially for self-relevant past experience recall [49], so VR environments could be a good option for memory studies [56, 57]. There is an ample amount of research focused on understanding and exploiting AM in VR [34]. For example, the recollection stimulation of autobiographical memory in VR has been suggested to be effective for the early-stage Alzheimer's patients' reminiscence rehabilitation therapy [5, 71, 87]. Mira et al. [52] suggested that the use of positive specific memories enhances mindfulness trait with an objective to promote highly adaptive emotion regulation. When participants wrote their positive AM while experiencing negative emotions and vice versa in VR, their reported general wellbeing was reported elevated [36]. The primary explanation for this insight was positive mood being able to update previous experience's emotional appraisal by developing an adaptive perspective of the past. Javier demonstrated the VR procedure to be a promising tool to activate positive mood consequently augment positive emotion regulation approaches in the short term [21].

Insights from these studies provide strong evidence suggesting a connection between AM and emotions in VR. However, the current literature on the effects of the emotional AM recall on elicited emotional dimensions (valence and arousal) in VR is still underdeveloped and would require further examination. Thus, to properly investigate the relationship between emotional AM recall in VR and perceived emotions, we replicated conventional laboratory based well-cited Autobiographical Memory Test (AMT) in VR.

2.2 Emotional Autobiographical Memory and Physiological Signals

According to Conway et al. [15], autobiographical memories (AM) consists of emotional content providing a meaningful and significant information during its recall. The neural correlation of AM recall process has been extensively examined through neuropsychological and neuroimaging studies using Electroencephalogram (EEG), and Functional Magnetic Resonance Imaging (fMRI) in various contexts such as in empathy [50], in depression [85, 88, 89], and dementia [23, 43, 81]. Cabeza et al. [8] linked the contribution of emotion and vividness of AM to the amygdala, and visual cortex of the brain respectively. Hur et al.'s fMRI study reported enhanced cognitive and physiological processing in people with social anxiety order through a VR-based exposure therapy where they were asked to report if the emotional words shown to them were self-relevant [29]. Even though fMRI provides a deeper understanding of AM recall process but due to non-mobile, large and complex setup, it becomes challenging to use it for mobile and interactive VR applications. On the other hand, setting up EEG requires a

specialist researcher and close proximity to the participant. In order to comply with COVID-19 pandemic social distancing restrictions, we chose to opt for non-invasive sensors that can be setup by the participants themselves through remote assistance and left EEG for future studies.

The relationship between emotional states and changing physiological responses has been well studied, especially using electrodermal activity (EDA), heart rate variability (HRV), respiration, skin temperature, electromyography (EMG), electroencephalogram (EEG), and pupil responses [3, 26]. This led to many studies specifically investigating connection between AM and emotional states using physiological measures [20, 65, 67]. Electrodermal Activity (EDA) and facial expressions have been used to investigate the emotional biases with the vivid elaboration and accessibility of AM using music-based retrieval cues [74]. While examining cardiac activities during AM recall in response to emotional words, Labouvie-Vief et al. [39] reported variations in physiological arousal. Emotional AM recall in an event narration task was investigated using EDA and HRV [67] and reported that the emotional memories were vividly narrated but not specific in detail as compared to the neutral ones. They reported that the recollection of emotional AM (with positive and negative valence) elicited an increase in mean EDA and decrease in HRV measures compared to AM with neutral valence. Furthermore, the mean EDA and HRV were equivalent (no significant difference) between positive and negative valence conditions. A study reported an increase in pupil diameter during AM recall but no significant effect in differentiating emotional AM from neutral ones [20]. These studies provide substantial evidence that the emotional AM recall accompanies with physiological changes, however these studies are conducted in conventional laboratory setting and not in VR.

There have been few research trying to validate the efficacy of autobiographical memory recall in VR. By comparing 2-D desktop video and the 360-degree videos in VR, Benjamin et al. [69] demonstrated that VR contributes to an autobiographical associative network and enhances the AM recall. Later, Kisker et al. used physiological information such as EEG to investigate the traditional laboratory-based and 3D-360 degree VR AM retrieval process and suggested that events retrieved in laboratory might not reflect real-life memory mechanism [33]. Neilson et al. [54] used EDA as a physiological measure to investigate the anxiety level by experiencing VR traumatic event consists of real-life involuntary autobiographical memories. The authors found an increase in EDA during event's most dramatic part and its correlation with emotional reactions and memories suggested it to be potential measure.

From past research, we can see that VR plays a crucial role in emotional AM retrieval and a combination of EDA, HRV, and pupil responses could provide useful insights about emotional AM recall. However, understanding the role of emotional AM these physiological responses specifically in VR is an under explored area. Therefore, we aim to fill this research gap by measuring emotional state during AM recall using physiological changes in an immersive and interactive 3D VR space and further extending it to detect this phenomenon with an objective to bring out its applicability to VR experiences.

2.3 Hypotheses

Based on literature review, we formulated the following hypotheses:

H1: Recalling emotional autobiographical memories in virtual reality elicits similar perceived emotional responses.

H2: Positive, negative, and neutral autobiographical memory recall in virtual reality have a significant effect on the physiological emotional states measured using electrodermal activity (EDA), heart rate variability (HRV), and pupil responses.

H3: The recall of emotional autobiographical memories can be reliably classified into binary-label (recall/no-recall), and multi-label (positive/ negative/ neutral) classes using physiological information such as EDA.

3 EXPERIMENTAL EVALUATION

In this research, we set out to understand how emotional autobiographical memory recall in VR impacts physiological signals using positive, negative, and neutral valence words. For this, we used a within-subject user study design with a total of 15 words shown in a randomized order with respect to the three valence categories i.e. positive (SAM Valence ratings > 4 out of 7), negative (SAM Valence ratings < 3 out of 7), and neutral (SAM Valence ratings between 3-4 out of 7). In order to reduce the carry-forward effect, we asked the participants to relax for 30 seconds before the start of each trial which is longer than '4 seconds', the time interval required to wear off the effect of emotions [19]. We decided on 30 seconds for the relax phase same as time provided for AM recall to keep the data points consistent for analysis. Later, the data collected during this relaxation phase was labelled as 'no AM recall'. We describe the system design, and the procedure in more detail below in this section.

3.1 System Design

The system primarily consists of an HTC Vive Pro Eye[1] Head Mounted Display (HMD), Shimmer3 GSR+[2] sensor for measuring EDA and HRV, and SRanipal SDK[3] to track eye-movements as shown in Figure 2.

The VR environment was developed in the Unity 3D game engine, with two main functions [4]. Firstly, we replicated the laboratory setup used by Haj et al. [20] where the participants were placed in a virtual room with white walls and asked to sit on a chair placed 40cm from one of the walls where the cue words were displayed (visualised in figure 1). The other function of the VRE was to collect EDA, eye gaze, and Photoplethysmography (PPG) physiological signals from the Shimmer3 and Vive eye tracker at a sampling rate of 128 Hz and 60 Hz respectively using the Lab streaming Layer (LSL) protocol while the participant was performing the main task.

For the main task, we implemented a modified Autobiographical Memory Test (AMT) [86] where the participants recalled any self-relevant past experience associated with the displayed cue word. We selected five emotional words each for neutral (hard work, election, huge, fork, and fast), positive (happy, energetic, loved, successful, and brave), and negative (hopeless, lonely, sad, afraid, and angry) valence emotions from previous emotional AMT research [31, 53]. For the subjective measures, we used the self-assessment manikin (SAM) [4] scale, consists of two 9-point sub-scales to get the perceived valence (rating 1: Unpleasant; 5: Neutral; 9: Pleasant), and arousal (rating 1: Calm; 5: Neutral; 9: Excited) ratings and Memory Experiences Questionnaire - Short Form (MEQ-SF) [41] to collect seven subjective AM characteristics such as vividness, accessibility, time perspective, sensory, visual, emotion, and valence. These questionnaires were integrated into the VRE both to avoid any negative impact on presence [70] and for the participants' comfort, which may effect their physiological state.

3.2 Participants

22 participants between the age of 20 to 36 (8 Females, 14 males, mean age = 27, SD age = 4) were recruited and screened for neurological or psychological disorders. This sample size was determined by considering the experimental design involving two independent variables (i.e. AM Recall and Emotional Words) with five levels measured within subjects (i.e. AM recall, No AM recall, positive, negative, and neutral). All of the participants had normal or corrected to normal vision to ensure sharp visibility on the HMD, and had experienced VR environments more than once. Two participants were excluded to avoid any negative mood-induced biases as they reported negative affect to Positive and Negative Affect Schedule (PANAS) questionnaire [83]. We attained ethics approval and participants' consent, and informed them that they could withdraw from the experiment anytime they want.

---

[1] https://www.vive.com/us/product/vive-pro-eye/overview/
[2] https://www.shimmersensing.com/products/shimmer3-wireless-gsr-sensor
[3] https://developer.vive.com/resources/vive-sense/sdk/vive-eye-and-facial-tracking-sdk/
[4] https://unity.com/

Fig. 2. The Experiment Setup: Participant in VR Lab, wearing HTC Vive Pro Eye VR HMD, and Shimmer3 GSR+ sensor recalling AM associated with the displayed word in VRE
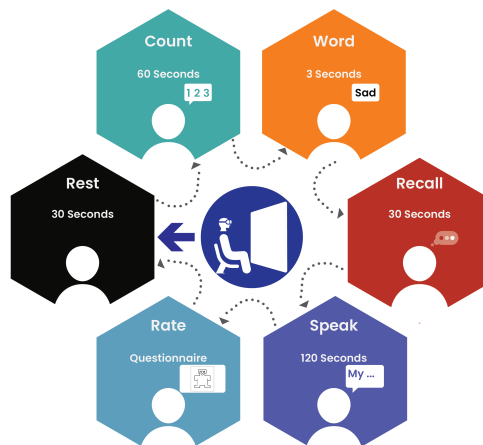

Fig. 3. Experiment Procedure: Participants wearing the VR HMD and Shimmer Sensor performing a total of 15 trials each in VR with the following sequence: rest (30 seconds), count(60 seconds), word (3 seconds), recall (30 seconds), speak (120 seconds), and rating subjective questionnaire.

3.3 Procedure

The participants were briefed about the experiment protocol on their arrival by the moderator and were asked to complete the pre-experiment questionnaire consisting of demography, past VR usage, history of epilepsy, and PANAS questionnaire. Next, the moderator confirmed with the participants if they had washed their hands in order to get clean EDA data and then asked them to put on the Shimmer3 GSR+ sensor on the non-dominant

hand and Vive HMD by showing the procedure from distance. Once the participant was ready, they were given a sample task to get familiar with the main task and VR environment.

We followed the procedure (shown in figure 3) with each trial consisted of four phases: relaxation, counting, recalling, and speaking. During the "relaxation" (No AM Recall) phase, participants were asked to relax for 30 seconds where they should not think about anything and keep their mind as clear as possible. This concluded with a bell sound after which they had to start counting as part of the "count" phase for the next 60 seconds which was indicated by another bell sound. Next, the cue word appeared for 3 seconds and they were asked to recall any past experience associated with that word for the next 30 seconds as part of the "recall" phase. The moderator instructed the participant to keep their dominant hand's thumb on the touchpad of the Vive controller when they are recalling the memory to tag the actual recall event. After that 30 seconds indicated by a bell sound, the "speak" phase was started where participants had to narrate their recalled experience for 120 seconds which was audio-recorded. At the end of each trial, they were asked to report their subjective experiences related to the recalled memory using the SAM scale, and the MEQ-SF in VR.

4 METHOD

In this section, we first describe the methods used to analyse recalled experiences, and physiological signals. Next, we introduce the machine learning workflow including data preparation, feature selection, classification algorithms, model evaluation, parameter tuning, and training procedures.

4.1 Analysis of Recalled Experiences

The audio recordings of the recalled memory experiences were transcribed automatically using otter.ai[5] and further edited manually to match the recordings. The transcripts were coded according to a coding scheme with variables of emotional valence (negative, positive, and neutral) and memory type (autobiographical and non-autobiographical). A transcribed experience was coded as non-autobiographical if it was not a specific experience and had no personal relevance (e.g., commentary on a witnessed event or other people). The coding scheme included two emergent codes from examining the transcripts: (1) if the experience contained 'mixed valence' (combination of more than one type of valence, e.g., negative and positive), and (2) if the experience had an overall valence that was 'inconsistent' with the intended valence of the cued word (e.g., experience coded as having negative valence for the neutral word, 'hardwork').

4.2 Physiological Analysis

First, any sudden motion peaks and general drift in EDA, PPG, and Pupil data were removed using visual inspection or an eye-balling technique. We used the Neurokit2 python package[6] for EDA and PPG processing and analysis. For EDA, we downsampled the data to 4 Hz, and decomposed the signals into fast-changing (Phasic) and slow-changing (Tonic) components [25]. Our analysis used the calculated total peak numbers (EDA-PN) and mean peak amplitudes (EDA-PA) of the EDA signals. We converted PPG data to HRV measures using Neurokit2. As our recall and no-recall data was for 30 seconds each, we considered only the measures relevant to ultra short-term HRV measurement norms (<5 minutes data) such as *RMSSD, SDNN, High Frequency relative power in normal units (HFnu)* (0.14—0.4 Hz).

For pupil responses and eye-movement analysis, we extracted measures related to pupil diameter and blinks. The pupil diameter was calculated from the eye tracker's "Pupil Size" data. We followed the pupil analysis pipeline for GazeR [22] and implemented it in Python [11]. The first step was "deblinking" the data by removing missing pupil size data (when values were -1.0) and removing data during low "Eye Openness" (when values were < 0.6).

---

[5]https://otter.ai/

[6]https://neurokit2.readthedocs.io/en/latest/index.html

Next, we applied a Hanning Window moving average filter (window size of 5) to smooth the data and applied cubic-spline interpolation. Finally, artefacts and outliers were filtered out using a Hampel filter (window size of 10, implementation code[7]) and a subtractive baseline correction was applied. Blink count and blink duration were calculated from the eye tracker's "Eye Openness" and "Timestamp" data.

4.3 Machine Learning Workflow

We extracted 12 EDA features (peak numbers (EDA-PN), amplitudes (EDA-PA), 10 statistical features (e.g., mean, median and maximum)) but by framing a sliding 10-seconds window data with 5-seconds overlapping to provide the most complete information possible to the machine learning model to predict accurately. We decided on sliding window size through hit and trial method to determine the smallest window size with best cross-validation accuracy. First, to understand if we can recognize when the participant was recalling AM, we considered a two-class "AM recall detection" problem with No AM recall (the state when no AM was recalled) and AM recall (the state when the participant saw the word and trying to recall self-relevant memory associated with the word) as classes. Then, to understand if we can recognize if the AM being recalled is positive, negative or neutral valence, we used three-class "Emotions in AM detection". Next, we considered this as a two-class and three-class recognition problem of differentiating between No AM recall (the state when no AM was recalled) and AM recall. We labelled the data with the target classes as 0-No AM recall and 1-AM recall for AM recall recognition. For Emotion in AM recognition, we labelled the target classes as 0-Neutral, 1-Positive, and 2-Negative, but modified the classes based on the analyzed recalled AM's emotion. For example, even if the cue word was of neutral valence, but if the recalled AM was positive, we assigned it as 1 for positive.

For feature reduction, we standardized the data and applied Principal Component Analysis (PCA) with a variance threshold to 95% as it helps in dimensionality reduction by selecting principal components that best characterises the dataset while keeping the information loss to the lowest. This also helped us in listing down the features that contributed to the relevant principle components with high variance. We randomly split our data into 70% training and 30% test sets per participant for personalized classifiers and from overall data for the generalized ones.

We evaluated the classification performance of 4 classifier algorithms namely, K-Nearest Neighbor (KNN), Random Forest (RF), Support Vector Classifier (SVC), Naive Bayes (NB) and lastly ensemble classifier (EC). For EC, we stacked a set of base-level estimators (KNN, RF, SVC, and NB) to get their average prediction but with reduced variance. For every machine learning algorithm, we trained two models, a generalized model that used the data from all the participants, and a personalized model that used the data from a specific participant. We selected the best model by performing 10-fold random stratified cross-validation on training data and used exhaustive grid search approach to identify the best hyper-parameters matching set.

5 RESULTS

This section reports on the results from the experiment data analysis into four parts. First, we report the results of our analysis of the recalled experiences. Then, we report on the subjective questionnaire data gathered from the participants. Next, we report the physiological measurements including EDA, HRV, and pupil responses. Finally, we describe our findings of detecting emotional autobiographical memory recall through machine learning algorithms.

Overall, our key findings are:
- Recalled experiences were mainly autobiographical (265 out of 287) with 22 having 'mixed valence' and 47 being 'inconsistent' with intended valence. Our analysis of physiological measures and machine learning algorithms were updated based on these results.

---

[7]https://towardsdatascience.com/outlier-detection-with-hampel-filter-85ddf523c73d

- Subjective SAM-Valence and MEQ-SF valence ratings corresponded to the intended valences with significant differences between all valences. There were significant differences in SAM-Arousal ratings between emotional (positive and negative valence) and neutral valences. Subjective MEQ-SF ratings showed significant difference in vividness between emotional (positive and negative valence) and neutral valences, in accessibility between negative and neutral valences, and in emotion ratings between all valences.
- Physiological measures: EDA peak number and amplitude were higher, HRV measures of RMSSD and HFnu were lower, pupil diameter and blink count were higher, while blink duration was lower in recall condition compared to no-recall condition.
- AM Recall Detection through machine learning attained generalized recognition accuracy of 77.1% and person-dependent accuracy of up to 95.1%. Emotion in AM Detection through machine learning attained generalized recognition accuracy of 50.3% and person-dependent accuracy of up to 70.7%.

5.1 Recalled Experiences

We collected 287 audio recordings of recalled memory experiences. Most of the experiences were autobiographical (n = 265). However, 22 were non-autobiographical, for example, *"...being loved is [a] very nice and powerful thing. Because it kind of gives you some solid ways to place [sic] [and] to engage with someone and also do something because you know it's something we're feeling (...)."* (P2, positive word, 'Loved'). This was more of a commentary on the feeling of being loved than an autobiographical experience of feeling loved. Only the physiological data associated to the autobiographical experiences were used in our analysis of physiological measures.
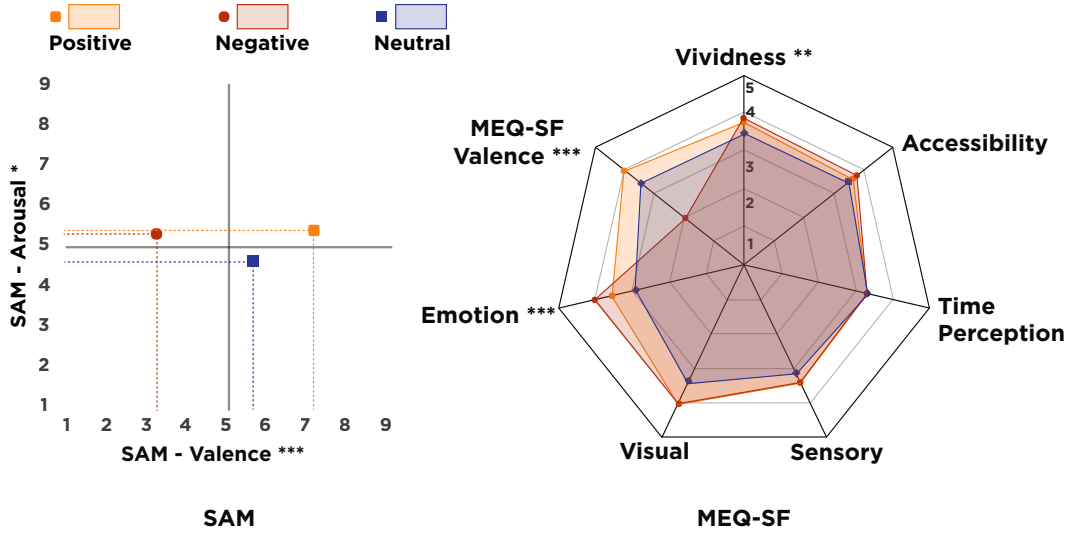
We found that 22 experiences had 'mixed valence'. For example, *"...It was a result announcement day from my high school (...) So I was so afraid before the result was announced (...) I ended up getting a decent mark and I was then happy about it [sic] (...)."* (P6, negative word, 'Afraid'). The valence indicated in this experience was negative at the start and positive in the end. The physiological data associated to these experiences were excluded from our analysis to avoid confounding the results.

There were 47 recalled experiences which were 'inconsistent' with the intended valence of the cued word. An example would be P7 describing a positive experience for 'Fast', a neutral word, *"...we used to go go-karting with people from work, and I enjoy how fast I'm driving this go-kart ... Though we're not supposed to crash into each other but it was still fun (...)."* These experiences were re-labelled to the coded valence before we conducted the physiological measures analysis.

5.2 Subjective Measures

Three hundred post-trial subjective responses (15 trials x 20 participants) were recorded and analysed. These responses includes the "Self-Assessment Manikin" (SAM) with valence and arousal as perceived emotional dimensions, and "Memory Experience Questionnaire - Short Form" (MEQ-SF) with vividness, accessibility, time perspective, sensory, visual, emotion, and valence dimension of the recalled autobiographical memory. First, we used Shapiro-Wilk test [72] to determine if the data is normally distributed i.e. p-value more than 0.05. We identified our subjective data to be non-parametric (not-normally distributed) and based on the within-subject experiment design with three emotional valence (positive, negative, and neutral) as repeated measures, we used Friedman test [73] to understand its effect. For the measures with Friedman test reporting main effect i.e. p-value < 0.05, we used Wilcoxon signed-rank test [14] for pairwise comparison between positive, negative, and neutral AM recall conditions.

5.2.1 Self-Assessment Manikin (SAM) Ratings. SAM - Valence and SAM - Arousal ratings were not normally distributed (Shapiro-Wilk test, $p<.05$). Friedman tests showed **significant main effect of emotional recall (i.e. positive, negative, and neutral) on valence** ($\chi^2(2)$=108.07, $p<.001$**), and arousal scale** ($\chi^2(2)$=10.23, p=.005). Post-hoc analysis with Wilcoxon signed-rank tests revealed **significant differences in valence ratings**

Fig. 4. [Left: Self-Assessment Manikin (SAM)] Average valence and arousal of the perceived emotional states during positive, negative, and neutral AM recall. [Right: Memory Experience Questionnaire - Short Form] Average vividness, accessibility, time perception, sensory, visual, emotion, and MEQ-SF valence associated with Positive, Negative, and Neutral AM. [Bottom Table] Statistical analysis results of SAM-Valence, Arousal, MEQ-SF vividness, emotions, and MEQ-SF valence indicating main and interaction effects.

between positive and negative (Z=14.62, p<.001), **positive and neutral** (Z=5.72, p<.001), and **negative and neutral** (Z=−10.69, p<.001). There were **significant differences in arousal ratings between positive and neutral** (Z=2.48, p=0.014), **and negative and neutral** (Z=2.45, p=.015), but not between positive and negative (Z=0.19, p=.844). The mean (standard deviation, SD) valence ratings for positive, negative, and neutral emotional recall were 7.04 (SD = 1.68), 3.3 (SD = 1.85), and 5.8 (SD = 2.03) respectively. Mean (SD) arousal ratings for positive, negative, and neutral recall were 5.37 (SD = 2.41), 5.35 (SD = 2.55), and 4.57 (SD = 2.34) respectively.

5.2.2 Memory Experience Questionnaire - Short Form (MEQ-SF). **Vividness** (1: not vivid, 5: very vivid): There was a **significant main effect on vividness** ($\chi^2(2) = 10.47, p = .005$). There were **significant differences between positive and neutral** (Z=2.17, p=.03), and **negative and neutral** (Z=2.96, p=.003), but no significant difference between positive and negative valences (Z=−0.30, p=.76). Mean (SD) ratings for positive, negative, and neutral were 3.78 (SD = 0.93), 3.82 (SD = 0.76), and 3.48 (SD = 0.89) respectively.

Accessibility (1: not accessible, 5: very accessible): There was a **significant main effect on accessibility** ($\chi^2(2) = 7.96, p = .018$). There were **significant differences between negative and neutral** (Z=2.87, p=.004), no significant difference between positive and negative (Z=−0.62, p=.53) and positive and neutral (Z=1.43, p=.15); Mean (SD) ratings for positive, negative, and neutral were 3.6 (SD = 1.12), 3.7 (SD = 0.86), and 3.4 (SD = 0.96) respectively.

Time Perspective (1: memory of event was vague, 5: memory of event was clear): There were no significant differences ($\chi^2(2)$=.75, p=.68) in time perspective ratings between valences for emotional recall. The mean (SD) ratings for positive, negative, and neutral recall were 3.34 (SD = 0.943), 3.33 (SD = 0.94), and 3.30 (SD = 0.965) respectively.

Sensory (1: memory did not involve strong sensory information, 5: memory involved strong sensory information): There were no significant differences ($\chi^2(2)$=1.53, p=.23) in sensory ratings between valences for emotional recall. The mean (SD) ratings for positive, negative, and neutral recall were 3.33 (SD = 0.84), 3.38 (SD = 0.83), and 3.20 (SD = 0.72) respectively.

Visual (1: memory was recalled from an observer's perspective, 5: memory was recalled from own perspective): There were no significant differences ($\chi^2(2)$=0.89, p=.64) in visual ratings between valences for emotional recall. The mean (SD) ratings for positive, negative, and neutral recall were 3.53 (SD = 0.85), 3.59 (SD = 0.86), and 3.5 (SD = 0.78) respectively.

Emotion (1: memory did not evoke intense emotions, 5: memory evoked intense emotions): There was a **significant main effect on emotion ratings** ($\chi^2(2) = 31.39, p < .001$). There were **significant differences between positive and negative valences** (Z=−2.02, p=0.04), **positive and neutral** (Z=3.52, p<.001), and **negative and neutral** (Z=5.89, p<.001). Mean (SD) ratings for positive, negative, and neutral were 3.37 (SD = 1.01), 3.6 (SD = 0.94), and 2.95 (SD = 1.00) respectively.

Valence (1: memory was negative, 5: memory was positive): There was a **significant main effect on valence ratings** ($\chi^2(2) = 123.63, p < .001$). There were **significant differences between positive and negative valences** (Z=17.11, p=< 0.001), **positive and neutral** (Z=4.42, p<0.001), and **negative and neutral** (Z=−12.85, p<.001). Mean (SD) ratings for positive, negative, and neutral were 4.16 (SD = 0.86), 2.1 (SD = 0.89), and 3.6 (SD = 0.97) respectively.

5.2.3 Relationship between Emotional AM (MEQ-SF) and Elicited Emotional States (SAM). In order to determine the relationship between the emotional states of AM i.e. MEQ-SF-Valence and elicited emotional states (SAM-Valence) during the AM recall, we performed Spearman's rank-order correlation test for non-parametric data. This reported a **strong positive correlation** ($\rho = 0.49, P$ <0.001). We also attempted to find any correlation between the MEQ-SF Emotion characteristic with SAM-Valence and SAM-Arousal emotional dimensions to determine any relationship between emotions associated with the memory and the elicited emotions while recalling those memories. Spearman's rank-order test reported a strong positive correlation between MEQ-SF Emotion and SAM-Arousal ($\rho = 0.401, P$ <0.001), but no correlation between MEQ-SF Emotion and SAM-Valence ($\rho = 0.005, P$ =0.96). We didn't find any correlation between MEQ-SF Valence and SAM-Arousal ($\rho = 0.139, P$ =0.190).

5.3 Physiological Measures

For physiological measures analysis, we first tested the normal distribution of the EDA, HRV, and Pupil data using Shapiro-Wilk test and found all of the features to be non-parametric (p<0.05). Next, we were interested

in understanding the impacts of emotional AM recall on physiology, i.e. two conditions - recall vs no-recall. Due to the non-parametric data and two conditions, we used Wilcoxon signed-rank test on physiological data. We were also interested in investigating the impact of positive, negative, and neutral AM recall on physiology. So, after testing the data normality using Shapiro-Wilk test, we found the data to be non-parametric. We used Friedman test for the repeated measures (positive, negative, and neutral) followed by Wilcoxon signed-rank test for pairwise comparison.

5.3.1 Electrodermal Activity (EDA) . The EDA features were not normally distributed (Shapiro-Wilk test, $p<.001$).
Wilcoxon signed-rank tests showed that **EDA-PN (Median = 4.0, SD = 17.2) and EDA-PA (Median = 0.15, SD = 2.05) in the recall condition were significantly higher than in the no-recall condition** (*EDA-PN*: Median = 2.0, SD = 11.35, $Z=-11.5$, $p<.001$; *EDA-PA*: Median = 0.08, SD = 2.3, $Z=-2.26$, $p=.024$).

Friedman tests showed a significant effect of emotional (positive, negative, and neutral) AM recall on *EDA-PA* ($\chi^2(2)=7.31$, $p=.025$) but no effect on *EDA-PN* ($\chi^2(2)=2.96$, $p=.23$). However, post-hoc analysis with Wilcoxon signed-rank tests and Bonferroni correction for *EDA-PA* showed no significant differences between negative and neutral ($Z=-2.15$, $p=.03$), positive and negative ($Z=0.82$, $p=.41$), and positive and neutral valences ($Z=-1.28$, $p=.20$); Negative: Median = 0.16, SD = 0.78; Neutral: Median = 0.21, SD = 2.89; and Positive: Median = 0.10, SD = 1.96.

5.3.2 Heart Rate Variability (HRV). The HRV features were not normally distributed (Shapiro-Wilk test, $p<.05$).
Wilcoxon signed-rank tests showed that **RMSSD (Median = 42.7, SD = 59.0) and HFnu (Median = 0.485, SD = 0.22) in recall condition were significantly lower than in the no-recall condition** (*RMSSD*: Median = 45.0, SD = 63.7, $Z=-4.80$, $p=.01$; *HFnu*: Median = 0.976, SD = 0.07 , $Z=-9.25$, $p<.001$). There was no significant difference in *SDNN* ($Z=-1.93$, $p=.63$) between recall (Median = 53.2, SD: 39.9) and no-recall (Median = 69.1, SD = 47.6). Friedman tests reported no significant effect between positive, negative, and neutral memory recall on *RMSSD* ($\chi^2(2)=.81$, $p=.66$), *SDNN* ($\chi^2(2)=.27$, $p=.87$),

and *HFnu* ($\chi^2(2)=.64$, $p=.72$); Positive (*RMSSD*: Median = 54.32, SD = 63.44; *SDNN*: Median = 53.82, SD = 30.7; *HFnu*: 0.49, SD = 0.23); Negative (*RMSSD*: Median = 67.96, SD = 77.3; *SDNN*: Median = 62.0; SD = 39.24; *HFnu*: 0.49, SD = 0.22); and Neutral (*RMSSD*: Median = 54.4, SD = 73.7; *SDNN*: Median = 73.2; SD: 36.6; *HFnu*: Median = 0.46, SD: 0.22).

5.3.3 Pupil Responses and Eye Movement. The pupil and eye movement features were not normally distributed (Shapiro-Wilk test, $p<.001$).
Wilcoxon signed-rank tests showed that the **mean pupil diameter (PD) and blink count were significantly higher in the recall condition compared to the no-recall condition**; Mean PD:$Z=-4.80$, $p<.001$, Recall Median = 3.23, Recall SD = 0.55, No-Recall Median = 3.11, No-Recall SD = 0.54; blink count: $Z=-4.875$, $p<.001$, Recall Median = 8.0, Recall SD = 6.53, No-Recall Median = 6.0, No-Recall SD = 5.55). **Mean blink duration for the recall condition was significantly lower than the no-recall condition**; $Z=-4.88$, $p<.001$, Recall Median = 0.188, Recall SD = 0.10, No-Recall Median = 0.22, No-Recall SD = 0.14).

A significant main effect of positive, negative, and neutral emotional recall on *blink count* ($\chi^2(2)=6.25$, $p=.044$) was reported by Friedman tests. No effect was found for *mean PD* ($\chi^2(2)=0.26$, $p=.77$) and *mean blink duration* ($\chi^2(2)=3.13$, $p=.21$).

Post-hoc analysis with Wilcoxon signed-rank tests on *blink count* showed no significant differences between positive and negative ($Z=-1.17$, $p=.24$), positive and neutral ($Z=-1.54$, $p=.13$), and negative and neutral valences ($Z=-0.18$, $p=.86$); Negative: Median = 9, SD = 6.9, Neutral: Median = 9, SD = 6.5, and Positive: Median = 8, SD = 6.4.

5.4 Machine Learning Detection of AM Recall and Emotion

5.4.1 AM Recall Detection. For the personalized models to recognize if the autobiographical memory is being recalled or not, we separated participant-wise data and followed the machine learning workflow to train and test the classifiers. After 10-fold cross-validation (CV) on the train data, we evaluated our models on test data with a total average score of 86.9% for all the participants' individual best models with a maximum of 95.1% and a minimum of 76.9% score. Out of 20 participants' models (individual AM recall detect CVS can be seen in table 1), 13 got a score between 80-90% with an average of 86.95%, 4 got a score above 90% with an average of 93.3%, and 3 got between 70-80% with an average of 77.8%. We also found that EC provided the best score in 17 participant's testing data whereas RF for 2, and KNN for 1 of the participants. For the generalized models, we used all the participant data for training and testing the classifiers. After 10-fold CV, EC performed better than other models with a CV Score of 77.1%. Through PCA on the overall training data, we found that 'Mean Absolute Deviation', 'Peak Number', and 'Maximum Absolute Amplitude' features were more important than others as these have larger absolute scores on the highly relevant components.

5.4.2 Emotion in AM (EAM) Detection. The 10-fold CV on the personalized classifier models to identify if the recalled experience is of positive, negative or neutral valence reported an overall average CV score of 70.68% with a maximum of 84.1% and minimum of 60.3%. As shown in table 1, 10 out of 20 participant's models performed between 70-80% with an average of 73.99% whereas 9 were between 60-70% and 1 above 80% with an average score of 65.52% and 84.1% respectively. For EAM detection, RF performed better than others for 12 participants, KNN for 7 and SVC for 1. KNN performed better for the generalized EAM detection technique with a CV score of 50.3%. 'Maximum Absolute Amplitude', 'Peak Number (EDA-PN)', and 'Amplitude (EDA-PA)' features reported higher absolute scores through PCA suggesting to be more important than others.

6 DISCUSSION

Investigating H1 (recalling emotional autobiographical memories in virtual reality elicits similar perceived emotional responses), we analyzed self-reported subjective responses i.e. SAM and MEQ-SF to understand the emotional AM and respective elicited emotional states. We first focused on learning if the emotional words used as stimuli helped in recalling emotional AM. The significant difference in recalled memory's perceived emotion and valence characteristics suggests that positive, negative, and neutral valence words helped in recalling corresponding emotional memories. The negative emotional memories were easily remembered as compared to neutral ones and were of stronger emotions as compared to both positive, and neutral words which resonates with the studies in conventional laboratory setting [51, 62, 77]. We also found that the emotional AM (i.e. positive, or negative words) are recalled vividly as compared to the the neutral words aligning with previous findings in non-VR studies [13]. The strong positive relationship between valence of the memory being recalled and elicited perceived valence suggest that positive AM recall tends to elicit positive emotional experience and vice-versa in virtual reality. These insights support H1 highlighting the notion that emotional past memories contribute in the emotion construction during the experience in virtual reality and could help in improving the emotion recognition classifiers if considered contextually.

The physiological analysis on EDA, HRV, and Pupil data was performed to address H2 (Positive, negative, and neutral autobiographical memory recall in virtual reality have a significant effect on the physiological emotional states measured using electrodermal activity (EDA), heart rate variability (HRV), and pupil responses). An increase in physiological arousal can be inferred from the higher EDA peak number and EDA mean peak amplitude [42, 76], lower HRV RMSSD and HRV HFnu [10, 27], and increase in pupil diameter and blink count [20] during emotional AM recall as compared to the no-recall condition. As we found a strong positive correlation between self-reported emotional arousal and emotions of the AM, indicating emotions associated with past memories affect perceived emotional intensity during the recollection, we can expect the increase in physiological

Table 1. Cross-Validation Score (CVS) of the participants for Autobiographical Memory Recall Detection (AM Detection), and Emotions in AM Recall (EAM Detection). [EC: Ensemble Classifier, KNN: K-nearest Neighbor, RF: Random Forest, SVC: Support Vector Classifier]

Participant ID	AM Detection		EAM Detection	
	CV Score (%)	Classifier	CV Score (%)	Classifier
1	88.2	EC	66.4	RF
2	77.9	EC	61.5	SVC
3	76.9	KNN	69.9	RF
4	93	EC	60.3	RF
5	87.6	EC	76.9	KNN
6	92.1	EC	71.8	RF
7	85.4	EC	74.6	KNN
8	85.5	EC	75.2	RF
9	93	RF	84.1	RF
10	78.6	RF	65.2	KNN
11	87.5	EC	72.1	KNN
12	88.2	EC	62.8	RF
13	95.1	EC	74.6	RF
14	88.2	EC	68.5	RF
15	86	EC	68.6	RF
16	84.3	EC	66.5	RF
17	87.7	EC	79.4	RF
18	89.1	EC	72.6	KNN
19	87.8	EC	70.3	KNN
20	86.2	EC	72.4	KNN
Generalized	77.1	EC	50.3	KNN

arousal. Additionally, an increase in mean EDA amplitude during emotional AM recall aligns with results from previous work [67]. However, we didn not find any significant change in physiology between positive, negative, and neutral emotional AM recall. This partially contradicts our subjective measure for arousal (SAM-A) and the physiological arousal results from previous work [30, 67] that showed significant differences between positive and neutral, and negative and neutral memories. A possible reason could be that the previous study [30] allowed a longer time for participants to recall (up to 90s) than our study (30s) allowing them to experience longer emotional experiences. Another reason for no significant change in physiology could be that the emotional words chosen were on the similar arousal scale with only valence scale varied even for neutral words. This should be further investigated by selecting the words with varied arousal emotional dimension. Overall, our study results supported H2 partially with emotional AM recall being able to show significant effect on physiological emotions, but failed to demonstrate significant difference between positive, negative, and neutral AM recall.

At last, to test H3 (The recall of emotional autobiographical memories can be reliably classified into binary-label (recall/no-recall), and multi-label (positive/ negative/ neutral) classes using physiological information such as EDA) we used basic supervised machine learning techniques. Through personalized binary-label classifier, we investigated if we can detect word-induced autobiographical memory recall using physiological information such as EDA. We achieved a cross-validation score (CVS) or accuracy of up to 95.1% with an overall average CVS of 86.9% and maximum participant's model providing a CVS between 80-90%. Our generalized Ensemble classifier

achieved a 77.1% CVS. This suggests that our method performs well enough to be considered an effective method for detecting if the participant is recalling an autobiographical memory after reading a word in Virtual Reality. Our results also proposed Mean Absolute Deviation, Peak Number (EDA-PN), and Maximum Absolute Amplitude EDA features to strongly contribute to the AM Recall detection for most of the participants' models.

Next, we explored the detection of emotional valence in autobiographical memory recall induced by text-stimuli in VR using multi-label classifier i.e. positive, negative, and neutral on EDA data. The personalized classification reported a CVS of up to 84.1% with 70.68% overall average CVS across all participants' models. A total of 19 out of 20 participant's models achieved between 60-80% CVS. Our generalized KNN classifier performed best with 50.3% CVS. These results indicate medium accuracy in predicting positive, negative or neutral EAM for personalized machine learning models. High absolute scores of Maximum Absolute Amplitude, Peak Number (EDA-PN), and Amplitude (EDA-PA) EDA features suggested being highly relevant in detecting whether the recalled autobiographical memory is of positive, negative or neutral valence in VR. These results support H3 suggesting the possibility of automatic recognition of emotional AM in virtual reality.

6.1 Summary of Contributions

The primary objectives of this study were to (1) investigate any relationship between recalled emotional autobiographical memory in VR and perceived emotional responses, (2) explore any relationship between emotional autobiographical memory in VR and physiological signals, and (3) determine if physiological signals can help in detecting when someone is recalling a self-relevant experience after reading text in VR and if it can also detect the emotional states associated with it. In order to investigate this, we asked three research questions. We sum up our answers to the questions based on our findings here:

RQ1: *What is the relationship between emotional autobiographical memory recall and elicited emotions in VR?*
A1: The strong positive relationship between valence of the memory being recalled and elicited perceived valence suggest that positive AM recall tends to elicit positive emotional experience and vice-versa in virtual reality.

RQ2: *How does the recalled emotional AM impacts the physiological signals such as EDA, HRV, and pupil responses in VR?*
A2: During AM recall in VR, we found that the EDA peak number (EDA-PN) and mean amplitude (EDA-PA) increased, HRV measures of RMSSD and HFnu decreased, pupil diameter and blink count increased, and blink duration decreased. However, we found no statistically significant differences in the physiological signals between the different valences.

RQ3: *How can we detect if the experience being recalled is autobiographical memory using physiological signals?*
A3: Higher and consistent performance of our method in detecting if the participant is recalling a autobiographical memory in VR which seems effective and promising. A significant effect of AM Recall on EDA-PN reported to answer RQ1 also supported our method. It was also one of the important features contributing to higher accuracy along with Mean Absolute Deviation, and Maximum Absolute Amplitude EDA features.

RQ3a: *How can we detect if the autobiographical memory being recalled in VR elicits* positive, negative or neutral valence emotional states?
A3a: This study allowed us to test the ability of our method to detect emotional states induced during the autobiographical memory recall through textual cues. It encourages the use of EDA as a modality to recognize the EAM during recall, which is demonstrated for the first time in VR as far as we know. EDA features such as Maximum Absolute Amplitude, Peak Number (EDA-PN), and in particular, Amplitude (EDA-PA) strongly contributed

to the EAM Detection models. EDA-PA being an important feature is also supported by the answer to RQ2 showing a significant effect of emotions associated with autobiographical memory being recalled. This concept paves the way for future investigations in detecting emotional states associated with past experiences recall in VR.

7 POTENTIAL APPLICATIONS

The automatic detection of whether the user is recalling their past information and the information about the associated emotions opens a plethora of applications to provide more personalized and empathetic systems as well as interactions in VR.

Language Learning System: A direct application of words-based emotional autobiographical memory recall detection could be a personalized second language learning system in VR that exploits the knowledge that a specific word has emotional past memory associated with it. The physiological sensors within the system could be further used to understand learners' cognitive and affective states (e.g. engagement and cognitive load during learning) and enable adaption of learning content accordingly [75]. It provides more empathetic and engaging interactions while promoting effective vocabulary acquisition.

Reflection and Therapy: Another application would be in supporting reflection — reviewing of past behaviour to enable learning from past experiences and reframing of self-identity. This would be beneficial as autobiographical memories are highly connected with self-identity. Systems could enable positive reframing of traumatic events through therapy. Previous research has used emotional AM recall in various therapies such as memory rehabilitation therapy for Alzheimer's patients, Eye Movement and Desensitization and Reprocessing (EMDR) therapy for people with emotional distress, and VR-based exposure therapy for social anxiety. With information on when the patient is recalling their past experience and its emotional state, these therapies can be designed to be adaptive. This also provides a feature to automatically emotion-tag the past experience implicitly while the patient is engaged in the virtual therapy session.

Reminiscence: Reminiscing in VR, which usually involves recalling positive memories, has been proven to reinforce the positive emotion and promote emotion regulation. This research can contribute towards a real-time adaptive VR system exploiting the directive function (i.e., using past experiences to decide on current and future activities). This vision also aligns with Pizolli's Personalized VR [60], that takes information about relevant life events of the user and uses it to render virtual content (e.g., objects) to provide a more relaxing experience, consequently enhancing well-being.

8 LIMITATIONS AND FUTURE WORK

One of the primary drawbacks in investigating the different emotional autobiographical memory recall was the absence of a wider range of different emotion triggering words. Previous research [79] has suggested that arousal plays an important role in AM retention, so for our future works, we would like to include higher emotional intensity stimulus cues such as images, video and 3D graphics. Although the results indicated that using EDA, HRV, pupil responses, and eye movements could provide promising results, we plan to expand on the sensors used by using Electroencephalogram (EEG) for memory recall as proposed by previous research [33]. This preliminary research is exploring emotional word cues as stimulus to retrieve AM in VR by replicating conventional laboratory based Autobiographical Memory Test (AMT). In order to further explore the potential of VR, we plan to use more immersive and interactive virtual environments towards identifying emotional AM. Our emotional autobiographical recall prediction model achieved comparatively less cross-validation score. One of the reason could be the imbalanced emotion data as we had to relabel the data based on the emotions of the recalled AM which were inconsistent with the intended emotion. Another reason could be less data resulting in inaccurate models. To deeply investigate this we would be collecting more data in future.

9 CONCLUSION

In this paper, we examined the effects of AM recall and emotional AM recall on physiological signals such as EDA, HRV, and pupil response in VR and investigated the automatic recognition of emotional AM recall. Specifically, we tested AM recall vs No recall and positive, negative, and neutral emotional AM recall in VR using an Autobiographical Memory Test (AMT) by showing emotional words in VR while recording their EDA, HRV and pupil data. For emotional words, we used 5 positive, negative, and neutral words each combining for a total of 15 trials for each participant. After every trial, we asked the participants to complete the Self-Assessment Manikin (SAM) scale and Memory Experience Questionnaire - Short Form (MEQ-SF) subjective questionnaires to collect their perceived emotional and memory experience information. Later, we modeled the AM recall and emotions in AM recall with the collected EDA data to develop recognition system.

Our results show that AM recall has a positive impact on EDA peak number, EDA mean peak amplitude, HRV Standard Deviation of NN intervals (SDNN), HRV relative High Frequency band power in normal units (HFnu), mean pupil diameter (mean PD), blink count, mean blink duration, maximum blink duration, and standard deviation (SD) blink duration compared to the condition when the participants were not recalling memories. Only maximum blink duration, and SD blink duration reported a significant effect of emotional AM recall i.e. positive, negative, and neutral emotional experiences. Our AM Detect system achieved a generalized accuracy of 77.1% and person-dependent accuracy of up to 95.1%. Generalized accuracy of Emotion in AM detection was 50.3% and person-dependent of up to 70.7% However, our results also imply the potential improvement points for the used emotional words to be of varying intensity. These issues will be addressed in the future work along with investigating more physiological sensors such as EEG.

REFERENCES

[1] Joseph Andreano, Kevin Liang, Lingjun Kong, David Hubbard, Brenda K Wiederhold, and Mark D Wiederhold. 2009. Auditory cues increase the hippocampal response to unimodal virtual reality. *Cyberpsychology & Behavior* 12, 3 (2009), 309–313.

[2] Dorthe Berntsen. 2002. Tunnel memories for autobiographical events: Central details are remembered more frequently from shocking than from happy experiences. *Memory & cognition* 30, 7 (2002), 1010–1020.

[3] Patricia J Bota, Chen Wang, Ana LN Fred, and Hugo Plácido Da Silva. 2019. A review, current challenges, and future possibilities on emotion recognition using machine learning and physiological signals. *IEEE Access* 7 (2019), 140990–141020.

[4] Margaret M Bradley and Peter 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.

[5] Lucie Bréchet, Christoph M Michel, Daniel L Schacter, and Alvaro Pascual-Leone. 2021. Improving autobiographical memory in Alzheimer's disease by transcranial alternating current stimulation. *Current Opinion in Behavioral Sciences* 40 (2021), 64–71.

[6] William F Brewer. 1994. Autobiographical memory and survey research. In *Autobiographical memory and the validity of retrospective reports*. Springer, 11–20.

[7] Roberto Cabeza, Steve E Prince, Sander M Daselaar, Daniel L Greenberg, Matthew Budde, Florin Dolcos, Kevin S LaBar, and David C Rubin. 2004. Brain activity during episodic retrieval of autobiographical and laboratory events: an fMRI study using a novel photo paradigm. *Journal of cognitive neuroscience* 16, 9 (2004), 1583–1594.

[8] Roberto Cabeza and Peggy St Jacques. 2007. Functional neuroimaging of autobiographical memory. *Trends in cognitive sciences* 11, 5 (2007), 219–227.

[9] Larry Cahill and James L McGaugh. 1995. A novel demonstration of enhanced memory associated with emotional arousal. *Consciousness and cognition* 4, 4 (1995), 410–421.

[10] Samantha WT Chan, Shardul Sapkota, Rebecca Mathews, Haimo Zhang, and Suranga Nanayakkara. 2020. Prompto: Investigating Receptivity to Prompts Based on Cognitive Load from Memory Training Conversational Agent. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 4 (2020), 1–23.

[11] S. W. T. Chan. 2022. *Augmenting Human Prospective Memory through Cognition-Aware Technologies*. Ph.D. Dissertation. The University of Auckland, Auckland, New Zealand.

[12] Elizabeth R Chrastil and William H Warren. 2015. Active and passive spatial learning in human navigation: acquisition of graph knowledge. *Journal of experimental psychology: learning, memory, and cognition* 41, 4 (2015), 1162.

[13] Christine Comblain, Arnaud D'Argembeau, and Martial Van der Linden. 2005. Phenomenal characteristics of autobiographical memories for emotional and neutral events in older and younger adults. *Experimental aging research* 31, 2 (2005), 173–189.

[14] William Jay Conover. 1999. *Practical nonparametric statistics*. Vol. 350. john wiley & sons.

[15] MA Conway. 1987. Verifying autobiographical facts. *Cognition* 26, 1 (1987), 39–58.

[16] Martin A Conway and Christopher W Pleydell-Pearce. 2000. The construction of autobiographical memories in the self-memory system. *Psychological review* 107, 2 (2000), 261.

[17] Arnaud D'Argembeau, Christine Comblain, and Martial Van der Linden. 2003. Phenomenal characteristics of autobiographical memories for positive, negative, and neutral events. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition* 17, 3 (2003), 281–294.

[18] Julia Diemer, Georg W Alpers, Henrik M Peperkorn, Youssef Shiban, and Andreas Mühlberger. 2015. The impact of perception and presence on emotional reactions: a review of research in virtual reality. *Frontiers in psychology* 6 (2015), 26.

[19] Sylvie Droit-Volet and Sandrine Gil. 2009. The time–emotion paradox. *Philosophical Transactions of the Royal Society B: Biological Sciences* 364, 1525 (2009), 1943–1953.

[20] Mohamad El Haj, Steve MJ Janssen, Karim Gallouj, and Quentin Lenoble. 2019. Autobiographical memory increases pupil dilation. *Translational neuroscience* 10, 1 (2019), 280–287.

[21] Javier Fernandez-Alvarez, Desirée Colombo, Carlos Suso-Ribera, Alice Chirico, Silvia Serino, Daniele Di Lernia, Azucena García Palacios, Giuseppe Riva, and Cristina Botella. 2021. Using virtual reality to target positive autobiographical memory in individuals with moderate-to-moderately severe depressive symptoms: A single case experimental design. *Internet Interventions* (2021), 100407.

[22] Jason Geller, Matthew B Winn, Tristian Mahr, and Daniel Mirman. 2020. GazeR: A package for processing gaze position and pupil size data. *Behavior research methods* (2020), 1–24.

[23] Matthew D Grilli, Aubrey A Wank, Matthew J Huentelman, and Lee Ryan. 2021. Autobiographical Memory Fluency Reductions in Cognitively Unimpaired Middle-Aged and Older Adults at Increased Risk for Alzheimer's Disease Dementia. *Journal of the International Neuropsychological Society* (2021), 1–11.

[24] Kunal Gupta, Sam WT Chan, Yun Suen Pai, Alexander Sumich, Suranga Nanayakkara, and Mark Billinghurst. 2021. Towards Understanding Physiological Responses to Emotional Autobiographical Memory Recall in Mobile VR Scenarios. In *Adjunct Publication of the 23rd International Conference on Mobile Human-Computer Interaction*. 1–5.

[25] Kunal Gupta, Ryo Hajika, Yun Suen Pai, Andreas Duenser, Martin Lochner, and Mark Billinghurst. 2020. Measuring human trust in a virtual assistant using physiological sensing in virtual reality. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 756–765.

[26] Kunal Gupta, Jovana Lazarevic, Yun Suen Pai, and Mark Billinghurst. 2020. Affectivelyvr: Towards vr personalized emotion recognition. In *26th ACM Symposium on Virtual Reality Software and Technology*. 1–3.

[27] Nis Hjortskov, Dag Rissén, Anne Katrine Blangsted, Nils Fallentin, Ulf Lundberg, and Karen Søgaard. 2004. The effect of mental stress on heart rate variability and blood pressure during computer work. *European journal of applied physiology* 92, 1 (2004), 84–89.

[28] Alisha C Holland and Elizabeth A Kensinger. 2010. Emotion and autobiographical memory. *Physics of life reviews* 7, 1 (2010), 88–131.

[29] Ji-Won Hur, Hyemin Shin, Dooyoung Jung, Heon-Jeong Lee, Sungkil Lee, Gerard J Kim, Chung-Yean Cho, Seungmoon Choi, Seung-Moo Lee, and Chul-Hyun Cho. 2021. Virtual Reality–Based Psychotherapy in Social Anxiety Disorder: fMRI Study Using a Self-Referential Task. *JMIR Mental Health* 8, 4 (2021), e25731.

[30] Steve MJ Janssen, Alicia Foo, Sheena N Johnson, Alfred Lim, and Jason Satel. 2021. Looking at remembering: Eye movements, pupil size, and autobiographical memory. *Consciousness and Cognition* 89 (2021), 103089.

[31] Carolyn H John. 1988. Emotionality ratings and free-association norms of 240 emotional and non-emotional words. *Cognition & Emotion* 2, 1 (1988), 49–70.

[32] Joanna Kisker, Thomas Gruber, and Benjamin Schöne. 2019. Behavioral realism and lifelike psychophysiological responses in virtual reality by the example of a height exposure. *Psychological research* (2019), 1–14.

[33] Joanna Kisker, Thomas Gruber, and Benjamin Schöne. 2020. Virtual reality experiences promote autobiographical retrieval mechanisms: Electrophysiological correlates of laboratory and virtual experiences. *Psychological Research* (2020), 1–17.

[34] Joanna Kisker, Thomas Gruber, and Benjamin Schöne. 2021. Experiences in virtual reality entail different processes of retrieval as opposed to conventional laboratory settings: A study on human memory. *Current Psychology* 40, 7 (2021), 3190–3197.

[35] Stanley B Klein. 2015. What memory is. *Wiley Interdisciplinary Reviews: Cognitive Science* 6, 1 (2015), 1–38.

[36] Artie Konrad, Simon Tucker, John Crane, and Steve Whittaker. 2016. Technology and reflection: Mood and memory mechanisms for well-being. *Psychology of well-being* 6, 1 (2016), 1–24.

[37] Sylvia D Kreibig. 2010. Autonomic nervous system activity in emotion: A review. *Biological psychology* 84, 3 (2010), 394–421.

[38] Peter Kuppens, Francis Tuerlinckx, Michelle Yik, Peter Koval, Joachim Coosemans, Kevin J Zeng, and James A Russell. 2017. The relation between valence and arousal in subjective experience varies with personality and culture. *Journal of personality* 85, 4 (2017), 530–542.

[39] Gisela Labouvie-Vief, Mark A Lumley, Elizabeth Jain, and Hillary Heinze. 2003. Age and gender differences in cardiac reactivity and subjective emotion responses to emotional autobiographical memories. *Emotion* 3, 2 (2003), 115.

[40] Maarten H Lamers and Maik Lanen. 2021. Changing Between Virtual Reality and Real-World Adversely Affects Memory Recall Accuracy. *Frontiers in Virtual Reality* 2 (2021), 16.

[41] Martina Luchetti and Angelina R Sutin. 2016. Measuring the phenomenology of autobiographical memory: A short form of the Memory Experiences Questionnaire. *Memory* 24, 5 (2016), 592–602.

[42] David T Lykken and Peter H Venables. 1971. Direct measurement of skin conductance: A proposal for standardization. *Psychophysiology* 8, 5 (1971), 656–672.

[43] Eleanor A Maguire, Dharshan Kumaran, Demis Hassabis, and Michael D Kopelman. 2010. Autobiographical memory in semantic dementia: A longitudinal fMRI study. *Neuropsychologia* 48, 1 (2010), 123–136.

[44] Katerina Mania, Shahrul Badariah, Matthew Coxon, and Phil Watten. 2010. Cognitive transfer of spatial awareness states from immersive virtual environments to reality. *ACM Transactions on Applied Perception (TAP)* 7, 2 (2010), 1–14.

[45] Katerina Mania, Andrew Robinson, and Karen R Brandt. 2005. The effect of memory schemas on object recognition in virtual environments. *Presence* 14, 5 (2005), 606–615.

[46] Javier Marín-Morales, Carmen Llinares, Jaime Guixeres, and Mariano Alcañiz. 2020. Emotion recognition in immersive virtual reality: From statistics to affective computing. *Sensors* 20, 18 (2020), 5163.

[47] Mara Mather. 2004. Aging and Emotional Memory. (2004).

[48] Mara Mather and Matthew Sutherland. 2009. Disentangling the effects of arousal and valence on memory for intrinsic details. *Emotion Review* 1, 2 (2009), 118–119.

[49] Kathleen B McDermott, Karl K Szpunar, and Shawn E Christ. 2009. Laboratory-based and autobiographical retrieval tasks differ substantially in their neural substrates. *Neuropsychologia* 47, 11 (2009), 2290–2298.

[50] Federica Meconi, Juan Linde-Domingo, Catarina S. Ferreira, Sebastian Michelmann, Bernhard Staresina, Ian A Apperly, and Simon Hanslmayr. 2021. EEG and fMRI evidence for autobiographical memory reactivation in empathy. *Human Brain Mapping* 42, 14 (2021), 4448–4464.

[51] Katherine R Mickley and Elizabeth A Kensinger. 2009. Phenomenological characteristics of emotional memories in younger and older adults. *Memory* 17, 5 (2009), 528–543.

[52] Adriana Mira, Daniel Campos, Ernestina Etchemendy, Rosa M Baños, and Ausiàs Cebolla. 2016. Access to autobiographical memory as an emotion regulation strategy and its relation to dispositional mindfulness. *Mindfulness & Compassion* 1, 1 (2016), 39–44.

[53] Andrew EP Mitchell. 2016. Phenomenological characteristics of autobiographical memories: responsiveness to an induced negative mood state in those with and without a previous history of depression. *Advances in cognitive psychology* 12, 2 (2016), 105.

[54] Niels Peter Nielsen, Sinué Salgado, and Dorthe Berntsen. 2020. Using virtual reality to examine emotional hotspots and intrusions in the trauma film paradigm. *Journal of Applied Research in Memory and Cognition* 9, 3 (2020), 370–380.

[55] Sezin Öner and Sami Gülgöz. 2018. Autobiographical remembering regulates emotions: A functional perspective. *Memory* 26, 1 (2018), 15–28.

[56] Xueni Pan and Antonia F de C Hamilton. 2018. Why and how to use virtual reality to study human social interaction: The challenges of exploring a new research landscape. *British Journal of Psychology* 109, 3 (2018), 395–417.

[57] Thomas D Parsons. 2015. Virtual reality for enhanced ecological validity and experimental control in the clinical, affective and social neurosciences. *Frontiers in human neuroscience* 9 (2015), 660.

[58] Pierre Philippot, Alexandre Schaefer, and Gwénola Herbette. 2003. Consequences of specific processing of emotional information: Impact of general versus specific autobiographical memory priming on emotion elicitation. *Emotion* 3, 3 (2003), 270.

[59] Rosalind W Picard. 2000. *Affective computing*. MIT press.

[60] Silvia Francesca Maria Pizzoli, Ketti Mazzocco, Stefano Triberti, Dario Monzani, Mariano Luis Alcañiz Raya, and Gabriella Pravettoni. 2019. User-centered virtual reality for promoting relaxation: an innovative approach. *Frontiers in psychology* 10 (2019), 479.

[61] Soujanya Poria, Erik Cambria, Rajiv Bajpai, and Amir Hussain. 2017. A review of affective computing: From unimodal analysis to multimodal fusion. *Information Fusion* 37 (2017), 98–125.

[62] Daniel Reisberg and Friderike Heuer. 2004. Memory for emotional events. (2004).

[63] Radiah Rivu, Ruoyu Jiang, Ville Mäkelä, Mariam Hassib, and Florian Alt. 2021. Emotion Elicitation Techniques in Virtual Reality. In *IFIP Conference on Human-Computer Interaction*. Springer, 93–114.

[64] James A Russell. 1979. Affective space is bipolar. *Journal of personality and social psychology* 37, 3 (1979), 345.

[65] Sinué Salgado and Osman Skjold Kingo. 2019. How is physiological arousal related to self-reported measures of emotional intensity and valence of events and their autobiographical memories? *Consciousness and Cognition* 75 (2019), 102811.

[66] David Sander. 2013. Models of emotion: The affective neuroscience approach. (2013).

[67] Alexandre Schaefer and Pierre Philippot. 2005. Selective effects of emotion on the phenomenal characteristics of autobiographical memories. *Memory* 13, 2 (2005), 148–160.

[68] Benjamin Schoene, Marlene Wessels, and Thomas Gruber. 2017. Differences between real-life and laboratory memory: Evidence from EEG and virtual reality. In *Psychophysiology*, Vol. 54. WILEY 111 RIVER ST, HOBOKEN 07030-5774, NJ USA, S128–S128.

[69] Benjamin Schöne, Marlene Wessels, and Thomas Gruber. 2019. Experiences in virtual reality: A window to autobiographical memory. *Current Psychology* 38, 3 (2019), 715–719.

[70] Valentin Schwind, Pascal Knierim, Nico Haas, and Niels Henze. 2019. Using presence questionnaires in virtual reality. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1–12.

[71] Ulrich Seidl, Ulrike Lueken, Philipp A Thomann, Josef Geider, and Johannes Schröder. 2011. Autobiographical memory deficits in Alzheimer's disease. *Journal of Alzheimer's disease* 27, 3 (2011), 567–574.

[72] Samuel Sanford Shapiro and Martin B Wilk. 1965. An analysis of variance test for normality (complete samples). *Biometrika* 52, 3/4 (1965), 591–611.

[73] Michael R Sheldon, Michael J Fillyaw, and W Douglas Thompson. 1996. The use and interpretation of the Friedman test in the analysis of ordinal-scale data in repeated measures designs. *Physiotherapy Research International* 1, 4 (1996), 221–228.

[74] Signy Sheldon, Kayla Williams, Shannon Harrington, and A Ross Otto. 2020. Emotional cue effects on accessing and elaborating upon autobiographical memories. *Cognition* 198 (2020), 104217.

[75] Priyashri K Sridhar, Samantha WT Chan, Yvonne Chua, Yow Wei Quin, and Suranga Nanayakkara. 2019. Going beyond performance scores: Understanding cognitive–affective states in Kindergarteners and application of framework in classrooms. *International Journal of Child-Computer Interaction* 21 (2019), 37–53.

[76] Priyashri K Sridhar, Samantha WT Chan, and Suranga Nanayakkara. 2018. Going beyond performance scores: understanding cognitive-affective states in kindergarteners. In *Proceedings of the 17th ACM conference on interaction design and children*. 253–265.

[77] Justin Storbeck and Gerald L Clore. 2005. With sadness comes accuracy; with happiness, false memory: Mood and the false memory effect. *Psychological Science* 16, 10 (2005), 785–791.

[78] Jennifer M Talarico, Dorthe Berntsen, and David C Rubin. 2009. Positive emotions enhance recall of peripheral details. *Cognition and emotion* 23, 2 (2009), 380–398.

[79] Jennifer M Talarico, Kevin S LaBar, and David C Rubin. 2004. Emotional intensity predicts autobiographical memory experience. *Memory & cognition* 32, 7 (2004), 1118–1132.

[80] Franca Tani, Carole Peterson, and Andrea Smorti. 2014. Empathy and autobiographical memory: Are they linked? *The Journal of genetic psychology* 175, 3 (2014), 252–269.

[81] Armelle Viard, Béatrice Desgranges, Vanessa Matuszewski, Karine Lebreton, Serge Belliard, Vincent de La Sayette, Francis Eustache, and Pascale Piolino. 2013. Autobiographical memory in semantic dementia: new insights from two patients using fMRI. *Neuropsychologia* 51, 13 (2013), 2620–2632.

[82] Andrea Vranić, Margareta Jelić, and Mirjana Tonković. 2018. Functions of autobiographical memory in younger and older adults. *Frontiers in psychology* 9 (2018), 219.

[83] David Watson, Lee Anna Clark, and Auke Tellegen. 1988. Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of personality and social psychology* 54, 6 (1988), 1063.

[84] Gerben J Westerhof and Ernst T Bohlmeijer. 2014. Celebrating fifty years of research and applications in reminiscence and life review: State of the art and new directions. *Journal of Aging studies* 29 (2014), 107–114.

[85] Matthew G Whalley, Michael D Rugg, and Chris R Brewin. 2012. Autobiographical memory in depression: an fMRI study. *Psychiatry Research: Neuroimaging* 201, 2 (2012), 98–106.

[86] J Mark Williams and Keith Broadbent. 1986. Autobiographical memory in suicide attempters. *Journal of abnormal psychology* 95, 2 (1986), 144.

[87] Jiaxin Xu and Baosheng Wang. 2020. Efficacy of VR-Based Reminiscence Therapy in Improving Autobiographical Memory for Chinese Patients with AD. In *International Conference on Applied Human Factors and Ergonomics*. Springer, 339–349.

[88] Kymberly D Young, Greg J Siegle, Jerzy Bodurka, and Wayne C Drevets. 2016. Amygdala activity during autobiographical memory recall in depressed and vulnerable individuals: association with symptom severity and autobiographical overgenerality. *American Journal of Psychiatry* 173, 1 (2016), 78–89.

[89] Kymberly D Young, Greg J Siegle, Vadim Zotev, Raquel Phillips, Masaya Misaki, Han Yuan, Wayne C Drevets, and Jerzy Bodurka. 2017. Randomized clinical trial of real-time fMRI amygdala neurofeedback for major depressive disorder: effects on symptoms and autobiographical memory recall. *American Journal of Psychiatry* 174, 8 (2017), 748–755.