

# Facial Electromyography-based Adaptive Virtual Reality Gaming for Cognitive Training

Lorcan Reidy  
University of Cambridge, UK  
lsr29@alumni.cam.ac.uk

Charles Nduka  
Emteq Ltd., UK  
charles@emteq.net

Dennis Chan  
University College London, UK  
dcfchan@doctors.org.uk

Hatice Gunes  
University of Cambridge, UK  
Hatice.Gunes@cl.cam.ac.uk

## ABSTRACT

Cognitive training has shown promising results for delivering improvements in human cognition related to attention, problem solving, reading comprehension and information retrieval. However, two frequently cited problems in cognitive training literature are a lack of user engagement with the training programme, and a failure of developed skills to generalise to daily life. This paper introduces a new cognitive training (CT) paradigm designed to address these two limitations by combining the benefits of gamification, virtual reality (VR), and affective adaptation in the development of an engaging, ecologically valid, CT task. Additionally, it incorporates facial electromyography (EMG) as a means of determining user affect while engaged in the CT task. This information is then utilised to dynamically adjust the game's difficulty in real-time as users play, with the aim of leading them into a state of flow. Affect recognition rates of 64.1% and 76.2%, for valence and arousal respectively, were achieved by classifying a DWT-Haar approximation of the input signal using kNN. The affect-aware VR cognitive training intervention was then evaluated with a control group of older adults. The results obtained substantiate the notion that adaptation techniques can lead to greater feelings of competence and a more appropriate challenge of the user's skills.

## CCS CONCEPTS

- **Human-centered computing** → **Interactive systems and tools**;
- **Computing methodologies** → **Virtual reality**; *Machine learning*.

## KEYWORDS

cognitive training; virtual reality; facial electromyography; affective computing; adaptive gaming.

### ACM Reference Format:

Lorcan Reidy, Dennis Chan, Charles Nduka, and Hatice Gunes. 2020. Facial Electromyography-based Adaptive Virtual Reality Gaming for Cognitive Training. In *Proceedings of the 2020 International Conference on Multimodal*

*Interaction (ICMI '20)*, October 25–29, 2020, Virtual event, Netherlands. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3382507.3418845>

## 1 INTRODUCTION

Cognitive training (CT) has garnered considerable attention due to the promising intervention outcomes it has provided for improving cognitive skills such as attention, problem solving, reading comprehension and information retrieval. Clinician-delivered or computerised CT has been used for children with attention problems (ADHD) [39] and engagement in intellectually stimulating lifestyle activities has been observed to help maintain cognitive function into later life, via enhancement of cognitive reserve [60]. The aim of CT therefore is to deliver such benefits to cognition as a targeted discrete intervention. However, to date, the numerous studies and CT products, often commercially marketed as "brain training" programmes or apps, have not been found to produce clear evidence of benefit to cognition [41]. This failure has been attributed to several factors. The first problem (P1) is that of adherence to the CT programme; with traditional CT programmes often having attrition rates exceeding 15% [66]. Prior studies have indicated that participant adherence can decrease with increasing intervention complexity and intensity [31], [14], with inhibition due to poor task performance being a strong predictor for training dropout [1]. These CT programmes are overly repetitive and fail to address the motivational deficits often characteristic of older people with memory difficulties [54]. Furthermore, it has been suggested that a VR format may increase training adherence (when compared to an on-paper equivalent task) in individuals with Mild Cognitive Impairment (MCI) and dementia patients [33]. The second problem (P2) is that the skills developed and the cognitive improvements made during these training programmes do not transfer to daily life. While the user might get better at playing the games, there is little evidence that these skills will generalise beyond that [56]. Existing dynamic difficulty adjustment (DDA) models have historically relied on the player's performance to infer what changes in difficulty are appropriate. However, Pagulayan et al. point out that a game's evaluation factor ought to be the affective experience provided by the gameplay, rather than user performance as in productivity software [44]. Mishra et al. were also of the opinion that the input channel for closed-loop games should include real-time data from player interactions and behaviour, rather than just player performance metrics [38].

The work presented in this paper introduces a novel CT programme designed to address these two problems, with an ultimate

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

ICMI '20, October 25–29, 2020, Virtual event, Netherlands

© 2020 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-7581-8/20/10.

<https://doi.org/10.1145/3382507.3418845>

goal of employing this CT environment for people with early stages of dementia to support cognitive training. In keeping with other approaches to CT, a game-based paradigm is used to enhance participant enjoyment and thus increase adherence (addressing P1). The use of VR to create simulated real-world environments within which the CT game is enacted helps overcome the transfer problem and facilitate extension of CT-generated gains to real life activities (addressing P2). Additionally, the design of the CT task incorporates an active control arm and an affective feedback loop that enables the dynamic adjustment of the game's difficulty (DDA), in real-time, based on the player's affective state [32] and their in-game performance (addressing P1). The affective feedback loop is based on measuring facial electromyography (EMG) signals related to the nonverbal facial behaviour of the participants, and training supervised machine learning models with user self-reported affect labels. The employment of facial EMG is due to three reasons: (i) other physiological measures such as the galvanic skin response (GSR) have been reported to provide extremely noisy signals in gaming where the user is using a controller and moving around during the game play [17]; (ii) it is widely acknowledged that the face is the primary means for communicating affective and cognitive mental states, including emotions, interest, agreement, comprehension, concentration and intentions [3], [13]; and (iii) computer-vision based analysis of facial expressions, despite its efficacy [53], [34], is not appropriate in a VR context as the user's face is mostly occluded by the head-mounted display.

## 2 RELATED WORK

### 2.1 Cognition, Memory and Affect

The bisection of memory into short-term-memory (STM) and long-term-memory (LTM), first proposed by Hebb [26], represents the core underlying division imposed in the taxonomy of cognitive systems. The unitary notion of STM can be partitioned to recognise a distinct cognitive system, working memory (WM), that is responsible for temporarily storing task relevant information available for manipulation. LTM can be divided into procedural (non-conscious) memory, which underpins skills/habits/conditioning, and declarative (consciously accessed) memory for storing facts and events [59]. Declarative memory can be further separated into semantic memory (SM), a store of facts about the world, and episodic memory (EM), the facility for re-experiencing events in the context in which they originally occurred [63]. EM is widely considered as a cognitive competence unique to humans. Unlike SM, it explicitly encodes spatial, contextual and temporal information.

Neurological disorders such as attention-deficit/hyperactivity disorder (ADHD) or dementia are associated with memory impairments. Numerous studies have shown deficits in short-term memory, long-term memory and coordination of multiple tasks resulting from such disorders (e.g. see [2]). Therefore, the VR cognitive training intervention developed for *this work incorporates both WM and EM training tasks*.

WM capacity can be tested through a variety of tasks. These tasks typically come in the form of a dual-task paradigm that combines a measure of memory span (a STM test that involves immediate recollection of an ordered list of items) with a simultaneous processing task (e.g. see [18]). It has also been argued that WM reflects

the ability to maintain multiple, task-relevant, pieces of information in the face of distracting irrelevant information [22]. The WM training task implemented in this work draws on both of these ideas. A wide variety of methods have been developed to assess EM capacity, but not all produce consistent results [11]. For the purpose of this work, the implemented EM training primarily focuses on spatial memory (SM) tasks, as they engage the same regions of the brain that EM requires [8]. SM tasks also proved to be a good fit for the VR paradigm, where ecologically valid scenarios could be presented to the user in an engaging manner.

Bennion et al. [4], in their study of the effect of emotion on memory, suggest that there is strong evidence to support the following hypotheses: emotion usually enhances memory; when it does not, its effect can be understood by the magnitude of elicited arousal (with arousal benefiting memory up to a point, but then having a detrimental influence); and when emotion facilitates the processing of information, it also facilitates the retention of that information. The general notion that arousal will enhance memory, up to a point, was reinforced by Yeh et al. in their research of the effects of negative affect on WM capacity [68]. They promote the idea of a game that appropriately challenges users in order to activate their attention, while avoiding negative emotional responses.

### 2.2 Affect and Gaming

Two domains, particularly relevant to this work, have shown promise in recent literature for the application of affective computing methods. The first is the domain of cognitive training, which is motivated by the strong relationship between emotions and cognitive performance [23]. The second is the domain of video games, where the interaction has been noted as a predominantly emotional one [67] and therefore susceptible to affective adaptation – i.e., dynamically changing the gameplay experience based on affective signals read from the player.

Affective gaming can be realised through the use of biofeedback techniques. However, for a game to be considered affective (and not simply a biofeedback game), it must exploit this biological information to propagate affective feedback [5]. That is, the game is an intelligent participant in the biofeedback loop. What distinguishes affective feedback from biofeedback is that the player is not deliberately controlling their physiological responses in order to influence gameplay.

There are numerous novel possibilities for emotive twists on conventional gameplay experiences. For example, Reynolds and Picard developed AffQuake [50], a modification to ID Software's Quake II that incorporated affective signals to alter gameplay in a variety of ways (e.g. in StartleQuake, when a player becomes startled, their avatar also becomes startled and jumps back). Valve Corporation have also experimented with similar modifications to their games: Half-Life 2 and Left 4 Dead 2 [7]. In these modifications, the player's stress level, measured as the electrical response of their skin, determines the pace of the gameplay. Compared to conventional gameplay, the use of VR in video games induces a greater degree of engagement and immersion in players. This heightened immersion, named presence (or the feeling of being there in the virtual world), has been reported to directly impact the affective

states experienced by the player during gameplay – i.e., high levels of presence induce more intense and vivid emotions [51].

### 2.3 Therapeutic Applications of Virtual Reality

In a systematic review of computerised cognitive training literature, Hill et al. concluded that computerised CT is a viable intervention for enhancing cognition in people with MCI [27]. Interestingly, they found that for individuals with dementia, the only clinically meaningful effect sizes were found in studies that utilised immersive technology such as VR or the Nintendo Wii [27]. These findings strongly support the idea that greater transference of cognitive performance to daily life could be achieved through training on ecologically valid tasks in VR.

Savulich et al. attempted to address the motivational deficits in older populations with memory impairments through a novel memory game on an iPad “Game Show” which targeted EM as its cognitive process [54]. The game was designed to have the same motivational properties typical of computer games (tutorials, stimulating music, progression, etc.) and employed DDA with in-game performance as a sole input. They conducted a randomised controlled trial with 42 patients (aged 45 and over) diagnosed with amnesic mild cognitive impairment, splitting the participants evenly into a training and control group. Their results showed significant EM improvements in the CT group, and suggested that gamification enhanced participant motivation.

Gamito et al. investigated the effectiveness of VR for neuropsychological rehabilitation [24] by developing a VR application for CT which incorporated attention and memory tasks resembling daily real-world activities. These tasks targeted WM, visuo-spatial orientation, selective attention, recognition memory and calculation. The training linearly increased in difficulty throughout the sessions. This was employed in a study consisting of twenty stroke patients, with a mean age of 55 years (SD = 13.5), who were randomly assigned between an intervention and control group. The results of their study generally supported the efficacy of VR-based interventions for CT, with significant benefits to memory and attention functions observed in the intervention group (with the exception that no significant improvement was seen in visual memory).

Optale et al. conducted a 6-month long randomised controlled pilot study investigating the efficacy of their VR training intervention in lessening cognitive decline and improving memory functions [42]. The VR tasks asked the user to navigate a simple environment and tested their capacity to recall the route they have taken and their orientation, with a gradual, linear, increase in the complexity of the stimuli. They recruited 36 elderly individuals (median age 80 years) with memory impairment and divided them into an experimental group and an active control group (which received music therapy sessions). Their results showed that participants in the experimental group exhibited improvement in general cognitive functioning and verbal memory, with the most significant effects observed in long-term memory.

Despite their promising results, none of the reviewed studies utilised affective feedback as part of their VR intervention, which is the main focus of this work. Other VR applications for CT have been proposed (e.g., [45], [20], [9]), however, these have not yet reached the point of carrying out studies with their target population.

### 2.4 Emotion Sensing from Facial EMG

Computer-vision based analysis of facial nonverbal behaviour (facial actions or facial expressions) is not appropriate in a VR context as the user’s face will be mostly occluded by the head-mounted display (HMD). Cohn and Ekman in [13] provide a detailed review about early studies that have used surface EMG to measure facial muscle activity in relation to emotion and the evidence found in terms of predictive correlation with self- and observer reported emotion. Therefore, this work utilises facial electromyogram (EMG), recorded from surface electrodes placed over regions of the user’s face, to measure signals related to the affective facial behaviour of the participants, using a novel device, Faceteq [35].

For extracting features from facial EMG, Jerritta et al. investigated the application of higher order statistics (HOS) to derive a set of facial EMG features for classifying Ekman’s six basic emotional states [21]. They used audio-visual (video clips) stimuli to induce emotional responses in participants, and employed a kNN classifier with PCA as a dimensionality reduction technique. Their results showed that the use of PCA prior to classification improved the classifier’s accuracy, achieving an average classification rate of 69.5% across the six basic emotions (anger, disgust, fear, happiness, sadness and surprise). Perusquia-Hernandez et al. [47] investigated the recognition of spontaneous vs. posed smiles, using spatial and temporal patterns of facial EMG. Due to the unbalanced nature of the collected data, they undersampled the majority class to match the minority class samples (as in [55]). The best classification results for this 2-class problem, obtained using spatial-temporal features with a Gaussian kernel SVM, range from 85.23% to 96.43% (across participants), using a 70/30 training/validation data split. Soon et al. developed an application for speech recognition based on facial EMG. Three participants were asked to say a series of numeric (spoken in Malay and English) and command words (spoken in English). Temporal features were extracted from a DWT-Haar approximation of the input signal. Four different classifiers were evaluated: Random Forest, Linear Discriminant Analysis, Naive Bayes and Multilayer Perceptron. Classification results were obtained through a cross-validation scheme with a 66/34 training/validation data split. Random Forest provided the best overall performance with temporal features achieving 64.7%, 49%, and 41.8% in Malay, English, and command words respectively.

## 3 STUDY DESIGN

### 3.1 Game Design

Two separate virtual environments were developed, a virtual supermarket and a virtual multi-room museum (see Fig. 1 and Fig. 2). These locales provided the setting for the WM and EM tasks respectively, and were selected to promote the ecological validity of the intervention, i.e., both environments are likely to be familiar to the older target population and, in the case of the supermarket, to reflect a daily activity. The underlying hypothesis was that by setting the tasks in highly immersive virtual re-creations of real-world environments and having users perform practical tasks (e.g. collecting products from a shopping list and interacting with displays in a museum) the acquired cognitive skills would better generalise to daily life.



**Figure 1: The custom virtual supermarket environment, in which WM tasks were carried out, and player interaction with this environment.**

In the initial phase of the work, a fixed difficulty framework was implemented for both the WM and EM tasks consisting of three difficulty levels (easy, medium and hard). This was designed with the purpose of eliciting a range of emotional responses from study participants and, thus, generating a balanced dataset. These difficulty levels differed in the cognitive load required from the participant (e.g. shorter/longer shopping lists in the supermarket, less/more display locations to remember in the museum).

Both tasks employ the same annotation and EMG logging scheme. EMG data is recorded from when the task starts. After every 45 seconds of gameplay (a time period arrived at through pilot tests) the recording is paused and written out to a log file (created for that segment). When this occurs, the game environment fades to black and the affective slider (see Fig. ) (implemented in VR to mitigate gameplay disruption), is displayed to the player. Players interact with the slider using a VR laser pointer and, when they are happy with their selection (which should best describe the average affect experienced by the player during the preceding 45 seconds of gameplay), press a confirmation button to append the arousal/valence values to the associated EMG log. Gameplay and EMG recording then resumes. This process repeats until the timer runs out.

Putze et al. performed a systematic investigation into the effects of interrupting the VR experience through a questionnaire either inside or outside of VR [49]. Their results showed that administering questionnaires in VR reduces the Break in Presence without affecting the self-reported player experience. This motivated our decision to collect the labels during gameplay. The use of the affective slider was motivated by its facility for expeditious annotation and its interpretability. We therefore decided that obtaining these labels inside VR will better represent the range of emotions experienced at different points during gameplay (while the experience is still fresh). Score tracking and a leaderboard (staples of gamification) were also included for both tasks.

A heads-up-display (HUD) was included that enabled the player to track their current score, time left and other task specific information. Together, the inclusion of these elements was intended to draw the player's focus away from the novelty of VR (thereby mitigating the expected positive bias in the dataset) onto their task performance. To further associate player performance and emotional response, audio-visual stimuli were added in response to correct (bell ringing sound and confetti explosion) and incorrect (buzzing sound and red X) answers.

**3.1.1 Working Memory Task.** The goal of the WM task is to find a (randomly generated) array of products in a virtual supermarket (see Fig. 1), placing each product in a shopping basket. The products



**Figure 2: The custom, multi-room, virtual museum environment, in which EM tasks are carried out. More rooms are unlocked as the difficulty level increases. Green arrow markers indicating which displays are to be remembered during the encoding phase. Bottom right: visual feedback for correct answer during the retrieval phase.**

are specified to the player at the start of each round through the HUD. Each product on the shopping list is displayed on the HUD (as an image and text description) for 1 second, with a 500 ms interval. Players are challenged to remember remaining items on the shopping list (stored in WM), while they actively search for each product. The number of products to be collected is determined by the difficulty level. The medium difficulty tasks users with finding 7 products, this is intended to be the most engaging and balanced difficulty for most users (based on Miller's magic number seven, plus or minus two [36]). The easy and hard difficulties task users with finding 2 and 12 products respectively. These difficulties were designed to increase the likelihood of inducing negative affect in users, i.e., calm-negative on easy (bored due to insufficient challenge) and energetic-negative on hard (frustrated due to excessive challenge). The specific number of products for each difficulty level was determined through pilot testing and feedback. For each correct item collected, 5 points are added to the player's score. Collecting an item that was not on the shopping list reduces the player's score by 4. Therefore, while the player is incentivised to carry out the task quickly, the priority is to ensure that no mistakes are made. The random generation of shopping lists promotes the task's replayability, maintaining the emphasis on short-term WM (rather than remembering the shopping lists from previous attempts) on repeated playthroughs.

**3.1.2 Episodic Memory Task.** The EM task takes place in a multi-room virtual museum environment (see Fig. 2). The core task is divided into two consecutive phases: encoding (storage of information, such that it can be distinguished from other distinct pieces of information) and retrieval (recognition of previously stored information) [64]. In the encoding phase, players are asked to search for one or more displays at randomly generated locations in the museum. Players interact with marked displays in this phase using a laser pointer. On doing so, the age of the display is shown to the player. This interaction can take place at a distance, allowing a greater degree of spatial context to be encoded. After all the marked



**Figure 3: Prototype Faceteq sensing HMD foam insert and how it is placed in the HTC Vive.**

displays have been interacted with, the game transitions to the retrieval phase removing the marked displays from the museum, and teleporting the player back to the museum entrance.

In the retrieval phase, players are tasked with placing a subset of the displays they interacted with during the encoding phase back in their original positions. The player uses the laser pointer to indicate where in the environment (from a selection of highlighted zones) they think it was located. On completion of the retrieval phase, a short bonus phase is initiated. Players are shown three displays they have interacted with and are asked which of them is the oldest/youngest. This textual (age) recall is not randomised, and players who can efficiently store the information in their LTM should perform better over repeated sessions.

### 3.2 Data Acquisition

After the study procedure and protocol was approved by Cambridge’s Department of Computer Science and Technology Ethics Committee, 18 participants (5 female and 13 male, ranging in age from 20 to 37) volunteered to engage in the EMG data acquisition study, by wearing the HTC Vive VR headset with Faceteq sensing HMD foam insert (see Fig. ). 6 of these participants engaged in a preliminary pilot study, while the data collected from the remaining 12 formed the final annotated facial EMG dataset.

The participants were first acquainted with the research goals of the study through an information sheet and verbal introduction. They were introduced to the meaning of arousal and valence, and shown the affective slider annotation scheme (see Fig. 3.1). The EMG recording sessions lasted for 3 minutes and 45 seconds, of which there were six in total (one per difficulty level, for both the WM and EM tasks). Half of the participants played through the difficulty levels in reverse order (hard-to-easy) to reduce the likelihood of the collected data being skewed positive (as in [55]). The hypothesis here is based on the concept of the difficulty curve, the idea that, for an optimal experience, a game’s difficulty should progress in a manner consistent with real-world skill acquisition (easy challenges during the cognitive stage, moderate challenges during the associative stage, and more difficult challenges during the autonomous stage) [6]. By delivering challenges to the player in a reversed order, it is expected that they will experience negative affect more frequently (e.g. frustration early on, and boredom towards the end).

Participants were given a two minute break between gameplay sessions, allowing them to return to a neutral affective state. During these breaks, participants were asked to give an affective label that

best summed up that session (using Russell’s circumplex model [52]). A short informal interview was conducted, after the EMG data collection and all other results from the study were recorded, with the following questions.

- Which of the two environments did you prefer spending time in?
- Which of the two tasks did you find more engaging?
- Did you experience any discomfort during the session and, if you have prior experience of VR, was the addition of the Faceteq sensor off-putting in any way?
- To what extent, if any, did the annotation scheme affect your gameplay experience?

### 3.3 Findings

The majority of participants (10/12) expressed a preference for the museum environment over the supermarket, with many responses indicating that the supermarket felt more mundane as it is an environment they are overly familiar with in the real world. This may point to a trade-off between ecological validity and engagement in the choice of cognitive training environment. Responses were evenly split when it came to task preference, with many stating they preferred the EM task as it had more gameplay variety, while others appreciated the more naturalistic interactions in the WM task. The response to the Faceteq sensor was positive. None of the participants indicated that they experienced any motion sickness or that the sensor was off-putting. Of the nine participants with prior VR experience, five participants responded that they weren’t aware of the sensor once they started playing, two participants responded that they there were aware of the sensor but it had no significant impact on their engagement, and two participants (self-identified regular VR users) stated that the ADC box attached to the back of the HMD (see Fig. 3.2) served as a counterweight to the front-heavy HTC Vive. Most participants (7/12) stated that the in-game annotation scheme was mildly disruptive to the gameplay experience, while others either found it did not affect their experience (3/12) or found it very disruptive (2/12).

## 4 SYSTEM EVALUATION

This evaluation aims to investigate whether and how the developed CT environment and the overall system (with its modules for sensing, feature extraction and affect recognition) work together.

### 4.1 EMG Feature Extraction

**4.1.1 Pre-processing.** Sousa and Tavares noted, in their review of EMG normalisation methods [58], that the voltage potential of surface EMG depends on several factors, varying between individuals and also over time within an individual. Baseline normalisation (removal) is a viable strategy to respond to these issues, having seen use in numerous studies on a variety of physiological signals (e.g., [70]). The user’s baseline was read during the first 45 seconds of gameplay as at this point the novelty of VR had diminished to some extent, and the activities during the early-game are typically less arousing (e.g. reading shopping list, and looking at displays).

**4.1.2 Feature Extraction.** After baseline normalisation, the input signal was processed using DWT [69]. The choice of mother wavelet

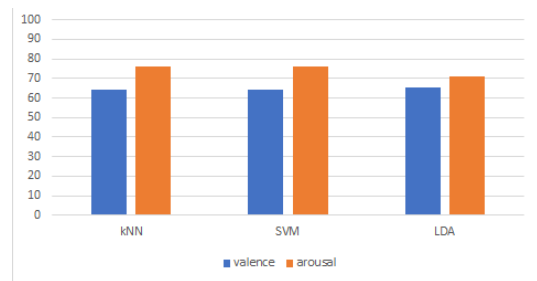
1. *Mean Absolute Value (MAV)* - Indicator of muscle contraction levels.
2. *Mean Absolute Deviation (MAD)* - The average distance of each data point from the mean.
3. *Zero Crossing (ZC)* - An approximate estimate of frequency domain features, defined by the number of times the signal crosses the y-axis.
4. *Slope Sign Change (SSC)* - Represents frequency information, defined as the number of changes between positive and negative slope among three consecutive segments.
5. *Variance (VAR)* - Uses the power of the EMG signal as a feature.
6. *Root Mean Square (RMS)* - Related to the constant force and non-fatiguing contraction in muscles.
7. *Waveform Length (WL)* - The cumulative length of the waveform over a given time segment.
8. *Standard Deviation (STD)* - Models dispersion of the data.
9. *Integrated EMG (IEMG)* - Represents signal power estimation.
10. *Willison Amplitude (WAMP)* - Number of times signal amplitude exceeds a given threshold, it is related to the firing of motor unit action potentials and muscle contraction.
11. *Simple Square Integral (SSI)* - Represents the energy of the signal.
12. *Modified Mean Absolute Value 1 (MMAV1)* - An extension to MAV using a weighting window function.
13. *Modified Mean Absolute Value 2 (MMAV2)* - Similar to MMAV1, but uses a smooth continuous weighting window function.
14. *Maximum Peak Value (MPV)* - Largest amplitude recorded in the segment.

**Figure 4: List of time and time-frequency domain features extracted from the DWT approximation of the signal.**

for signal approximation was informed by the work of Phinyomark et al. [48]. They found that, for the purpose of denoising, *coif5*, *Haar (db1)*, *bior1.1* and *rbio1.1* are the most suitable. A preliminary evaluation with our dataset showed that each of these wavelets resulted in very similar classification improvements (around +5% to +7% accuracy depending on the classifier). Therefore, going forward, the presented results are based on the DWT-Haar approximation of the EMG signal due to its efficient computation. A significant number of studies, in the domain of facial EMG classification, have shown temporal features to be the most informative [25, 30, 47, 57]. Based on these findings, the time and time-frequency domain features were extracted from the DWT approximation of the signal. These are shown in Fig. 4 (see [61] for mathematical definitions). Extracting this many features, from eight EMG channels, results in a high dimensional ( $8 * 14 = 112$ ) dataset.

**4.1.3 Feature Selection.** Due to the dimensionality of the dataset, it was considered pertinent to include a feature selection step prior to classification. The strategy employed here was inspired by the work of Clerico et al. [12], who utilised the minimal-redundancy-maximal-relevance criterion (mRMR) [46] to select the best features in an EMG affective gaming context. mRMR attempts to find optimal features, based on mutual information, through forward selection. The greatest classification improvement was achieved by selecting the best 30 features, identified by mRMR (around +3% to +5% accuracy depending on the classifier). These features were distributed among different muscle groups with the top 30 ranking made up of 10 features from users' eyes, 9 from their mouth, 7 from their eyebrows, and 3 from their corrugator supercillii.

**4.1.4 Findings.** SSC was the most common feature in the ranking, being extracted from all muscle groups bar the corrugator supercillii. SSC, extracted from the right eye sensor, was also computed to be the second most informative feature in the ranking. This was accompanied by MMAV1 extracted from the right mouth sensor (1st) and ZC extracted from the left mouth sensor (3rd), in a top three that scored significantly higher (by a factor of at least 3) than



**Figure 5: Classification accuracies for kNN, SVM and LDA for positive valence vs. negative valence and high arousal vs. low arousal. Results are obtained using leave-one-subject-out cross-validation strategy.**

the remaining 27 features. Interestingly, extracted RMS features were considered relatively uninformative despite it being regularly cited as one in facial EMG emotion recognition research (e.g. [25, 30]). Though deviations in expected results, such as this, may be attributed to the fact that physical expressions of affect in peoples' faces are likely to be impacted by the VR headset.

## 4.2 Affect Classification

**4.2.1 Classification.** First the original (-1 to +1 continuous) valence/arousal labels were truncated into one of four emotion labels: energetic-positive (high valence, high arousal); calm-positive (high valence, low arousal); energetic-negative (low valence, high arousal); calm-negative (low valence, low arousal). This allows emotion recognition to be framed as a 4-class classification problem. Three classifiers, which have been utilised to varying degrees of success in existing facial EMG literature [25, 30, 47, 57], SVMs (with Gaussian kernel); kNN (with various values for *k*); and LDA were evaluated using the (subject-independent) leave-one-subject-out (LOSO) cross-validation strategy. The best classification results for each classifier can be seen in Fig. 5. The kNN (*k* = 4) offered the best arousal classification rate (76.2%), and the best classification rate on the combined valence/arousal 4-class classification problem (68.8%). LDA yielded the best valence classification rate (64.1%), but fell behind kNN on the 4-class classification problem (65.9%). The Gaussian SVM generally underperformed, though by investigating different kernels and further tuning hyperparameters, this could be improved in future work.

**4.2.2 Findings.** The most noticeable discrepancy between the labels acquired in-game and those acquired post-game, was a consistently lower annotated value for arousal in the post-game interview. This manifested as a significantly lower classification accuracy for arousal when using the post-game labels as the four classes (around -15% to -19% depending on the classifier), while valence classification remained comparable. This suggests that the primary benefit of acquiring affective labels in-game is a more reliable estimate of the intensity of emotions. Placing participants in a reversed difficulty group had the desired effect of inducing negative affect more frequently (resulting in a more balanced dataset), with that group accounting for approximately 61% of energetic-negative and 67% of calm-negative annotations. A notable difference between how our classification results were arrived at compared to those

discussed in similar works (section 3.3), is that these results were computed using the (subject-independent) leave-one-subject-out (LOSO) cross-validation strategy.

## 5 INTERVENTION EVALUATION

The goal of this evaluation was to gather qualitative feedback from the target population (older adults) to examine the potential benefits of utilising affective adaptation for CT and gather insights for defining future research steps.

### 5.1 Affective Feedback Loop Integration

The system integrated here relies on both affect sensing and player performance as a data point, due to its exhibited value in existing DDA solutions and to offset the (classification) inaccuracies in the model (i.e. to avoid fixing what is not broken [28]). The number of difficulty levels in this new adaptive version of the game was increased from three (easy-medium-hard) to ten (10-point scale). This allows for more subtle transitions in difficulty to avoid the player becoming overly conscious of the adaptation (and potentially feeling ‘cheated’ by it [28]). Every 45 seconds (in place of the annotation interface from the previous study) the player’s affect is classified in real-time based on incoming facial EMG signals. The following DDA rules, encompassing both player affect and performance, govern how the difficulty is adapted:

*[Calm-Negative + Perfect Score]*: Increment difficulty by 2;  
*[Calm-Negative + Imperfect Score]* or *[Positive Valence + Perfect Score]*: Increment difficulty by 1;  
*[Positive Valence + Imperfect Score]*: No change in difficulty;  
*[Negative Score]* or *[Energetic-Negative + Imperfect Score]*: Decrement difficulty by 1;  
*[Energetic-Negative + Negative Score]*: Decrement difficulty by 2.

This rule set was arrived at following a short testing period with pilot participants. For instance, in the *[Calm-Negative + Perfect Score]* case, it is hypothesised that the player is not enjoying the game as valence is negative, and is likely bored as they are in a non-energetic state, and the score is perfect, so the level is increased by two. Instead, *[Positive Valence + Perfect Score]* leads to a smaller increment of one in order to gradually maximise the cognitive challenge provided by the game, while minimising the risk of disrupting the player’s engagement (as indicated by the positive valence score). In the best-case execution, the adaptation is intended to lead players to a flow state [16], where they are faced with tasks that they have a chance of completing through application of their skills. In addition to being a signifier of high engagement, being in a state of flow has been shown to improve cognitive performance [23].

### 5.2 Evaluation Study

The protocol for the evaluation study was largely similar to the previous study, with a few notable exceptions. 6 participants (4 female and 2 male, ranging in age from 60 to 100), with no history of cognitive impairment, volunteered to engage in this evaluation study. The recruitment of this older control group was facilitated by the researchers from The University of Cambridge’s Department

of Neuroscience. None of the participants played video games with any degree of regularity and only one had prior experience of VR. Participants started by completing a battery of standardised cognitive tests [10, 15, 37, 40, 43, 62, 65] (administered by the Clinical Neuroscience researchers). This enabled accurate characterisation of the sample and the investigation of correlations between the standardised tests and the new VR paradigm. The time taken to administer these tests averaged at about 1 hour.

Participants were given an explanation of the task goals and time to practice in the VR environment prior to starting the session proper. They played both an adaptive and non-adaptive (linearly increasing difficulty) version of the game (without being told which version is which), in two, fifteen minute, gameplay sessions (7 minutes and 30 seconds for both WM and EM tasks). To mitigate any order effects bias in the evaluation of the adaptive and non-adaptive versions, half of the participants played the adaptive version first, while the other half played the non-adaptive version first. Participants’ subjective experience (i.e. immersion, engagement, and flow) with the VR training intervention was evaluated using the in-game and post-game components of the game experience questionnaire (GEQ) [29].

Each session was concluded with an informal interview that took place after all other results from the study were recorded. The contents of this interview was the same as in the first study, with the exception of an additional question on the impact of the in-game affective annotation during the first study. Participation in the study lasted for about 2 hours and 15 minutes on average (including breaks). Immediately after playing each version (adaptive/non-adaptive) of the game, participants reported on their feelings of competence, sensory and imaginative immersion, flow, tension, challenge, negative affect, and positive affect by completing the *in-game module* of the GEQ. Each of these categories is represented by a series of sub-components in the questionnaire, a numeric value is then computed for each by averaging across their sub-component values. The accumulated responses can be found in Table 1. Finally, when both gameplay sessions were complete, the participants filled out the *post-game module* of the GEQ [29]. This gave participants the opportunity to think and reflect on the experience as a whole. The accumulated results, similarly calculated by averaging across their sub-components in the questionnaire, are provided in Table 2.

### 5.3 Findings

Looking at Table 1, while the response to both versions of the game can generally be described as positive, there are a few noteworthy differences. The two standout differences are the increased feeling of competence and the decreased feeling of challenge while playing the adaptive version of the game. The noteworthy increase in competence is particularly encouraging as it relates to one of the key deficiencies identified with existing cognitive training interventions, the drop-off in user engagement [56]. In their research of intrinsic motivation, Deci and Ryan argue that structures that enable feelings of competence during action can enhance intrinsic motivation for that action [19]. This increased feeling of competence, brought about through affective adaptation, highlights the potential of adaptive techniques in motivating users to engage with cognitive training interventions. However, specific aspects related

PID	P1		P2		P3		P4		P5		P6		Mean	
	Ad	NAd	Ad	NAd	Ad	NAd	Ad	NAd	Ad	NAd	Ad	NAd	Ad	NAd
Game Version														
Competence	2	1	3	2.5	2.5	0.5	3	2	2	2	3	2	2.58	1.67
Sensory and Imaginative Immersion	4	3	4	4	3	2	3	3	2	3	4	4	3.33	3.17
Flow	3.5	3.5	4	3.5	3	2	3	3	2	3	3.5	2	3.17	2.83
Tension	2	3.5	0.5	0.5	0	2	0	0.5	2	0	0.5	1.5	0.83	1.33
Challenge	4	4	2.5	3	3	4	3.5	4	1	3	2	3.5	2.67	3.58
Negative Affect	1.5	3	0	0.5	0	1	0	0	0.5	0	0	0	0.33	0.75
Positive Affect	2.5	1.5	3.5	3.5	3	2	2.5	2	1.5	3	3	2.5	2.67	2.42

**Table 1: In-game GEQ [29] module responses for each participant (P) represented with participant ID (PID), for Adaptive (Ad) and Non-Adaptive (NAd) versions of the game. Values are on a scale of 0 (not at all) to 4 (extremely), and were calculated by averaging across their respective components.**

PID	P1	P2	P3	P4	P5	P6	Mean
Positive	2.4	2.6	1.4	1.6	2	2.8	2.13
Negative	0.33	0	0	0	0	0.33	0.11
Tired	1	0.5	0	2.5	0	0	0.67
R2R	1	1	0.33	0.33	0.33	1.66	0.78

**Table 2: Post-game GEQ module responses for each participant (P) on a scale of 0 (not at all) to 4 (extremely), calculated by averaging across their respective components. R2R refers to Return to Reality.**

to intrinsic motivation were not included in the questionnaires employed in our studies. For more insightful conclusions on intrinsic motivation, relevant aspects should be investigated explicitly in future studies.

While the drop in challenge is not unequivocally positive, the decrease to a more neutral value in the adaptive version, along with the slight increase in flow (which describes a state of high engagement), suggests that participants are being met with more appropriate challenges that they can overcome using their skills [16] (potentially explaining the slight increase in positive affect, and decrease in negative affect while playing the adaptive version). The largely positive responses provided for the *post-game module* of the GEQ as seen in Table 2, in conjunction with those recorded by the *in-game module*, are promising indicators that gamification and VR can play a role in increasing engagement with cognitive interventions in older adults.

Another area of interest for this evaluation was to what extent the WM and EM tasks, implemented in VR, engaged the intended cognitive abilities of the participants. This was examined by looking for correlations between how participants performed (relative to each other) in the standardised tests and the VR tasks. Positive correlations, calculated using Spearman’s rank-order correlation ( $\rho$ ), were found between the performance rankings of participants in the VR paradigm and closely related standardized tests. Most notably, high positive correlations were found between the Trail Making Test [62], which examines executive functioning (a superset of WM), and the WM task in VR ( $\rho=+0.60$ ), and between the 4 Mountains test (a short SM test) and the EM task in VR ( $\rho=+0.74$ ).

In the post-session interview, all 6 participants stated their preference for the museum environment (finding the supermarket more

mundane). The overall consensus was, however, that the environments they would prefer would match those in the real-world. 5 participants preferred the EM task, stating that while they found it to be more complex, the greater variety it offered was a motivating factor for them to return to it and improve. This may point to task variety being an important factor in maintaining user engagement in cognitive training. All 6 participants responded that they felt no motion sickness during the study. 4 out of the 6 participants indicated that they felt no facial discomfort, while 2 participants, who wore glasses throughout the study, felt a bit of pressure on their face towards the end. This was likely a result of the slightly thicker face cushion used with the Faceteq prototype.

## 6 SUMMARY AND CONCLUSION

This work investigated the development of an affect-aware VR game for cognitive training using facial EMG signals for affect classification. Classification rates of 64.1% and 76.2%, for valence and arousal respectively, were achieved through a combination of DWT-Haar filtering, temporal feature extraction, feature selection, and kNN classification. The promise of DDA in the development of more engaging cognitive training was substantiated through a small-scale user study with older adults.

The qualitative feedback garnered over the course of the study pointed to a notable increase in feelings of competency and participants being more appropriately challenged. Participant feedback relating to both the adaptive and non-adaptive versions of the game was largely positive. This response lends credence to the notion that gamification and VR are viable tools for improving engagement in cognitive training with older adults. The findings here should be qualified by reiterating an inherent limitation of the study. Six participants is a small user group for an evaluation study and should be expanded in future research to fully determine the veracity of these findings.

## ACKNOWLEDGMENTS

This work has been partially supported by the EPSRC (grant ref. EP/R030782/1).

## REFERENCES

- [1] D.C. Arbiv and N. Meiran. 2015. Performance on the antisaccade task predicts dropout from cognitive training. *Intelligence* 49 (2015), 25–31.



- [2] A.D. Baddeley, S. Bressi, S.Della Sala, R. Logie, and H. Spinnler. 1991. The decline of working memory in Alzheimer's disease: a longitudinal study. *Brain* 114 (1991), 2521–2542.
- [3] Simon Baron-Cohen, Angel Riviere, Masato Fukushima, Davina French, Julie Hadwin, Pippa Cross, Catherine Bryant, and Maria Sotillo. 1996. Reading the Mind in the Face: A Cross-cultural and Developmental Study. *Visual Cognition* 3 (1996), 39–59.
- [4] K.A. Bennion. 2013. Oversimplification in the study of emotional memory. *J. Int. Neuropsychological Society* 19, 9 (2013), 953–961.
- [5] D. Bersak. 2001. Biofeedback using an Immersive Competitive Environment. In *Online Proceedings of the Designing Ubiquitous Computing Games Workshop, Ubicomp*.
- [6] B. Bostan and S. Ögüt. 2009. Game challenges and difficulty levels: lessons learned from RPGs. In *International Simulation and Gaming Association Conference*.
- [7] S. Bouchard, F. Bernier, É. Boivin, B. Morin, and G. Robillard. 2012. Using biofeedback while immersed in a stressful videogame increases the effectiveness of stress management skills in soldiers. *PLoS one* (2012).
- [8] N. Burgess, E.A. Maguire, and J. O'Keefe. 2002. The human hippocampus and spatial and episodic memory. *Neuron* 35 (2002), 625–641.
- [9] G. Caggianese, A. Chirico, G. De Pietro, L. Gallo, A. Giordano, M. Predazzi, and P. Neroni. 2018. Towards a virtual reality cognitive training system for mild cognitive impairment and Alzheimer's disease patients. In *Proc. 32nd Int. Conf. Adv. Inf. Netw. Appl. Workshops*, Vol. 8. 663–667.
- [10] D. Chan, L.M. Gallaher, K. Moodley, L. Minati, N. Burgess, and T. Hartley. 2016. The 4 mountains test: a short test of spatial memory with high sensitivity for the diagnosis of pre-dementia Alzheimer's disease. *J. Vis. Exp* (2016).
- [11] L.G. Cheke and N.S. Clayton. 2013. Do different tests of episodic memory produce consistent results in human adults? *Learning & Memory* 20, 9 (2013), 491–498.
- [12] A. et al. Clerico. 2016. Biometrics and classifier fusion to predict the fun-factor in video gaming. In *IEEE Conference on Computational Intelligence and Games (CIG), Santorini*. 1–8.
- [13] J. F. Cohn and P. Ekman. 2005. Measuring Facial Action by Manual Coding, Facial EMG, and Automatic Facial Image Analysis. In *Handbook of Methods in Nonverbal Behavior Research*, J. A. Harrigan, R. Rosenthal, and K. Scherer (Eds.). Oxford Univ. Press, Oxford, UK, 9–64.
- [14] N. Coley, T. Ngandu, J. Lehtisalo, H. Soininen, B. Vellas, and E. Richard. 2019. Adherence to multidomain interventions for dementia prevention: data from the FINGER and MAPT trials. *Alzheimer's and Dementia* 15 (6) (2019), 729–741.
- [15] H. Tuokko D. Crockett. 2008. Cued recall and memory disorders in dementia. *Journal of Clinical and Experimental Neuropsychology* 11, 2 (2008), 278–294.
- [16] M. Csikszentmihalyi. 2014. *Toward a psychology of optimal experience*. Springer.
- [17] Gábana Arellano D., Tokarchuk L., and Gunes H. 2016. Measuring Affective, Physiological and Behavioural Differences in Solo, Competitive and Collaborative Games. In *Proc. of Intelligent Technologies for Interactive Entertainment 2016 (INTEENTAIN), Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, Poppe R., Meyer JJ., Veltkamp R., and Dastani M. (Eds.). Vol. 178. Springer, Cham, 184–193.
- [18] M. Daneman and P.A. Carpenter. 1980. Individual differences in working memory and reading. *Journal of Verbal Learning & Verbal Behavior* 19, 4 (1980), 450–466.
- [19] E.L. Deci and R.M. Ryan. 1985. *Intrinsic Motivation and Self-Determination in Human Behavior*. Plenum, New York.
- [20] G.M. Doniger, M.S. Beeri, and A. Bahar-Fuchs et al. 2018. Virtual reality-based cognitive-motor training for middle-aged adults at high Alzheimer's disease risk: A randomized controlled trial. *Alzheimer's and Dementia: Translational Research and Clinical Interventions* 4 (2018), 118–129.
- [21] P. Ekman. 1992. An Argument for Basic Emotions. *Cognition and Emotion* 6 (1992), 169–200.
- [22] R.W. Engle, S.W. Tuholski, J.E. Laughlin, and A.R. Conway. 1999. Working memory, short-term memory, and general fluid intelligence: a latent-variable approach. *Journal of Experimental Psychology: General* 128, 3 (1999), 309–331.
- [23] D. Gabana, L. Tokarchuk, E. Hannon, and H. Gunes. 2017. Effects of valence and arousal on working memory performance in virtual reality gaming. In *International Conference on Affective Computing and Intelligent Interaction*. 36–41.
- [24] P. Gamito, J. Oliveira, C. Coelho, D. Morais, P. Lopes, J. Pacheco, R. Brito, F. Soares, N. Santos, and A.F. Barata. 2015. Cognitive training on stroke patients via virtual reality-based serious games. *Disability and Rehabilitation* 5 (2015), 1–4.
- [25] M. Hamed, S.H. Salleh, C.M. Ting, M. Astaraki, and A.M. Noor. 2018. Robust Facial Expression Recognition for MuCI: A Comprehensive Neuromuscular Signal Analysis. *IEEE Transactions on Affective Computing* 9, 1 (2018), 102–115.
- [26] D.O. Hebb. 1949. *The Organization of Behavior*. John Wiley and Sons, New York.
- [27] N.T. Hill, L. Mowszowski, S.L. Naismith, V.L. Chadwick, M. Valenzuela, and A. Lampit. 2017. Computerized cognitive training in older adults with mild cognitive impairment or dementia: a systematic review and Meta-Analysis. *Am. J. Psychiatry* 174, 94 (2017), 329–334.
- [28] R. Hunicke. 2005. The case for dynamic difficulty adjustment in games. In *Proc. of ACM SIGCHI Int. Conf. on Advances in computer entertainment technology*. 429–433.
- [29] W. A. Ijsselstein, Y. A. W. de Kort, and K. Poels. 2007. The Game Experience Questionnaire. *Technische Universiteit Eindhoven* (2007).
- [30] S. Jerritta, M. Murugappan, K. Wan, and S. Yaacob. 2014. Emotion recognition from facial emg signals using higher order statistics and principal component analysis. *Journal of the Chinese Institute of Engineers* 37, 3 (2014), 385–394.
- [31] L.C. Lam, W.C. Chan, A.W. Fung T. Leung, and E.M. Leung. 2015. Would older adults with mild cognitive impairment adhere to and benefit from a structured lifestyle activity intervention to enhance cognition?: A cluster randomized controlled trial. *PLoS One* 10(3) (2015).
- [32] C. Liu, P. Agrawal, N. Sarkar, and S. Chen. 2009. Dynamic Difficulty Adjustment in Computer Games Through Real-Time ANXIETY-Based Affective Feedback. *Int. J. Hum. Comput. Interaction* 25 (2009), 506–529.
- [33] V. Manera, E. Chapoulie, J. Bourgeois, R. Guerchouche, R. David, J. Ondrej, G. Drettakis, and P. Robert. 2016. A Feasibility Study with Image-Based Rendered Virtual Reality in Patients with Mild Cognitive Impairment and Dementia. *PLoS One* 11 (3) (2016), 118–129.
- [34] Brais Martinez, Michel F. Valstar, Bihan Jiang, and Maja Pantic. 2019. Automatic Analysis of Facial Actions: A Survey. *IEEE Transactions on Affective Computing* 10, 3 (2019), 325–347.
- [35] I. Mavridou, J. T. McGhee, M. Hamed, M. Fatoorechi, A. Cleal, E. Ballaguer-Balester, G. Cox, and C. Nduka. 2017. FACETEIQ interface demo for emotion expression in VR. In *Proceedings of IEEE Virtual Reality (VR)*. 441–442.
- [36] G.A. Miller. 1956. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review* 63, 2 (1956), 81–97.
- [37] D. Mioshi, K. Dawson, J. Mitchell, R. Arnold, and J.R. Hodges. 2006. The Addenbrooke's Cognitive Examination Revised (ACE-R): a brief cognitive test battery for dementia screening. *Int. J. Geriatric Psychiatry* 21, 11 (2006), 1078–1085.
- [38] J. Mishra, J. A. Anguera, and A. Gazzaley. 2016. Video games for neuro-cognitive optimization. *Neuron*. 90 (2) (2016), 214–218.
- [39] A. L. Moore, D. M. Carpenter, T. M. Miller, and C. Ledbetter. 2018. Clinician-delivered cognitive training for children with attention problems: effects on cognition and behavior from the ThinkRx randomized controlled trial. *Neuropsychiatric disease and treatment* 14 (2018), 1671–1683.
- [40] H.E. Nelson. 1982. *The National Adult Reading Test*. NFER-Nelson, Windsor, England.
- [41] Gates N.J., Vernooij R.W.M., Di Nisio M., Karim S., March E., Martínez G., and Rutjes A.W.S. 2019. Computerized cognitive training for preventing dementia in people with mild cognitive impairment (Review). *Cochrane Database of Systematic Reviews, Issue 3. Art. No.: CD012279*. (2019).
- [42] G. Optale, C. Urgesi, and V. Busato. 2010. Controlling memory impairment in elderly adults using virtual reality memory training: a randomized controlled pilot study. *Neurorehabilitation and Neural Repair* 24 (4) (2010), 348–357.
- [43] P.A. Osterrieth. 1944. Test of copying a complex figure: contribution to the study of perception and memory. *Archives de Psychologie* 30 (1944), 206–356.
- [44] R. J. Pagulayan, K. Keeker, D. Wixon, R. L. Romero, and T. Fuller. 2002. User-centered design in games. In *The human-computer interaction handbook*. Vol. 5. 883–906.
- [45] E. Pedrolí, S. Serino, M. Stramba-Badiale, and G. Riva. 2018. An Innovative Virtual Reality-Based Training Program for the Rehabilitation of Cognitive Frail Patients. *Pervasive Computing Paradigms for Mental Health* 207 (2018), 62–66.
- [46] H. Peng, F. Long, and C. Ding. 2005. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27, 8 (2005), 1226–1238.
- [47] M. Perusquía-Hernández, M. Hirokawa, and K. Suzuki. 2017. Spontaneous and posed smile recognition based on spatial and temporal patterns of facial EMG. In *International Conference on Affective Computing and Intelligent Interaction (ACII)*. 537–541.
- [48] A. Phinyomark, C. Limsakul, and P. Phukpattaranont. 2010. Optimal wavelet functions in wavelet denoising for multifunction myoelectric control. *ECTI Transactions on Electrical Eng., Electronics, and Communications* 8, 1 (2010), 43–52.
- [49] Susanne Putze, Dmitry Alexandrovsky, Felix Putze, Sebastian Höffner, Jan David Smeddinck, and Rainer Malaka. 2020. Breaking The Experience: Effects of Questionnaires in VR User Studies. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI '20). Association for Computing Machinery, New York, NY, USA, 1–15. <https://doi.org/10.1145/3313831.3376144>
- [50] C.J. Reynolds and R.W. Picard. 2004. AffQuake.
- [51] G. Riva. 2007. Affective interactions using virtual reality: the link between presence and emotions. *CyberPsychology & Behavior* 10 (2007), 45–56.
- [52] J. Russell. 1980. A circumplex model of affect. *Journal of Personality and Social Psychology* 39 (1980), 1161–1178.
- [53] E. Sariyanidi, H. Gunes, and A. Cavallaro. 2015. Automatic Analysis of Facial Affect: A Survey of Registration, Representation, and Recognition. *IEEE Trans.*

- Pattern Analysis and Machine Intelligence* 37 (6) (June 2015), 1113–1133.
- [54] G. Savulich, T. Piercy, J. Suckling, C. Fox, J.B. Rowe, J.T. O'Brien, and B.J. Sahakian. 2017. Cognitive Training Using a Novel Memory Game on an iPad in Patients with Amnesic Mild Cognitive Impairment (aMCI). *International Journal of Neuropsychopharmacology* 20 (8) (2017), 624–633.
- [55] I. Shumailov and H. Gunes. 2017. Computational Analysis of Valence and Arousal in Virtual Reality Gaming using Lower Arm Electromyograms. In *Int. Conf. on Affective Computing and Intelligent Interaction (ACII)*.
- [56] D. J. Simons, W. R. Boot, N. Charness, S. E. Gathercole, C. F. Chabris, D. Z. Hambrick, and E. A. Stine-Morrow. 2016. Do 'Brain-Training' Programs Work? *Psychol Sci Public Interest*. 17, 3 (2016), 103–186.
- [57] M.W. Soon, M.I.H. Anuar, M.H.Z. Abidin, A.S. Azaman, and N.M. Noor. 2017. Speech recognition using facial sEMG. In *IEEE International Conference on Signal and Image Processing Applications (ICSIPA), Kuching*. 1–5.
- [58] A. Sousa and J. Tavares. 2012. Surface electromyographic amplitude normalization methods: A review. In *Electromyography: New Developments, Procedures and Applications*. 85–102.
- [59] L.R. Squire. 2004. Memory systems of the brain: a brief history and current perspective. *Neurobiol. Learn. Mem* 82 (2004), 171–177.
- [60] Y. Stern. 2012. Cognitive reserve in ageing and Alzheimer's disease. *Lancet Neurol*. 11, 11 (2012), 1006–1012.
- [61] D. Tkach, H. Huang, and T.A. Kuiken. 2010. Study of stability of time-domain features for electromyographic pattern recognition. *Journal of Neuroengineering and Rehabilitation* 7, 1 (2010), 1–13.
- [62] T.N. Tombaugh. 2004. Trail Making Test A and B: normative data stratified by age and education. *Arch. Clin. Neuropsychol* 19 (2004), 203–214.
- [63] E. Tulving. 1983. *Elements of episodic memory*. Oxford University Press.
- [64] W. Wang, B. Subagdja, A.-H. Tan, and J.A. Starzyk. 2012. Neural Modeling of Episodic Memory: Encoding Retrieval and Forgetting. *IEEE Trans. Neural Netw. Learn. Syst* 23, 10 (2012), 1574–1586.
- [65] D. Wechsler. 1958. *The measurement and appraisal of adult intelligence*. Williams & Wilkens, Baltimore, Md.
- [66] T. Wykes, V. Huddy, C. Cellard, S.R. McGurk, and P. Czobor. 2011. A meta-analysis of cognitive remediation for schizophrenia: methodology and effect sizes. *Am. J. Psychiatry* 168 (2011), 472–485.
- [67] N. Yee. 2007. Motivations for Play in Online Games. *CyberPsychology & Behavior* 9, 6 (2007), 772–775.
- [68] Y. Yeh. 2015. How stress influences creativity in game-based situations: Analysis of stress hormones, negative emotions, and working memory. *Computers & Education* 81 (2015), 143–153.
- [69] X. Zhang, Y. Wang, and Ray P.S. Han. 2010. Wavelet Transform Theory and its Application in EMG Signal Processing. In *Proc. of IEEE International Conference on Fuzzy Systems and Knowledge Discovery*. Vol. 5. 2234–2238.
- [70] M. Zivanovic and M. González-Izal. 2013. Simultaneous powerline interference and baseline wander removal from ECG and EMG signals by sinusoidal modeling. *Medical Engineering and Physics* 35, 10 (2013), 1431–1441.