A Machine Learning Approach to Automatic Phobia Therapy with Virtual Reality

Oana Bălan, Alin Moldoveanu, and Marius Leordeanu

1 Phobias: Statistics, Effects, and Treatment

Phobia is a type of anxiety disorder manifested through an extreme, uncontrolled, and irrational fear that appears when the subject is exposed to certain stimuli—a naturalistic situation, the presence of people, animals, or objects. There are different types of phobias, such as *agoraphobia* – fear of crowds or open spaces, *social phobias*—fear of speaking in public, meeting people of higher authority, eating or using the telephone in front of others, and *specific phobias*—caused by various objects and situations (World Health Organization 2017). Social phobias affect all age categories, but the onset is usually in the adolescence (95% begin before the age of 20). In what concerns the sex categories, women are more affected than men (Olesen 2015). Also, anxiety disorders are more common in women—4.6% at the world level, compared to 2.6% in men. As for specific phobias, they occur at least once in a lifetime for 15–20% of the world's population (Olesen 2015). They have the following prevalence at the world level: acrophobia (fear of heights)—7.5%, arachnophobia (fear of spiders)—3.5%, aerophobia (fear of flying)—2.6%, astraphobia (fear of lightning and thunder)—2.1%, dentophobia (fear of dentist)—2.1% (Nation Wide Phobias Statistics 2019). Some phobias are connected: for example, acrophobia is related to fear of elevators and fear of flying (Muris et al. 1999). Specific phobias usually appear in childhood and prolong throughout the entire life (Olesen 2015). For acrophobia, various researchers supported hereditary and *nonassociative* factors in the development of this anxiety disorder, as the subjects were unable to account for a height-related experience triggering acrophobia. The subjects from the control group did not develop acrophobia, even if they have been exposed to heights (Menzies and Clarke 1993, 1995). Besides, Poulton et al. (1998) and Poulton and

O. Bălan (✉) · A. Moldoveanu · M. Leordeanu
Politehnica University of Bucharest, Bucharest, Romania
e-mail: oana.balan@cs.pub.ro

I. Opris et al. (eds.), *Modern Approaches to Augmentation of Brain Function*,
Contemporary Clinical Neuroscience,
https://doi.org/10.1007/978-3-030-54564-2_27

Menzies (2002) showed that the lowest incidence of acrophobia was encountered for those who suffered heights-related injuries in childhood. Moreover, Menzies and Parker (2001) reported that the non-phobic subjects had the highest incidence of traumatic falls, without affecting their perception on heights. Many other studies supported the non-conditioning theory (Field et al. 2001; Graham and Gaffan 1997; Withers and Deane 1995) and claimed that phobias emerge as a result of other experiences that cannot be recalled or consciously brought into memory.

A phobia crisis causes both physical and emotional symptoms. Among its physical manifestations, we account for high heart rate, sweating, tremor, rapid breathing, or dizziness. On the other hand, the emotional symptoms could include anxiety attacks and difficulty in controlling one's emotional state despite intense efforts. The treatment for phobias comprises of medication (antianxiety and antidepressive drugs), in-vivo exposure in a controlled environment, Cognitive Behavioral Therapy (CBT), and virtual exposure. In 1958, Wolpe (1958) developed a technique called "systematic desensitization", based on deep mental and muscular relaxation. In 1977, Bandura (1977) proposed the "self-efficacy" theory that relies on one's confidence and personal judgment about the ability of overcoming the stressful stimuli. Another model of therapy was "reinforced practice", based on a continuous practice and improvement of the responses to certain therapeutic elements, such as the attitude toward stimuli, feedback to the therapist, and self-control (Leitenberg 1976). Ritter (1969) introduced the "contact desensitization" therapy, where the patient was assisted by the therapist who held his hand or arm during exposure. The desensitization method provided good results when the therapist behaved warmly or not with the patient (Morris and Magrath 1979) and even when the therapist was not present in the room—a tape recorder played the instructions for treatment (Baker et al. 1973). CBT is a strategy that encourages the subjects to change their attitude toward the aversive experience by replacing negative thoughts with positive ones. Only 23% of the people suffering from phobias seek treatment, especially medication and CBT. The study of Steinman and Teachman (2014) showed that CBT has the same rate of success for treating acrophobia as in-vivo exposure to heights.

2 Virtual Reality in Phobia Therapy

Virtual reality (VR) was used since the 1990s in phobia therapy. It benefits from some practical advantages such as a better control of the exposure, possibility to render situations that are not easily accessible, ability to provide stimuli of lower or higher magnitudes than in real-world settings (Choi et al. 2001), higher comfort for both the patient and the therapist, confidentiality, friendly environment, suitable especially for those who do not possess imaginative skills (Coelho et al. 2009). The idea of developing virtual worlds for training purposes dates back to the 1940s, when the American government invested in flight simulators in the context of the Second World War (Littman 1996). *Virtual Reality Exposure Therapy* systems (VRETs) emerged in 1996 when North and North (1996) observed that a flight

simulating a virtual environment produced fear responses that were not associated with motion sickness.

According to Garcia-Palacios et al. (2002), 90% of respondents preferred VR exposure than in-vivo exposure for arachnophobia therapy. In Garcia-Palacios et al. (2001), over 80% of patients opted in favor of virtual exposure for acrophobia therapy. VRET systems offer similar results in the posttreatment assessments, comparable to those provided by CBT, with strong real-life impact and good stability of results in time (Opris et al. 2012).

In the virtual environment described in Hodges et al. (1995), there have been designed three situations to be used in acrophobia therapy: an elevator, a balcony, and a bridge. The participants were randomly divided into two groups: a treatment and a control group. The subjects from the treatment group used virtual therapy and they were free to spend as much time as they wanted in various sessions. The subjects from the control group received no treatment; they were only subjected to two evaluations after 7 weeks. The results of the study have shown that the VR-based treatment was as effective as traditional therapy (Hodges et al. 1995).

The quantitative meta-analysis performed in (Parsons and Rizzo 2008) highlighted that VRET has potential in treating anxiety and certain phobias, including acrophobia. The results of a VRET-based study are presented in Shiban et al. (2015), where the issue of Return of Fear (ROF) after successful treatment was thoroughly approached. All participants completed both a VR test and an in-vivo Behavioral Avoidance Test (BAT). The results of a meta-analysis demonstrated that VRET can produce significant behavior changes in real-life situations that support its applicability in treating specific phobias (Morina et al. 2015).

A comparison between an Augmented Reality (AR) and a VR system including acrophobic scenarios is presented in Carmen Juan and Perez (2010). There were no significant differences regarding the therapy results. In AR, the participants could see their hands, feet, and the scene is real, while in VR, all of these are simulated (Buna et al. 2017).

Nowadays, new and sophisticated VR devices have emerged. Their low price makes them affordable, so that they can be successfully used to build immersive VR environments for treating certain phobias, including acrophobia (Buna et al. 2017). The Climb (Robertson 2016) is a game that can be played on the Oculus device (Oculus Rift n.d.) with input from the Xbox gamepad. Richie's Plank Experience (n.d.) is a game for HTC Vive (n.d.) employing a customizable real plank replicated in the virtual environment. C2Phobia (n.d.) treats acrophobia. The player can exit on the balcony, take a transparent elevator, or move from one building to another using walkways with low walls, ropes, or without any protection. In the Stim Response Virtual Reality system (2BIOPAC n.d.), the events from VR and the physiological data are synchronized in real time and the scenes are adapted according to the player's biophysical output. A component of this system is VR-Acrophobia, which is fully modular and customizable, so that the therapist can create and recreate various scenes. The Virtual Reality Medical Center (VRMC) (Virtual Reality Medical Center n.d.) uses 3D computer simulation, biofeedback, and CBT to treat phobias and anxieties.

The above-mentioned systems demonstrate the advantages of using VR for treating phobias. However, they can only be used under medical surveillance, with guidance from a physician or psychologist.

3 Our Main Contributions

Our scientific contributions are presented in more detail in the next sections of the chapter. We also briefly enumerate them here:

1. We perform a comparison of several machine learning techniques (Support Vector Machine, Linear Discriminant Analysis, Random Forest, and *k*-Nearest Neighbors and 4 deep neural networks with different numbers of layers and neurons per layer), with and without feature selection, for classifying the six basic emotions (anger, joy, surprise, disgust, fear, and sadness). We classified the emotion of fear in two ways: first, a binary classification called the 2-level paradigm (0—no fear and 1—fear) and secondly, the 4-level paradigm (0—no fear, 1—low fear, 2—medium fear, 3—high fear).
2. We introduce the stages of development and evaluation of a virtual environment for treating acrophobia that relies on gradual exposure to stimuli, accompanied by physiological signals monitoring in a pilot experiment which involved the participation of 4 acrophobic subjects. Then, we present the design and development of a VR environment for acrophobia therapy in a naturalistic scenario—a mountain landscape (Fig. 1).
3. We introduced a novel approach toward using an intelligent virtual therapist that recognizes human emotions based on biophysical signals, provides encouragement, gives advice, changes his voice parameters, and adapts the scenario according to the subject's affective state.
4. We design a novel method for reducing in-game artifacts which consists in recognizing artifact patterns in the signals recorded during gameplay sessions, by

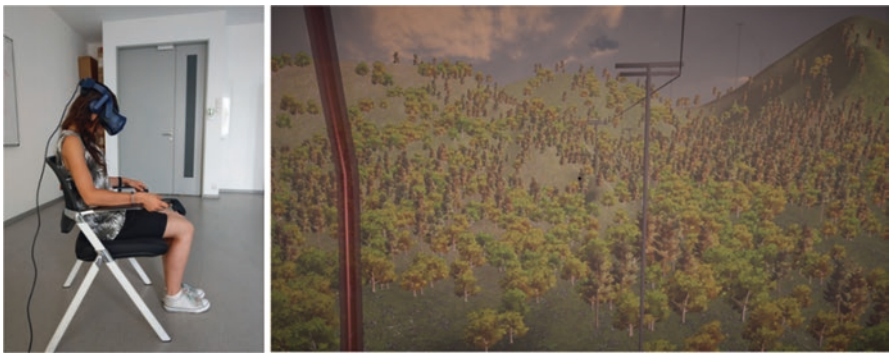

Fig. 1 User playing and the VR game—view from the cable car

aligning the biophysical data segments corresponding to the moments when the users performed head/hand/body movements with the artifacts recorded during a reference procedure.
5. Lastly, we introduce an approach for estimating respiration rate which consists in placing two HTC Vive trackers on the chest and on the back of the subjects and measure the distance between them. This distance varies during breathing—increases while inhaling and decreases during exhaling.

4 Emotion Models

Various emotion models have been issued throughout the years. The *discrete model*, proposed by Paul Ekman, consists of six basic emotions: sadness, happiness, disgust, anger, fear, and surprise (Ekman et al. 1969). The most well-known model for emotion classification is the *bipolar model* (Russell 1979). It considers two orthogonal dimensions, *arousal and valence*. Arousal ranges from "not excited" to "excited", while valence extends from "negative" to "positive". A third dimension, *dominance*, indicates how much the subject is in control over his emotions. Each emotion can be described as a combination of these three dimensions. For instance, fear is characterized by low valence, high arousal, and low dominance (Demaree et al. 2005). The *approach-withdrawal model* takes into account the motivating factor of emotions, reflecting the tendency to reach or reject a certain stimulus or situation (Davidson et al. 1990).

5 Biophysical Data

Biophysical data analysis is a more objective method of interpreting and assessing human emotions, compared to questionnaires or subjective ratings (Toth 2015). However, if used together, a wider perspective on the modality in which people decode the affective states can be obtained. According to Steimer (2002), fear causes a defensive behavior. The human body responds differently to fear, in an either active (high heart rate, increased sweat production, cortical activation) or passive modality (low pulse and respiration rate) (Kometer et al. 2010).

5.1 Galvanic Skin Response

Galvanic Skin Response (GSR) or *Electrodermal Activity* refers to a change in sweat glands activity or skin conductance, measured by electrodes applied on the skin. GSR has two components—a tonic (Skin Conductance Level—SCL) and a phasic one (Skin Conductance Response—SCR, a measure of arousal to stimuli, reflected

in changes in the sympathetic nervous system's level of activation). Fear is mapped by an increase in the production of sweat and, consequently, in skin conductance (DiMeglio 2015). GSR has been intensively used in psychophysiological experiments, with high rates of success—in Healey (2009) and Fleureau et al. (2012) it has been the main classification factor for emotions, while in Westerink et al. (2009), the changes in GSR have been in line with the changes in arousal and also a comfortable type of measurement for the users, reliable in discriminating fear from other negative affective states (AlZoubi et al. 2012). The typical baseline values are around 0.03 and 0.05 microSiemens, while threatening stimuli produce a raise to around 2–3 microSiemens or extreme values of 8 microSiemens (Braithwaite et al. 2015). The subjects who watched a scary 2D video measured 8.05 microSiemens (1.57% ± 12.10 increase from baseline) while those who viewed a horror virtual reality video recorded on average 11.814 microSiemens (4.26% ± 6.31 increase from baseline) (Kometer et al. 2010).

5.2 Blood Volume Pulse

Blood volume pulse reflects the changes in the volume of the blood vessels and is recorded by a photoplethysmography (PPG) device, a noninvasive optical sensor that determines changes in the light absorption density of the skin (Agrafioti et al. 2012). PPG has been used in various experiments as a reliable estimator of emotional changes (Eum et al. 2011; Gouizi et al. 2011; Walter et al. 2013), being usually attached to the ear lobe or to the finger. Its values are converted into heart rate, measured in beats per minutes (bpm). High values of heart rate, over 90–100 bpm, indicate fear and anxiety (Wen et al. 2014; Rainville et al. 2006). The average heart rate was 80 bpm when the subjects watched a scary 2D video with a 6.97% ± 12.74 increase from baseline and 77.8 bpm with a 3.49% ±12.09 increase from baseline for those who watched a horror VR video (Kometer et al. 2010).

5.3 Electroencephalography

Electroencephalography (EEG) is a technique of recording and interpreting the electrical activity of the brain using electrodes placed on the scalp.

In the brain, fear is perceived first by the amygdala and then goes through the hypothalamus and midbrain (Quirk 2013). The right lobe mediates withdrawal, while the left side of the brain is involved in appetitive emotions and approach (Mauss and Robinson 2009). Phillips et al. (2003) pointed to a 2-way circuit for emotion regulation: a ventral one (including the amygdala, responsible for the identification of stimuli emotional significance) and a dorsal one (including the hippocampus, responsible for the regulation of affective states and behavior). In Petrantonakis and Hadjileontiadis (2009) and Chanel et al. (2011), it has been found

that EEG was more reliable in fear classification than other biophysical features. EEG is commonly very susceptible to outside noise, especially body movements and artifacts introduced by the recording devices, but advanced filtering methods have emerged in order to remove them and obtain clearer signals. In a recent work (Cudlenco et al. 2020), it has been shown that EEG could also be used to predict the semantics of the visual input perceived by the human subject, even though the prediction is highly accurate only when combined with deep visual features directly extracted from the image.

The alpha waves (8–12 Hz) are neural oscillations that originate from the occipital lobe, being a reflection of the relaxation state of the individual, with high amplitudes when he has his eyes closed. Moreover, it has been demonstrated that the alpha waves are a marker of functional inhibition of the brain areas, involved in attentional processes (low alpha activity in the regions that are processing information and high alpha activity in the regions that are not involved in the current task) and anticipation of upcoming stimuli (Horschig et al. 2014).

When the subject performs mental processes, a phenomenon called *alpha blocking* occurs, which is reflected in a decrease of alpha amplitudes (Scott 1976). The cognitive states and the level of alpha waves are inversely related. The alpha waves have their origin in the occipital cortex and advance to the frontal lobes, the most evolved area of the brain, responsible for emotion, consciousness, and behavior. Usually there is a balance of alpha activation between the two hemispheres, but this balance impairs when emotional stimuli are provided. According to the approach/withdrawal model of *frontal alpha asymmetry* (Davidson 1993), left frontal activation corresponds to a positive approach to stimuli, while right frontal brain activation indicates negative affective responses (Bos 2006; Trainor and Schmidt 2003; Jones and Fox 1992; Canli et al. 1998). Both left and right activation correspond to low alpha levels.

The beta waves (13–30 Hz) are neural oscillations indicating wakefulness and consciousness, with average amplitudes around 20–200 μV. High levels of the beta waves indicate anxiety, alert, and fear (Arikan et al. 2006). In Kometer et al. (2010), for the beta band, horror virtual reality gameplay led to an increase of 33 μV from baseline.

The ratio of slow waves to fast waves (SW/FW) has a negative correlation with fear (Schutter and Van Honk 2005; Putman et al. 2010). There was a statistically significant reduction in the SW/FW ratio (delta/beta and theta/beta) in the left frontal lobe in an experiment where the EEG data has been recorded from a single electrode (Cheemalapati et al. 2016).

As suggested by Brouwer et al. (2015), body movements, mental states, subtle movements of sensors and wires are confounding factors that can affect the estimation of cognitive or affective states from neurophysiological signals. It is advisable to correctly detect and remove artifacts from the classification analysis and perform experiments where little movement of the body or recording devices is involved.

5.4 Biophysical Data and Virtual Reality

A Magnetic Resonance Imaging (MRI) experiment showed an increased level of emotional responses in the amygdala when VR stimuli have been presented to the subjects, compared to 2D videos (Dores et al. 2014). Coelho et al. (2008) found that movement in an acrophobia-simulated virtual environment conducted to a more realistic behavior of the subjects, similar to what has been observed in in-vivo exposure. In Costa et al. (2014), EEG and GSR data have been collected in real time while the users played an acrophobia-oriented game with a CAVE device. A VR system for treating stress-related disorders has been developed in Brouwer et al. (2011). Stress was induced by depicting a scenario simulating a bomb explosion, while associative stress was measured by immersing again the user in the scene after a period of time. Associative stress has been related to EEG mid-frontal alpha asymmetry and to an increase in heart rate variability.

6 Machine Learning Techniques for Emotion Classification

Emotion classification has been performed using various machine learning and feature selection algorithms in psychophysical experiments. In Koelstra et al. (2012), Fisher's linear discriminant was used for feature selection and the Naïve Bayes classifier for discriminating into low/high valence, arousal, and liking, with accuracies of 62%, 56%, and 55%. Atkinson and Campos (2016) used the minimum-Redundancy Maximum-Relevance (mRMR) method for feature selection and Support Vector Machines (SVM) for binary classification into low/high valence and arousal, with an accuracy of 73% for both. The study has been performed by extracting and processing the EEG features from the DEAP database. Yoon and Chung (2013) used the Pearson correlation coefficient (PCC) for feature extraction and a probabilistic classifier based on the Bayes theorem for resolving the binary classification problem of low/high valence and arousal discrimination, with an accuracy of 70% for both. Similarly, emotion recognition has been done based on the EEG data from the DEAP dataset. A similar approach is presented in Naser and Saha (2013), where the SVM algorithm conducted to an accuracy of 66%, 64%, and 70% for classifying valence, arousal, and liking into the low and high groups. By applying the SVM technique on the EEG features, a classification accuracy of 62 and 69% has been achieved during a music-induced affective states evaluation experiment where the users were required to rate their currently perceived emotion in terms of valence and arousal (Daly et al. 2015). Liu and Sourina (2013) conducted two experiments in which visual and audio stimuli have been used to evoke emotions. The SVM classifier, having as input Fractal Dimension Features, statistical and Higher Order Crossings extracted from the EEG signals provided the best accuracy of 53% for recognizing eight emotions—happy, surprised, satisfied, protected, angry, frightened, unconcerned, and sad. A comparative study of four machine

learning methods showed that SVM offered the best accuracy—85%, followed by Regression Tree—83% for the classification of five types of emotions—anxiety, boredom, engagement, frustration, and anger into three categories—low, medium, and high (Liu et al. 2005). Soleymani et al. (2009) obtained an accuracy of 63% for differentiating three classes of emotions—calm, positive excited, and negative excited using a Bayesian classification method. A more complex SVM-based algorithm did not show improvements, compared to the Bayesian technique. In the case of binary classification into low/high valence, arousal, and liking, using EEG signals, the accuracy rates were 55%, 58%, and 49% with SVM. Having as input features the peripheral physiological responses, the classification accuracies recorded 58%, 54%, and 57% (Koelstra et al. 2010). Based on the MAHNOB dataset and using the SVM algorithm, Wiem and Lachiri (2017) reached a classification accuracy of 68% for valence and 64% for arousal when discriminating into the low/high groups and 56%, respectively, 54% for classifying into three groups. The most relevant features were the electrocardiogram and the respiration volume. In Alhagry et al. (2017), a deep learning method based on the Long-Short Term Memory (LSTM) networks was used for classifying low/high valence, arousal, and liking based on the EEG raw data from the DEAP dataset (Koelstra et al. 2012), with accuracies of around 85%. Jirayucharoensak et al. (2014) trained a deep neural network implemented with a stacked autoencoder based on the hierarchical feature learning approach. The input features were the power spectral densities of the EEG signals from the DEAP database, which were selected using Principal Component Analysis (PCA). The subjective ratings from 1 to 9 have been divided into three levels and mapped into “negative”, “neutral”, and “positive” for valence and into “passive”, “neutral”, and “active” for arousal. They were finally classified with an accuracy of 49% for valence and 46% for arousal.

7 Our Machine Learning Approach to Classifying the Six Basic Emotions

The evaluation of the users’ emotional states is fundamental in VRET systems. In order to address the issue of emotion classification based on biophysical signals in terms of maximum accuracy and feature selection efficiency, we performed a comparison of several machine learning and deep learning techniques applied on the data from the DEAP database (Koelstra et al. 2012). The renowned DEAP database contains physiological recordings (GSR, PPG, skin temperature, breathing rate, electromyogram, data from 32 EEG channels) and subjective ratings from 32 subjects who watched 40 short videos eliciting various emotions. The participants were required to rate each video in terms of valence, arousal, and dominance on a scale from 1 to 9. By *combining the discrete model of emotions and the three-dimensional continuous model* (Ekman et al. 1969), we classified each of the six basic emotions into two groups—positive (the existence of emotion) and negative (lack of emotion),

using four classical machine learning techniques (Support Vector Machine—SVM, Linear Discriminant Analysis—LDA, Random Forest—RF, and k-Nearest Neighbors—kNN) and four deep neural networks with different numbers of layers and neurons per layer, with and without feature selection. The feature selection algorithms were: Fisher score, Principal Component Analysis (PCA), and Sequential Forward Selection (SFS). Classification has been done based on the physiological data and subjective ratings of valence, arousal, and dominance from the DEAP database. From the EEG data, we extracted the Petrosian Fractal Dimension, Higuchi Fractal Dimension, and Approximate Entropy. The machine learning and deep learning algorithms have been trained and cross-validated, having as input the biophysical data and as output, two possible conditions: *0—negative or lack of emotion* and *1—positive or emotion*. The deep neural networks have been cross-validated using the *k*-fold and leave-one-subject-out methods, while for the machine learning techniques, the data have been divided into 70% training data and 30% test data. In the case of leave-one-subject-out, each classifier has been trained on the data of 31 subjects and tested on the data of the 32th user. Figure 2 presents the distribution of each of the six basic emotions—sadness, happiness, disgust, anger, fear, and surprise across the valence-arousal-dominance axis in the 3-dimensional continuous model of valence-arousal-dominance, as proposed by Russell and Mehrabian (1977).

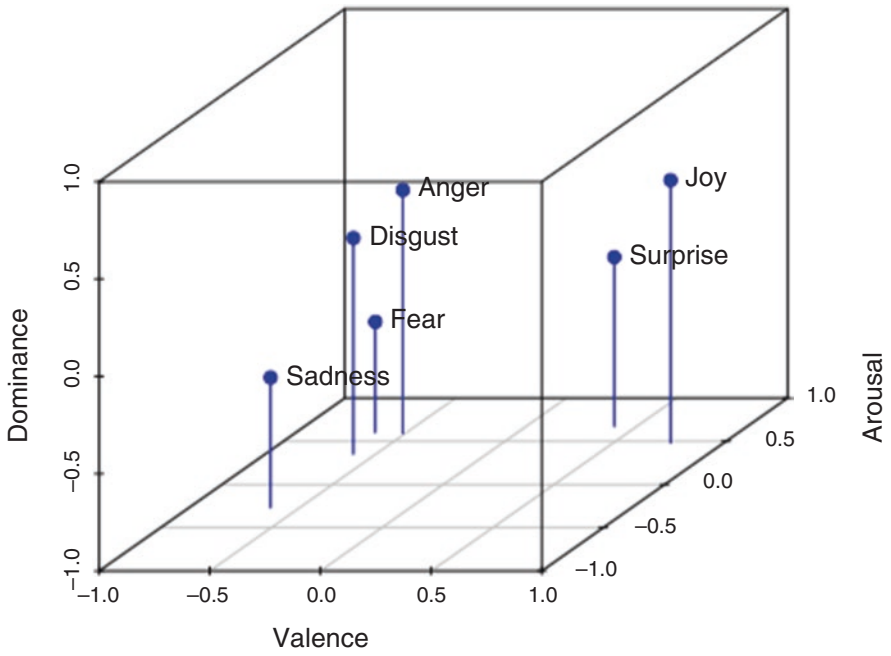

Fig. 2 The six basic emotions in the valence-arousal-dominance model

As the valence, arousal, and dominance dimensions have been rated on a scale from 1 to 9, we considered the following correspondences for each of the six basic emotions in terms of condition 1 (positive or the existence of emotion):

(a) Anger—low valence ([1; 5)), high arousal ([5; 9]), dominance in the interval [6;7],
(b) Joy—high valence ([5; 9]), high arousal ([5; 9]), dominance in the interval [6;7],
(c) Surprise—high valence ([5; 9]), high arousal ([5; 9]), dominance in the interval [4; 5],
(d) Disgust—low valence ([1; 5)), high arousal ([5; 9]), dominance in the interval [5; 6],
(e) Fear—low valence ([1; 5)), high arousal ([5; 9]), dominance in the interval [3; 4],
(f) Sadness—low valence ([1; 5)), low arousal ([1; 5)), dominance in the interval [3; 4].

The classification results showed that the highest F1 cross-validation scores were:

(a) Anger—Petrosian and Higuchi Fractal Dimension extracted from the EEG signals and peripheral biophysical data, using SVM (98.02%),
(b) Joy—kNN using Petrosian EEG values and peripheral data (87.9%),
(c) Surprise—kNN with raw EEG values and peripheral data (85.01%),
(d) Disgust—kNN with Petrosian EEG values and peripheral data (95%),
(e) Fear—kNN with raw EEG values and peripheral signals (90.75%),
(f) Sadness—SVM with Higuchi Fractal Dimensions extracted from the EEG signals and peripheral data (90.8%).

The *k*fold method provided higher F1 scores than Leave-One-Subject-Out. Without feature selection, kNN has provided the highest F1 scores in 13 cases, followed by Random Forest (seven times) and SVM (four times). With feature selection, kNN has provided the highest F1 scores in 12 of the tested cases, Random Forest in seven cases, SVM in five cases, and LDA only once. SFS has been selected two times and Fisher score 14 times. The most important classification features were:

(a) Anger—trapezius electromyography (EMG) and respiration rate,
(b) Joy—GSR and zygomaticus EMG,
(c) Surprise—GSR and FC1,
(d) Disgust—vertical and horizontal electrooculography (EOG),
(e) Fear—vertical and horizontal EOG, zygomaticus EMG, and activation of the frontal cortex (FC1, F4),
(f) Sadness—the left prefrontal cortex (FC1 and FP1).

The results obtained for classifying into two classes (the existence of emotion—positive condition and lack of emotion—negative condition) are higher than those obtained in the literature for classifying into low/high valence and arousal: 62%/56% (Koelstra et al. 2012), 73% (Atkinson and Campos 2016), 70% (Yoon and Chung 2013), 85% using the Long-Short Term Memory algorithm (all using the data from

the DEAP database), 66%/64% (Daly et al. 2015), 62%/69% (Liu and Sourina 2013), 55%/58% (Soleymani et al. 2009), 68%/54% using the data from the MAHNOB database (Wiem and Lachiri 2017). A thorough description of the experiment, methods, and results, including a comparison between the cross-validation F1 scores achieved using the *k*fold and leave-one-subject-out techniques is provided in Bălan et al. (2020a).

8 Our Machine Learning Approach to Fear Level Classification

Using the same machine learning, deep learning, and feature selection algorithms applied on the biophysical recordings and subjective ratings from the DEAP database, we classified the emotion of *fear* in two ways: first, a binary classification called the *2-level paradigm (0—no fear and 1—fear)* and secondly, the *4-level paradigm (0—no fear, 1—low fear, 2—medium fear, 3—high fear).* Considering the emotion dimensions from the 3-dimensional continuous model of emotions, fear was characterized by *low valence, high arousal, and low dominance*. The recordings have been assigned to either the 0—no fear or 1—fear group (in the case of the 2-level fear evaluation paradigm) or to the 0, 1, 2, or 3 classes (for the 4-level paradigm), considering the subjective ratings of valence, arousal, and dominance from the DEAP dataset. We applied the unsupervised K-means clustering algorithm on the data from DEAP and achieved a prediction accuracy of 87% for the 2-level evaluation modality. This means that 87% of the ratings proposed for the *fear* or *no fear* classes by the theory of low valence/high arousal/low dominance have been classified in the same cluster by the k-means technique. We used for training and cross-validation not only the raw EEG values, but also the peripheral signals. The EEG recordings have been decomposed into Power Spectral Densities of the alpha, beta, and theta frequencies, Petrosian Fractal Dimensions, Higuchi Fractal Dimension, and Approximate Entropy. *The highest F scores have been obtained by using the Random Forest Classifier—89.96%, having as input EEG Higuchi Fractal dimensions and peripheral data for the 2-level fear evaluation modality and 85.33% for the 4-level fear evaluation modality, both without feature selection. The most important classification features were the raw, alpha, and beta values in the left frontal hemisphere, GSR, and respiration rate*. We computed the difference in spectral power between the right and the left frontal hemispheres, for the alpha and theta frequency bands. We noticed that this difference increases with fear, being higher for the *medium fear* and *high fear* condition than for the *no fear* and *low fear* conditions. There was a higher level of alpha and theta activation in the left side of the brain. Moreover, the intensity of the left central and right frontal beta waves was directly associated with fear onset. In addition, a positive correlation between fear and the ratio of slow to fast waves has been observed, for both the 2-level and 4-level evaluation modalities. The purpose of this research approach was not just to

classify the data into low/high valence/arousal/dominance, but also to combine these emotion dimensions and define a complex emotion such as fear. The results we obtained were similar to Alhagry et al. (2017), who achieved a classification accuracy of 85% by training and testing a Long-Short Term Memory network using raw EEG values. In the paper published in the Sensors journal in April 2019, we provide a full presentation of the research, results, and comparison with similar studies (Bălan et al. 1738).

9 Fear Level Classification in a VRET System for Acrophobia Therapy

In an experiment performed during June–August 2018, we trained and tested two classifiers: C1, which determines the patient's *current level of fear* and C2, which estimates *the next scenario of exposure* in a VR-based game for treating acrophobia. For this, we have collected biophysical data (EEG in the alpha, beta, and theta frequency ranges, GSR, and HR) from four subjects suffering from acrophobia, aged 22–50, in both in-vivo and virtual conditions. The subjects have been exposed to heights at the first, fourth, and sixth floors of a building, at 4 m, 2 m, and a few centimeters away from the balcony's railing. Besides, they played a VR-based game where they had to collect bronze, silver, and gold coins at the ground level and on terraces at the first, fourth, and sixth floor, as well as on the building's rooftop (Fig. 3).

During each trial of a session, the participants had to rate their perceived level of fear on a scale from 0 to 10 (the 11-choice scale), where 0 represents no fear at all and 10 stands for a high level of anxiety. The ratings on the 11-choice scale have been divided into the 2-choice and 4-choice scales. In the 2-choice scale, 0 means *relaxation* and 1 means *fear*. In the 4-choice scale, 0 stands for *relaxation*, 1—*low fear*, 2—*medium fear*, 3—*high fear* (Table 1).

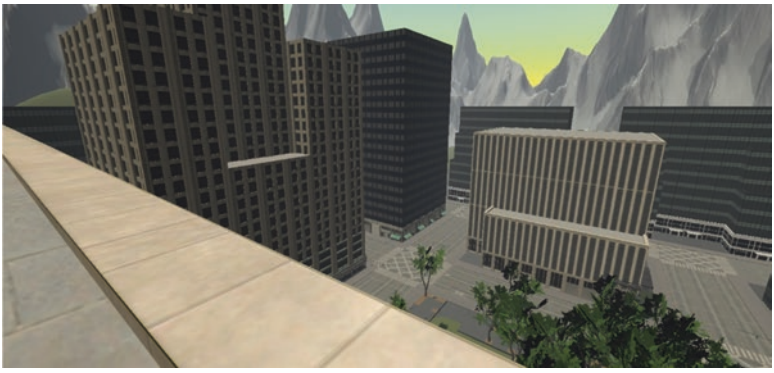

Fig. 3 Our computer game for automatic acrophobia treatment

Table 1 Fear level classification scales

11-Choice-scale	4-Choice-scale	2-Choice-scale
0	0 (relaxation)	0 (relaxation)
1	1 (low fear)	
2		
3		
4	2 (medium fear)	1 (fear)
5		
6		
7		
8	3 (high fear)	
9		
10		

The data recorded during this preliminary experiment have been used for training classifier C1, on the 2-choice, 4-choice, and 11-choice scales. Classifier C1 received as input the EEG, GSR, and HR data and provided as output the perceived fear level. We have performed a comparative study of various classic machine learning and modern deep learning techniques as classification models: k Nearest Neighbors, Linear Discriminant Analysis, Random Forest, Support Vector Machine (with and without feature selection using the Sequential Forward Selection algorithm), and four types of deep neural networks—DNN_Model_1: 3 hidden layers, with 150 neurons on each hidden layer, DNN_Model_2: 3 hidden layers, with 300 neurons on each hidden layer, DNN_Model_3: 6 hidden layers, with 150 neurons on each hidden layer, DNN_Model_4: 6 hidden layers, with 300 neurons on each hidden layer. Our purpose was to automatically adapt the exposure scenarios according to the user's level of fear. For instance, if the patient is anxious and feeling as losing control of his emotional reactions, the level of exposure should be lowered. On the contrary, if he is in a relaxed state, the level of exposure should be increased. Classifier C2 automatically determines the next level of exposure, by taking into account the physiological data (EEG, GSR, and HR) and a parameter called target fear level (FL_t), computed using the current level of fear (FL_{cr}). FL_{cr} is determined using classifier C1. We have used the following formulas for the 2-choice and 4-choice scales:

2-Choice scale	4-Choice scale
If $FL_{cr} = = 0$ then $FL_t = 1$ If $FL_{cr} = = 1$ then $FL_t = 0$	If $FL_{cr} = = 0$ or $FL_{cr} = = 1$ then $FL_t = FL_{cr} + 1$ If $FL_{cr} = = 2$ then $FL_t = FL_{cr}$ If $FL_{cr} = = 3$ then $FL_t = FL_{cr} - 1$

For classifier C2, we have used the same machine learning and deep learning algorithms as for C1 in our comparative study. For testing the accuracy of both classifiers, the acrophobic subjects have been required to play the VR game two times. Each session had a number of 10 trials. The game started at the ground floor where

they had to collect bronze, silver, and gold coins, rate their perceived level of fear for ground truth acquisition and then, based on the FL_{cr} parameter estimated by C1 in real time and the computed Fl_t, classifier C2 determined the next level of the game where the players should be taken to. Classifier C1 has been cross-validated on the training dataset using the *k*fold method ($k = 10$) and tested on the test dataset obtained in the second experiment. On the other hand, classifier C2 has been only cross-validated on the training dataset acquired in the second experiment. As for now, we did not define a method for evaluating the test accuracy of C2. In the future, we will perform an experiment with a larger number of people and evaluate the therapeutic procedure. Thus, the users will play the VR game several times, across a certain number of days and then they will be exposed in real-world conditions to see whether their fear of heights has diminished. We consider that only by a final in-vivo exposure we can assess the efficiency of the VR therapy. We have used a *user-dependent* and a *user-independent* modality for assessing the classifiers' accuracy. In the case of the user-dependent modality, each classifier has been trained and tested on the same data—for each subject, on his own recordings. As for the user-independent modality, for each subject, each classifier has been trained on the data of the other three participants and tested on the recordings of the current subject. The highest cross-validation and test accuracies are presented in Table 2.

The results showed a very high cross-validation accuracy on the training set and good test accuracies, ranging from *42.5 to 89.5%*. For the 2-choice scale, the highest accuracy has been obtained by DNN_Model_4 (79.12%) for the player-independent

Table 2 Highest cross-validation and test accuracies

Method	**C 1**				
	2-Choice scale		**4-Choice scale**		**11-Choice scale**
	Cross-validation	**Test**	**Cross-validation**	**Test**	**Cross-validation**
Player-independent	kNN 99.5% RF 99.25%	DNN_ Model_4 79.12%	kNN 99% RF 99%	kNN 52.75%	kNN 98.25% RF 99%
Player-dependent	kNN 99.5% RF 99.75%	SVM 89.5%	kNN 99% RF 99.25%	SVM 42.5%	kNN 98.25%RF 99%
	C 2				
	2-Choice scale		**4-Choice scale**		**11-Choice scale**
	Cross-validation	**Test**	**Cross-validation**	**Test**	**Cross-validation**
Player-independent	RF 99.75%	–	RF 100%	–	RF 100%
Player-dependent	RF 99.75%	–	RF 99.75%	–	RF 100%

modality and SVM (89.5%) for the player-dependent modality. For the 4-choice scale, the highest accuracies were obtained using kNN (52.75%, player-independent modality) and SVM (42.5%, player-dependent modality).

The Radom Forest classifier adds the benefit of computing *feature importance*—how important is that feature for reducing impurity across the decision trees. For classifier C1, the most important features were *GSR, HR, and the EEG values in the beta frequency range, closely followed by the alpha and theta power spectral densities*. These findings are comparable to the results from other experiments (Arikan et al. 2006; Kometer et al. 2010). For C2, the most significant feature resulted to be $\mathbf{FL_t}$. It had a high importance index and also has been selected on all three fear estimation scales (2-choice, 4-choice, and 11-choice), for both the user-dependent and user-independent modalities.

Our results are comparable to those obtained by Liu et al. (2009), who reached a classification accuracy of 78% in a game where dynamic difficulty adjustment depended on simple "if" clauses and not on an automatic computation. Chanel et al. (2011) obtained an accuracy of 63% for classifying three classes of emotions in a study where 20 subjects played a Tetris game on three difficulty levels. In Hu et al. (2018), a convolutional deep neural network was used to classify fear ratings on a scale from 1 to 4. The EEG data of 60 subjects have been recorded while playing the Ritchie's Plank Experience VR game, with a classification accuracy of 88.77%. The system described in Šalkevicius et al. (2019) was used in the therapy of fear of public speaking. The GSR, blood volume pulse, and skin temperature of 30 subjects have been recorded and the current level of anxiety has been classified into four classes: low, mild, moderate, and high, using the SVM algorithm. The fusion of all three types of biophysical signals provided a classification accuracy of 86.3%.

10 Acrophobia Game in Naturalistic Landscape

During 2019, we refined the VRET system for acrophobia therapy, considering Jerald's statement: *We must create VR experiences with both emotion and logic*. For this, we adopted the Human-Centered Machine Learning approach that takes into account human interests in designing Machine Learning algorithms, making Machine Learning more useful and usable. According to this theory, humans and machines not only cooperate, but also adapt to each other—humans are able to alter the behavior of the machines and the machines modify human goals (Jerald 2016).

The VR game is rendered via the HTC Vive Head Mounted Display and contains a mountain environment with three scenes: a walk by foot—incorporating a path, a transparent platform across a canyon and a bridge, a ride by cable car (Fig. 4), and one by ski lift.

There are ten stops throughout each ride, where the user is asked to rate his fear level, valence, arousal, and dominance using the Self-Assessment Manikin, on a scale from 1 to 5 (Fig. 5).

Fig. 4 Our acrophobia computer game: view from the cable car

Fig. 5 Our acrophobia computer game: arousal rating

Also, the player has to give the answer to some short mathematical exercises. Logical thinking decouples cortical activation in the right brain hemisphere which is responsible with emotional processing. In this way, the subject begins to feel more detached from the anxiety-provoking experience, relaxes, and gains confidence. At this stage, in order to explore the environment and spend as much as time as possible immersed, the player is required to collect some small objects (stars, diamonds, and coins) that appear randomly and disappear as he fixes his gaze toward them. At any time, he can make use of some assistive elements—he can pause the game and listen to his favorite song, read an inspirational quote, and look at a nice

picture. These elements are configured personally for each user apart and saved in his database profile.

We will perform a series of experiments to evaluate the efficacy of the VR environment in treating acrophobia. The participants will need to fill in the Heights Interpretation Questionnaire (Steinman and Teachman 2011), Visual Height Intolerance Severity Scale (Huppert et al. 2017), and Acrophobia Questionnaire (Cohen 1977). Then, they will be divided into three groups: Low, Medium, and High acrophobia. Also, they will initially pass through a mathematical test to evaluate their skills and divide them into Novice, Medium, and Expert. Based on their skill, in the game they will receive different numbers of simple, medium, and complicated exercises: Novice—3 simple exercises, Medium—1 simple and 2 medium, 1 medium and 2 complicated. The human-centered approach is ensured by having a virtual environment with a high level of realism in a real-world context (mountain site), with a scenario that is receptive to the player's needs—provides means of relaxation and exploration tasks, has a reward system, combines emotions with logical activity, applies the constructivist learning theory stating that knowledge and skills acquisition are gained by linking a new experience to a previous one, and the possibility of transferring the cognitive and emotional acquisition from the virtual to the real world (Bălan et al. 2019).

As the EEG recording device is rather cumbersome and difficult to be applied on the head when using the HTC Vive glasses at the same time, we chose to record solely peripheral biophysical data: GSR, HR, and Respiration Rate (RR). In order to determine the respiration rate, we have placed two HTC Vive trackers on the chest and on the back of the users and then, during breathing, measured the distance between these two trackers, normalized the values between $[-1; 1]$, applied several filters for smoothing the signal, and counted the number of peaks in the signal which represented the respiration rate.

We will perform a baseline recording when the user stands still in a relaxed position, for a time period of 3 min. As *artifacts identification* is an important step in obtaining clean physiological data, we have developed a method for artifacts reduction, which consists in recognizing artifact patterns in the signals recorded during gameplay sessions, by aligning the biophysical data segments corresponding to the moments when the users performed head/hand/body movements with the artifact signals recorded during a reference procedure. Physiological responses that are not correlated with the content of the game and the emotional responses generated in it make signal analysis very difficult. We define an artifact as *any misleading or confusing alteration in physiological data that appears as a result of external action such as head, hand, or body movements, being unrelated to the emotional effects that specific stimuli or the object under observation exert upon the user* (Balan et al. 2019). For validating this procedure, we have performed an experiment with five healthy subjects, aged 24–50. At first, we recorded a set of reference artifact measurements for each user, in order to acquire the physiological pattern (GSR, HR, and RR) for each artifact: *deep breath, head movement to the left, head movement to the right, head movement up, head movement down, click with the right hand on the HTC Vive controller, and right hand raise*. These are the common artifacts that can

Fig. 6 Acrophobia game: indication to move the head down during the VR game

occur during gameplay in VR. In the second phase of the experiment, each user played the VR game for acrophobia therapy and took the ride by cable car. During each of the ten stops, they were required to perform one of the tasks mentioned above (Fig. 6).

During the analysis step, we have aligned the reference artifacts to the biophysical data recorded during gameplay. Also, we have mapped the reference artifacts onto the gameplay data segments that start before and after the recorded timestamps, with one or two steps before and after and computed the Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) for both GSR and HR. The results showed that the bias is lower on the perfectly aligned biophysical segments than on those located one or two steps before and after, for all seven types of tested artifacts. However, the results were not statistically significant in a Student t-test for independent means. Deep breath was the most relevant artifact introduced in the analysis, more prominent than the rest of the artifacts. During the VR experiment, different artifacts of breathing, head, hand, or body movements can be encountered. If the head movements do not produce significant artifacts and raising the right hand is not a frequent event during the VR game, breathing is a physiological artifact that must be taken into account to avoid overestimating the skin conductance responses during the experiments.

11 Intelligent Virtual Therapist for Acrophobia

We propose a novel and effective approach in which we replace the human therapist with a virtual one, called RoboTher. RoboTher has the appearance of a female avatar and a feminine voice as well. We choose to use a female voice because it is usually perceived as helping, not commanding (Borkowska and Pawlowski 2011), we

are more familiar with a female voice because it is usually associated with the maternal presence (Lee and Kisilevsky 2014), the female voice is more pleasant, being processed in the same auditory area as music (Sokhi et al. 2005), and is perceived as offering more confidence than the male voice due to its higher pitch (Re et al. 2012).

RoboTher automatically identifies the users' emotional states (degree of relaxation or anxiety), compared to a physiological (GSR and HR) baseline recording, 3 min long, performed under resting conditions. RoboTher provides encouragement and is able to change its voice parameters—pitch, tempo, and volume according to the users' emotional states. It provides means of relaxation in the game, by inviting the player to take a break and listen to his favorite song, read his favorite quote, or look at a photo he likes. If these relaxation modalities are ineffective and the stress level is still high, the virtual therapist lowers the degree of exposure. Five subjects played the VR game (ride by cable car) under two conditions: control (one session, without RoboTher's assistance) and three sessions with assistance from RoboTher. We computed the difference, in percent, between the HR and GSR baseline values and the current ones during each game trial. A form of biofeedback has been provided to the subjects as bars of changeable colors (green, yellow, orange, red) for both GSR and HR that appeared in the left top corner of the visual field during the game sessions. If the percent was lower than 10%, the color was green, for a percent in the interval [10%; 40%], the color was yellow, [40%; 70%]—orange, higher than 70%—red (Fig. 7).

The Robot Interactions (RIs), which refer to the statements made by RoboTher and the changes in the voice parameters are presented in Table 3.

Fig. 7 Biofeedback automatically provided in the VR game

Table 3 Intelligent robot interactions based on HR and GSR signals

Robot interaction	Phrase	Voice parameters		
		Pitch (%)	Tempo (%)	Volume (%)
RI1	"Good job! Keep going!"	+10	+10	+10
RI2	"Enjoy and relax for a while"	0	0	0
RI3	"Calm down and relax"	−10	−10	−10
RI4	"You are too tense. Take a deep breath and try to relax more"	−20	−20	−20

Table 4 RoboTher responses based on different Gameplay situations

Situation no.	Condition		Procedure	
	GSR color	HR color	Robot interaction	Change game level (only after the relaxation modalities are provided)
Situation1	Green	Green	RI1	No
Situation2	Green	Yellow	RI1	No
Situation3	Yellow	Green	RI1	No
Situation4	Yellow	Yellow	RI1	No
Situation5	Green	Orange	RI2	−1 level
Situation6	Orange	Green	RI2	−1 level
Situation7	Yellow	Orange	RI2	−1 level
Situation8	Orange	Yellow	RI2	−1 level
Situation9	Orange	Orange	RI3	−2 levels
Situation10	Orange	Red	RI3	−2 levels
Situation11	Red	Orange	RI3	−2 levels
Situation12	Green	Red	RI3	−2 levels
Situation13	Red	Green	RI3	−2 levels
Situation14	Yellow	Red	RI3	−2 levels
Situation15	Red	Yellow	RI3	−2 levels
Situation16	Red	Red	RI3	−3 levels

We identified 16 situations, corresponding to all possible combinations of HR and GSR colors. The Robot Interactions and the alteration in game levels (the automatic adaptation of scenario exposure) during gameplay are presented in Table 4.

The purpose of the therapy is to maintain the player within the green and yellow areas for both HR and GSR parameters throughout the entire game session.

For instance, in Situation5 to Situation8, one of GSRcolor or HRcolor is Green/Yellow and the other is Orange, which means that the subject tends to become anxious. RoboTher plays RI2 and then presents randomly either the user's favorite image, song, or quotation for 20 s. After these 20 s, the subject's emotional state is evaluated again. If it falls into Situation1–Situation4, RI1 appears and he may continue the game from there. If it falls into Situations 5–8, the player is taken to the previous level, so the level of exposure decreases with 1. If it falls into Situations 9–15, the level of exposure decreases with 2 and for Situation 16, decreases with 3.

The results showed that the subjects succeeded to finish the game quicker in the last game session with assistance from RoboTher. In the last session of the experiment, for all users, the most frequent situation was Situation1 (44%), followed by Situation2 (21%), Situation3 (17%), Situation6 (5%), Situation13 (9%), and Situation15 (4%). Both skin conductance parameters and heart rate decreased at the end of the 3 days of gameplay (from 1.68 to 0.9 uS for GSR and from 77.34 to 75.17 bpm for HR), compared to the control condition where the subjects initially played the game without support from RoboTher. These results were statistically significant in a paired-samples *t*-test (Bălan et al. 2020b).

12 Limitations of the Current Research

One limitation of the current research was the small number of subjects. The relatively small training and testing data size imposes a limit on the usage of modern deep neural networks, which usually need a much larger training set (at least in the order of tens of thousands) in order to generalize well and avoid overfitting. This might explain why in our experiments the more classical machine learning approaches, such as Support Vector Machines and Random Forests, achieved the best accuracy. The combination of HMD and EEG device was cumbersome to be worn on the head, causing serious discomfort to the volunteering patients for which we tested the system. Therefore, in the future we will explore VR-based smart therapy systems without the use of EEG. We also plan to extend the research with a larger number of participants (over 15), which is expected to improve the effectiveness of the deep learning approach and validate the effects of the VR therapy in real-world settings to see whether the level of acrophobia indeed decreased. Also, we will provide an introductory session that would familiarize the subjects with the perception of VR environments, apply questionnaires related to immersion and quality of presentation of the VR environment, give more attention to other GSR and HR features, such as inter-beat variability. Biofeedback can be provided not only as color bars, but also as changing elements in the environment—clouds and darker sky when the user is experiencing stress or clear, sunny weather when she or he is relaxed.

13 Limits of Current Phobia Therapy Systems

In a survey in which 19 psychologists from Romania replied, only two mentioned that they use VR to treat phobias: one is using the C2Care application (C2Phobia n.d.), and the other is using 3D mobile applications from Google Play. In the AcTiVity system (Activity System n.d.), exposure adjustment is not determined by the physiological recorded data. The biophysical data are solely recorded for analysis, as in the case of VRMC (Virtual Reality Medical Center n.d.), where

physiological monitoring with visual feedback is used for acknowledging the patients when they are feeling stressed and not. In the system we envision, the next level of game exposure will be selected either by the psychologist, the user, or adjusted automatically, by an artificial intelligence algorithm, according to the patient's biophysical data. The psychologists appreciated that the most urgent phobias to treat are aerophobia (fear of flying by plane), fear of public speaking, claustrophobia, and agoraphobia. On a scale from 1 to 5, the most useful is to integrate an analysis tools addressed to the therapist, allowing him to analyze the patient's performance and evolution (4.5/5), followed by relaxation techniques (4.4/5), awareness techniques (4.25/5), games and rewards (4.1/5).

14 Proposed Future System

We aim to develop a system for phobia therapy that relies on gradual exposure in the Virtual Reality (VR), accompanied by physiological signals monitoring (pulse, electrodermal activity, and respiration rate) and real-time visual bio-feedback. The system can be used in the presence of the therapist or at home, for the patients who suffer from a mild phobia condition. During the therapy, the scenes from the virtual environment can be changed by the psychologist, the user himself or automatically, by a virtual therapist who adapts the scenario exposure based on the biophysical data recorded. Here, the human therapist is replaced by a virtual one with the shape of a game avatar, who offers support and encouragement to the patient. It directly interacts with the user and changes its voice parameters—pitch, tempo, and volume and facial expressions—according to the patient's emotional state. It identifies the current fear level and provides three modalities of relaxation—by determining the user to look at a favorite picture, listen to an enjoyable song, or read an inspirational quote. If the relaxation modalities fail to be effective, the virtual therapist automatically lowers the level of exposure according to a set of rules. The set of rules are part of an artificial intelligence future model, most likely trained using reinforcement learning and unsupervised learning techniques (Sutton and Barto 1998; Kallenberg et al. 2016; Erhan et al. 2010; Croitoru et al. 2019; Leordeanu et al. 2016) combined with either classical machine learning or deep neural networks, using biophysical user data and emotional ratings of valence/arousal/dominance (the emotion dimensions). A control panel allows introducing new patients, managing existing ones, recording sessions, replaying them, and generating statistics. An important aspect is the patients' gaze direction. We will record where the users are looking during the therapy and correlate it with the emotional state. Thus, our approach is at the confluence of psychology, artificial intelligence, and computer vision.

14.1 Methods and Instruments of Investigation

As *methods and instruments of investigation*, we will use game design, Virtual Reality integration, biophysical sensors that record electrodermal activity, heart rate and respiration rate (Shimmers Multisensory n.d.), artificial intelligence—classic machine learning and deep learning techniques (depending on the amount of training data available) for training and testing two classifiers: one that estimates the user's current emotional state (fear level) and one that determines the next exposure scenario according to the estimated fear level. Due to the strong limitations in supervised training signal and ground truth information (which usually comes from doctors), our intelligent system will learn and improve by itself during sessions, based on different self-supervised and reinforcement learning strategies, which we will explore. The combination of multiple sensors and actual user interaction with the VR system will enable the effectiveness of automatic self-training of the intelligent phobia therapist. The virtual therapist having the appearance of a female avatar will provide encouragement or invite the user to relax. The avatar will say expressions like "Well done! Keep going!" or "Calm down and try to relax". The virtual therapist's voice parameters—pitch, volume, tempo, and facial expression will change according to the user's emotional state. It selects relaxation modalities to provide the user during the game—a favorite song, image, or quote and then, based on the artificial intelligence models, estimates fear level and determines the next exposure scenario—whether the user will increase or decrease the level of exposure. During the game, the user is offered bio-feedback. The differences (in percent) between the current biophysical values and the baseline ones (recorded during a 3-min resting state) are presented as bars colored in green, yellow, orange, and red. Thus, the user can visualize his emotional state in a comprehensive way and struggle to relax in order to change the bars' color to green or yellow throughout the game session.

14.2 Potential Risks

The potential risks come from the very strong interdisciplinary nature of the project, combining psychology, medicine, engineering, and computer science. A potential risk resides in the uncontrolled reactions of the patient during exposure—motion sickness, anxiety, and inadaptation to VR. They will be minimized by pilot-testing the designed scenes using non-phobic persons or people who suffer from a mild phobic condition. The patients will be allowed to exit the virtual environment at any time. During the tests, a research assistant will monitor the users and ensure that they are feeling comfortable and safe. We will interact with psychologists in order to identify the potential risks that will be minimized by designing, implementing, and testing as many solutions as possible for each of the system's components. We will make sure that the virtual environment contains efficient gamification elements that engage the user in the therapy. The rendering quality will be enhanced by

employing the most recent commercial technology. Perturbations and noises caused by the signal recording devices, disconnections, hardware, and physical limitations will be overcome by using advanced devices, monitoring the recording procedure, filtering, and post-processing the data. Another risk could be the slow learning rate of the intelligent system. As modern advanced machine learning algorithms significantly improve as training data size and diversity increases, it is possible that a long time will pass until sufficient data are captured for optimum performance. However, as the current system, trained on limited data, already shows good performance, we could realistically expect that future versions will improve considerably, once we have access to more data and develop more advanced self-supervised and reinforcement learning strategies for training a more powerful automated therapist.

14.3 High Future Gains in Phobia Therapy

The ability to bring the power of current computer systems and combine sophisticated technologies (such as virtual reality and artificial intelligence) with modern medicine, in order to treat various phobias and improve brain function, could have a tremendous positive impact in improving human life. Our proposed smart VR phobia therapist, with excellent initial results, even in the case of limited data, is a solid proof that automated technology will eventually improve current medicine and psychology practice. Doctors will be able to design more efficient treatments in combination with such smart robotic assistants. Time is on our side, as more powerful AI and VR algorithms and systems are created every day. As our understanding of artificial intelligence and the human brain will improve, along with our multidisciplinary experience and knowledge, the high gain of combining high tech with human sciences is almost certain. The only real question remaining is when should we expect current research to grow into a mature and well-trusted technology. Nevertheless, as history showed us time and again, research ideas with great potential to benefit human life, become a reality sooner rather than later.

Acknowledgements The work has been funded by the Operational Programme Human Capital of the Ministry of European Funds through the Financial Agreement 51675/09.07.2019, SMIS code 125125, UEFISCDI project 1/2018, and UPB CRC Research Grant 2017. This work has also been funded in part through UEFISCDI, from EEA Grants 2014–2021, project number EEA-RO-NO-2018-0496.

References

2BIOPAC (n.d.). https://www.biopac.com/application/virtual-reality/
Activity System (n.d.). https://www.unitylab.de/
Agrafioti F, Hatzinakos D, Anderson AK (2012) ECG pattern analysis for emotion detection. IEEE Trans Affect Comput 3(1):102–115

Alhagry S, Aly A, Reda A (2017) Emotion recognition based on EEG using LSTM recurrent neural network. Int J Adv Comput Sci Appl 8:355–358

AlZoubi O, D'Mello SK, Calvo RA (2012) Detecting naturalistic expressions of non-basic affect using physiological signals. IEEE Trans Affect Comput 3(3):298–310

Arikan K, Boutros NN, Bozhuyuk E, Poyraz BC, Savrun BM, Bayar R et al (2006) EEG correlates of startle reflex with reactivity to eye opening in psychiatric disorders: preliminary results. Clin EEG Neurosci 37:230–234

Atkinson J, Campos D (2016) Improving BCI-based emotion recognition by combining EEG feature selection and kernel classifiers. Expert Syst Appl 47:35–41

Baker BL, Cohen DC, Saunders JT (1973) Self-directed desensitization for acrophobia. Behav Res Ther 11(1):79–89

Bălan O, Moise G, Moldoveanu A, Leordeanu M, Moldoveanu F (1738) Fear level classification based on emotional dimensions and machine learning techniques. Sensors 2019:19

Bălan O, Moise G, Moldoveanu A, Leordeanu M, Moldoveanu F (2019) Challenges for ML-based emotion recognition systems in medicine. A human-centered approach. In: CHI'19 extended abstracts, may 4–9, 2019, Glasgow, Scotland, UK. ACM. ISBN: 978-1-4503-5971-9/19/05

Balan O, Moldoveanu A, Petrescu L, Moise G, Cristea S, Petrescu C, Moldoveanu F, Leordeanu M (2019) Sensors system methodology for artefacts identification in virtual reality games. ISAECT, Rome

Bălan O, Moise G, Petrescu L, Moldoveanu A, Leordeanu M, Moldoveanu F (2020a) Emotion classification based on biophysical signals and machine learning techniques. Symmetry 12:21

Bălan O, Cristea Ș, Moise G, Petrescu L, Moldoveanu A, Moldoveanu F, Leordeanu M (2020b) RoboTher—an assistive robot for acrophobia therapy in virtual reality. Submitted to ICRA

Bandura A (1977) Self-efficacy—toward a unifying theory of behavioral change. Psychol Rev 84(2):191–215

Borkowska B, Pawlowski B (2011) Female voice frequency in the context of dominance and attractiveness perception. Anim Behav 82(1):55–59

Bos DO (2006) EEG-based emotion recognition. The Influence of Visual and Auditory Stimuli University

Braithwaite JJ, Jones R, Rowe M, Watson DG (2015) A guide for analysing electrodermal activity (EDA) and skin conductance responses (SCRs) for psychological experiments. University of Birmingham, UK: Selective Attention and Awareness Laboratory

Brouwer AM, Neerincx MA, Kallen V, van der Leer L, ten Brinke M (2011) EEG alpha asymmetry, heart rate variability and cortisol in response to virtual reality induced stress. J Cyber Ther Rehabil 4:27–40

Brouwer A-M, Zander TO, van Erp JBF, Korteling JE, Bronkhorst AW (2015) Using neurophysiological signals that reflect cognitive or affective state: six recommendations to avoid common pitfalls. Front Neurosci 9:136. https://doi.org/10.3389/fnins.2015.00136

Buna P, Gorskia F, Grajewskia D, Wichniareka R, Zawadzkia P (2017) Low–cost devices used in virtual reality exposure therapy. Proc Comput Sci 104:445–451

C2Phobia (n.d.). https://www.c2.care/en/c2phobia-treating-phobias-in-virtual-reality/

Canli T, Desmond JE, Zhao Z, Glover G, Gabrieli JDE (1998) Hemispheric asymmetry for emotional stimuli detected with fMRI. Neuroreport 9(14):3233–3239

Carmen Juan M, Perez D (2010) Using augmented and virtual reality for the development of acrophobic scenarios. Comparison of the levels of presence and anxiety. Comput Graph 34:756–766

Chanel G, Rebetez C, Betrancourt M, Pun T (2011) Emotion assessment from physiological signals for adaptation of game difficulty. IEEE Trans Syst Man Cybern A Syst Hum 41(6):1052–1063

Cheemalapati S, Adithya PC, Valle MD, Gubanov M, Pyayt A (2016) Real time fear detection using wearable single channel electroencephalogram. Sensor Netw Data Commun 5:140. https://doi.org/10.4172/2090-4886.1000140

Choi YH, Jang DP, Ku JH, Shin MB, Kim SI (2001) Short-term treatment of acrophobia with virtual reality therapy (VRT): a case report. Cyberpsychol Behav 4(3):349–354

Coelho CM, Santos JA, Silva C, Wallis G, Tichon J, Hine TJ (2008) The role of self-motion in acrophobia treatment. Cyberpsychol Behav 11(6):723–725
Coelho CM, Waters AM, Hine TJ, Wallis G (2009) The use of virtual reality in acrophobia research and treatment. J Anxiety Disord 23:563–574
Cohen DC (1977) Comparison of self-report and behavioral procedures for assessing acrophobia. Behav Ther 8:17–23
Costa JP, Robb J, Nacke LE (2014) Physiological acrophobia evaluation through in vivo exposure in a VR CAVE. In: 2014 IEEE games media entertainment
Croitoru I, Bogolin SV, Leordeanu M (2019) Unsupervised learning of foreground object segmentation. Int J Comput Vision (IJCV) 127(9):1279–1302
Cudlenco N, Popescu N, Leordeanu M (2020) Reading into the mind's eye: boosting automatic visual recognition with EEG signals. Neurocomputing 386:281–292
Daly I, Malik A, Weaver J, Hwang F, Nasuto S, Williams D, Kirke A, Miranda E (2015) Identifying music-induced emotions from EEG for use in brain computer music interfacing. In: 6th affective computing and intelligent interaction
Davidson RJ (1993) Cerebral asymmetry and emotion: conceptual and methodological conundrums. Cognit Emot 7:115–138
Davidson RJ, Ekman P, Saron C, Senulis J, Friesen WV (1990) Approach/withdrawal and cerebral asymmetry: emotional expression and brain physiology, I. J Pers Soc Psychol 58:330–341
Demaree HA, Everhart DE, Youngstrom EA, Harrison DW (2005) Brain lateralization of emotional processing: Historical roots and a future incorporating "dominance". Behav Cogn Neurosci Rev 4:3–20
DiMeglio C (2015) Fear feedback loop: creative and dynamic fear experiences driven by user emotion, Master Thesis, Rochester Institute of Technology
Dores AR, Barbosa F, Monteiro L, Reis M, Coelho CM, Ribeiro E, Castro-Caldas A (2014) Amygdala activation in response to 2D and 3D emotion-inducing stimuli. PsychNol J 12(1–2):29–43
Ekman P, Sorenson ER, Friesen WV (1969) Pan-cultural elements in facial displays of emotions. Science 164:86–88
Erhan D, Bengio Y, Courville A, Manzagol PA, Vincent P, Bengio S (2010) Why does unsupervised pre-training help deep learning? J Mach Learn Res 11:625–660
Eum YJ, Jang EH, Park BJ, Choi SS, Kim SH, Sohn JH (2011) Emotion recognition by responses evoked by negative emotion. In Engineering and Industries (ICEI), 2011 international conference on IEEE. pp 1–4
Field AP, Argyris NG, Knowles KA (2001) Who's afraid of the big bad wolf: a prospective paradigm to test Rachman's indirect pathways in children. Behav Res Ther 39:1259–1276
Fleureau J, Philippe G, Huynh-Thu Q (2012) Physiological-based affect event detector for entertainment video applications. IEEE Trans Affect Comput 3(3):379–385
Garcia-Palacios HG, Hoffman S, Kwong See A, Botella C (2001) Redefining therapeutic success with virtual reality exposure therapy. Cyberpsychol Behav 4(3):341–348. http://online.liebertpub.com/doi/abs/10.1089/109493101300210231
Garcia-Palacios A, Hoffman H, Carlin A, Furness TA, Botella C (2002) Virtual reality in the treatment of spider phobia: a controlled study. Behav Res Ther 40:983–993
Gouizi K, Reguig F, Maaoui C (2011) Analysis physiological signals for emotion recognition. In: WOSSPA international workshop, IEEE. pp 147–150
Graham J, Gaffan EA (1997) Fear of water in children and adults: etiology and familial effects. Behav Res Ther 35(2):91–108
Healey J (2009) Affect detection in the real world: recording and processing physiological signals. In: Affective computing and intelligent interaction and workshops, 2009. ACII 2009. 3rd international conference on IEEE. pp 1–6
Hodges LF, Kooper R, Meyer TC, Rothbaum BO, Opdyke D, de Graaff JJ, Williford JS, North MM (1995) Virtual environments for treating the fear of heights. Computer 28(7):27–34

Horschig JM, Zumer JM, Bahramisharif A (2014) Hypothesis-driven methods to augment human cognition by optimizing cortical oscillations. Front Syst Neurosci 8:119. https://doi.org/10.3389/fnsys.2014.00119

HTC Vive (n.d.). https://www.vive.com/eu/

Hu F, Wang H, Chen J, Gong J (2018) Research on the characteristics of acrophobia in virtual altitude environment. In: Proceedings of the 2018 IEEE international conference on intelligence and safety for robotics, Shenyang, China, August, 24–27

Huppert D, Grill E, Brandt T (2017) A new questionnaire for estimating the severity of visual height intolerance and acrophobia by a metric interval scale. Front Neurol 8:211. https://doi.org/10.3389/fneur.2017.00211

Jerald J (2016) The VR book: human-centered design for virtual reality. ACM

Jirayucharoensak S, Pan-Ngum S, Israsena P (2014) EEG-based emotion recognition using deep learning network with principal component-based covariate shift adaptation. Sci World J 2014, article ID 627892, 10 pages

Jones NA, Fox NA (1992) Electroencephalogram asymmetry during emotionally evocative films and its relation to positive and negative affectivity. Brain Cogn 20(2):280–299

Kallenberg M, Petersen K, Nielsen M, Ng AY, Diao P, Igel C, Vachon CM, Holland K, Winkel RR, Karssemeijer N, Lillholm M (2016) Unsupervised deep learning applied to breast density segmentation and mammographic risk scoring. IEEE Trans Med Imaging 35(5):1322–1331

Koelstra S, Yazdani A, Soleymani M, Muhl C, Lee J-S, Nijholt A, Pun T, Ebrahimi T, Patras I (2010) Single trial classification of EEG and peripheral physiological signals for recognition of emotions induced by music videos. Brain Inform Ser Lect Notes Comput Sci 6334(9):89–100

Koelstra S, Muehl C, Soleymani M, Lee J-S, Yazdani A, Ebrahimi T, Pun T, Nijholt A, Patras I (2012) DEAP: a database for emotion analysis using physiological signals. IEEE Trans Affect Comput 3:18–31

Kometer H, Luedtke S, Stanuch K, Walczuk S, Wettstein J (2010) The effects virtual reality has on physiological responses as compared to two-dimensional video. University of Wisconsin School of Medicine and Public Health, Department of Physiology

Lee GY, Kisilevsky BS (2014) Fetuses respond to father's voice but prefer mother's voice after birth. Dev Psychobiol 56(1):1–11

Leitenberg H (1976) Behavioral approaches to treatment of neuroses. In: Leitenberg H (ed) Handbook of behavior modification and behavior therapy. Prentice-Hall, Englewood Cliffs NJ

Leordeanu M, Radu A, Baluja S, Sukthankar R (2016) Labeling the features not the samples: efficient video classification with minimal supervision. In: Thirtieth AAAI conference on artificial intelligence. AAAI

Littman MK (1996) Alternative meanings through the world of virtual reality. In: Vandergrift K (ed) Mosaics of meaning: enhancing the intellectual life of young adults through story. Scarecrow Press, Lanham, pp 425–455

Liu Y, Sourina O (2013) EEG databases for emotion recognition. In: 2013 international conference on cyberworlds (CW)

Liu C, Rani P, Sarkar N (2005) An empirical study of machine learning techniques for affect recognition in human-robot interaction. In: International conference on intelligent robots and systems, IEEE

Liu C, Agrawal P, Sarkar N, Chen S (2009) Dynamic difficulty adjustment in computer games through real-time anxiety-based affective feedback. Int J Hum Comput Interact 25(6):506–529

Mauss IB, Robinson MD (2009) Measures of emotion: a review. Cognit Emot 23(2):209–237

Menzies RG, Clarke JC (1993) The etiology of fear of heights and its relationship to severity and individual response patterns. Behav Res Ther 31(4):355–365

Menzies RG, Clarke JC (1995) The etiology of acrophobia and its relationship to severity and individual-response patterns. Behav Res Ther 33(7):795–803

Menzies RG, Parker L (2001) The origins of height fear: an evaluation of neoconditioning explanations. Behav Res Ther 39:185–199

Morina N, Ijntema H, Meyerbroker K, Emmelkamp PMG (2015) Can virtual reality exposure therapy gains be generalized to real-life? A meta-analysis of studies applying behavioral assessments. Behav Res Ther 74:18–24
Morris RJ, Magrath KH (1979) Contribution of therapist warmth to the contact desensitization-treatment of acrophobia. J Consult Clin Psychol 47(4):786–788
Muris P, Schmidt H, Merckelbach H (1999) The structure of specific phobia symptoms among children and adolescents. Behav Res Ther 37:863–868
Naser DS, Saha G (2013) Recognition of emotions induced by music videos using DT-CWPT. In: Medical informatics and telemedicine (ICMIT), 2013 Indian conference IEEE
Nation Wide Phobias Statistics (2019). https://blog.nationwide.com/common-phobias-statistics/
North MM, North SM (1996) Virtual psychotherapy. J Med Virtual Real 1:28–32
Oculus Rift (n.d.). https://www.oculus.com/rift/
Olesen J (2015) Phobia statistics and surprising facts about our biggest fears. http://www.fearof.net/phobia-statistics-and-surprising-facts-about-our-biggest-fears/
Opris D, Pintea S, Garcia-Palacios A, Botella C, Szamoskozi S, David D (2012) Virtual reality exposure therapy in anxiety disorders: a quantitative meta-analysis. Depress Anxiety 29:85–93
Parsons TD, Rizzo AA (2008) Affective outcomes of virtual reality exposure therapy for anxiety and specific phobias: a meta-analysis. J Behav Ther Exp Psychiatry 39(3):250–261
Petrantonakis PC, Hadjileontiadis LJ (2009) EEG-based emotion recognition using hybrid filtering and higher order crossings. In: Affective computing and intelligent interaction and workshops, 2009. ACII 2009. 3rd international conference on IEEE. pp 1–6
Phillips ML, Drevets WC, Rauch SL, Lane R (2003) Neurobiology of emotion perception II: implications for major psychiatric disorders. Biol Psychiatry 54:515–528
Poulton R, Menzies RG (2002) Non-associative fear acquisition: a review of the evidence from retrospective and longitudinal research. Behav Res Ther 40:127–149
Poulton R, Davies S, Menzies RG, Langley JD, Silva PA (1998) Evidence for a non-associative model of the acquisition of a fear of heights. Behav Res Ther 36(5):537–544
Putman P, Van PJ, Maimari I, Vander WS (2010) EEG Theta/Beta ratio in relation to fear-modulated response-inhibition, attentional control, and affective traits. Biol Psychol 83:73–78
Quirk GJ (2013) Fear. Neuroscience in the 21st century: from basic to clinical. pp 2009–2026
Rainville P, Bechara A, Naqvi N, Damasio AR (2006) Basic emotions are associated with distinct patterns of cardiorespiratory activity. Int J Psychophysiol 61(1):5–18
Re DE, O'Connor JJ, Bennett PJ, Feinberg DR (2012) Preferences for very low and very high voice pitch in humans. PLoS One 7(3):e31353
Ritchie's Plank Experience (n.d.). http://store.steampowered.com/app/517160/Richies_Plank_Experience/
Ritter B (1969) Treatment of acrophobia with contact desensitization. Behav Res Ther 7(1):41–45
Robertson A (2016) The climb turns virtual reality acrophobia into an extreme sport. https://www.theverge.com/2016/4/28/11526150/crytek-the-climb-vr-oculus-rift-review
Russell JA (1979) Affective space is bipolar. J Pers Soc Psychol 37(3):345–356
Russell JA, Mehrabian A (1977) Evidence for a three-factor theory of emotions. J Res Pers 11(3):273–294
Šalkevicius J, Damaševičius R, Maskeliunas R, Laukienė I (2019) Anxiety level recognition for virtual reality therapy system using physiological signals. Electronics 8:1039
Schutter DJ, Van Honk J (2005) Electrophysiological ratio markers for the balance between reward and punishment. Cogn Brain Res 24:685690
Scott D (1976) Understanding EEG: an introduction to electroencephalography. Duckworth. pp 18–32
Shiban Y, Schelhorn I, Pauli P, Mühlberger A (2015) Effect of combined multiple contexts and multiple stimuli exposure in spider phobia: a randomized clinical trial in virtual reality. Behav Res Ther 71:45–53
Shimmers Multisensory (n.d.). https://www.shimmersensing.com/products/shimmer3-wireless-gsr-sensor

Sokhi DS, Hunter MD, Wilkinson ID, Woodruff PW (2005) Male and female voices activate distinct regions in the male brain. NeuroImage 27(3):572–578
Soleymani M, Kierkels J, Chanel G, Pun T (2009) A Bayesian framework for video affective representation. In: Proceedings of the international conference on affective computing and intelligent interaction. pp 1–7
Steimer T (2002) The biology of fear and anxiety-related behaviors. Dialogues Clin Neurosci 4(3):231–249
Steinman SA, Teachman BA (2011) Cognitive processing and acrophobia: validating the heights interpretation questionnaire. J Anxiety Disord 25:896–902
Steinman SA, Teachman BA (2014) Reaching new heights: comparing interpretation bias modification to exposure therapy for extreme height fear. J Consult Clin Psychol 82(3):404–417. PMID: 24588406
Sutton RS, Barto AG (1998) Introduction to reinforcement learning, vol 2(4). MIT Press, Cambridge
Toth V (2015) Measurement of stress intensity using EEG, BSc Thesis
Trainor LJ, Schmidt LA (2003) Processing emotions induced by music. Cognitive neuroscience of music (Oxford). 317p
Virtual Reality Medical Center (n.d.). http://www.vrphobia.com/aboutus.htm
Walter S, Kim J, Hrabal D, Crawcour SC, Kessler H, Traue HC (2013) Trans-situational individual-specific biopsychological classification of emotions. IEEE Trans Syst Man Cybern Syst 43(4):988–995
Wen WH, Liu GY, Cheng NP, Wei J, Shangguan PC, Huang WJ (2014) Emotion recognition based on multi-variant correlation of physiological signals. IEEE Trans Affect Comput 5:126–140
Westerink J, Ouwerkerk M, de Vries GJ, de Waele S, van den Eerenbeemd J, van Boven M (2009) Emotion measurement platform for daily life situations. In: Affective computing and intelligent interaction and workshops, 2009. ACII 2009. 3rd international conference on IEEE. pp 1–6
Wiem MBH, Lachiri Z (2017) Emotion classification in arousal valence model using MAHNOB-HCI database. Int J Adv Comput Sci Appl Ijacsa 8(3)
Withers RD, Deane FP (1995) Origins of common fears: effects on severity, anxiety responses and memories of onset. Behav Res Ther 33(8):903–915
Wolpe J (1958) Psychotherapy by reciprocal inhibition. Stanford University Press, Palo Alto, CA
World Health Organization (2017) Depression and other common mental disorders: global health estimates. World Health Organization, Geneva. License: CC BY-NC-SA 3.0 IGO.PS
Yoon HJ, Chung SY (2013) EEG-based emotion estimation using Bayesian weighted-log-posterior function and perceptron convergence algorithm. Comput Biol Med 43(12):2230–2237