5-15-2019

AUTOMATIC ADAPTATION OF EXPOSURE INTENSITY IN VR ACROPHOBIA THERAPY, BASED ON DEEP NEURAL NETWORKS

Oana Balan
*University POLITEHNICA of Bucharest, Faculty of Automatic Control and Computers, oana.balan@cs.pub.ro*

Gabriela Moise
*University Petroleum-Gas, gpmoire@gmail.com*

Alin Moldoveanu
*University POLITEHNICA of Bucharest, Faculty of Automatic Control and Computers, alin.moldoveanu@cs.pub.ro*

Florica Moldoveanu
*University POLITEHNICA of Bucharest, Faculty of Automatic Control and Computers, florica.moldoveanu@cs.pub.ro*

Marius Leordeanu
*University POLITEHNICA of Bucharest, Faculty of Automatic Control and Computers, marius.leordeanu@cs.pub.ro*

Follow this and additional works at: [https://aisel.aisnet.org/ecis2019_rp](https://aisel.aisnet.org/ecis2019_rp)

Recommended Citation


This material is brought to you by the ECIS 2019 Proceedings at AIS Electronic Library (AISeL). It has been accepted for inclusion in Research Papers by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.
AUTOMATIC ADAPTATION OF EXPOSURE INTENSITY IN VR ACROPHOBIA THERAPY, BASED ON DEEP NEURAL NETWORKS

Research paper

Oana Bălan, University POLITEHNICA of Bucharest, oana.balan@cs.pub.ro
Gabriela Moise, Petroleum-Gas University of Ploiești, gmoise@upg-ploiesti.ro
Alin Moldoveanu, University POLITEHNICA of Bucharest, alin.moldoveanu@cs.pub.ro
Marius Leordeanu, University POLITEHNICA of Bucharest, marius.leordeanu@cs.pub.ro
Florica Moldoveanu, University POLITEHNICA Bucharest, florica.moldoveanu@cs.pub.ro

Abstract

This paper proposes a real-time Virtual Reality game for treating acrophobia that automatically tailors in-game exposure to heights to the players’ individual characteristics – affective state and physiological features. The elements of novelty are the automatic estimation of fear and the prediction of the next game level based on the electroencephalogram (EEG) and biophysical data – Galvanic Skin Response (GSR) and Heart Rate (HR). Two neural networks have been trained with the data recorded in an experiment where 4 subjects have been in-vivo and virtually exposed to various heights. In order to test the validity of the approach, the same users played the acrophobia game, using two modalities of expressing fear level. After completing a game level, the EEG and biophysical data were averaged and one neural network estimated the current fear score, while the other predicted the next game level. A measure of similarity between the self-estimated fear level during a game epoch and the fear level predicted by the first neural network showed an accuracy rate of 73% and 42% respectively for the two modalities of expressing fear level. 3 out of 4 users succeeded to obtain a fear level of 0 (complete relaxation) in the final game epoch.

Keywords: Virtual Reality, Gamification, Deep learning, Acrophobia, Emotion Recognition
1 Introduction

Even if the mental health disorders are widely under-reported, there some studies that indicate their worldwide high incidence. According to (Ritchie and Roser, Institute of Health Metrics and Evaluation, World Health Organization (WHO) Global Health Observatory (GHO)), in 2016, an estimation of the number of people with any mental and substance use disorders was about 1.1 billion, from which 275 million were affected by anxiety disorders. The prevalence per country varies between 2.5% and 6.5% and the highest incidence was found in the US, Canada, West and North Europe, Australia, North Africa and South America. Women are more affected than men, with about 4.5% (Ritchie and Roser, 2019). The number of people living with anxiety disorders increased in 2015 with 14.9% since 2005, the estimation being around 264 million persons (WHO, 2017).

Specific phobia is a type of anxiety or fear-related disorder, as classified in the ICD-11 for Mortality and Morbidity Statistics (ICD-11-MMS). Specific phobia is defined as a marked and excessive fear or anxiety that consistently occurs when exposed to one or more specific objects or situations (e.g., proximity to certain animals, flying, heights, closed spaces, sight of blood or injury) and that is out of proportion to actual danger (ICD, 2018). In terms of statistics related to phobias, there are studies which estimate that 15-20% of the world’s population experience specific phobias at least once in the lifetime (Olesen, 2015). The most common phobias concern heights and animals (Eaton, 2018). The following results have been obtained in a study which involved 22 countries between 2001 and 2011: the cross-national lifetime and 12-month prevalence rates of specific phobia were, respectively, 7.4% and 5.5%, being higher in females (9.8% and 7.7%) than in males (4.9% and 3.3%) and higher in high and higher-middle income countries than in low/lower middle income countries (Wardenaar et al, 2017).

Phobias are generally treated with medication and/or psychotherapy. A successful type of psychotherapy is Cognitive-Behavioural Therapy (CBT) - with two methods: cognitive and exposure (behavioural) therapy. Exposure therapy consists in gradual exposure to anxiety eliciting objects or situations, in the presence of a therapist. Virtual Reality is an emergent technology which begins to be adopted more often in phobias therapy. It simulates worlds full of anxiety-producing stimuli and exposes the patients to them in a safe and controlled manner.

In this paper we propose a Virtual Reality (VR) game for treating acrophobia, using a real-time adaptation of in-game height exposure. 7.5% of the world’s population suffers from acrophobia. 10% of the U.S. population and 14% of the people from U.K. are afraid of heights. Thus, we aim to find a solution based on VR for treating this prevalent anxiety disorder. Using physiological signals (heart rate and galvanic skin response originating from the peripheral nervous system and EEG, stemming from the central nervous system), we feed two Deep Neural Networks (DNNs) in order to estimate the subject’s current fear level and to predict the game level to be played next. In order to validate our method, we performed an experiment with 4 acrophobic users and observed a high correspondence between the fear level predicted by the neural network and the self-estimated subjective anxiety scores. In addition, 3 out of 4 users succeeded to relax and obtained a fear level of 0 in the last game epochs, concluding that the game levels have been adjusted according to the subjects’ emotional state, a fundamental aspect in acrophobia treatment. The presented experiment is part of a series of experiments taking place within a project whose goals is developing a VR system for treating various phobias.

The paper is organized as follows: chapter 2 presents the most relevant VR-based systems for phobia therapy, chapter 3 introduces the relationship between emotions and biophysical data, chapter 4 details the machine learning techniques used for emotion recognition, chapter 5 describes our approach for heights exposure adaptation based on deep neural networks, chapter 6 presents the game design, experimental procedure and research results while finally, chapter 7 shows the conclusions and discusses future work directions.
2 Virtual Reality for Phobia Therapy

Behavioural therapy may be difficult for patients who have problems in imagining scenes full of anxiety producing stimuli and/or who are too afraid to be exposed in real situations. The first use of VR techniques in exposure therapy was reported by the Human-Computer Interaction Group at Clark Atlanta University in November 1992 (North et al., 1997). The first pilot experiments in using Virtual Reality Exposure Therapy (VRET) were conducted by North’s team for specific phobia treatment: fear of heights, flying, public speaking and fear of being in certain situations (North et al., 1997). Since then, more studies have been undertaken and the results showed that VRET is highly effective and preferred by the patients. In 2 experiments described in (Garcia-Palacios et al., 2001) more than 80% of the subjects (81% and 89%, respectively) chose VRET instead of in-vivo exposure therapy. Also, VRET offered valuable results in the post-treatment assessments, comparable with traditional behavioural therapy (Opris et al., 2012).

One of the largest experiments for acrophobia treatment using VR was performed between October 2017 and February 2018 by a team led by Prof. Freeman from University of Oxford, Department of Psychiatry (Freeman et al., 2018). Using a software application called Now I Can Do Heights, the team proved that immersive VR technologies are highly effective for reducing fear of heights. The procedure did not involve the presence of a therapist, as he was replaced by a virtual coach (Freeman et al., 2018). More examples of VRET are provided by Levski: Bravemind (University of Southern California, Institute for Creative Technologies), VR-based therapeutic solutions for hospital patients developed by Cedars-Sinai, VR therapy for patients with fear of heights, elevators, thunderstorms, flying and public speaking offered by Duke Psychiatry and Behavioural Sciences, Richie’s Plank Experience, CityScapes and Landscapes offered by Samsung, Limelight developed by Virtual Neuroscience Lab, Gide Meditation VR developed by Cubicle Ninjas, Relax Soothe Sleep: The Nap App created by Virtually Better, Inc., Limbix VR, Psious.

Gamification elements have been integrated in VRET. The Climb, designed for the Oculus device uses as input the Xbox gamepad, the Rift’s head tracking and motion tracking, while Richie’s Plank Experience for HTC Vive (HTC Vive) uses a customizable real plank replicated in the virtual environment (Robertson, 2016).

C2Phobia (C2Phobia) gradually exposes the users to different heights, from the first to the 15th floor of a skyscraper. Stim Response Virtual Reality offers a wide range of VR worlds, which are changed using the players’ biophysical responses and VR events (2BIOPAC). Acrophobia Therapy with Virtual Reality (ActiVity-System) (Schafer, 2015) – uses an Oculus Rift device to render the 3D scenes. The participants who played the game using an avatar related positively to the approach of the game and even tried to control the system through physical interaction with their bodies. The Stim Response Virtual Reality system (2BIOPAC) offers of wide range of VR worlds. The events from VR and physiological data are synchronized in real-time, while the scenes are changed based on the player’s biophysical responses. The Virtual Reality Medical Center (VRMC) used simulation technologies for anxiety and phobia alleviation and educational purposes. VRMC treats patients suffering from panic attacks, specific phobias such as agoraphobia, social phobia, claustrophobia, arachnophobia, fear of flying, fear of driving, fear of thunderstorms, fear of public speaking, using Virtual Reality-enhanced Cognitive Behavioural Therapy (VR-CBT).

In this paper, we continue our previous work (Bălan et al., 2018, Bălan et al., 2019) and propose a VRET system for treating acrophobia, in which patients’ data - HR, GSR and EEG are used for fear evaluation and automatic change of VR scenarios. Deep Neural Networks (DNNs) are used for fear classification and automatic height exposure estimation. As far as we know, there is no VRET system for treating acrophobia based on physiological data and machine learning techniques.
3 The Relationship between Emotions and Biophysical Data

Emotions are classified using the Circumplex Model of Affects proposed by Russell (Russell, 1979), which consists of two orthogonal emotion dimensions, namely arousal and valence. Arousal ranges from “not excited” to “excited”, while valence, from “positive” to “negative”. A third dimension, dominance, indicates the degree of control the subject possesses over his emotions. Usually, fear is characterized by low valence, high arousal and low dominance (Demaree et al., 2005).

The approach-withdrawal model, on the other hand, suggests that the right side of the brain mediates withdrawal-based emotions, while activation in the left cortical area is correlated with approach (or appetitive) mental state changes.

Galvanic Skin Response (GSR) is a reflection of skin conductance/ resistance change, measured by electrodes applied on the distal phalanges of the index and middle fingers. GSR is a response of the sympathetic nervous system, along with heart rate. Fear is characterized by an increase of sweat production and, in consequence, of skin conductance (DiMeglio, 2015, Healey, 2009, Fleureau et al., 2012, Westerink et al., 2009). Moreover, GSR proved to be efficient in discriminating fear from other negative emotions (AlZoubi et al., 2012). In what concerns heart rate, fear can produce an increase of over 40 bpm from baseline, exceeding the tachycardic threshold of 100-120 bpm (Komet et al., 2010).

Electroencephalography (EEG) measures brain activity by recording the signals originating from the central nervous system. According to the approach/withdrawal model of frontal alpha asymmetry (Davidson, 1993), left frontal activation, corresponding to low levels of alpha waves (8-12 Hz) indicate a tendency of approach, while, on the other hand, right frontal brain activation (low levels of alpha) elicits negative affective responses (Bos, 2006, Trainor & Schmidt, 2003, Jones & Fox, 1992, Canli et al., 1998). High levels of beta waves (13-30 Hz) indicate anxiety, alert and fear (Arikan et al., 2006, Komet et al., 2010).

4 Machine Learning for emotion recognition

Emotions play an important role in human communication and interaction. The ability to recognize and differentiate emotions is specific to humans. However, in the last decades, several approaches for automatic identification of emotions have emerged in Emotion Recognition Systems, most of which are using Machine Learning techniques. The most used feature selection algorithms are: Sequential Forward Selection (SFS), Principal Component Analysis (PCA), ANOVA, Fisher’s linear discriminant and correlation-based feature selection. The most popular classification techniques are: k-Nearest Neighbours (kNN), Bayesian Networks, Regression Trees, Support Vector Machine (SVM), Linear Discriminant Analysis (LDA) and artificial networks.

Soleymani et al (Soleymani et al., 2009) used both SVM and a Bayesian framework for classifying different values of arousal and valence into 3 classes – calm, positive excited and negative excited. The Bayesian classifier produced an accuracy of 64% and SVM with linear kernel, which is identical to linear regression, 56%. In (Koelstra et al., 2012), the Fisher linear discriminant was used for feature selection and a Gaussian naïve Bayes classifier for discriminating EEG and peripheral signals into low/high valence, arousal and liking with accuracies of 61%, 64% and 61%. In (Koelstra et al., 2010), the SVM algorithm conducted to an accuracy of 59%/52%/49% using the Power Spectral Density as EEG feature extraction method and 59%/56%/49% with Common Spatial Pattern components extracted from the EEG signals. In the same study, the classification rates for peripheral physiological signals using the SVM classifier and FCBF feature selection method were 54%/59%/58%. In (Sourina and Liu, 2013), an emotion assessment of EEG data from which Fractal Dimension and Higher Order Crossings features have been extracted conducts to a classification performance of 53.7% for the recognition of up to 8 emotions and 87% for the recognition of 2 emotions, using 4 electrodes. For affective elicitation, in these experiments the users have been presented emotional-stimulating videos, as well as images and sounds from the IAPS and IADS databases (Sourina and Liu, 2013). Chanel et al (Chanel et al., 2011) proposed a method for adapting game levels difficulty, according to the user’s emotional states – boredom, engagement and anxiety. The best accuracy has been obtained by employing ANOVA as feature
selection method and LDA as classifier. Marin-Morales et al (Marin-Morales et al, 2018) designed four virtual immersive environments with varying levels of colour, illumination and geometry with the purpose of eliciting the 4 possible combinations of arousal-valence from the Circumplex Model of Affects. EEG and ECG relevant features have been extracted using the PCA algorithm and the SVM classifier predicted with an accuracy of 71% and 75% along the valence, respectively, arousal dimensions. Although many machine learning techniques have been employed for classification, SVM remains the most popular method, together with PSD features extracted from the frequency bands of the EEG signal (Nafjan et al, 2017). For feature dimensionality reduction, SFS appears to be the most adopted approach (Bontchev, 2016).

Recently, deep learning approaches emerged in the field of emotion recognition. Zheng and Lu (Zheng and Lu, 2015) employed Deep Belief Networks (DBF) for recognizing three emotional levels—positive, neutral and negative from differential entropy features extracted from EEG signals. The DBN model’s accuracy (86.08%) exceeds that of shallow models (SVM – 83.99%, LR- 82.7% and KNN- 72.6%). In the study presented in (Jirayucharoensak et al, 2014), a deep network with a stacked autoencoder is used to discriminate 3 levels of valence and arousal from 32-channel PCA dimensionally reduced EEG data. The classification rate is 53.42% and 52.05%. With SVM, the classification is 41.12% and 39.02%.

Alhagry et al (Alhagry et al, 2017) used a Long-Short Term Memory network to classify raw EEG signals from the DEAP database into low/high arousal, valence and linking with accuracies of 85.65%, 85.45% and 87.99%. In human-centric emotion recognition and affective assessment experiments, classification accuracy depends on the context of the experiment, pursued objectives, methodology, biophysical data recording procedure, number of users, structure and cleanliness of training dataset, feature extraction methods, cross-validation approach and classifier statistical power & parameters tuning.

Classification accuracy depends on the context of the experiment, pursued objectives, methodology, biophysical data recording procedure, number of users, structure and cleanliness of training dataset, feature extraction methods, cross-validation approach and classifier statistical power & parameters tuning. Our approach detaches from these previous research methods, as it adds elements of novelty and originality that consist in an automatic prediction of the next game difficulty level using a trained deep neural network.

5 The Acrophobia VRET Game. A Deep Neural Networks Approach

In the proposed VR system, a game level is characterized by a degree of exposure to a certain height. This level is selected in real-time to ensure that the player faces a challenging scenario, without forcing him into an extreme situation. We recorded in real-time the EEG and biophysical data of the players and used two DNNs: one for fear level classification (DNN1) and one for determining the next level of the game, according to the desired level of fear (DNN2). This sequence of events is included in a game epoch. We call an epoch the execution of the game at a certain level.

Figure 1 presents the system’s architecture and workflow. The user interacts with the Acrophobia VRET and the Virtual Game. Using the Data Acquisition Module, his EEG, HR and GSR signals are collected, pre-processed and transferred to the Database Management System (DBMS). At each game epoch, the user provides his self-assessed fear level of the current game level, called Subjective Unit of Distress (SUD). The SUD is used to determine the prediction accuracy of DNN1. The EEG, HR and GSR data are fed to DNN1 to determine the current fear level. Based on the EEG, biophysical data and desired fear level, DNN2 predicts the game level to be played next.

For fear level prediction, we used two different fear level scales. For the 2-choices scale, 0 represents relaxation and 1 stands for fear. For the 4-choices scale, 0 is mapped to complete relaxation, 1 to low fear, 2 to moderate fear and 3 to a high level of anxiety.

In order to determine the next game level to be played, we used the following approach: we considered n ordered game levels, each level corresponding to a degree of height exposure: l₀, l₁, … lₙ₋₁. The user starts playing the first level of the game (l₀). The current game level is lₙ. During the game, the EEG
and biophysical data are recorded. When a level is completed, the biophysical averaged values are computed and using the trained DNN1, the current fear level (fl_{cr}) is determined. To achieve a gradual and appropriate exposure to height, the next desired fear level (fl_{d}) is calculated.

**Data Acquisition Module**

**Virtual Game**

**Acrophobia VRET**

**DBMS**

**DNNs (DNN1 & DNN2)**

![System architecture and workflow](image)

**Figure 1. System architecture and workflow**

There are two situations: one that considers the 2-choices scale, another that considers the 4-choices scale.

For the 2-choices scale we applied the formulas:

1. if fl_{cr} = 0 then fl_{d} = 1
2. if fl_{cr} = 1 then fl_{d} = 0

Thus, if the user experiences no fear at all (fl_{cr} is 0), we want to move him to a more challenging level, so the desired fear level will output 1, which means a certain level of fear. On the other hand, if the player experiences fear (fl_{cr} is 1), then he may find himself in a too difficult situation during height exposure and then we reduce the desired fear level to 0.

For the 4-choices scale we applied the formulas:

3. if fl_{cr} = 0 or fl_{cr} = 1 then fl_{d} = fl_{cr} + 1
4. if fl_{cr} = 2 then fl_{d} = fl_{cr}
5. if fl_{cr} = 3 then fl_{d} = fl_{cr} - 1

If the user records complete relaxation (fl_{cr} is 0) or a low level of fear (fl_{cr} is 1), then we want to move him to a more difficult level, corresponding to a higher intensity of height exposure. Thus, we calculate the desired fear level to be one level higher than the current one. If the player currently experiences complete relaxation, we want him in the next game level to experience a low level of fear (fl_{d} will be 1). If in the current game level he feels low fear, we want him to go through a medium anxiety intensity in the next game level (fl_{d} will be 2). Moreover, if fl_{cr} is 2, corresponding to a medium level of fear, we maintain the desired fear level to this score, as it means that the player is neither too relaxed nor too anxious and the game level is challenging enough to ensure a motivating and exciting gameplay experience with appropriate height exposure. On the other hand, in the situation when the current fear level has a value of 3, pointing to extreme fear, then the desired fear level will be reduced to 2 – medium fear level – so that the prediction algorithm will take the player to a lower game level where height exposure will be adequate to meet his emotional characteristics.

The desired fear level and biophysical data are inputs for the second deep neural network (DNN2) and a game level (l_{pr}) is predicted to be played next by the user. Consequently, the user plays the predicted
level of the game and his EEG and physiological data are recorded. DNN1 determines again a new general fear level and DNN2 predicts the next game level to be played. The process goes on until a total predefined number of epochs is reached. Figure 2 presents the game workflow for the 4-choices scale.

![Game Workflow Diagram]

**Figure 2. Description of game workflow for the 4-choices scale**

### 6 Method and experimental results

We performed an experiment in which 4 volunteer subjects - 3 women and 1 man, aged 21-49, who have previously been informed about the purpose of the experiment and signed a consent form, played the acrophobia game while their EEG and biophysical data have been recorded. The in-game height level exposure was predicted in real-time by DNN2, based to the user’s fear score (estimated by DNN1), EEG, HR and GSR data.

#### 6.1 The DNN models and their cross-validation accuracies

In order to train the deep neural networks DNN1 and DNN2, we performed some preliminary experiments in which we gradually exposed the subjects to different heights, in both the real and virtual world. In the real-world, some baseline measurements have been performed during complete relaxation, as well as at the first, fourth and sixth floors of a building, at about 4m, 2m and a few centimetres away from the balcony’s railing. Each user has been in-vivo exposed to these height levels twice, before and after the virtual exposure. EEG and biophysical data have been recorded (GSR and HR), as well as the user’s perceived level of fear, the Subjective Unit of Distress (SUD). The SUD was recorded on the 11-choices scale, where the subject had to indicate his self-assessed fear level on a scale from 0 to 10, where 0 represents complete relaxation and 10 stands for extreme fear, anxiety and panic attack. In the case of the in-vivo experiments, the SUD was reported verbally to the researcher assisting the experiment, while for the virtual exposure, the user indicated the SUD by pointing a virtual laser with the controller on a panel that appeared in front of him in the virtual environment.

For recording EEG data, we used the Acticap Xpress Bundle (Acticap Xpress Bundle) device with 16 dry electrodes, where the ground and reference electrodes have been attached to the ears. The dry electrodes have better conductance and are more comfortable for the patient than the wet ones. The following positions have been used, according to the 10/20 system: FP1, FP2, FC5, FC1, FC2, FC6, T7, C3, C4, T8, P3, P1, P2, P4, O1, O2. We recorded the alpha, beta and theta log-normalized powers for all channels, as well as the ratio of the theta to the beta powers (slow waves/fast waves). Electrodermal activity
and heart rate values have been recorded using the GSR unit of the Shimmers Multi-Sensory device (Shimmer Sensing) that was attached to the subject’s left hand.

In the virtual environment, the patients had to collect coins of different colours (bronze, silver and gold) at the ground level, first, fourth, sixth and eighth floors of a building. For providing an immersive and interactive experience, the game has been integrated with the HTC Vive head-mounted display. The player perceived the environment via the virtual glasses, while the interaction was guaranteed by using the controllers – clicking on the floor at various positions by pressing the central button of the controller ensured teleportation, while the coins could be grabbed by pressing and releasing the trigger button. Each user played the game twice. In both virtual and real-world conditions, we totaled a number of 63 trials per patient. These trials resulted in a dataset of 25 000 entries (or data vectors) on average for each patient, recorded at intervals of 65 ms. The datasets were saved in .csv files and used for training DNN1 and DNN2. Thus, although we have a small number of users, we still benefit from a large training dataset to be fed to the networks in order to create reliable prediction models.

Our first goal was to create an accurate and reliable deep neural network model (DNN1) for estimating fear level, based on the recorded EEG and biophysical data. For training DNN1, we used the data recorded during the 63 trials in both the real-world and virtual environment. The input features were the EEG log-normalized powers of all the channels, GSR and HR values, whereas the output feature was the fear level (or SUD), on the 11-choices (values from 0 to 10), 2-choices (values of 0 or 1) and 4-choices (values from 0 to 3) scales. The 2-choices and 4-choices scales have been obtained by grouping together the values from the 11.choices scale. For the 11-choices scale, 0 corresponded to complete relaxation, values from 1-3 to small levels of fear, from 4-7 to medium levels of fear and 8-10 to high anxiety. For the 4-choices-scale, the previous scale values have been grouped together, so that finally 0 corresponded to complete relaxation, 1 to small, 2 to medium, 3 to high level of fear. On the 2-choices-scale, the 4-choices-scale values have been also grouped together, in 0 (for the first two levels, meaning relaxation) and 1 (for the third and fourth levels, meaning fear) values. This grouping has been done in order to improve categorization and classification in the neural networks.

Using the TensorFlow (Tensor Flow Python Framework) deep learning framework backend, we created four Keras (Keras library) sequential models for binary and multi-class classification: Model_1, Model_2, Model_3 and Model_4.

Model_1 has 3 hidden layers, with 150 neurons on each layer. Model_2 has 3 hidden layers, with 300 neurons on each layer. Model_3 has 6 hidden layers, with 150 neurons on each layer. Model_4 has 6 hidden layers, with 300 neurons on each layer. For all the models, the hidden layers used the Rectified Linear Unit (RELU) activation function. For the 2-choices scale, we used the Sigmoid activation function in the output layer and the binary crossentropy loss function. However, for the 4-choices and 11-choices scales, we used the Softmax activation function in the output layer, logarithmical categorical crossentropy loss function and one-hot-encoding that creates 4 or 11 output values (correspondingly 4 or 11 neurons), one for each class. The largest output value will be taken as the class predicted by the model. The models also employ the efficient Adam gradient descent optimization algorithm. Prior to training, the data has been standardized, to reduce it to zero mean and unit variance. The Keras Classifier received as arguments a number of 1000 epochs for training and a batch size of 20. The neural network model has been evaluated on the training data using the KFold method from the scikit-learn library (Scikit Learn Python Library).

We ran the evaluation using a 10-fold cross-validation procedure 10 times and saved the weights of the network in .hdf5 files, together with the corresponding accuracies. Finally, the model version with the highest accuracy has been selected and further used in the experiment. This procedure has been repeated for each user, so that we obtained four personalized fear estimation DNN1 models for each user, for the 2-choices, 4-choices and 11-choices fear scales.

Our second goal was to define a deep neural network model (DNN2) to predict the next game level. For training DNN2, the neural network received as inputs the EEG, GSR, HR and SUD values, while the output represented an encoding of the height where these physiological values have been recorded – 0 for ground floor, 1 for the first floor, 2 for the fourth floor, 3 for the sixth floor and 4 for the eighth floor.
Similarly, using TensorFlow and Keras, we created four multilayer perceptron sequential models—Model_1, Model_2, Model_3, Model_4, RELU activation function for the hidden layers, Softmax activation function for the output layer, logarithmical categorical crossentropy loss function and one-hot-encoding that creates 5 output values (correspondingly 5 neurons), one for each estimated class. Model_1 has 3 hidden layers, with 150 neurons on each layer. Model_2 has 3 hidden layers, with 300 neurons on each layer. Model_3 has 6 hidden layers, with 150 neurons on each layer. Model_4 has 6 hidden layers, with 300 neurons on each layer. The Keras Classifier received the same arguments and the 10 times evaluation with the 10-fold cross-validation procedure was similar to DNN1. The procedure has been repeated for each user, so that we obtained a personalized height exposure (game level) DNN2 model for each user, for both the 2-choices, 4-choices and 11-choices scales. Similarly to DNN1, the model version with the highest accuracy has been selected and further used in the experiment. The maximum cross-validation accuracies obtained for the models are presented in Table 1.

<table>
<thead>
<tr>
<th>DNN1</th>
<th>Model</th>
<th>2-choices scale</th>
<th>4-choices scale</th>
<th>11-choices scale</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Model_1</td>
<td>95.03</td>
<td>87.94</td>
<td>85.09</td>
</tr>
<tr>
<td></td>
<td>Model_2</td>
<td>95.51</td>
<td>90.49</td>
<td>79.48</td>
</tr>
<tr>
<td></td>
<td>Model_3</td>
<td>94.43</td>
<td>86.32</td>
<td>74.27</td>
</tr>
<tr>
<td></td>
<td>Model_4</td>
<td>94.57</td>
<td>88.28</td>
<td>80.45</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DNN2</th>
<th>Model</th>
<th>2-choices scale</th>
<th>4-choices scale</th>
<th>11-choices scale</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Model_1</td>
<td>98.40</td>
<td>98.67</td>
<td>98.75</td>
</tr>
<tr>
<td></td>
<td>Model_2</td>
<td>98.72</td>
<td>98.50</td>
<td>98.65</td>
</tr>
<tr>
<td></td>
<td>Model_3</td>
<td>97.45</td>
<td>97.82</td>
<td>98.50</td>
</tr>
<tr>
<td></td>
<td>Model_4</td>
<td>97.37</td>
<td>97.77</td>
<td>98.17</td>
</tr>
</tbody>
</table>

Table 1. Maximum cross-validation accuracies for DNN1 and DNN2

Besides deep neural networks with the 4 architectural models from above, we also trained our data using the Linear Discriminant Analysis classifier. We obtained a cross-validation accuracy of 87% for the 2-choices model, 71% for the 4-choices model and 64% for the 11-choices model.

6.2 The Game Development

The VR-based game has been developed using the Unity game engine (Unity Game Engine) and written entirely in the C# programming language. It has integrated connectivity with the OpenVibe (OpenVibe) application for collecting EEG signals and with Shimmers Capture (Shimmers Capture C# API) for recording GSR and HR data. The Shimmers Capture application has been modified according to our game integration needs, thanks to the availability of its C# API. The recordings were synchronized in real-time using Lab Stream Layer (LSL) (Lab Stream Layer). The game starts with the user placed on the ground floor (Figure 3). After he collects three coins (bronze, silver and gold), the application remains in standby for a few seconds, the EEG and biophysical data are averaged and the fear level & next game level prediction processes take place in the background. Thus, the corresponding Python scripts for testing are selected, according to the current user and condition (2-choices scale or 4-choices scale). The 11-choices scale has not been integrated yet. First, the script for fear level estimation (DNN1) is called and predicts the current fear level. The desired fear level is calculated based on the formulas described in Chapter 5. Secondly, the script for next game level (or next height exposure level) estimation is called (DNN2), predicting the level where the user should be taken in the next gameplay epoch.
We introduced some gamification elements, i.e. the challenge of collecting coins, as it adds interactivity and purpose to gameplay. The coins are placed at gradual distances from the building’s balcony railing, so that for collecting the golden one it requires the user to bend over the railing and forcefully catch a glimpse of the view (Figure 4).

For each epoch, the averaged EEG, GSR and HR values are stored in log files, together with the predicted fear and game levels. Prior to entering the DNNs, the data is denoised and pre-processed. As the recording devices introduce noise, interrupt temporarily, disconnect or malfunction, we applied a method called “last good value”. For instance, if the HR value at a moment of time is invalid (a negative or a very big number, which is a clear sign of failure), we replace it with the last good value recorded at a previous timestamp (let’s say 86 bpm). If the device malfunction from the beginning, we initialize the last good value with 4.5 microVolts² for the EEG log-normalized power, 1 microSiemens for GSR and 75 bpm for HR. We applied this method because our application runs in real-time and we are not able to manually interfere for inspecting, interpolating or removing the noisy data. Even though we used advanced and expensive sensory devices, the drawback of being unable to fully rely on the recording tools still persists.

Moreover, for each epoch, we saved in separate log files the EEG alpha, beta, theta, GSR and HR values, recorded at intervals of 65 milliseconds. They are saved in both unprocessed and processed (denoised) format, being useful for further experimentation and analysis.

6.3 Experiment and results

Our 4 subjects played the game twice – once using the 2.choices and once using the 4.choices model. Each session contained a number of 10 game epochs. After the user finished one epoch and succeeded in collecting the three coins, he was required to report the perceived fear level for that particular trial (the SUD). A menu appeared on the screen and the answer was given by pointing to the value corresponding to the current self-estimated SUD. Consequently, his biophysical data was saved, together with the current SUD and DNN1 & DNN2 started to run in the background for establishing the current fear level and the next game level where the player should be automatically taken. The purpose of collecting self-estimated SUDs was to validate the accuracy of DNN1. DNN1 predicted the current fear level based on a neural network model created using the data from the previous experimentation and a measure of certifying its faultlessness was by comparing its output with the fear level perceived and acknowledged by the users directly during gameplay – the SUD. We called this parameter validation accuracy. The validation accuracies for each model are presented in Table 2.
The validation accuracy of the LDA classifier for the same data is 60% for the 2-choices scale and 21% for the 4-choices scale. We conclude that Model_1 provided the best training cross-validation and test validation accuracies for DNN1 – for the 2-choices scale, a cross-validation accuracy of 95.03% and validation accuracy of 72.90%. For the 4-choices scale, the values are of 87.84% and 41.89%. As we have not designed yet a method for validating whether the game levels predicted by DNN2 in the experiment are appropriately determined, we do not have a test set for DNN2 and thus we could not calculate its validation accuracy. The only modality used for assessing the efficiency of the proposed approach was by comparing the SUDs reported by the subjects during gameplay with the fear score predicted by DNN1.

The test set is small, containing 10 records for each game session played, this is probably the reason why the validation accuracy did not reach a higher value, especially for the 4-choices scale. Moreover, User1 recorded a low validation accuracy for both the 2-choices and the 4-choices scale, whereas the other users obtained a validation accuracy of over 75%. Due to the poor test results of User1, the average validation accuracy for all the users dropped to the values of approximately 73%, respectively 42%, as presented in Table 2. Without taking into account the data from User1, the validation accuracy of DNN1 is 85% for the 2-choices scale and 60% for the 4-choices scale.

The game levels varied throughout gameplay according to the fear scores, with good results for 3 out of 4 users who recorded a level of fear of 0 (complete relaxation) in the final gameplay epoch. The fourth subject, who suffered from a more severe form of acrophobia, recorded a fear level of 2 in the final gameplay epoch.

A Dynamic Difficulty Adjustment (DDA) of game levels based on the affective state information was proposed by Liu et al (Liu et al, 2009), with prediction accuracy of 78%. Our method conducted to a comparable accuracy. However, their adjustment was based on some simple “if” clauses, not on an advanced prediction method, as ours. Chanel et al (Chanel et al, 2011) tried to adapt the game difficulty levels to the players’ emotional states (boredom, engagement and anxiety). Without feature selection, the best classifiers obtained an accuracy of 55% for peripheral signals and 48% for EEG (LDA, followed by SVM). After the fusion of the two signal categories, their accuracy increased to 63%. We conclude that our classifiers performed equally good, with accuracies of 73% and 42%. The results are promising, but in order to demonstrate the strengths of our method, more experiments need to be done and with a larger number of subjects.

### Table 2. Validation accuracy for DNN1

<table>
<thead>
<tr>
<th>Model</th>
<th>Validation accuracy for DNN1 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2-choices scale</td>
</tr>
<tr>
<td>Model_1</td>
<td>72.90</td>
</tr>
<tr>
<td>Model_2</td>
<td>68.73</td>
</tr>
<tr>
<td>Model_3</td>
<td>62.45</td>
</tr>
<tr>
<td>Model_4</td>
<td>54.12</td>
</tr>
</tbody>
</table>

Conclusions and future work

This paper presented a real-time deep learning-based approach for treating acrophobia in the virtual environment. Two complex neural networks that have been trained with the subjects’ data from an experimental procedure where they have been in-vivo and virtually exposed to different heights. The same users participated in an experiment where they were required to play the game, have their EEG and biophysical signals recorded, report the perceived fear level, but advance to the next game automatically, based on the output provided by the two neural networks. The validation accuracy, defined as the measure of similarity between the fear level estimated by the first deep neural network and the fear level reported subjectively by the user was 73% and 42%. The game levels varied throughout gameplay according to the relaxation/anxiety scores, with good results for 3 out of 4 users who recorded 0 level of
fear in the final gameplay epoch. The main challenges are represented by the instability of the sensory recording devices that sometimes fail to connect, introduce noise or errors in the data. It is very important to have clean data, for both training and testing the neural network models, as they can influence the prediction accuracy. In this phase, offline and online pre-processing and denoising is an essential, indispensable step.

Relying on the promising obtained results, we will continue to extend the research for other types of phobias. Moreover, we will try to use other machine learning techniques, in order to determine the best solutions and make a comparison between a totally automatic approach and a human-centred approach. In addition, we will perform more experiments with a larger number of users and do real-world tests in order to validate the efficiency of the VR treatment and see whether their acrophobic condition has indeed improved.

Acknowledgment

This work has been funded by UEFISCDI proiect 1/2018, UPB CRC Research Grant 2017 and UEFISCDI project PN-III-P1-1.1-TE-2016-2182.

References


Keras Library. URL: https://keras.io/ (visited on 20/11/2018).


