Predicting Real Fear of Heights Using Virtual Reality

Giuseppe Boccignone, Davide Gadia, Dario Maggiorini,
Laura Anna Ripamonti, and Valentina Tosto
{firstname.lastname}@unimi.it
University of Milan
Milan, Italy

ABSTRACT

Every year, in Europe alone, hundreds of workers die by falling from high height. This number could be greatly reduced by means of better training and quick detection of individuals with issues toward work at height. Workers proving to be less suited for the job can be subject to more intensive training or recruited for different positions. Unfortunately, the early detection of workers unsuited for working at height involves specialized personnel and expensive equipment to recreate a stressful environment. In this paper we propose a methodology to predict fear of heights by means of a virtual reality environment. We demonstrate that a 3D virtual environment is feasible for the prediction and give guidelines about meaningful physiological parameters useful for detection.

CCS CONCEPTS

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

KEYWORDS

virtual reality, skills assessment, affective computing, unsupervised machine learning

ACM Reference Format:

Giuseppe Boccignone, Davide Gadia, Dario Maggiorini,, Laura Anna Ripamonti, and Valentina Tosto. 2021. Predicting Real Fear of Heights Using Virtual Reality. In *Conference on Information Technology for Social Good (GoodIT '21), September 9–11, 2021, Roma, Italy*. ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3462203.3475882

1 INTRODUCTION

Every year, work-related accidents are causing a considerable number of deaths. The World Health Organization is reporting, for European Union alone, 5144 and 4712 casualties in 2017 and 2018 respectively [23]. Among these accidents, a fair share (around 25% [9]) is cause of death due to fall from high height. As a consequence, the correct training of workers has always been of paramount importance for many companies. Unfortunately, standard training is requiring specialized personnel and the setup of a very expensive environment in order to put the trainee under elevation stress.

In these last years, *Virtual Reality* (VR) technology proved to be a viable alternative to training in a real setup [20]. By using VR, it is possible to put a trainee under stress by providing a visual stimulus of the working environment without recreating a real dangerous situation. Moreover, in a virtual environment it is possible to monitor physiological parameter and correlate them with the stress level to achieve a better understanding of the situation.

In this paper we propose an evaluation methodology based on the correlation between stress and physiological parameters to assess the attitude of an individual to work at height. This test can identify the best suited subjects to train for working at height while other worker can undergo specialized training or be assigned to less dangerous tasks in safer environments.

We use a VR environment to stress the user and evaluate her suitability to work at height. In particular, we use the VR application *Richie's Plank Experience* to expose test subjects to a simulated height. In the simulation, the user is supposed to walk on a virtual plank suspended in the void while wearing an oculus device. During the experiment, a sensor is also used to record physiological parameters in real time. These physiological parameters are correlated with affect state dynamics along the simulation.

Affect denotes the mental counterpart of bodily sensation and affective features, such as valence and arousal, capture what a given instance of experience feels like [3]. Valence refers to the feeling of pleasure or displeasure; arousal refers to a feeling of activation or sleepiness. It is worth remarking that in the literature concerning the computational modelling of emotions, the term "affect" is often used interchangeably with that of "emotion" but they should not be confused; emotions are constructed from affect, emotional events being specific instances of affect that are linked to the immediate situation and involve intentions to act [3]. Indeed, the approach presented here deals with affect. However, in what follows, markedly when discussing related work, we will occasionally adopt such relaxed convention for sake of simplicity.

Under such circumstances and the peculiar experimental setup of this work, gauging the affect state of participants is assumed to be effective in order to assess a potential acrophobia (fear of heights) or basophobia (fear of falling). To evaluate affect, a feedback form based on the *Pleasure (Valence) - Arousal - Dominance* (PAD) [19] affect state model is used.

The remainder of this paper is organized as follows: in Sec. 2 related work is discussed while in Sec. 3 our virtual plank experiment is described in detail. Collected data is analyzed in Sec. 4 and observed results are discussed in Sec. 5. Section 6 concludes the paper and proposes future work.

2 RELATED WORK

A wide spectrum of studies are already present in literature about emotional response analysis in virtual environments. These studies focus mainly on how to analyze physiological response and how different environments can stimulate distinct emotions.

With respect to physiological response analysis, the general approach is to stress the user and then correlate physiological parameters such as heart rate and blood pressure variation to the proposed virtual situation.

In [18] a test subject is placed first inside an elevator and then on an aerial moving platform, from which she is supposed to jump off. Authors claim that, despite an increase of heart rate, blood pressure and hydrocortisone (also known as cortisol) levels remain constant while on the platform. Moreover, the hydrocortisone level decreases when inside the elevator.

In [1], instead, the test subject is asked to traverse a grid of ice blocks with the risk of falling down. During the experiment user movements are recorded alongside skin conductance level and facial electromyography. Authors found that a risk-averse behavior was more evident in participants with an high neuroticism personality profile. Moreover, these users made also more frequent Risk Assessments than the average.

Authors of [15] measure cardiovascular and cortisol reactivity to the VR equivalent of a *Trier Social Stress Test* (TSST). In a TSST the participant is asked to hold a speech and to do an arithmetic task in front of an audience to reproduce stress in laboratory settings. Virtual reality was used to recreate a virtual audience to the presenter and, for the proposed case, results resembled those obtained in prior studies using a real-life TSST.

Other contributions in literature, such as [8] focus on understanding if a VR environment is capable to generate the right psychophysiological condition for an effective exposure treatment. In [8], authors found that VR exposure does provoke psychophysiological arousal, especially in terms of electro-dermal activity, making it feasible for cognitive behavioral therapy.

Another branch of research is focusing instead in understanding the emotional state induced by a virtual environment. The general idea is to devise and perform a reliable classification of the emotional state in the user and correlate it with a virtual experience.

The *Affective Virtual Reality System* (AVRS) [17] is a system to elicit emotions using a VR environment. AVRS detect the level of arousal (i.e., the autonomic nervous system stimulation) through measurement of the heart rate and using the *Self-Assessment Manikin* (SAM) technique. Using SAM, a stylized figure representing the intensity of the affect dimension must be selected on a scale or grid. This way, SAM can overcome problems related to the social and cultural context linked to the adjective used to describe the level of arousal. In the paper, the levels of arousal solicited by the same video in VR and via standard screen are compared. Authors claim that the average arousal level is higher in VR when the scene is depicting happiness or fear, while there is no significant difference for other emotions. In particular, for fear, the difference seems to depend on the negativity of the scene and the VR visual quality. Distaste seems to be the only emotion stronger on screen, even if the heart rate did not show significant changes between the two technologies.

Similar to AVRS, the EMMA project [21] proposes the development of a *Mood Induction Procedure* using VR (VR-MIP) to elicit emotions such as sadness, joy, anxiety, and relax in experimental subjects. Users have been exposed to virtual environments populated with objects capable to trigger an emotion. The goal was to observe relevant changes in the mood of the users during the simulation. During the experiments, the virtual environment proposed to the user was neutral at first, but then it got modified with respect to the emotion that should be reached. In this project the intensity of the emotion has been evaluated using the combination of three feedback forms: *Visual Analogue Scale* (VAS), *ITC-Sense of Presence Inventory* (ITC-SOPI), and *Reality Judgement and Presence Questionnaire* (RJPJ). Results demonstrated that the four environments (one for each emotion) were actually able to induce a mood change in the user. Differently from other contributions, in [21] anxiety is kept well separated from fear. This is due to the definition from Barlow [2] stating that "anxiety is a diffuse, objectless apprehension" while fear is an emotional response provoked by a "present" and specific threat.

In [6] authors demonstrated that observed geometric shapes and daylight illumination are in correlation with heart rate and skin conductance. Different building façade patterns have been considered. Each pattern induced light with a specific geometric diffusion (irregular, regular, and "venetian style") inside a virtual space. Experimental results hinted that participants found the space to be more interesting when illuminated using an irregular pattern. The calculation of Spearman's coefficient [24] showed a statistically significant negative correlation between mean heart rate change and interest.

3 THE PLANK EXPERIMENT

In this section we are going to describe the setup and how the experiments have been performed in detail.

As already mentioned in Sec. 1, we are using *Richie's Plank Experience* as a virtual environment to stress the user with a simulated height. To perform the test in the virtual space the user is first required to take a virtual elevator riding upward for a long time. This is to give the feeling that an extremely high floor has been reached. When the elevator doors open an altitude view of a modern cityscape is presented. The doors open on nothing but for a small plank protruding in the void for around one meter (see Fig. 1). The user is supposed to walk in a real (and safe) environment to reach the end of the plank. In order to add realism, we also place a piece of wood on the floor for the user to walk on (see Fig. 2). The experiment ends as soon as the user reaches the end of the plank.

Each experiment is composed by three steps:

- **Preparation**. During this step the user is required to sit and see a soothing video in order to stabilize her physiological data. The elevator phase during the simulation is supposed to suffice to the same function. Unfortunately, we discovered that it was not long enough for some users, so we decided to reinforce this step with an additional video. While watching the video, the user is also wearing the physiological sensor and a calibration is performed.
- **Simulation**. In this step the user is equipped with the oculus rift visor and the simulation is started. This step is taking

Figure 1: The plank protruding out form the elevator door.

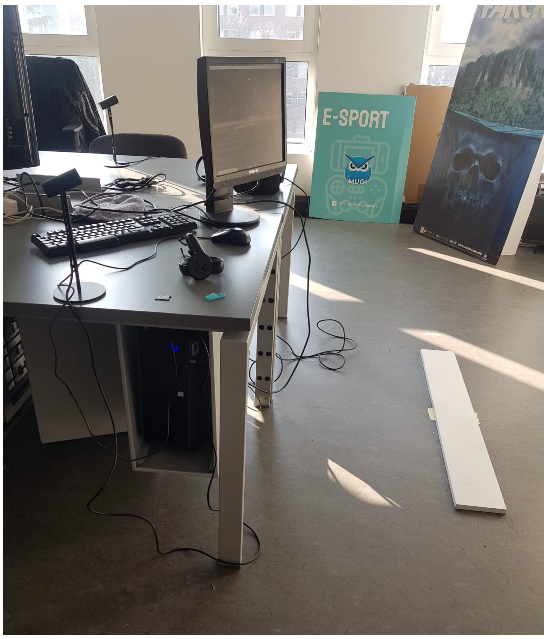

Figure 2: The physical plank setup in our lab.

one to four minutes to complete, depending on the user's immersion.
- **Self evaluation**. In this step the user, using an affect annotation tool [4], is evaluating her experience.

During the experiment, the user is recorded on camera. This video, merged with another video recorded in first person from inside the virtual environment, is used to collect the affect measurements.

The Experiment has been performed on 33 volunteer students; 28 males and 5 females in the age range between 22 and 30 years. All of them reported to be in good health: no cardiovascular pathologies, no anxiety disorders, and no neurological alterations. Moreover, in order to be accepted as volunteer each student must not had used the Richie's Plank Experience application before.

3.1 Measuring Physiological Parameters

To collect physiological data we use the *E4* sensor wristband from *Empatica*. While walking on the virtual plank, the physiological sensors will collect information about heart rate and *Electro-Dermal Activity* (EDA). We focus on these two values because they are strictly connected with the arousal, as demonstrated by Leyner et al. in [16].

While the importance of heart rate can be quite straightforward, some more explanation is required for EDA. EDA is measuring the skin conductance variation caused by sweating. It is possible to identify two kinds of EDA: tonic and phasic. The tonic EDA, measuring the skin electrical resistance, is an indicator of changes in the nervous system activation while the phasic EDA is detecting sharp signal changes linked to emotional stimuli as outlined in [12].

The sensor collects data from the user during all the simulation step. When the simulation is over, all data is downloaded from the wristband via bluetooth link for later processing.

In our case filtering the sensor output was not required because, due to the excellent quality of the wristband and the very controlled environment during the experiments, the collected data was already clean enough for processing. Examples of collected values for EDA and heart rate can be seen in Fig. 3.

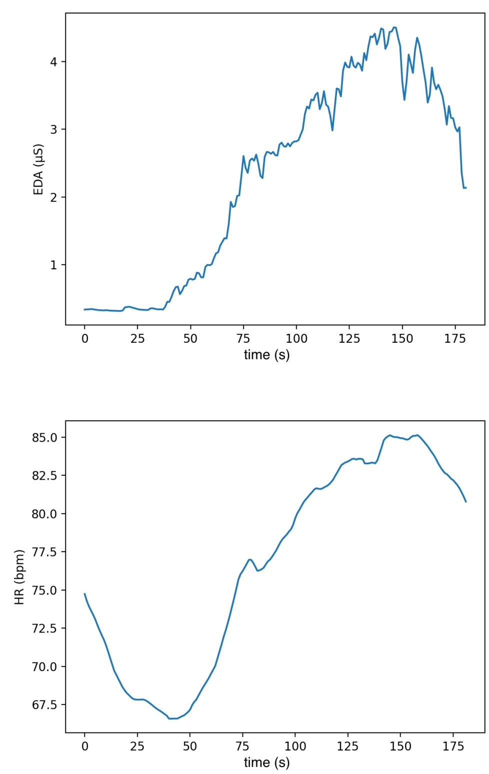

Figure 3: Sample values for EDA and heart rate.

3.2 Measuring Affect

In order to reconstruct the complete status of the user during the virtual experience, collecting physiological parameters is not enough. We also must collect the affective states. In our experiments, affective states are detected through user's self-evaluation. As we said, in this experiment we use the PAD model proposed by Mehrabian et al. [19] (see Fig. 4). This model provides a way to map all

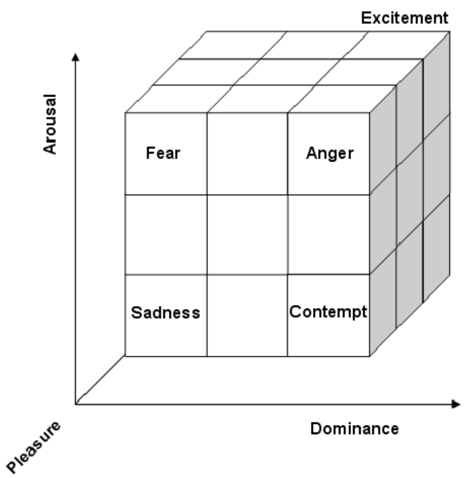

Figure 4: Pleasure (Valence) - Arousal - Dominance (PAD) model of affect status.

possible states in a 3D space. The three dimensions of this space are *valence (pleasure)*, *arousal*, and *dominance*. Briefly, the model augments the basic valence/arousal dimensions with a dominance-submissiveness dimension. We collect intensity information with SAM, already described in Sec. 2, using a grid (see Fig. 5).

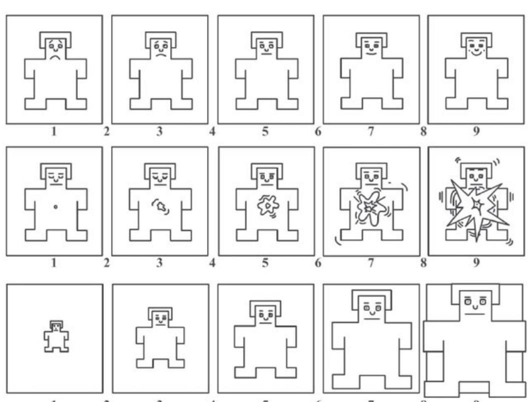

Figure 5: SAM grid: first row is valence, second is arousal, and last row is dominance.

After walking on the virtual plank, the user is required to provide a self assessment about the levels of arousal and valence felt during the experiment. To provide the assessment, we ask the user to annotate the video recorded during the experiment with an online tool named *Dimensional ANnotation Tool for Emotions* (DANTE) [4] (see Fig. 6). While replaying the video, the user can input in real

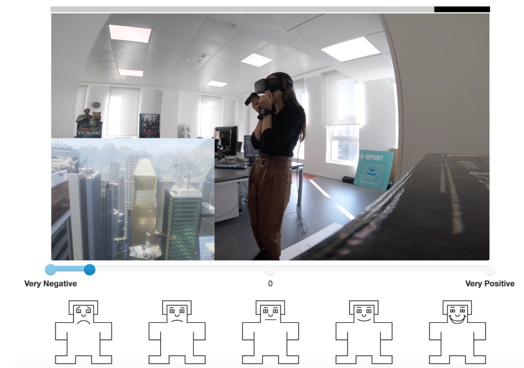

Figure 6: Valence assessment using SAM.

time the levels of arousal and valence. The video timestamps allow us to map each affect state to the right position in the data sequence collected by the sensor.

4 ANALYSIS

To analyze data collected during the experiments, we use a unsupervised machine learning approach to classify test subjects based solely on available information. This approach has been required because we want to keep the proposed solution as general as possible. Using a supervised approach could have been more efficient both in term of computational time and accuracy but, at the same time, it would be binded to a specific training method or philosophy. Moreover, using an unsupervised approach allows to easily adapt the methodology to other attitudinal evaluations where, most important, different groupings/clusters might emerge.

The analysis activity has been divided into three steps: selection, clustering, and statistical evaluation.

4.1 Selection

In the first step, we select descriptors for EDA and heart rate. What we are looking for is a set of functions to describe the collected data as a vector. To select these descriptors we have to devise a set of general features for both data series. After a long and detailed analysis of the current literature, we selected a total of 30 descriptors. Unfortunately, the detailed discussion of each descriptor do not fall into the goal of the current paper and would not fit in the allowed space. For this reason, we will limit ourselves to provide a list here.

With respect to EDA, we selected Shannon's Entropy and other 14 descriptors leveraging on [22] from Shukla et al. These descriptors are: mean, standard deviation, kurtosis ([7]), asymmetry, peaks number, total and mean amplitude, total and mean signal raise time, subtended area, the three Hjiorth parameters, and the spectral density.

On the other hand, when dealing with heart rate, we define other 15 descriptors based on *heart rate variability* (HRV), which is a recognized as an indicator of physiological stress. Four of the

15 descriptors are based on statistics: mean, standard deviation, minimum and maximum HRV. Seven are based on the *inter-beat interval*: IBI, N-N median and range, SDNN, SDSN, RMSSD, and NN50. Three indicators are in the frequency domain: spectral density, hi-frequency variance (in HRV), and low-frequency variance (in HRV). The last indicator is a geometric one: the triangular index.

On this sets we must apply a selection technique to identify the best features to be used in the clustering step. This is a standard approach in Machine Learning to avoid the learning model to suffer from overfitting. Overfitting means that the we adopt a number of parameters too large with respect to the collected samples; thus, reducing performances and accuracy.

To select the most relevant descriptors we use the *Spectral Feature Selection* (SPEC) algorithm. SPEC estimates a descriptor relevance based on its consistency with the spectrum of a similarity matrix calculated using the *Radial Basis Function* (RBF) [5]. In simpler words, the more relevant descriptors are those exposing a uniform behavior between clusters.

In our case, using SPEC, we are able to reduce the number of descriptors from 30 to ten. The ten selected descriptors are: Kurtosis, mean signal raise time, standard deviation of HRV, SDNN, N-N range, RMSSD, NN50, low-frequency variance, spectral density, and the triangular index.

As a final remark, it is important to remember that different selections may lead to a different set of clusters in the next step.

4.2 Clustering

In this stage, the goal is to identify similar users and classify them in groups named clusters.

As already said, we are using an unsupervised approach, precisely we infer clusters from unlabeled information [14]. At the most general level, a clustering procedure to group N data points into K clusters relies upon (*i*) a model for representing each point as a vector of informative descriptors, or features, in a D-dimensional space, (*ii*) a metric to calculate the similarity between pairs of elements, (*iii*) the actual clustering algorithm, (*iv*) a set of parameters summarising cluster information and (*v*) eventually a validation step. The features selected to represent the original data should be maximally informative so to obtain clusters that are suitable to unveil the hidden structure of observable data. As to the metric, this can be often shaped in the form of a distance defined over the D-dimensional vector space.

In particular, the clustering algorithm is in charge to identify groups of points that are close together in the feature D-dimensional space (clusters), while maximising the distance between clusters. Once the clusters have been identified, a concise description can be provided for each cluster; in what follows, we exploit clustering techniques that can be framed in probabilistic terms (e.g., mixture of gaussians); thus, one such description can be given via the model parameters (e.g, the gaussian mean) so far inferred. Clearly, the clustering output needs to be evaluated against a set of quality indexes. Eventually, if such comparison raises critical issues, then a model revision should be taken into account, which can lead to an iterative refinement of cluster computation.

In our specific case, to identify clusters, we apply techniques well known in literature such as *K-means* and *Expectation Maximization*. K-means is an iterative method: it refines the clusters at each iteration using the concept of centroid as the center of mass for each cluster. This method tries to find a stable configuration for a preset number of centroids such as that each centroid is the closest one for all points belonging to its cluster. The clustering process terminates when the centroids movement after the iteration is lower than a given threshold. Expectation Maximization, instead, is designed to tune the parameters of a mixture of gaussian distributions. With this method, we look for the gaussian parameters that are maximizing the probability to generate all the points in the D-dimensional space.

4.3 Statistical Evaluation

In this third step the clustering is evaluated from a statistical standpoint; i.e., we have to check if the clusters are distinct enough in the D-dimensional space to bring us enough information.

Usually, in statistics, to test whether the null hypothesis of equal test statistic of two or more measurements can be rejected, an analysis of the variance is performed; this is called the ANOVA [13] test. The ANOVA test is typically conducted via p-values, namely the probability of obtaining test results at least as extreme as the results actually observed, under the assumption that the null hypothesis is correct; if $p < .05$, the result is deemed statistically significant and the null hypothesis is rejected; otherwise, the the null hypothesis is retained.

In our experiment, to identify statistical differences between clusters, we consider each cluster as a set of (D-dimensional) points and use MANOVA, a multivariate extension of ANOVA, to perform the analysis. MANOVA adopts a linear multivariate regression model that can be used to measure the correlation between groups of one or more independent variables. We can use MANOVA to evaluate the output of the clustering step by means of four test statistics which are quite common in literature: *Wilks' lambda*, *Pillai's trace*, *Hotelling-Lawley trace*, and *Roy's greatest root*.

The MANOVA test is applied on the clusters obtained from all the combinations described so far: with all the descriptors or just one of the selections, with K-mean or Expectation Maximization, and requiring to evaluate from 2, 3, or 4 clusters. Among all the results, the most promising proved to be the combination of 2 clusters using K-means and the descriptors selection for EDA. As a matter of fact, this was the only combination for which the MANOVA test rejected the null hypothesis of the clusters being similar in terms of test statistic equality, with any test statistic ($p < 0.05$, always).

5 RESULTS

As already mentioned, the best combination to classify the users seems to be by using only the descriptors for EDA, to assume two clusters, and to perform clustering using K-mean.

First of all, the division in two clusters seems to be confirming the goal of this research: it is actually feasible to divide the test subjects into two categories: suited and unsuited to work at height. Second, the electro-dermal activity seems to be a better indicator compared to heart rate, which is in contrast with the usual belief that heart rate is the primary indicator for physiological stress in every situation.

To draw some more practical conclusions, we have to compare the users belonging to each cluster with the video (annotated with SAM) recorded during the experiment and the data collected by the sensor. Among the 33 test subject, 26 provided valid data for the experiments; the other 7 were not able to complete the simulation due to being too scared from the virtual environment. With respect to the 26 valid tests, 8 users has been assigned to cluster 1 (unsuited for working at height) and 18 to cluster 2 (suited for working at height). All users in cluster 2 reported a medium-low level of arousal and quite steady physiological data; for this reason, here we will focus on cluster 1.

By examining the videos of users belonging to cluster 1, we could observe a lot of similarities such as the way they moved on the plank and how they looked around in the virtual environment. In particular, two of them exposed visible traits of fear and confusion during the simulation. Moreover, all users of cluster 1 reported a very high level of arousal (> 0.8 in the range $[0, 1]$).

Considering now the physiological data collected by the sensor, the EDA values have been steadily increasing during the simulation for all users of cluster 1. On the contrary, for the heart rate, we can observe two distinct behaviors: for some users the heart rate raised sharply and stood there for the whole time, for the others it decreased smoothly during the simulation. Our best hypothesis is that these behaviors are the result of how each individual is dealing with fear. Anyway, this can be the reason why the HRV indicators have not been so relevant for the clustering. As a matter of fact, the users reporting a sharp increase in the heart rate presented also a bit more clearly their fear in the recording.

Finally, to analyze the effectiveness of each indicator, it is possible to use a scatter plot matrix. A scatter plot matrix is a matrix of scatter plots used to visualize bivariate relationships between combinations of variables. Each plot in the matrix provides a graphical representation of the relationship between a pair of variables. In our case, by means of a matrix with scatter plots of clusters distribution for the EDA indicators, it is possible to observe that the best indicators are peaks number, total signal raise time, and the Hjiorth parameters for activity and complexity. Basing on this observation, it could be possible to re-run the analysis using just these four indicators and obtain similar results. Of course, this will be subject for future research.

6 CONCLUSION AND FUTURE WORK

In this paper we proposed a methodology to predict fear of heights by means of a virtual reality environment. Data has been collected with real experiments and then processed using an unsupervised machine learning approach for clustering. We demonstrate that a 3D virtual environment is actually feasible for the early prediction of fear of heights. Our methodology could actually split the users into two groups: suited and unsuited to work at elevations. Moreover, we discovered that, for the detection, electro-dermal activity is a better indicator rather then heart rate. Our analysis is also hinting to the fact that four features of the Electro-Dermal Activity (peaks number, total signal raise time, and the Hjiorth parameters for activity and complexity) could be enough to identify the two groups.

In the future we plan to extend this research by using more physiological information such as body temperature or blood pressure and by applying the same methodology to early detection for other kinds of irrational fears such as arachnophobia, autophobia, or trypophobia. Moreover, it could be also possible to apply this same methodology for early diagnosis of learning disorders [10, 11].

REFERENCES

[1] C. Baker, R. Pawling, and S. Fairclough. 2020. Assessment of threat and negativity bias in virtual reality. *Scientific Reports* 10 (Oct. 2020), 10 pages.
[2] D.H. Barlow. 2004. *Anxiety and its disorders: The nature and treatment of anxiety and panic.* Guilford press.
[3] Lisa Feldman Barrett. 2017. The theory of constructed emotion: an active inference account of interoception and categorization. *Social cognitive and affective neuroscience* 12, 1 (2017), 1–23.
[4] G. Boccignone, D. Conte, V. Cuculo, and R. Lanzarotti. 2017. AMHUSE: a multimodal dataset for HUmour SEnsing. In *Proceedings of the 19^{th} ACM International Conference on Multimodal Interaction.* 438–445.
[5] M. D. Buhmann. 2000. Radial basis functions. *Acta numerica* 9 (2000), 1–38.
[6] K. Chamilothori, G. Chinazzo, R. de Matos, P. João, E. Dan-Glauser, J. Wienold, and M. Andersen. 2018. *Perceived interest and heart rate response to façade and daylight patterns in Virtual Reality.* Technical Report. Academy of Neuroscience for Architecture.
[7] C. A. Gill D. N.Joanes. 1998. Comparing measures of sample skewness and kurtosis. *Journal of the Royal Statistical Society: Series D (The Statistician)* 47, 1 (1998), 183–189.
[8] J. Diemer, A. Mühlberger, P. Pauli, and P. Zwanzger. 2014. Virtual reality exposure in anxiety disorders: Impact on psychophysiological reactivity. *The World Journal of Biological Psychiatry* 15, 6 (2014), 427–442.
[9] EuroSafe. 2016. *Injuries in the European Union, Summary on injury statistics 2012-2014* (6th ed.). EuroSafe, Amsterdam.
[10] O. Gaggi, G. Galiazzo, C. E. Palazzi, A. Facoetti, and S. Franceschini. 2012. A Serious Game for Predicting the Risk of Developmental Dyslexia in Pre-Readers Children. In *Proceedings of the 21^{st} International Conference on Computer Communication Networks.* 1–5.
[11] O. Gaggi, C. E. Palazzi, M. Ciman, G. Galiazzo, S. Franceschini, M. Ruffino, S. Gori, and A. Facoetti. 2017. Serious Games for Early Identification of Developmental Dyslexia. *Computers in Entertainment* 15, 2, 1–24.
[12] A. Greco, G. Valenzao, and E. P.Scilingo. 2016. Modeling for the Analysis of the EDA. In *Advances in Electrodermal Activity Processing with Applications for Mental Health.* Springer, 19–33.
[13] P. K. Ito. 1980. 7 robustness of anova and manova test procedures. *Handbook of statistics* 1 (1980), 199–236.
[14] A. K. Jain, M. N. Murty, and P. J. Flynn. 1999. Data clustering: a review. *ACM computing surveys (CSUR)* 31, 3 (1999), 264–323.
[15] P. Jönsson, M. Wallergård, K. Österberg, Å Marie Hansen, G. Johansson, and B. Karlson. 2010. Cardiovascular and cortisol reactivity and habituation to a virtual reality version of the Trier Social Stress Test: A pilot study. *Psychoneuroendocrinology* 35, 9 (2010), 1397–1403.
[16] D. Leiner, A. Fahrs, and H. Früh. 2012. EDA positive change: A simple algorithm for electrodermal activity to measure general audience arousal during media exposure. *Journal on Communication Methods and Measures* 6, 4 (2012), 237–250.
[17] D. Liao, W. Zhang, G. Liang, Y. Li, J. Xie, L. Zhu, X. Xu, and L. Shu. 2019. Arousal evaluation of VR affective scenes based on HR and SAM. In *2019 IEEE MTT-S International Microwave Biomedical Conference (IMBioC)*, Vol. 1. IEEE, 1–4.
[18] M. Martens, A. Antley, D. Freeman, M Slater, P. J. Harrison, and E. Tunbridge. 2019. It feels real: physiological responses to a stressful virtual reality environment and its impact on working memory. *Journal of Psychopharmacology* 33, 10 (2019), 1264–1273.
[19] A. Mehrabian. 1996. Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament. *Current Psychology* 14, 4 (1996), 261–292.
[20] G. Quadrio, A. Bujari, O. Gaggi, and C. E. Palazzi. 2019. New Interactive VR/AR Media: A Challenge Worth to Address. In *Proceedings of the 5^{th} EAI International Conference on Smart Objects and Technologies for Social Good.* 128–129.
[21] R.M.Baños, C. Botella, V. Liaño, B. Guerrero, B. Rey, and M. Alcañiz. 2004. Sense of presence in emotional virtual environments. In 7^{th} *Annual International Workshop on Presence.* 156–159.
[22] J. Shukla, M. Barreda-Angeles, J. Oliver, G. C.Nandi, and D. Puig. 2019. Feature Extraction and Selection for Emotion Recognition from Electrodermal Activity. *IEEE Transactions on Affective Computing* (2019).
[23] World Health Organization 2020. *Number of deaths due to work-related accidents.* World Health Organization. Retrieved Feb 24, 2021 from https://gateway.euro.who.int/en/indicators/hfa_457-4071-number-of-deaths-due-to-work-related-accidents/
[24] J. H. Zar. 1972. Significance testing of the Spearman rank correlation coefficient. *J. Amer. Statist. Assoc.* 67, 339 (1972), 578–580.