Emotional Virtual Reality Stroop Task: Pilot Design

Deniz Mevlevioğlu
d.mevlevioglu@cs.ucc.ie
University College Cork
Cork, Ireland

Sabin Tabirca
tabirca@cs.ucc.ie
University College Cork
Cork, Ireland

David Murphy
d.murphy@cs.ucc.ie
University College Cork
Cork, Ireland

Figure 1: The Virtual Reality waiting room environment

ABSTRACT

Anxiety-inducing and assessment methods in Virtual Reality has been a topic of discussion in recent literature. The importance of the topic is related to the difficulty of getting accurate and timely measurements of anxiety without relying on self-report and breaking the immersion. To this end, the current study utilises the emotional version of a well-established cognitive task; the Stroop Color-Word Task and brings it to Virtual Reality. It consists of three levels; congruent which is used as control and corresponds with no anxiety, incongruent, which corresponds with mild anxiety and emotional, which corresponds with severe anxiety. This pilot serves two functions. The first is to validate the effects of the task using biosignal measurements. The second is to use the bio signal information and the labels to train a machine-learning algorithm. The information collected by the pilot will be used to decide what types of signals and devices to use in the final product, as well as what algorithm and time frame will be better suited for the purpose of accurately determining the user's anxiety level within Virtual Reality without breaking the immersion.

KEYWORDS

biosensors, biosignals, VR, emotional stroop, anxiety, GSR, PPG, EEG

ACM Reference Format:

Deniz Mevlevioğlu, Sabin Tabirca, and David Murphy. 2021. Emotional Virtual Reality Stroop Task: Pilot Design. In *27th ACM Symposium on Virtual Reality Software and Technology (VRST '21), December 8–10, 2021, Osaka, Japan.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3489849.3489952

VRST '21, December 8–10, 2021, Osaka, Japan

ACM ISBN 978-1-4503-9092-7/21/12.
https://doi.org/10.1145/3489849.3489952

1 INTRODUCTION

Anxiety inducing methods have been long discussed in the field of affective computing. When it comes to using cognitive tasks to measure arousal levels, it's important to think of how stress affects cognitive performance. According to Yerkes-Dodson law [4], intermediate stress leads to optimum performance when too much or too little stress leads to poor performance. Based on this, we aim to develop a system to get classify arousal levels within Virtual Reality (VR).

This system is a VR application of an emotional variant of the well-known Stroop Color-Word task (SCWT) [6]. SCWT is a commonly-used and validated cognitive task asking the person taking the test to name the colour of the ink in words [3]. It generally has two stages, non-interference and interference, where words match the colour and not, respectively (fig 2) [3]. There are similar applications of this task within Virtual Reality, and they show similar results to real-world applications of the task [1]. In this study, we add a third stage, where the words are negatively valenced English words from Affective Norms for English Words (ANEW) [2].

2 DESIGN

2.1 Experience Design

The scene is built in Unity3d and takes place in a waiting room, modelled to look like a modern clinic. For the administration of the Stroop task, the user is taken in front of a big screen and an elevated platform to input their responses (fig 2). The structure of the flow of the task can be seen in table 1.

2.2 Biosignals

The Shimmer device will be used to measure the skin conductivity level (SCL). For half of the users, the optical pulse attached to the

Table 1: Flow of the experiment

Instructions + Relaxation	Prep	Congruent	Break	Prep	Incongruent	Break	Emotional
1 Minute	30 sec (10 questions, 3 seconds each)	2.5 Minute (50 questions, 3 seconds each)	15 sec	30 sec (10 questions, 3 seconds each)	2.5 Minute (50 questions, 3 seconds each)	15 Sec	2.5 Minute (50 questions, 3 seconds each)

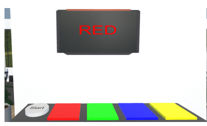

Figure 2: Stage 1

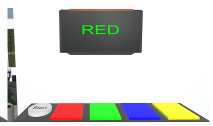

Figure 3: Stage 2

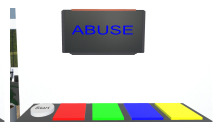

Figure 4: Stage 3

Figure 2: The 3 stages of the Stroop task

Shimmer device will determine the photoplethysmogram (PPG), using which the heart rate (HR) and heart rate variability (HRV) will be derived. For the other half, an electrocardiogram (ECG) will be used to measure the HR and HRV. Generally, ECG is more reliable than PPG, but also bulkier, which might decrease the immersion and make the user enjoy the experience less. The two will be compared in terms of how comfortable the user is and how accurate the assessment is. Myndplay brainband will be used to measure electrical brain activity. The reason this device was chosen is its superior mobility and ease of being used in conjunction with a head-mounted display (HMD). Half the users will be administered a single dry node and half a single wet node. Wet nodes provide better signals but are less comfortable, so the two of them will be compared in terms of signal quality, the accuracy of the final algorithm and comfort. Head movement information will be received from the HMD.

2.3 Participants and procedure

15 participants are to be selected from university students on a voluntary basis. The participants will receive an information sheet and be asked to fill in a consent form if they agree. Then, they will receive a short questionnaire including demographics (age, gender, exposure to VR) and State-trait anxiety inventory (STAI) to assess general anxiety. The participants will then be prepared for the experience. The experimenter will put on the biosensors and HMD to make sure they fit well and the user doesn't have any problems with them. The instructions on the experience are included in the experience, so there will be no extra explanation to make sure all the participants receive the same information. When the experiment is over, the experimenter will take off all the equipment and the participant will receive a second questionnaire for feedback on the experience and immersion, which will be measured using Igroup Presence Questionnaire (IPQ) [5]. The experience will be delivered using the HTC Vive HMD and a computer with a GeForce GTX Titan X graphics card and Intel i7-5820k processor. HTC Vive earphones will be used for the audio.

3 STRESS CLASSIFICATION

Classification of stress will be done through supervised learning using the stages from the Stroop Task to label the data. The inputs are going to be the HR measured in beats per minute (bpm), HRV (R-R interval), SCL measured in µS, breathing rate and alpha1, alpha2, beta1, beta2, theta1, theta2 band signals from the EEG.

There will be multiple classification methods used to be able to extract the best inputs and the most efficient classifier. Simpler classification models such as K-Nearest Neighbor(KNN), Support Vector Machines (SVM) and decision trees (DT) will be compared to neural networks, both simple (3 layers) and more complex deep neural networks. The difference in accuracy between deep neural networks and other classifiers will be measured to find a simple yet accurate solution to the problem to make sure that the efficiency is the highest. For the inputs, those that are less bulky and make the user more comfortable and immersed are going to be given priority while making sure not to decrease accuracy significantly. Timeframes of 1 second and 3 seconds will be compared to find the best accuracy as well. In an application where the stressor is constant and reactions to new stimulus are not being measured, 3 seconds is a short enough time to measure stress without it affecting the user's experience, so the accuracy will be prioritised in this case. The accuracy difference between 2-level (no-anxiety and anxiety) and 3-level (no anxiety, mild anxiety and severe anxiety) will also be tested.

4 CONCLUSION

This pilot study serves two main purposes. The first is to test the usability of the materials. We want to make sure that the hardware is not too uncomfortable or hard to use and adapt to for the users. We also want to validate our task, to see if it elicits the predicted response of increased stress per stage. We do that by using self-report measurements and biosensor information. We believe that this will keep the bias to a minimum by combining different validation methods.

The second is to find the best combination of signals, devices, levels (2 or 3) and classifiers for the classification of stress models. This is very important to us to make sure that the system is as comfortable as possible without compromising on accuracy.

ACKNOWLEDGMENTS

This publication has emanated from research supported in part by a Grant from Science Foundation Ireland under Grant number 18/CRT/6222.

REFERENCES

[1] Christina M Armstrong, Greg M Reger, Joseph Edwards, Albert A Rizzo, Christopher G Courtney, and Thomas D Parsons. 2013. Validity of the Virtual Reality Stroop Task (VRST) in active duty military. *Journal of Clinical and Experimental Neuropsychology* 35, 2 (2013), 113–123. https://doi.org/10.1080/13803395.2012.740002

[2] Margaret M Bradley and Peter J Lang. 1999. *Affective norms for English words (ANEW): Instruction manual and affective ratings.* Technical Report. Technical report C-1, the center for research in psychophysiology https://doi.org/10.1.1.306.3881

[3] Charles J Golden and Shawna M Freshwater. 1978. Stroop color and word test. (1978). https://doi.org/10.3389/fpsyg.2017.00557

[4] Karl Halvor Teigen. 1994. Yerkes-Dodson: A law for all seasons. *Theory & Psychology* 4, 4 (1994), 525–547. https://doi.org/10.1177/0959354394044004

[5] Jacinto Vasconcelos-Raposo, Maximino Bessa, Miguel Melo, Luis Barbosa, Rui Rodrigues, Carla Maria Teixeira, Luciana Cabral, and António Augusto Sousa. 2016. Adaptation and validation of the Igroup presence questionnaire (IPQ) in a Portuguese sample. *Presence* 25, 3 (2016), 191–203. https://doi.org/10.1162/PRES_a_00261

[6] J Mark G Williams, Andrew Mathews, and Colin MacLeod. 1996. The emotional Stroop task and psychopathology. *Psychological bulletin* 120, 1 (1996), 3. https://doi.org/10.1037/0033-2909.120.1.3