Same Same but Different: Exploring the Effects of the Stroop Color Word Test in Virtual Reality

Romina Poguntke, Markus Wirth, Stefan Gradl

HAL Id: hal-02544574
https://inria.hal.science/hal-02544574

Submitted on 16 Apr 2020

Same Same But Different: Exploring the Effects of the Stroop Color Word Test in Virtual Reality

Romina Poguntke[1], Markus Wirth[2], and
Stefan Gradl[2]

[1] University of Stuttgart, Germany {`firstname.lastname`}`@vis.uni-stuttgart.de`
[2] Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Germany
{`firstname.lastname`}`@fau.de`

Abstract. Virtual reality (VR) is used for different trainings e.g. for pilots, athletes, and surgeons. Dangerous and difficult situations are often focused in such simulations in VR, targeting to learn how to perform well under stress. However, there has been little work on understanding stress perception in VR compared to the real-world situation. In this paper we present an investigation of how users experience a stressful task in VR compared to in a classic office environment. Specifically, we investigate the subjective stress experience and physiological arousal with 15 participants performing the Stroop color word test either on a regular desktop screen, in VR, or in VR requiring head movements. Our findings suggest that stressful tasks are perceived less stressful when being performed in VR compared to the real environment as long as there is no additional stress factor, such as head movement involved. Our work indicates that it may be valuable to transfer stressful tasks, currently done in traditional office environments into VR.

Keywords: Virtual Reality · Stress · Human-Computer Interaction.

1 Introduction and Background

Virtual reality (VR) is becoming increasingly popular for researchers in different domains and for various purposes, e.g. rehabilitation [6, 18, 25], sports [24], and gaming [5]. Moreover VR has been proven to be an important tool for simulating stressful and cognitively demanding scenarios in e.g. military trainings [3, 5], pilot flight trainings [13, 14], other dangerous working environments, such as petrol refinery [7]. Aiming to train soldiers, flight pilots, and also firefighters stress elicitation in VR has pointed towards an important research gap. To not interrupt immersion in the VR scenario and elicit stress at the same time, there is a need for effectively stress inducing tasks that can be applied in VR. Aiming to elicit stress in VR, Legkov et al. [12] asked participants to react on approaching objects in a space shooter scenario. Measuring the Galvanic Skin Response (GSR) and the subjective stress level, they found that their dual task paradigm increased physiological arousal and affected certain stress dimensions. Jönsson et al. [10] validated if the Trier Social Stress Test (TSST) [11], a standardized

and validated task mainly used for laboratory studies, performed in VR also induces stress. They found that the cortisol level in the participants' saliva had increased by 88% and also the heart rate and heart rate variability showed a solid indication of experienced social stress. Based on these results, stress induction in VR has been studied [1, 4, 17]. However, there has been little work on the role of stress in VR. Particularly the investigation of the effects of stress elicitation in virtual environments compared to classic desktop based scenarios has been neglected so far. VR represents a promising approach to be applied as a training technique for learning to deal with heavy stressors. To address this research gap, we explore the transferability of stressful tasks from office to virtual environments using the Stroop color word test [21], a well established stress elicitation task [26], to observe stress perception in VR.

With this work, we contribute the first investigation of transferring it from its classic desktop screen version into VR. Through the stress assessment based on a three dimensional structure comprising engagement, worry, and distress, and the evaluation of heart rate variability, we provide first insights in the exploration of stressful tasks in VR and prospective effects of transferring these tasks from reality into virtual environments.

2   Implementation

Subsequently, we will explain what the Stroop color word test is and how we implemented it in our three experimental conditions.

2.1   Stroop Color Word Test

Among the classical version of the Stroop color word test [21] and several distinct versions that exist, we took the Stroop color word version for this work presenting both, the congruent and incongruent condition (cf. Figure 2). This task is commonly used in HCI [26] and intensifies physiological reactions [20], moreover it can be easily implemented in VR which is an advantage when considering the transferability of a stress inducing task. While 'congruent' means that words are displayed in the color that they signify, 'incongruent' refers to
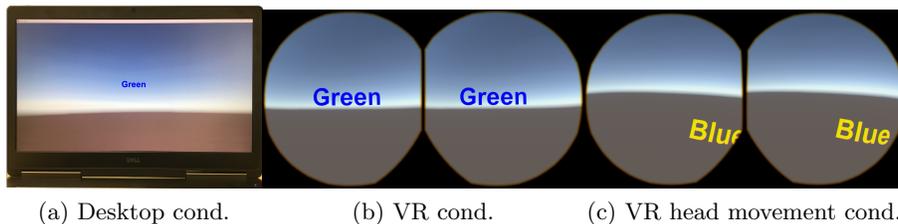

(a) Desktop cond.   (b) VR cond.   (c) VR head movement cond.

Fig. 1: Depicts the three conditions compromising our independent variable.

the incongruence between the color of the word and in its actual meaning. For example, in the congruent condition the word 'red' is being presented in red and in the incongruent conditions it is painted in blue (cf. Figure 1). Hereby the participant's task is to name the word's color and neglect the color that the word is designating.

2.2   Implementation of the Stroop Test in VR

For our user study setup, we utilized the HTC Vive virtual reality system connected to a PC consisting of a 3.7 GHz Intel Xeon processor, 16 GB RAM and a GeForce GTX 970 graphics card with a display size of 17,3 inch (Full-HD, 1.920x1080) (cf. Figure 1a). To render the stimuli of the Stroop color word test, we employed the Unity 3D engine[3]. For each condition, two sequences of 120 randomly selected Stroop items were generated at the beginning. Randomization was based on the subject ID, thus an individual sequence was generated for each participant. To get an equal distribution of Stroop items among all participants, two initial static buckets containing all items were created: one for the incongruent and one for the congruent test. The incongruent bucket contained each possible color combination ten times, the congruent bucket each stimulus 30 times. For each participant, the sequence of Stroop items that was presented, was randomly drawn from these buckets until they were empty. For the desktop screen and the VR condition, all stimuli were displayed in the centered field of view. For the VR head movement condition, the Stroop items were displayed at a pseudorandom position in the field of view of the participants in the virtual reality. This means that a random position was selected in either the left/right (50 degrees from the center field of view) or lower/upper (50.5 degrees from the center field of view) hemisphere of the subject with the constraint that no hemisphere could be selected two subsequent times.

---

[3] https://unity3d.com/

RANDOMIZED CONDITIONS:	120 Stimuli - 3 sec each intermitted by 1 sec					120 Stimuli - 3 sec each intermitted by 1 sec
Desktop	REST	PRAC	INCONGRUENT	REST	PRAC	CONGRUENT
VR						
VR - hm	5 min	1 min	8 min	5 min	1 min	8 min

Fig. 2: Study design showing the sequence of trials including the three conditions (desktop screen, VR, VR requiring head movements). We let the participants perform a practice trial before each of the two trials (incongruent, congruent) started, resulting in four trials in total preceded by a five minutes resting period.

3   User Study

For our user study we invited 15 participants and randomly split them into three equally sized groups, resulting in an in-between-subject design. Whereas one group experienced the Stroop color test in VR and another group performed head movements in VR, the third group conducted the test on a regular desktop screen positioned on an office desk. We randomized the sequence of three conditions for each participant according to Latin square.

3.1   Measures

As independent variables, we designed three different conditions how the Stroop color word test should be performed: (a) sitting in front of a desktop screen and all stimuli presented in the centered field of view, (b) in VR and the stimuli also presented in the centered field of view, and (c) in VR but the stimuli appeared as described above in the visual field of the participant. The latter condition was introduced based on the observations that the vestibular system being involved in head movements, was found to have an influence on the susceptibility to motion sickness [19]. Thus, we were interested if requiring head movements to accomplish the task would increase stress perception. As dependent variables we assessed the subjectively perceived stress level employing the Short Stress State Questionnaire (SSSQ) [8]; also used by Legkov et al. [12] to observe stress reactions induced in VR. Likewise, we recorded physiological data, i.e. heart rate (HR) and heart rate variability (HRV) to monitor if the physiological arousal corresponding to stress also varies among our conditions as could be shown in related work for VR experiences in general [2, 18]. For recording physiological data we used the Nexus Kit 4 by MindMedia[4].

3.2   Procedure and Data Collection

Before we started with the evaluation, participants were explained the study's purpose and procedure. After giving their written consent, we asked each one to place the three gel electrodes connected to a two channel ExG sensor for assessing HR and HRV to themselves; meaning the negative (black) electrode was attached at the right collar bone, the positive (red) electrode on the midaxial line on the lateral aspect of the chest, and the ground electrode near the right leg on the chest. Inspired by previous study designs [22], we specified the sequence of trials as depicted in Figure 2. Before the initial resting phase of five minutes, baseline measurements started, each participant was asked to fill in the first part ('At the moment') of the Short Stress State Questionnaire [8] for assessing the baseline stress level. Then, there was a one minute practice phase; by this we ensured that everyone understood the task. It was followed by an eight minute incongruent phase. During that time the participants were presented 120 words, either on the desktop screen or in VR, for 3 seconds intermitted by one second. After another

---

[4] https://www.mindmedia.com/de/produkte/nexus-4/

five minutes of resting and a one minute practice trial, the congruent phase lasted eight minutes followed, presented in the same manner as the incongruent condition. When the last Stroop color test trial was completed, all participants were asked to complete the second part of the Short Stress State Questionnaire referring to their stress perception 'During the task'.

Fig. 3: Participant wearing the HTC Vive VR glasses performing the Stroop test in VR while measuring heart rate and heart rate variability.

3.3   Participants

Our sample consisted of 15 participants ($M = 23.5$, $SD = 2.5$ years), among these were seven females – in each group at least two females. The majority (ten people) were VR novices, while three stated to have some VR experience and two others said they had been using VR "a lot". Recruiting our participants via university mailing lists and personal acquisition, we had eleven students, one PhD student, one teacher and two engineers among our sample. The experimental procedure had been approved by the ethics committee of our institution.

4   Results

For the physiological data analysis, we removed the first and last 30 seconds of the baseline and the experimental (incongruent, congruent) sessions to avoid primacy effects. Prior to the statistical calculations, we normalized the data according to each participant's baseline values. We focused on the physiological measures HR and HRV using the standard formula provided by the manufacturers processing software [5] for HRV value calculation.

---

[5] https://www.mindmedia.com/en/products/biotrace-software/

	HR		HRV	
	incongr.	con.	incongr.	con.
Desktop	1.00	1.00	1.12	1.26
	(0.02)	(0.06)	(0.37)	(0.66)
VR	1.04	1.01	0.99	1.25
	(0.23)	(0.22)	(0.60)	(0.45)
VR-hm	1.08	1.04	0.93	1.05
	(0.07)	(0.06)	(0.36)	(0.56)

Table 1: Table shows means and standard deviations of the incongruent and congruent trials for heart rate and heart rate variability according to the three different conditions. While high values indicate increased arousal for HR, lower values are associated with high arousal for HRV [23].

Physiological Measures During the incongruent trial, participants showed slightly higher HR values in the VR condition and also for the head movement condition, compared to performing the Stroop test in front of a desktop screen (cf. Table 1). For the congruent trial, the results were similar to those for the incongruent one, but overall the values were little lower as can be obtained from Table 1. Regarding the HRV, we observed that during the incongruent condition in VR, participants had lower HRV values in both conditions, namely VR and VR requiring head movements, indicating physiological arousal (cf. Table 1). In front of the desktop screen, the HRV was higher with an average of 1.12. Again, these results are similar to the HRV values recorded during the congruent trial showing that this trial resulted in lower physiological arousal (cf., Table 1).

Performance As performance measures, we recorded the error rate. Most errors in the incongruent condition were made in VR requiring head movements ($M = 1.4$, $SD = 2.2$) followed by the Stroop at a desktop screen ($M = 1.0$, $SD = 1.0$) and in VR-only the average errors were 0.6 ($SD = 1.3$). In the congruent condition there occurred only one error in VR requiring head movements.

Subjective Measures For the SSSQ results, we calculated the difference between the pre-test and post-test SSSQ scores ((post-score − pre-score) / pre-score standard deviation) [15] resulting in a so-called change score. Hereby positive scores signify a higher stress rating after the task was accomplished and negative change scores mean that stress was higher before the task. Since the SSSQ has an underlying three factorial structure divided into Disstress, Worry, and Engagement [9] we present the results according to these factors in Figure 4. The desktop screen condition induced the most distress with an average change score of 12.45 ($SD = 12.90$), while for the Stroop test performed in VR, the participants felt almost not stressed at all ($M = -0.57$, $SD = 1.27$). When head movements in VR were required, the distress increased to an average score of 3.50 ($SD = 9.37$). Referring to the sub-dimension worry, there was no change when the Stroop test was performed in front of a desktop screen ($M = 0.02$,
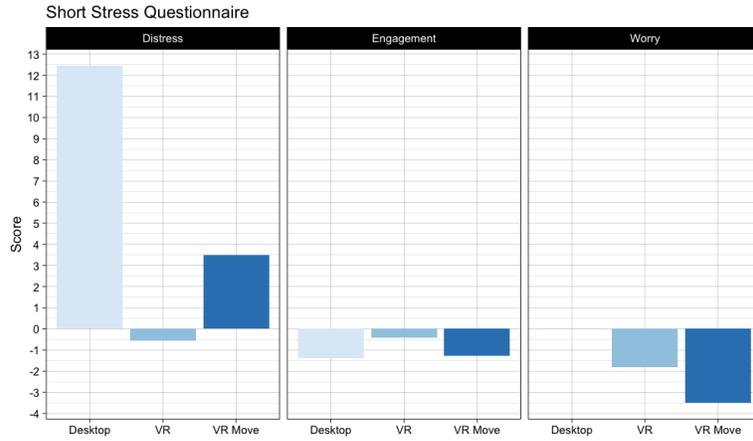
Fig. 4: Results from the SSSQ [8] according to its three dimensions *Engagement*, *Disstress*, and *Worry* showing that the Stroop test in VR does not have been perceived as stressful in VR as in the desktop screen condition.

$SD = 7.39$). And for both VR conditions we observed that worry was even de-creasing compared to the baseline measurement, namely to an average of -1.81 in VR ($SD = 1.93$) and to -3.48 ($SD = 3.78$) when head movements were required in VR. For engagement there was no difference in the VR condition (M =-0.42, $SD = 3.65$) and for the desktop screen condition (M =-1.40, $SD = 4.51$) as well as the head movement requiring one (M =-1.28, $SD = 2.48$), there were only minor changes signifying that engagement was lower after the Stroop test.

Inferential Statistical Analysis Since our data was not normally distributed, which is required for parametric tests, we used non-parametric tests. Thus, we performed a Kruskal-Wallis test aiming to reveal differences among our three conditions. No significant results could be found here. We further investigated correlations between the different variables. For this, we used the Spearman rank coefficient which is also robust against outliers in our data. We found a strongly positive correlation between the stress assessing SSSQ overall score and it's two underlying dimensions distress ($r = .862, p = .000$) and worry ($r = .601, p = .018$).

5   Discussion

Our results show that the participants subjectively perceived the task on a desk-top screen as the most stressful ($M = 12.45$), whereas the same task in VR has been rated almost not stressful at all with an average of -0.57 signifying that there has been almost no difference between the stress level before and after the task. The participants rated the VR condition requiring head movements

more stressful ($M = 3.50$), which suggests that the involvement of motor skills acts as an additional stress factor. Whereas the standard deviation of distress perception in VR had been low with an average of 1.27, it was exceeded by the two other conditions. Particularly in the desktop screen variant, there were two participants for whom there was almost no change between the distress level before and after the Stroop test, while three other participants felt enormously stressed having a change score of 14.14 and respectively 19.80. These differences underline the subjective perception of stress that is challenging [26]. Moreover, our results are supported by prior work [12] where the SSSQ was applied and distress increased with the presentation of a stressful task, while worry decreased. In contrast, during the two most stressful conditions, desktop screen and VR requiring head movements, engagement was low after performing the Stroop test ($M = -1.40$, $M = -1.28$) indicating that stress dominated then. This is further strengthened by the correlational analysis revealing that engagement was the only dimension of the SSSQ that didn't correlate significantly with its total score and thus insufficiently reflected the participant's stress perception. Again, in VR there was almost no difference before and after the task ($M = -0.42$). For worry we found almost no difference when being performed on a desktop screen ($M = 0.02$). In the VR conditions it even decreased after finishing the task ($M = -1.81$, $M = -3.48$) what can be explained with a feeling of relief after having accomplished the test. Regarding the physiological data we recorded, the results show that our participants had lower arousal values in HR ($M = 1.00$) and HRV ($M = 1.12$) during the desktop screen task. While arousal had been mild but slightly higher in the VR condition for HR ($M = 1.04$), and HRV ($M = 0.99$), there was a greater rise in HR ($M = 1.08$) and respectively a decrease in HRV ($M = 0.93$). These findings show that the participants experienced higher physiological arousal in both VR conditions, what is supported by the results for the subjective measures. However, performing motor skills seems to increase only the subjective stress perception but does not affect physiological arousal. Thus, the Stroop color word test seems not suitable for inducing stress in participants when it is transfered into VR. To successfully evoke subjectively perceived stress, there is the requirement of moving the head as an additional factor. This is in line with the findings from research on the reason for why head-mounted displays (HMDs) used for VR are causing visual stress. Mon-Williams et al. [16] stated the that vertical gaze angle is a crucial factor and that therefore the HMD needs to be placed in the correct vertical position for each user individually. Consequently, the presentation of the stimuli in the virtual space shifted on both, the x- and y-axis, could have provoked a level of stress in the user that is perceived only subjectively.

Although we believe, that this piece of work yields important insights in the perception and the transfer of stress in VR, we have to acknowledge that due to our limited number of participants, future work should repeat this experiment involving a greater sample so that the observed tendencies can be verified statistically. Nevertheless, our results show that inducing stress in VR cannot

be adopted on a one to one basis for VR and thus could benefit from further investigations, particularly focusing on the design of stressful tasks for VR.

6   Conclusion and Future Work

In this paper, we explored whether a stressful task can be transferred into VR. The results show that participants felt higher distress and lower engagement when the test was performed in the office environment compared to the VR condition. Likewise, the involvement of motor skills in the virtual environment led also to higher distress and lower engagement, what could only be observed in the subjective data. Hence, our findings suggest that the Stroop color word test is not suitable for inducing stress when being performed in VR and when being adopted one to one. To successfully evoke subjectively felt stress, e.g. as part of an VR flight simulation scenario to practice reactions under pressure, an additional requirement is needed, e.g. to perform motor skills. Consequently future work should focus on the exploration and determination of suitable motor skill tasks in VR to elicit stress. Through the initial exploration of the transferability of stressful tasks into VR, we believe to provide a valuable starting point for further investigations in the underlying mechanics, to ultimately design effective training scenarios for VR.

Acknowledgements

References

1. Annerstedt, M., Jö, P., Wallergård, M., Johansson, G., Karlson, B., Grahn, P., Hansen, Å.M., Währborg, P.: Inducing physiological stress recovery with sounds of nature in a virtual reality forest - results from a pilot study. Physiol Behav **118**, 240 – 250 (2013)
2. Calvert, S.L., Tan, S.L.: Impact of virtual reality on young adults' physiological arousal and aggressive thoughts: Interaction versus observation. J Appl Dev Psychol **15**(1), 125 – 139 (1994)
3. Council, N.R.: Modeling and simulation: Linking entertainment and defence (1997)
4. Diemer, J., Mühlberger, A., Pauli, P., Zwanzger, P.: Virtual reality exposure in anxiety disorders: Impact on psychophysiological reactivity. World J Biol Psychiatry **15**(6), 427–442 (2014)
5. Fong, G.: Adapting cots games for military simulation. In: Proc 2004 Int Conf Virtual Reality Continuum Applications Industry. pp. 269–272. VRCAI '04, ACM, New York, NY, USA (2004)
6. Gradl, S., Wirth, M., Zillig, T., Eskofier, B.M.: Visualization of heart activity in virtual reality: A biofeedback application using wearable sensors. In: Proc 15th Int Conf Wearable Implantable Body Sensor Networks. pp. 152–155 (March 2018)
7. Haller, M., Kurka, G., Volkert, J., Wagner, R.: omvra safety training system for a virtual refinery. In: Proc ISMCR 99 Topical Workshop Virtual Reality Advanced Human Robot Systems. pp. 291–298. ISMCR '99 (1999)
8. Helton, W.S.: Validation of a short stress state questionnaire. Proc Hum Factors Ergonomics Society Annual Meeting **48**(11), 1238–1242 (2004)

9. Helton, W.S., Näswall, K.: Short stress state questionnaire. Eur J Psychol Assess **31**(1), 20–30 (2015)
10. Jönsson, P., Wallergård, M., Österberg, K., Hansen, r.M.e.a.: Cardiovascular and cortisol reactivity and habituation to a virtual reality version of the trier social stress test: A pilot study. Psychoneuroendocrino **35**(9), 1397 – 1403 (2010)
11. Kirschbaum, C., Pirke, K.M., Hellhammer, D.H.: The 'trier social stress test' - a tool for investigating psychobiological stress responses in a laboratory setting. Neuropsychobiology **28** (1993)
12. Legkov, P., Izdebski, K., Kärcher, S., König, P.: Dual task based cognitive stress induction and its influence on path integration. In: Proc 23rd Symp Virtual Reality Software Technology. pp. 41:1–41:5. ACM, New York, NY, USA (2017)
13. Loftin, R.B., Kenney, P.: Training the hubble space telescope flight team (1995)
14. Markov-Vetter, D., Moll, E., Staadt, O.: Evaluation of 3d selection tasks in parabolic flight conditions: Pointing task in augmented reality user interfaces. In: Proc 11th Int Conf Virtual Reality Continuum Applications Industry. pp. 287–294. VRCAI '12, ACM, New York, NY, USA (2012)
15. Matthews, G., Joyner, L., Gilliland, K., Campbell, S., Falconer, S., Huggins, J.: Validation of a comprehensive stress state questionnaire: Towards a state 'big three.'. In: Personality Psychology Europe. pp. 335–350. Tilburg University Press (1999)
16. Mon-Williams, M., et al., A.P.: Gaze angle: a possible mechanism of visual stress in virtual reality headsets. Ergonomics **41**(3), 280–285 (1998)
17. Montero-López, E., Santos-Ruiz, A., García-Ríos, M.C., Rodríguez-Blázquez, R., Pérez-García, M., Peralta-Ramírez, M.I.: A virtual reality approach to the trier social stress test: Contrasting two distinct protocols. Behavior Research Methods **48**(1), 223–232 (Mar 2016)
18. Moore, K., Wiederhold, B.K., Wiederhold, M.D., Riva, G.: Panic and agoraphobia in a virtual world. CyberPsychology Behavior **5**(3), 197–202 (2002)
19. Paillard, A., Quarck, G., Paolino, F., Denise, P., Paolino, M., Golding, J.F., Ghulyan-Bedikian, V.: Motion sickness susceptibility in healthy subjects and vestibular patients: effects of gender, age and trait-anxiety. J Vestibular Research **23**(4, 5), 203–209 (2013)
20. Renaud, P., Blondin, J.P.: The stress of stroop performance: physiological and emotional responses to color-word interference, task pacing, and pacing speed. Int J Psychophysiol **27**(2), 87 – 97 (1997)
21. Stroop, J.R.: Studies of interference in serial verbal reactions. J Exp Psychol **18**(6), 643–662 (1935)
22. Svetlak, M., Bob, P., Cernik, M., Kukleta, M.: Electrodermal complexity during the stroop colour word test. Auton Neurosci: Basic Clinical **152**, 101–107 (2010)
23. Tarbell, S.E., Millar, A., Laudenslager, M., Palmer, C., Fortunato, J.E.: Anxiety and physiological responses to the trier social stress test for children in adolescents with cyclic vomiting syndrome. Auton Neurosci: Basic Clinical **202**, 79 – 85 (2017)
24. Wirth, M., Gradl, S., Poimann, D., Schaefke, H., Matlok, J., Koerger, H., Eskofier, B.M.: Assessment of perceptual-cognitive abilities among athletes in virtual environments: Exploring interaction concepts for soccer players. In: Proc 2018 Designing Interactive Sys Conf. pp. 1013–1023. ACM, New York, NY, USA (2018)
25. Yeh, S.C., Rizzo, A., Zhu, W., Stewart, J., McLaughlin, M., Cohen, I., Jung, Y., Peng, W.: An integrated system: Virtual reality, haptics and modern sensing technique (vhs) for post-stroke rehabilitation. In: Proc ACM Symposium Virtual Reality Software Technology. pp. 59–62. VRST '05, ACM, New York, NY, USA (2005)
26. Zhai, J., Barreto, A.: Stress recognition using non-invasive technology. Proc Int Florida Artificial Intelligence Research Society Conf. pp. 395–400 (2006)