Identifying Unique Physiological Indicators of Virtual Reality Sickness

Olivia Fox Cotton, Kevin Durkee, Justin Morgan, Sarah Meyer, Sheila Galbreath
Aptima, Inc.
Fairborn, OH
{ofox, kdurkee, jmorgan, smeyer, sgalbreath}@aptima.com

Leanne Hirshfield, G.S. Rajshekar Reddy, Cara A. Spencer, Gavin Zimmerman
University of Colorado Boulder
Boulder, CO
{Leanne.Hirshfield, rajshekar.gudasubhash, cara.spencer, gavin.zimmerman}@colorado.edu

Gabriella Severe-Valsaint, Ada Mishler, Michael Natali
Naval Air Warfare Center, Training Systems
Orlando, FL
{gabriella.severe-valsaint.civ, ada.d.mishler.civ, michael.w.natali.mil}@us.navy.mil

Brennan D. Cox
Naval Medical Research Unit Dayton
Wright-Patterson AFB, OH
Cox.brennan@gmail.com

ABSTRACT

Navy pilots rely heavily on simulator technologies to train and develop critical skills. It is important to evaluate the impact of new training technologies as they emerge. While virtual reality (VR) is a promising tool for immersive training, VR sickness is common, leading to a range of disorientation, oculomotor, and nausea symptoms (Kennedy et al., 1993). The goal of this paper is to assess the feasibility of deriving physiological signatures of VR sickness. Identification of physiological signatures specific to VR exposure provides the foundation for both future predictive models of VR sickness and potential augmentation strategies to prevent or ameliorate the symptoms of VR sickness, increasing the utility of VR as an immersive training tool. While prior research has demonstrated physiological changes during VR exposure (Martin et al., 2020), we also consider the impacts of workload variations and individual differences. Our study utilizes a setup that resembles a Navy flight training environment in which participants perform various flight maneuvers that may induce VR sickness (e.g., barrel rolls and 180° turns). We ask whether this VR training induces detectable changes in human physiology, and if so, what relationships exist between physiological measures and the previously validated Virtual Reality Sickness Questionnaire (VRSQ). Further, the Motion Sickness Susceptibility Questionnaire (MSSQ) was administered to assess individual differences in susceptibility as a possible influence on VR sickness and the NASA Task Load Index (NASA-TLX) was administered to control for any changes in workload across trials that may also induce physiological changes. We hypothesized that heart rate, respiration, and blink rate would show strong associations with self-report measures of VR sickness but our findings did not support this hypothesis.

ABOUT THE AUTHORS

Olivia Fox Cotton is a Scientist at Aptima, Inc. with a background in the fields of human factors psychology, cognitive neuroscience, and cognitive systems engineering. She holds an MS in human factors and industrial organizational psychology from Wright State University, and a BA in psychology from Clemson University. She is currently pursuing her doctorate at Wright State University where she performs research in cognitive neuroscience.

Kevin Durkee, Principal Scientist and Lead Scientist at Aptima, Inc., has a technical background in applied human factors issues that directly impact worker performance and safety, such as cognitive workload, vigilance, situational awareness, and decision making. Mr. Durkee received an MA in human factors and applied cognitive from George Mason University, and a BA in psychology from DePaul University.

Dr. Leanne Hirshfield is an Associate Research Professor in the Institute of Cognitive Science at the University of Colorado (CU) Boulder. Hirshfield directs the research in the System-Human Interaction with NIRS & EEG (SHINE)

Lab at CU. She is also a co-lead of the research for the NSF Institute of Student-AI Teaming. Dr. Hirshfield's specialties are in the areas of human-computer interaction, machine learning, and cognitive science using non-invasive and physiological measurements.

Justin Morgan is a Research Intern at Aptima, Inc., and a human factors graduate student at Wright State University. He is currently studying automation, mental workload, and situational awareness (in the context of aviation) with a background in cognitive neuroscience.

Sarah Meyer is a Research Intern at Aptima, Inc. She attended Bowling Green State University where she earned a BS in psychology. She has served as a research assistant for multiple research teams in cognitive and clinical psychology. She is passionate about applying findings from the psychological sciences to help people become happier, healthier and more efficient, and hopes to continue her education in psychology by pursuing a graduate degree.

Sheila Galbreath is a Research Intern at Aptima, Inc. She has expertise in both medicine and engineering, and has earned an MA in biomedical engineering from Wright State University. Her research interests include medical technology development and physiological effects in training.

LCDR Brennan D Cox, Ph.D. is Deputy Director of the Naval Aerospace Medical Research Laboratory at Naval Medical Research Unit – Dayton, OH, where he manages the activities of 75 scientists across a broad portfolio addressing challenges within the aerospace operational environment concerning human health, safety, readiness, and performance. He currently serves as Technical Representative for the NATO Human Factor and Medicine Science and Technology Board on the Assessment of Factors Impacting Cybersickness, responsible for establishing a common experimentation framework for investigating cybersickness and developing mitigation strategies.

Gabriella Severe-Valsaint is a Senior Research Psychologist in the Human Performance Science and Technology Branch at the Naval Air Warfare Center Training Systems Division (NAWCTSD), in Orlando, Florida. Mrs. Severe-Valsaint received a MS in Industrial & Organizational Psychology and her BS in Psychology from the University of Central Florida. Mrs. Severe-Valsaint has 13 years of government service executing basic research programs, performing training analyses, and facilitating training effectiveness evaluations across various military platforms. Currently, Mrs. Severe-Valsaint's work focuses on assessing immerging technology training capabilities, specifically, extend reality's impact on pilot learning as well as evaluating curriculum modifications for aviation training.

Dr. Ada Mishler is a Research Psychologist at NAWCSTD with experience in the design, development, and evaluation of virtual training programs as well as basic training research. She currently specializes in the evaluation of virtual flight trainers and educational programs that incorporate virtual flight trainers. She holds a PhD in Human Factors and Cognitive Psychology from the University of Central Florida and a BS in Psychology from Colorado State University.

LCDR Michael Natali, Ph.D. serves as the Deputy, Air Warfare Training Development Integrated Project Team Lead for PMA-205, Naval Air Training Systems and Ranges, and is stationed at the Naval Air Warfare Center Training Systems Division (NAWCTSD) in Orlando, FL. At NAWCTSD, LCDR Natali manages acquisition, research, development, training, and evaluation for programs enhancing aviator performance in training and in the fleet, identifying emerging technologies to integrate into aircraft systems and processes ensuring the continued dominance and lethality of Naval aviation.

G S Rajshekar Reddy is a Masters student in the Computer Science department at the University of Colorado Boulder and a Graduate Research Assistant at SHINE Lab. He previously worked as a Research Assistant at the Indian Institute of Science. His research focuses on immersive technologies such as Virtual and Augmented Reality and determining effortless interaction mediums for these technologies that lessen cognitive processing and improve decision-making abilities.

Cara Spencer received two BS degrees from UC San Diego in Cognitive Science and Human Developmental Sciences. Afterwards, she accepted a Research Associate position at NASA's Behavioral Health and Performance Laboratory through the contracting company JES Tech. There she gained data science skills and specialized in various wearable and operational performance measures. She is currently a Computer Science PhD student at CU Boulder under Leanne Hirshfield where she is continuing her work with unobtrusive human performance measurement.

Gavin Zimmerman is currently pursuing a BS in Computer Science at University of Colorado, where he works in Leanne Hirshfield's lab.

Identifying Unique Physiological Indicators of Virtual Reality Sickness

Olivia Fox Cotton, Kevin Durkee, Justin Morgan, Sarah Meyer, Sheila Galbreath
Aptima, Inc.
Fairborn, OH
{ofox, kdurkee, jmorgan, smeyer, sgalbreath}@aptima.com

Leanne Hirshfield, G.S. Rajshekar Reddy, Cara A. Spencer, Gavin Zimmerman
University of Colorado Boulder
Boulder, CO
{Leanne.Hirshfield, rajshekar.gudasubhash, cara.spencer, gavin.zimmerman}@colorado.edu

Gabriella Severe-Valsaint, Ada Mishler, Michael Natali
Naval Air Warfare Center, Training Systems
Orlando, FL
{gabriella.severe-valsaint.civ, ada.d.mishler.civ, michael.w.natali.mil}@us.navy.mil

Brennan D. Cox
Naval Medical Research Unit Dayton
Wright-Patterson AFB, OH
Cox.brennan@gmail.com

BACKGROUND

The US military has been long considered on the bleeding edge of technological advancement that allows for the maintenance of land, sea, and air superiority. Naval aviation's use of virtual reality (VR) training environments is one example of an emerging technology with the potential to provide a tactical advantage. VR training can provide the ability to mimic a variety of different conditions, can be safer than real life training exercises, and can be more cost effective (Bos et al., 2021). Bos and colleagues (2021) also report that VR training has the potential to increase training effectiveness by allowing for targeted and optimized training that repeats only the tasks within a larger training scenario that require extra practice. Benefits to trainee performance are seen in a variety of domains when VR training is used, including surgical training (Hart & Karthigasu, 2007) and training for novice drivers (Agrawal et al., 2017). Looking specifically at aviation training, multiple studies have provided empirical support for the claim that the use of VR training environments improves student pilot performance beyond that of legacy training syllabi (e.g., McCoy-Fisher et al., 2019; Severe-Valsaint et al., 2022). Perceptions of VR training and its efficacy also tend to be positive with Instructor Pilots (IPs) reporting an expectation that VR-trained student naval aviators (SNAs) would be more capable of successfully responding to unexpected events (Mishler et al., 2022).

While the benefits of VR training are well documented, so is the prevalence of VR sickness. VR sickness refers to the occurrence of a set of symptoms that are similar to motion sickness but result from exposure to VR technology that produces patterns of sensory cues that deviate from normal perceptual experiences (McCauley & Sharkey, 1992). While estimates of the rates of VR sickness vary from study to study depending upon the age of the VR technology and the specific VR experience, they range from about 33% of participants reporting VR sickness at the low end (Stone, 2017) and 100% of participants reporting VR sickness at the upper end (Davis et al., 2015). The symptoms reported following VR technology exposure typically cluster into the categories of nausea, oculomotor, and disorientation symptoms (Bos & Lawson, 2021), although some researchers have also found support for a two-factor model including just nausea and oculomotor factors (Bouchard et al., 2007). Oculomotor symptoms typically include eye fatigue, headache, eyestrain, difficulty focusing, difficulty concentrating, fullness of head, and blurred vision, and are often suspected to be caused by accommodation-convergence conflicts that arise in the visual system while using VR technology (Bos & Lawson, 2021). In natural vision, the accommodation of the lenses of the eyes for focal distance and vergence eye movements for directing gaze location are coupled. Accommodation-convergence conflicts result when this link is broken, and the eyes must converge to focus on nearby virtual objects and diverge for distant virtual objects while the lens accommodation remains constant. In contrast, disorientation symptoms are likely experienced in response to visual-vestibular conflicts that arise when motion cues are visually perceived without the accompanying vestibular sensations (Bos & Lawson, 2021).

This high prevalence of VR sickness is problematic and constitutes a barrier to the widespread adoption of new VR technologies that could aid military training. Impacts to trainee health are unpredictable given individual differences in severity of symptoms (Howard & Van Zandt, 2021). Additionally, differences in the timeline for the onset of symptoms vary from person to person as well as between different exposure experiences (Reason & Graybiel, 1970). In addition to health concerns, Reed and colleagues (2007) suggest that VR sickness can also lead to trainee dropout at rates in excess of 50%, constituting a massive potential loss for the military training community. For those trainees who do not dropout, VR sickness can compromise training performance and reduce the efficiency of training (Reed et al., 2007).

In order to learn more about VR sickness, researchers have employed many self-report measures (e.g., Motion Sickness Susceptibility Questionnaire, Golding,1998; Virtual Reality Sickness Questionnaire, Kim et al., 2018, etc.), but these are susceptible to the same issues that plague self-report measures more generally and limit their utility (e.g., inconsistent reporting, response bias, lack of real-time assessments, etc.; De Winter, 2014). Capturing physiological data during VR training is one approach to addressing both the limits of self-report measures and the challenges of individual differences. Many physiological signals have been studied in association with VR sickness. For example, heart rate typically increases along with VR sickness and the perception of nausea (e.g., Nalivaiko et al., 2015; LaCount et al., 2011). Researchers have shown that respiration rate tends to decrease as VR exposure time increases (e.g., Gavgani et al., 2017), and a larger change in respiration rate relative to baseline is correlated with higher scores on self-report VR sickness measures (Kim et al., 2005). The conductance levels measured via galvanic skin response tends to increase with exposure time (e.g., Gavgani et al., 2017). Many eye tracking measures are also informative of VR sickness. Blink rate increases with immersion duration and is correlated with the oculomotor discomfort subscale of the VRSQ (e.g., Dennison et al., 2016). Other technologies including electroencephalography (EEG) have also shown reliable relationships with VR sickness (Kim et al., 2005), but as they are more difficult to field operationally, such technologies were not a focus of the present effort.

Consideration of physiological signals is an important step towards a deeper understanding of the mechanisms of VR sickness, but many studies still fail to consider the role of cognitive workload. The phenomenon of physiological responses induced by changes to cognitive workload is well established in the literature (e.g., Ranchet et al., 2017). Therefore, controlling for levels of cognitive workload in VR sickness studies is important for ensuring a clear understanding of the physiological data. By studying the physiological response to VR sickness and controlling for workload, we can provide the information necessary to begin designing novel techniques that help reduce the severity and/or occurrence of VR sickness and support the development of augmentation strategies. In service of this goal, we created a VR flight training scenario intended to induce detectable changes in human physiology, and explored what relationships exist between physiological measures and self-report measures. It was hypothesized that heart rate, blink rate, and respiration in particular will show strong correlations with the self-report measure of VR sickness.

METHOD

Participants

Forty nine participants (85.71% male, 14.29% female) with a mean age of 32.69 (*SD*=12.20) were recruited from the Wright Patterson Air Force Base (WPAFB) community and included active duty and retired military members, civilians, and college students. All participants were at least 18 years old, native English speakers, and had experience playing video games. 23% of the participants had prior aircraft flight experience and 75.5% had prior VR experience.

Study Design

Figure 1. A screen capture of the flight scenario showing the green virtual gate.

This study was designed to induce changes in VR sickness and cognitive workload through a series of VR flight simulation tasks that vary in complexity. The VR flight scenario was built and customized in Lockheed Martin's Prepar3D flight simulator and Steam VR was used to incorporate the VR headset. In the scenario, participants were tasked with flying a Boeing T-45C jet through a series of virtual gates (Figure 1) and, depending on the trial, execute particular flight maneuvers as seen in Table 1. The trials were designed in consult with flight experts and contained flight maneuvers that were suspected to increase the likelihood of VR sickness. Trial 1 was suspected to be the least visually complex requiring the fewest head movements, and presumably least likely to induce VR sickness. As trials progressed, the suspected visual complexity, anticipated head and eye movements, and assumed likelihood of VR sickness increased. Since each successive trial was intended to increase VR sickness beyond that of the prior trial, trial order remained consistent for all participants. Each trial included six gates to which participants needed to navigate and the scenario maintained constant weather conditions characterized by sunny, clear skies and no winds throughout all trials. The duration of each trial depended upon how quickly the participant was able to navigate to each sequential gate; average trial time was 5 minutes 52 seconds (*SD*=33 seconds). At the conclusion of each trial, the participants completed two questionnaires to assess VR sickness and cognitive workload. These questionnaires were administered within the VR environment and participants responded using predetermined buttons on the flight controls. While performing these tasks, participants were outfitted with several physiological sensors, described below.

Table 1. Description of Flight Tasks

Trial No.	Trial Name	Description
1	Straight Flight	Flying in a straight-line path and passing through gates with little to no change in altitude
2	Minor Turns	Performing minor turns that range from 30 to 90 degrees between gates with slight altitude changes
3	Major Turns	Performing 180-degree turns between gates that require large head movements to track gates. Virtual arrows help guide participants
4	Loops	Performing 5 vertical inside loop maneuvers when prompted by audio tone and making 180-degree turns between gates
5	Barrel Rolls	Performing 5 barrel roll maneuvers when prompted by audio tone and making 180-degree turns between gates
6	In/Out Focusing	Reading and verbally reporting current airspeed, altitude, and heading off in-cockpit instruments when prompted by audio tone and performing 180-degree turns

Measures and Equipment

Self-reported Measures. To assess participants' cognitive workload, the mental demand subscale from the NASA-TLX (Hart & Staveland, 1988) was administered following each trial. The Virtual Reality Sickness Questionnaire (VRSQ), which asks participants to rate nine symptoms in severity on a four-point scale (Kim et al., 2018), was also completed following each trial. Symptoms include general discomfort, fatigue, eyestrain, difficulty focusing, headache, fullness of the head, blurred vision, dizziness, and vertigo. Participants completed the Motion Sickness Susceptibility Questionnaire (MSSQ), an assessment that asks participants to rate using a four-point scale how nauseated they felt in the past during exemplar scenarios (Golding, 1998). Finally, participants completed the Motion Sickness Assessment Questionnaire (MSAQ) that results in an overall measure of motion sickness experienced across all trials (Gianaros et al., 2001).

Equipment. The HP Reverb G2 Omnicept Edition headset was used for displaying the VR environment to participants and the physical setup is seen in Figure 2. This headset includes onboard physiological sensors including eye tracking and photoplethysmography (PPG). Participants also wore a Polar H10 heart rate monitor and a Zephyr Bioharness respiration monitor. The T-45C aircraft was controlled using Thrustmaster Warthog HOTAS and rudder pedals.

Figure 2. A simulator seat, flight controls, and HP Reverb G2 Omnicept headset were used.

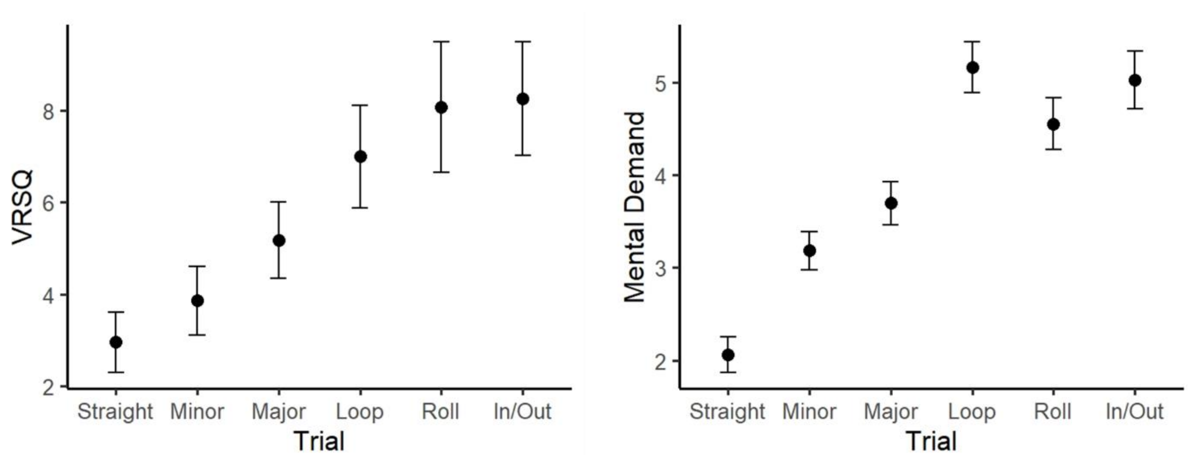
Experimental Procedures

This study was approved by the University of Colorado Institutional Review Board (IRB). Participants provided informed consent to participate and completed a demographics questionnaire and the MSSQ. Next, participants were outfitted with physiological sensors and viewed an instructional video that previewed the flight scenario they would be performing. The video explained the flight controls, the six types of flight tasks and how to perform each maneuver, how the virtual gates worked, and the VR-based questionnaires that would be completed after each trial (VRSQ and NASA-TLX). Following the video, participants completed a baseline data collection lasting five minutes during which the participants were instructed to sit still and were not given an active task to perform. Following this baseline was a simple flight tutorial that allowed them to orient to the simulator and practice with the flight controls. This tutorial was similar to the Straight Flight trial, but participants were allowed to practice any of the other flight maneuvers between gates as they saw fit. Next, the flight scenario with all six trials began. Following each trial, participants completed the VRSQ and the NASA-TLX before the next trial. After completing all trials and removing the equipment, participants completed the MSAQ before being debriefed.

RESULTS

Manipulation Check

Before proceeding with planned analyses, we first performed a manipulation check to ensure that participants were experiencing VR sickness of any magnitude in any trial using the MSAQ self-report data. A one sample *t*-test was performed to compare the reported level of VR sickness across all trials (MSAQ total score) against the population mean. The mean value of the MSAQ (M=18.94, SD=8.06) was significantly higher than the population mean ($t(48)$=6.65, p<0.00), confirming that our flight task manipulation did indeed induce VR sickness in our sample.



(a) (b)

Figure 3. (a) Average VRSQ scores, and (b) Average NASA-TLX scores. Error bars denote standard error.

VR Sickness and Cognitive Workload

We performed a one-way repeated measures ANOVA to test the effect of flight trial on VR sickness and cognitive workload. The results of Mauchly's test indicated that the assumption of sphericity was violated for both VRSQ scores (χ^2 (14) = 110.11, p < .01, ε = .47) and NASA-TLX scores (χ^2 (14) = 86.61, p < .01, ε = .62), so a Greenhouse-Geisser correction was used. After this correction, the results show that VRSQ scores (F(2.34, 112.47) = 13.95, p < .001) and NASA-TLX scores (F(3.11, 149.14) = 65.74, p < .001) were significantly affected by the trials. Average scores for the VRSQ and NASA-TLX across trials are shown in Figure 3.

Physiological Indicators

Prior to performing any statistical analyses, the physiological data were average across trials and each physiological measure was baseline adjusted in order to remove the effects of individual differences. These baseline-adjusted averages were then used in a one-way repeated measures ANOVA, and averages of each measure can be seen across trials in Figure 4. We found that flight trial significantly affected heart rate (F(2.71, 130.17) = 65.74, p < .001), blink rate (F(3.72, 178.36) = 65.74, p < .001), and respiration (F(3.45, 165.35) = 65.74, p < .001), but the assumption of sphericity was again violated (χ^2 (14) = 97.06, χ^2 (14) = 39.60, χ^2 (14) = 42.40, respectively). Degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity for heart rate (ε = .54), blink rate (ε = .74), and respiration (ε = .69).

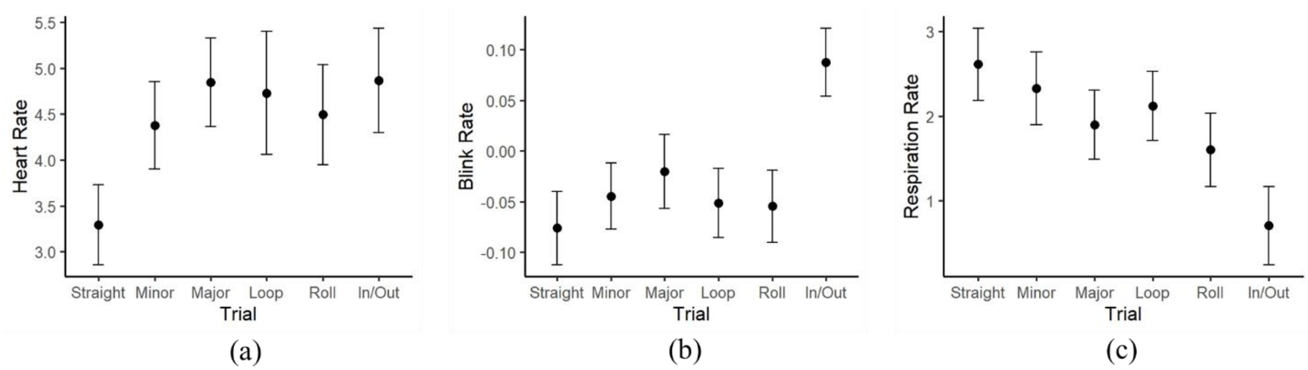


Figure 4. (a) Average heart rate across trials, (b) average blink rate across trials, and (c) average respiration rate across trials, with error bars denoting standard error.

In addition, we conducted Pearson's correlations between the self-report measures and the physiological measures. The results are presented in a correlation matrix in Table 2. We observed a strong correlation between the self-report measures of VR sickness and cognitive workload (r (47) = 0.45, p < .001). However, we found no significant correlations between the VRSQ and the physiological measures. There was a significant negative correlation between the NASA TLX and respiration rate (r (47) = -0.14, p < .05). We also observed significant correlations between heart rate and both blink rate (r (47) = 0.18, p < .001) and respiration (r (47) = 0.12, p < .05).

Table 2. Correlation Matrix of Self-Report and Physiological Measures

	VRSQ	NASA-TLX	Heart Rate	Blink Rate	Respiration
VRSQ	1.00				
NASA-TLX	0.45***	1.00			
Heart Rate	0.06	0.07	1.00		
Blink Rate	0.05	0.09	0.18***	1.00	
Respiration	-0.10	-0.14*	0.12*	-0.06	1.00

p < 0.05*, p < 0.01**, p < 0.001***

Susceptibility

In order to investigate individual differences in susceptibility, the MSSQ was administered prior to the study. We calculated a Pearson's correlation between the MSSQ and the MSAQ and found a significant correlation (r (47) = 0.37, $p < .008$).

DISCUSSION

One of the key contributions of this effort was our attempt to design a reusable VR flight simulation scenario that is capable of inducing changes in VR sickness and cognitive workload in both novice participants and trained pilots. The testbed comprising this scenario and the accompanying simulator equipment will provide the opportunity for a substantial quantity of additional future research studies in this area. Further, this VR flight scenario was used to validate the utility of physiological measures as potential indicators VR sickness. The results of the statistical analysis may support the conclusion that varying levels of complexity in VR tasks affects both VR sickness and cognitive workload, although additional research addressing study limitations is required. Our physiological data findings were partially consistent with prior research. The heart rate results were consistent with prior research in that we saw an increasing trend across trials, but inconsistent with prior findings that heart rate is positively correlated with VRSQ scores (Kim et al., 2005). The respiration rate results were partially consistent with prior research, as we saw a drop in respiration rate with time in the simulator (Gavgani et al., 2017). However, we did not observe a statistically significant correlation between VRSQ scores and respiration rate change relative to baseline. Similarly, our blink rate measure did increase with time in the simulator as reported in prior research, but the relationship between VRSQ and baseline-adjusted blink rate was not significant (Dennison et al., 2016). Finally, we considered potential individual differences in susceptibility to motion sickness, and a significant correlation supported the conclusion that prior susceptibility to motion sickness is likely to influence VR sickness.

Study Limitations

In an attempt to account for individual differences in physiology, we baseline-adjusted our physiological data. However, we did notice some peculiarities in our baseline data when viewing the continuous time series data over the course of the baseline trials, particularly in our ocular measures. For this baseline, participants wore VR helmets and were instructed to sit still, looking at a loading screen but not engaged in an active task. It is possible that participants were still acclimating to the VR environment in the early minutes of the baseline and we captured that process, artificially increasing our baseline values. In contrast, it is also possible that our baseline was too unconstrained and lead to a wide variation in behavior from participant to participant during the baseline.

Furthermore, while the scores for the MSAQ did show that overall a significant number of our participants experienced VR sickness, average scores were low (M=5.88, SD=7.49) and the dataset included many individual trials in which the VRSQ scores were 0, suggesting potential range restriction. Due to safety considerations, we limited the number and intensity of flight maneuvers experienced by participants and it is likely that we were too conservative in our efforts to induce VR sickness across all participants. A higher frequency and severity of VR sickness could lead to different findings with respect to the physiological indicators. Additional piloting could help to identify the upper limit of flight maneuvers that could be included in future studies to increase the frequency and severity of VR sickness.

Finally, our manipulation of complexity was confounded with duration of VR exposure and imposed potential order effects. Previous research suggests that as the length of time spent in a VR environment increases, so does VR sickness (Dużmańska et al., 2018). In future studies, a control group could be recruited to repeat the Straight Flight trial six times rather than progressing through all six flight trials to eliminate this issue. Further, since our study involved a gradual induction of VR sickness across time that prevented counterbalancing, order effects cannot be discounted.

Conclusions and Future Work

The results of the present study discussed thus far constitute a promising start towards understanding how task complexity impacts VR sickness, cognitive workload, and physiological response during VR exposure. However, we have additional analyses planned for the present dataset to continue exploring these concepts, as the analyses presented do not cover all the metrics in the dataset. First, we intend to consider additional potential physiological indicators including heart rate variability, conductance levels via galvanic skin response data, and additional eye tracking metrics such as pupil diameter, number of fixations, and number of saccades. Our current dataset also includes screen capture recordings of each flight trial for each participant. We plan to analyze these videos for visual complexity using an open-source tool (AUVANA; Alghamdi et al., 2017) to objectively quantify the differences in the experimental scenario tasks. Third, given our significant findings related to motion sickness susceptibility, we will consider statistical methods for accounting for these differences, possibly including the division of the dataset into groups based on susceptibility and running a mixed-effects ANOVA that treats susceptibility as a between-groups factor. Beyond continued analysis of the existing dataset, the current study did not assess the impact of existing medical conditions (i.e., recurrent migraines) on likelihood of VR sickness, which could be of interest. Additionally, it would be useful to conduct similar future research with a participant sample comprised entirely of trained pilots and a scenario design that increases the number of flight maneuvers.

ACKNOWLEDGEMENTS

This work was supported by the U.S. Navy under Contract No. N68335-18-C0133. The views, opinions, and/or findings contained in this paper are those of the authors and should not be construed as an official position, policy or decision of the U.S. Navy.

REFERENCES

Agrawal, R., Knodler, M., Fisher, D. L., & Samuel, S. (2017, September). Advanced virtual reality based training to improve young drivers' latent hazard anticipation ability. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 61, No. 1, pp. 1995-1999). Sage CA: Los Angeles, CA: SAGE Publications.

Alghamdi, E., Velloso, E., & Gruba, P. (2021). AUVANA: An automated video analysis tool for visual complexity.

Bos, J. E., & Lawson, B. D. (2021). Symptoms and measurement. In *Guidelines for Mitigating Cybersickness in Virtual Reality Systems*. (Publication No. NATO STO-TR-HFM-MSG-323). https://www.sto.nato.int/publications/STO%20Technical%20Reports/STO-TR-HFM-MSG-323/$$TR-HFM-MSG-323-ALL.pdf

Bos, J., Lawson, J.A., Rigato, P., & Secci, A. (2021). Introduction. In *Guidelines for Mitigating Cybersickness in Virtual Reality Systems*. (Publication No. NATO STO-TR-HFM-MSG-323). https://www.sto.nato.int/publications/STO%20Technical%20Reports/STO-TR-HFM-MSG-323/$$TR-HFM-MSG-323-ALL.pdf

Bouchard, S., Robillard, G., & Renaud, P. (2007). Revising the factor structure of the Simulator Sickness Questionnaire. *Annual review of cybertherapy and telemedicine*, *5*(Summer), 128-137.

Davis, S., Nesbitt, K., & Nalivaiko, E. (2015). Comparing the onset of cybersickness using the Oculus Rift and two virtual roller coasters. Proceedings of the 11th Australian Conference on Interactive Entertainment (IE2015), Sydney, Australia, 27-20 January, 3-14.

De Winter, J. C. (2014). Controversy in human factors constructs and the explosive use of the NASA-TLX: a measurement perspective. *Cognition, technology & work*, *16*(3), 289-297.

Dennison, M. S., Wisti, A. Z., & D'Zmura, M. (2016). Use of physiological signals to predict cybersickness. *Displays*, *44*, 42-52. doi: 10.1016/j.displa.2016.07.002.

Dużmańska, N., Strojny, P., & Strojny, A. (2018). Can simulator sickness be avoided? A review on temporal aspects of simulator sickness. *Frontiers in Psychology*, *9*, 2132. doi: 10.3389/fpsyg.2018.02132

Gavgani, A.M., Hodgson, D. M., & Nalivaiko, E. (2017). Effects of visual flow direction on signs and symptoms of cybersickness. *PloS One*, *12*(8), e0182790. doi: 10.1371/journal.pone.0182790

Gavgani, A. M., Nesbitt, K. V., Blackmore, K. L., & Nalivaiko, E. (2017). Profiling subjective symptoms and autonomic changes associated with cybersickness. *Autonomic Neuroscience*, *203*, 41-50. doi: 10.1016/j.autneu.2016.12.004

Gianaros, P. J., Muth, E. R., Mordkoff, J. T., Levine, M. E., & Stern, R. M. (2001). A questionnaire for the assessment of the multiple dimensions of motion sickness. *Aviation, Space, and Environmental Medicine*, *72*(2), 115.

Golding, J. F. (1998). Motion sickness susceptibility questionnaire revised and its relationship to other forms of sickness. *Brain Research Bulletin*, *47*(5), 507-516.

Hart, R., & Karthigasu, K. (2007). The benefits of virtual reality simulator training for laparoscopic surgery. *Current Opinion in Obstetrics and Gynecology, 19*(4), 297-302.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in Psychology*, *52*, 139-183.

Howard, M. C., & Van Zandt, E. C. (2021). A meta-analysis of the virtual reality problem: Unequal effects of virtual reality sickness across individual differences. *Virtual Reality*, *25*(4), 1221-1246.

Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993). Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology*, *3*(3), 203-220.

Kim, Y. Y., Kim, H. J., Kim, E. N., Ko, H. D., & Kim, H. T. (2005). Characteristic changes in the physiological components of cybersickness. *Psychophysiology, 42*(5), 616-625.

Kim, J., Chung, C.Y.L., Nakamura, S., Palmisano, S., & Khuu, S.K. (2015). The Oculus Rift: a cost-effective tool for studying visual-vestibular interactions in self-motion perception. *Frontiers in Psychology, 6*. doi: 10.3389/fpsyg.2015.00248

Kim, H. K., Park, J., Choi, Y., & Choe, M. (2018). Virtual reality sickness questionnaire (VRSQ): Motion sickness measurement index in a virtual reality environment. *Applied Ergonomics*, *69*, 66-73. doi: 10.1016/j.apergo.2017.12.016

LaCount, L. T., Barbieri, R., Park, K., Kim, J., Brown, E. N., Kuo, B., & Napadow, V. (2011). Static and dynamic autonomic response with increasing nausea perception. *Aviation, space, and environmental medicine*, *82*(4), 424-433.

McCauley, M. E., & Sharkey, T. J. (1992). Cybersickness: Perception of self-motion in virtual environments. *Presence: Teleoperators & Virtual Environments*, *1*(3), 311-318.

McCoy-Fisher, C., Mishler, A., Bush, D., Severe-Valsaint, G., Natali, M., & Riner, B. (2019). *Student naval aviation extended reality device capability evaluation* (Report No. NAWCTSD-TR-20190-001). Fort Belvoir, VA: Defense Technical Information Center.

Mishler, A., Severe-Valsaint, G., Natali, M., Seech, T., McCoy-Fisher, C., Cooper, T., & Astwood, R. (2022). *Project Avenger training effectiveness evaluation* (Report No. NAWCTSD Public Release 22-ORL006). Fort Belvoir, VA: Defense Technical Information Center.

Nalivaiko, E., Davis, S. L., Blackmore, K. L., Vakulin, A., & Nesbitt, K. V. (2015). Cybersickness provoked by head-mounted display affects cutaneous vascular tone, heart rate and reaction time. *Physiology & Behavior*, *151*, 583-590. doi: 10.1016/j.physbeh.2015.08.043

Ranchet, M., Morgan, J. C., Akinwuntan, A. E., & Devos, H. (2017). Cognitive workload across the spectrum of cognitive impairments: A systematic review of physiological measures. *Neuroscience & Biobehavioral Reviews, 80*, 516-537.

Reason, J. T., & Graybiel, A. (1970). Changes in subjective estimates of well-being during the onset and remission of motion sickness symptomatology in the slow rotation room. *Aerospace medicine*, *41*(2), 166–171.

Reed, N., Diels, C., & Parkes, A. M. (2007). Simulator sickness management: Enhanced familiarisation and screening processes. In *Proceedings of the First International Symposium on Visually Induced Motion Sickness, Fatigue, and Photosensitive Epileptic Seizures (VIMS2007),*156-162.

Severe-Valsaint, G., Mishler, A., Natali, M., Astwood, R., Seech, T., & McCoy-Fisher, C. (2022). *Training effectiveness evaluation of an adaptive virtual instructor for naval aviation training* (Report No. NAWCTSD Public Release 22-ORL044). Fort Belvoir, VA: Defense Technical Information Center.

Stone III, W.B. (2017). Pyschometric evaluation of the Simulator Sickness Questionnaire as a measure of cybersickness. PhD Dissertation, Iowa State University.