Advances in Transportation, Vol. 60, 2022, 732–739
https://doi.org/10.54941/ahfe1002511

Mental Workload Assessment in Virtual Reality Environment Based Ship Navigating Operations

Yu Zhang[1], Chongbin Li[1], Quan Ding[1], and Jie Chen[2]

[1]School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an 710049, China
[2]Wuhan Second Ship Design and Research Institute, Wuhan 430205, China

ABSTRACT

With the development of virtual reality technology and human-computer interaction technology, workload assessment in a virtual reality environment become a reality. Using virtual reality technology to design digital models to simulate products and verify related solutions can reduce costs, shorten the development cycle, and improve the quality of product design. The goal of this paper is to verify the effectiveness of workload assessment in a virtual reality environment. In this study, the ship console system is an example to conduct a comparative experiment in the virtual prototype and physical prototype. Workload assessment indicators include task performance (task completion rate, task completion time), eye tracking measures (gaze entropy, pupil change rate and fixation rate of AOI, etc.), and subjective scales (NASA_TLX). According to a ship console's size, display interface, and environment, a digital model and the experimental environment were built in Unity engine. The data were collected to verify the consistency of the experimental results of the VR prototype and the physical prototype. There was no significant difference in task completion rate, NASA_TLX scores, fixation rate of AOI, pupil change rate, gaze entropy, and blink rate between the two prototype systems. The results show that the VR prototype can be used to replace the physical prototype for workload assessment to some extent.

Keywords: Human-computer interaction, Mental workload assessment, Virtual reality, Eye-tracking, NASA-TLX

INTRODUCTION

Workload is used to measure the availability or acceptability of the man-machine system index reflects the ability of the operator in the process of completing tasks in a specific environment and conditions. It includes physical workload and mental workload, also known as cognitive workload (Lebiere, Anderson and Bothell, 2001). Over the past 40 years, many different methods have been developed for quantitative measurement of workload, mainly including three principal measurements: subjective measures such as self-reporting, Cooper-Harper scale, and NASA-TLX (Hart and Staveland, 1988); Behavior measures are used to evaluate the performance of the operator in tasks, such as primary task (Boles et al., 2007) and secondary task method (Ogden, Levine and Eisner, 1979); And physiological measures, such as EEG (Arico et al., 2015), eye-tracking (Das, Maiti and Krishna, 2020), HRV (Gagnon, Tremblay and Dehais, 2014), GSR (Kajiwara, 2014) and other physiological measures. Corresponding criteria are also proposed for the evaluation basis of various measurement methods, including five criteria: sensitivity, diagnosticity, intrusiveness, implementation requirements, and operator acceptance (Eggemei et al., 1991). In recent years, researchers tend to focus on the measurement of mental workload and adopt a variety of methods to comprehensively measure and reflect the changes in workload in real-time.

Virtual Reality is a digital technology that creates a three-dimensional Virtual world through computer programs. In recent years, with the development and maturity of VR technology, it has been applied in many fields, including scene display, design verification, training, and so on. Based on VR technology in the design, evaluation and validation have been widely applied, including many companies such as Boeing, Volkswagen, and GM (Kulkarni et al., 2011). It is a good way to reduce the use of physical prototypes, save costs; shorten the development cycle and improve the quality of product design. In recent years, some researchers have applied VR technology for training and evaluating human factors and ergonomics such as shunting training (Tschoerner et al., 2021).

Eye-tracking is a sensor technology for measuring an individual's eye positions and eye movement based on the optical tracking of corneal reflections to evaluate visual attention. The use of eye-tracking measures has increased in recent years, making eye trackers a non-obtrusive and non-distracting tool for assessing mental workload. Eye movement could provide continuous, moment-to-moment measures of workload (Ahlstrom and Friedman-Berg, 2006). Three types of eye related measurements are commonly investigated to assess mental workload: eye movement (fixations and saccade), pupillometry, and blinks (Di Nocera et al., 2015).

Recently, the integration of eye-tracking into VR headsets has substantially in-creased the scope of experimental settings. Testing scenarios are no longer bound by factors, such as time, safety, and budget, that would prevent conducting certain experiments (Mirault et al., 2020). With the further development of VR related technologies involving artificial intelligence, voice and audio processing, and multi-sensor technology, complex system simulation will be better. The objective of this paper is to verify the effectiveness of workload assessment in a VR environment by taking ship navigation consoles as a study case.

METHODOLOGY

Workload Measures

Workload reflects two aspects: objective task performance and demands for operator's ability. Task-related factors involved in information flow, multitasking, difficulty, and duration (time pressure) affect workload (Adler and Benbunan-Fich, 2015). Three types of measurements are investigated to assess workload in this research : 1) task performance: task completion rate, task completion time); 2) eye-tracking measures (see Table 1): gaze entropy, pupil change rate, fixation rate of AOI, and blink rate; 3)

Table 1. Eye-tracking measures.

Measure	Description	Definition
Gaze entropy	Refers to quantitative approaches that have been used to assess visual scanning behavior during engagement in tasks with high visual demand (measured in bits)	Calculated using Shannon's entropy formula: $H_g(X) = -\sum p(x,y) \cdot \log_2 p(x,y)$ where p(x,y) is the probability of the subject's gaze falling in the (x,y) position of the visual field for a given sample.
Pupil change rate	Measures of pupil size and reactivity (Measured in pixels or millimeters)	$x' = \frac{x - \min(x)}{\max(x) - \min(x)}$ where x' is the rate of pupil chance, x is real-time pupil diameter; maximum pupil diameter of the subject; max (x) is maximum pupil diameter of the subject; min (x) is the minimum pupil diameter of the subject.
Fixation rate	State of a gage that is focused (fixated) on an object	The ratio of the AOI fixation duration to the total fixation duration of all regions
Blink rate	Measures of partial or full eye closure	$BLR = \frac{n_{winks}}{t_{i+n_{winks}} - t_i} \times 60$ Where $t_{i+n_{winks}} - t_i$ is a target TOI and n_{winks} is the number of the subject's winks in the TOI

subjective measures: NASA_TLX, which is based on six scales including mental demand, physical demand, temporal demand, performance, effort, and frustration level.

Experimental System Design

Two different experimental environments were designed to correspond to the virtual environment and the physical operation environment respectively, to verify whether the virtual environment and the physical environment have a good consistency. According to relevant size standards, 3D modeling of the ship console was carried out, and the model was imported into Unity engine (unity 2019.4.2f1) for material rendering and lighting environment setting. Virtual buttons, joysticks, display interfaces, and other human-computer interaction functions consistent with real objects were endowed by C# programming. The ship navigation console model was reconstructed and optimized. Two different experimental environments were designed to correspond to the VR prototype and the physical prototype respectively. VR eye tracker, VR locator, digital gloves, and other equipment are connected to the host running virtual reality simulation system (see Figure 1). The physical prototype that can simulate ship dynamics and interface interaction combined with an eye tracker to build the physical simulation experiment platform

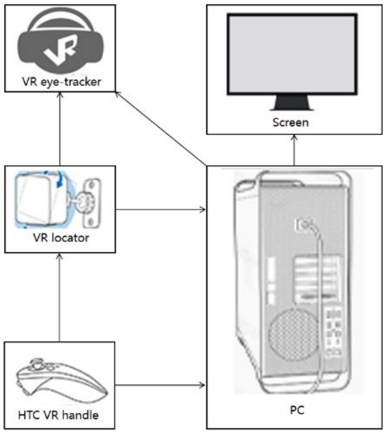

Figure 1: VR simulation platform.

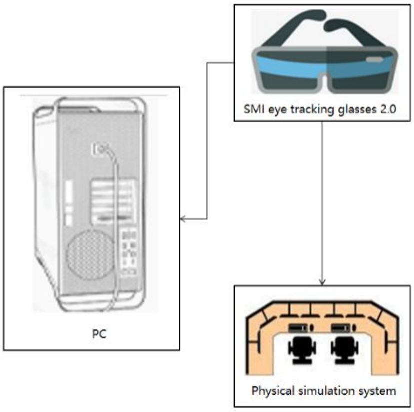

Figure 2: Physical simulation platform.

(see Figure 2). In addition, software and hardware for recording data include video recorder and iMotions, etc.

Experimental Task Design

Firstly, subjects proceeded directly to set up the devices and undergo one practice session. Three test sessions then were conducted.

1) Session #1 is to simulate abnormal parameter conditions including 6 subtasks. Abnormal parameters appeared on the display interface randomly, which lasted for 10 seconds. The subjects were required to find abnormal events within the duration.
2) Session #2 is to simulate command tasks including 4 subtasks. The subjects need to follow the oral command of the experimenter to complete the corresponding operations and report to the experimenter after observing the parameter changes and reaching the command requirements. Each subtask takes about 3-4 minutes, with an interval of one minute between each subtask.

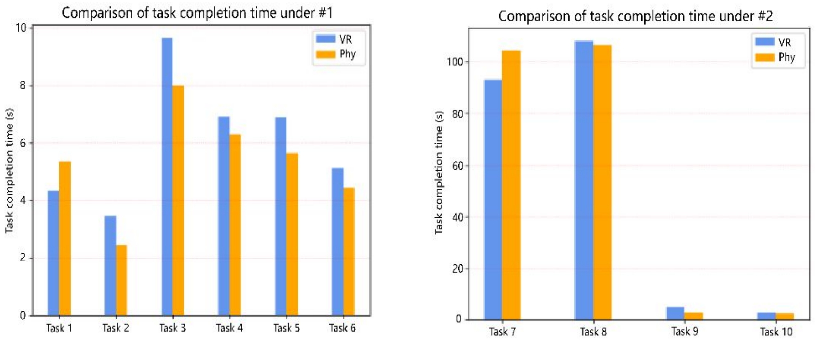

Figure 3: Comparison of the completion times of session #1 and #2 on VR platform and physical platform.

Table 2. Comparative statistics of task performance in two experimental systems.

Measures	Session	VR		Phy		F	P
		Mean	StDev	Mean	StDev		
Completion time	#1	8.321	8.115	5.536	7.824	1.190	0.280
	#2	112.208	103.372	130.458	123.928	0.611	0.439
Completion rate	#1	0.9622	0.07496	0.9244	0.08960	3.114	0.097
	#2	1.0000	0.0000	1.0000	0.0000	/	/

Participates

Nine subjects were tested in this experiment. Group 1 was 4 HFE experts, who had mastered the knowledge of human factors engineering but were not familiar with the task flow of ship operation. Group 2 consists of 3 ship designers who have experience in ship design and are familiar with the interface and operation mode of the ship console. Group 3 was two professional operators who have experience in ship operation and are familiar with the interface, task procedure, and abnormal situations of the ship operation.

RESULTS AND DISCUSSION

Comparison of Task Performance Between VR Platform and Physical Platform

In terms of average completion time, there was a high consistency between the physical platform and the VR platform. The completion time of the physical platform is generally lower than the VR platform (see Figure 3). There was no significant difference in task completion time and completion rate between the two platforms (see Table 2).

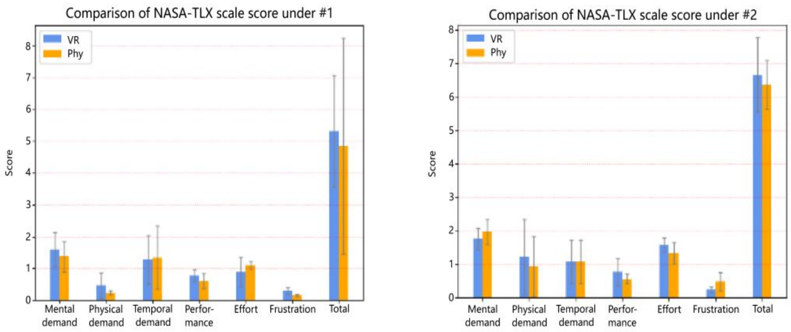

Figure 4: Comparison of NASA-TLX scores of session #1 and #2 on VR platform and physical platform.

Table 3. Comparative statistics of subjective responses in two experimental systems.

Measures	Session	VR		Phy		F	P
		Mean	StDev	Mean	StDev		
Mental demand	#1	1.5878	0.73138	1.3822	0.69036	0.019	0.892
	#2	1.7633	0.56862	1.9656	0.61295	0.001	0.975
Physical demand	#1	0.4678	0.63259	0.2333	0.27518	2.052	0.171
	#2	1.2189	1.05814	0.9256	0.94834	0.201	0.660
Temporal demand	#1	1.2856	0.87082	1.3478	0.99524	0.473	0.502
	#2	1.0778	0.80973	1.0811	0.80463	0.059	0.811
Performance	#1	0.7789	0.43825	0.6078	0.48033	0.030	0.865
	#2	0.7700	0.64327	0.5622	0.38124	1.644	0.218
effort	#1	0.8922	0.68412	1.1111	0.33773	4.953	0.041
	#2	1.5767	0.45741	1.3411	0.55627	1.719	0.208
Frustration level	#1	0.3000	0.31733	0.1600	0.19660	0.495	0.492
	#2	0.2556	0.29241	0.4856	0.52403	1.735	0.206
Total	#1	5.3100	1.32000	4.8400	1.84225	1.093	0.311
	#2	6.6644	1.05533	6.3622	0.85400	1.878	0.189

Comparison of NASA-TLX Scores Between VR Platform and Physical Platform

The mental demand of the physical platform is slightly lower than that of the VR platform in session #1 and session #2, but there is no significant difference (see Figure 4 and Table 3). From the perspective of the influencing factors of workload, the physical demand of the physical platform is slightly lower, and the VR platform is prone to visual fatigue.

Comparison of Eye Tracking Measures Between VR Platform and Physical Platform

Gaze entropy on the VR platform is generally higher than that on the physical platform, which is caused by the fact that subjects need to saccade more frequently due to the smaller field of vision on the VR platform (see Figure 5).

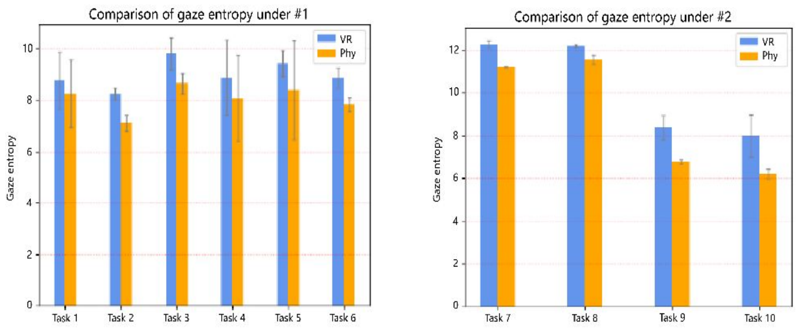

Figure 5: Comparison of gaze entropy of session #1 and # 2 on VR platform and physical platform.

Table 4. Comparative statistics of eye movement metrics in two experimental systems.

Measures	Session	VR		Phy		F	P
		Mean	StDev	Mean	StDev		
Fixation rate of AOI	#1	0.2526	0.10364	0.2567	0.10360	0.000	0.983
	#2	0.2380	0.06509	0.2424	0.06299	0.014	0.906
Pupil change rate	#1	0.621	0.058	0.607	0.093	1.337	0.253
	#2	0.582	0.080	0.643	0.084	0.061	0.806
Gaze entropy	#1	9.171	1.017	8.005	1.004	0.110	0.742
	#2	10.962	2.408	10.119	2.686	1.082	0.304
Blink rate	#1	18.5489	6.26473	18.4467	6.57050	0.004	0.948
	#2	18.4000	9.99379	18.4389	9.91380	0.004	0.949

There were no significant differences in the fixation rate of AOI, pupil change rate, gaze entropy, and blink rate of the VR platform and the physical platform (see Table 4). The results show that there was a high consistency of the experimental results of the VR platform and the physical platform. It suggests that the VR platform can replace the physical platform for complex system workload assessment to some extent.

CONCLUSION

This paper proposes a new method of workload assessment based on eye-tracking measures in a VR environment. There were no significant differences in task performance, NASA-TLX scales, and eye movement measures, it can be considered that workload assessment in a VR environment is feasible to replace physical prototype to some extent. Eye-tracking is the only physiological measurement method used in this paper, and other physiological measurement methods such as GSR, HRV, and so on can be added in future work. With the further development of VR technology, more convenient and practical human-computer interaction approaches applied in the VR environment will be gradually updated. These technological advances are more conducive to the simulation of a complex system based on virtual reality technology, and the workload assessment methods would be improved.

REFERENCES

Adler, R.F. and Benbunan-Fich, R. (2015) 'The Effects of Task Difficulty and Multitasking on Performance', Interacting with Computers, 27(4), pp. 430–439. doi:10.1093/iwc/iwu005.

Ahlstrom, U. and Friedman-Berg, F.J. (2006) 'Using eye movement activity as a correlate of cognitive workload', International Journal of Industrial Ergonomics, 36(7), pp. 623–636. doi:10.1016/j.ergon.2006.04.002.

Arico, P. et al. (2015) 'Reliability over time of EEG-based mental workload evaluation during Air Traffic Management (ATM) tasks', in 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC).Milan: IEEE, pp. 7242–7245. doi:10.1109/EMBC.2015.7320063.

Boles, D.B. et al. (2007) 'Predicting Dual-Task Performance With the Multiple Resources Questionnaire (MRQ)', Human Factors: The Journal of the Human Factors and Ergonomics Society, 49(1), pp. 32–45. doi:10.1518/001872007779598073.

Das, S., Maiti, J. and Krishna, O.B. (2020) 'Assessing mental workload in virtual reality based EOT crane operations: A multi-measure approach', International Journal of Industrial Ergonomics, 80, p. 103017. doi:10.1016/j.ergon.2020.103017.

Di Nocera, F. et al. (2015) 'Mental workload assessment using eye-tracking glasses in a simulated maritime scenario', Proceedings of the Human Factors and Ergonomics Society Europe, Groningen, The Netherlands, pp. 14–16.

Eggemeier, F. T., Wilson, G. F., Kramer, A. F., & Damos, D. L. (1991). General considerations concerning workload assessment in multi task environment. In D.L. Damos(Ed.). In Multiple task performance (pp. 207–216). Taylor & Francis.

Gagnon, J.-F., Tremblay, S. and Dehais, F. (2014) 'Using near infrared spectroscopy and heart rate variability to detect mental overload', Behavioural Brain Research, 259, pp. 16–23. doi:10.1016/j.bbr.2013.10.042.

Hart, S.G. and Staveland, L.E. (1988) 'Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research', in Hancock, P.A. and Meshkati, N. (eds) Advances in Psychology. North-Holland (Human Mental Workload), pp. 139–183. doi:10.1016/S0166-4115(08)62386-9.

Kajiwara, S. (2014) 'Evaluation of driver's mental workload by facial temperature and electrodermal activity under simulated driving conditions', International Journal of Automotive Technology, 15(1), pp. 65–70. doi:10.1007/s12239-014-0007-9.

Kulkarni, A. et al. (2011) 'Virtual prototyping used as validation tool in automotive design', in. 19th International Congress on Modelling and Simulation, Modelling and Simulation Society of Australia and New Zealand (MSSANZ), Inc. doi:10.36334/modsim.2011.A5.kulkarni.

Lebiere, C., Anderson, J.R. and Bothell, D. (no date) 'Multi-Tasking and Cognitive Workload in an ACT-R Model of a Simplified Air Traffic Control Task', p. 9.

Mirault, J. et al. (2020) 'Using virtual reality to study reading: An eye-tracking investigation of transposed-word effects', Methods in Psychology, 3, p. 100029. doi:10.1016/j.metip.2020.100029.

Ogden, G.D., Levine, J.M. and Eisner, E.J. (1979) 'Measurement of Workload by Secondary Tasks', Human Factors: The Journal of the Human Factors and Ergonomics Society, 21(5), pp. 529–548. doi:10.1177/001872087902100502.

Tschoerner, B. et al. (2021) 'Human factors evaluation in VR-based shunting training', The Visual Computer, 37(12), pp. 3063–3076. doi:10.1007/s00371-021-02251-x.