A Study on Visual Workload Components: Effects of Component Combination and Scenario Complexity on Mental Workload in Maritime Operation Tasks

Ye Deng[1], Yuexuan Wang[2], Manrong She[2], Yijing Zhang[2], and Zhizhong Li[2(✉)]

[1] China Institute of Marine Technology & Economy, Beijing, People's Republic of China
dengye@cimtec.net.cn
[2] Department of Industrial Engineering, Tsinghua University, Beijing 100084, People's Republic of China
wyx_thul4@qq.com, {shemanrong,ldlzyj}@l63.com, zzli@tsinghua.edu.cn

Abstract. Mental workload is a key factor in influencing human performance of maritime tasks among which visual tasks dominate. Through a lab experiment, this study examined the effects of visual workload components, their combinations and scenario complexity on mental workload. 20 participants were recruited to perform visual tasks on a simulated platform. Results showed that some combinations of visual components significantly increased the participants' mental workload. Scenario complexity was found to have significant interaction effects with workload component combinations. The results indicate that because of their negative effects on task performance, some combinations of workload components should be specially avoided in task design.

Keywords: Mental workload · Visual task component · Scenario complexity · Maritime operation

1 Introduction

In safety-critical industries, mental workload is considered as a key factor for operator performance [1]. Based on the theory of resource demand and supply, mental workload is defined as the relationship between the mental resources demanded by a task and those resources to be supplied by an operator [2]. In a review of maritime accident reports, researchers illustrated how the 13 out of 31 accidents was caused by either too high or too low level of mental workload [3]. During the system design, usability testing and operator training, it is important not only to measure the level of mental workload, but also to predict how much mental workload an operator would perceive while performing the tasks [4].

The Visual/Auditory/Cognitive/Psychomotor (VACP) model is a popular method to predict mental workload. VACP was first developed for a military lightweight

helicopter system [5]. To adopt the VACP method, researchers first break the mission into phases, segments, functions, tasks, and the performance elements for each task. A performance element is usually composed of a verb and an object. On the basis of multiple resource theory, workload components are categorized into four channels (i.e. visual, auditory, cognitive and psychomotor channels). Through expert judgment, a standard rating scale for workload components was proposed. For example, the component "detect" in the visual channel is rated 1.0 (for the detailed rating scale, please refer to [4, 5]. For each performance element, researchers identify its relevant workload components and sum up the scores by each channel. In this way, the workload for each task is quantified in terms of the four channels. The score is determined by the nature of the task, and is independent of individual factors.

The VACP method is not without potential flaws. As we know, mental workload is influenced not only by the demand/resource balance, but also by other factors such as time pressure, scenario complexity, individual experience and ability [6, 7]. Under a high complexity scenario (e.g. more targets on the display) or under high time pressure, an operator may perceive higher mental workload even if they are executing the same task as before [8]. In addition, when using the scale, the scores are sum up for a specified time period, without considering whether the workload components in this period are presented in series or parallel, meaning that combination of workload components (multi-task) is not differentiated from individual workload components. In the current study, we provided a list of visual workload components for maritime operation tasks, and investigated whether different components and their combinations would influence operator's mental workload, and whether the influence would change under different scenario complexity levels.

2 Visual Workload Components for Maritime Operations

Through mission segmentation and expert judgment, we revised McCracken and Aldrich's workload components [5] to make them applicable for maritime operation tasks. The process of developing maritime workload components would not be detailed in this paper. We focused on the workload components from the visual channel. The visual workload components are shown in Table 1. Two components in Mccracken and Adrich's model, "scan/search/monitor" and "read", were replaced by "retrieve" and "compare" as described in Table 1.

3 Experiment

The experiment was conducted to investigate whether different workload components and combinations of them would influence operator's mental workload, and whether the influence would be different under different scenario complexities. The maritime operation tasks were designed to meet the study requirements of different workload components. Participants executed the tasks using a simulated maritime platform. Their performance data were recorded for later analysis.

Table 1. Description of visual workload components.

Workload component	Description
Retrieve	Look through the information, retrieve the needed information, obtain the parameter values
Compare	Compare different pieces of information visually to select the proper one(s)
Detect	Detect the appearance and disappearance of an object in the field of vision
Discriminate	Discriminate the change (e.g. color, shape, value, location) of an object
Search	Search a certain object in the field of vision
Check	Check the state of equipment/parameter, check the action/command to be performed
Track	Track/follow an object in order to find out any abnormalities

3.1 Participants

Twenty undergraduates from Tsinghua University (10 males and 10 females) were recruited as participants. Their average age was 20 with the standard deviation of 2.8. All the participants had normal or correct-to-normal vision. The participants were informed of the experimental details and voluntarily signed the consent form.

3.2 Experimental Platform

A simulated maritime operation system was developed. The interface is shown in Fig. 1. The targets were represented with different colors (i.e. red, blue, yellow and green) and different shapes (i.e. square, circle and triangle). The system log data and the participant's performance data were recorded and exported to text files.

3.3 Independent Variables

There were two independent variables in the experiment: scenario complexity and combination of workload components. Both of them were within-subject variables. Scenario complexity was defined as the number of targets on the maritime display. The three levels of scenario complexity were low (10 targets on the display), medium (20 targets on the display), and high (30 targets on the display). The combination of workload components included detect, discriminate, search, check, track, and the combination of any two of above. The combination of search and check were excluded because the "search" and the "check" tasks could not be executed at the same time. Therefore, there were 14 combinations of workload components. Each participant was required to perform tasks in all cases. A case meant a single workload component or the combination of two workload components combined with a level of scenario complexity. The participant was supposed to repeat the task for 10 trials in a case.
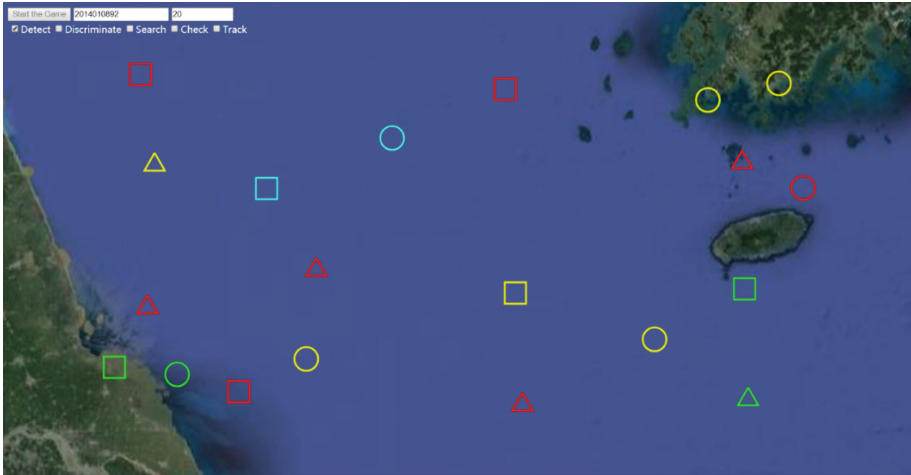
Fig. 1. The maritime interface (Color figure online)

3.4 Tasks and Procedure

At the beginning of the experiment, the participants filled out the demographic information. The experimenter introduced the purpose of the experiment, the tasks to be performed, and the usage of the simulated platform. The participants then practiced with the platform for about 10 min to get familiar with the whole experiment. The experimenter would answer any questions during the practice.

Before the experiment, the participant adjusted the seat height and the distance to the computer screen. Every time the participant completed the task in a case, he/she rested for 15 s.

During the formal experiment, the experimenter first entered a scenario complexity level in the system, and the display would present the corresponding number of targets to the participant. The tasks to be performed were as follows.

- Detect. Within the time limit, new targets randomly appeared on the display. The participant was expected to detect the newly appeared target and click on it as quickly as possible. If the participant did not respond to the new target within 5 s, the system would record it as time-out.
- Discriminate. Within the time limit, an existing target on the display changed its color. The participant was asked to click on the target as soon as he/she noticed the color change. If the participant did not respond to the target within 5 s, the system would record it as time-out.
- Search. The system randomly presented searching-relevant questions for the participant to answer.
- Check. The interface was divided into six areas. The participant was asked to decide whether the total number of targets within an area exceeded a certain value.

- Track. A new target appeared and randomly moved on the display. If the new target ran into an exiting one, the participant was asked to click on the existing target that was ran into. If the participant did not respond within 5 s, the system would record it as time-out.

The error rate and the response time were recorded for all the above tasks.

3.5 Dependent Variables

Performance data could reflect the level of mental workload. Although workload is not the only factor that influences operator performance [9], higher workload is believed to contribute to worse performance [6]. In the experiment, the error rate and the response time were recorded as two dependent variables [10].

4 Results

4.1 Effects of Single Workload Component and Scenario Complexity

Table 2 summarizes the descriptive statistics of single workload component and scenario complexity on error rate and response time.

Table 2. Descriptive statistics of single workload component and scenario complexity.

Workload component	Scenario complexity	Error rate			Response time		
		N	Mean	SD	N	Mean (ms)	SD (ms)
Detect	Low	20	0.065	0.0813	20	1288	147.8
	Medium	20	0.120	0.1281	20	1327	221.9
	High	20	0.185	0.1461	20	1270	246.8
Discriminate	Low	20	0.300	0.1654	20	1450	307.0
	Medium	20	0.450	0.2013	20	1436	245.9
	High	20	0.520	0.1361	20	1487	260.6
Search	Low	20	0.040	0.0821	20	4365	1183
	Medium	20	0.050	0.1100	20	6573	1521
	High	20	0.110	0.1373	20	8994	2923
Check	Low	20	0.017	0.0745	20	1574	492
	Medium	20	0.025	0.0816	20	1667	698
	High	20	0.033	0.1026	20	1924	774
Track	Low	20	0.015	0.0489	20	808	254.4
	Medium	20	0.040	0.1146	20	773	183.0
	High	20	0.035	0.0587	20	838	247.2

The error rate data violated the assumption of normality. The nonparametric Kruskal-Wallis test was used. The effect of scenario complexity on error rate was significant with the "detect" ($p = 0.020$) and "discriminate" ($p < 0.001$) tasks, but was not significant

with the "search" ($p = 0.267$), "check" ($p = 0.954$), and "track" ($p = 0.559$) tasks. For the "detect" and "discriminate" tasks, the higher scenario complexity resulted in higher error rate. Under the same scenario complexity, the error rates among different workload components were significantly different (all $p < 0.001$). Under the medium scenario complexity level (i.e. 20 targets on the display), Mann-Whitney's U test was used for post hoc analysis. The post hoc results are shown in Table 3. The error rate exhibited a significant difference between the "detect" task and any of the other four tasks. The same could be said between the "discriminate" task and any of the other four tasks. The error rates between two of the "search", "check", and "track" tasks were not significantly different.

The natural logarithm of the response time data satisfied the normality and homogeneity assumptions, and thus ANOVA was used. The main effects of workload component ($F = 64.70$, $p < 0.001$) and scenario complexity ($F = 3.73$, $p = 0.025$) were significant. The interaction effect between workload component and scenario complexity was also significant ($F = 3.73$, $p = 0.025$). Tukey's method was used for multiple comparisons. Results showed that the differences between low and high scenario complexity levels and between medium and high scenario complexity levels were significant. Except for between "detect" and "discriminate" and between "discriminate" and "check", the differences between any other two of the workload components were significant.

Table 3. Post Hoc analysis (Mann-Whitney's U Test) for error rate

Workload component 1	Workload component 2	W	p
Detect	Discriminate	248.50	<0.001
Detect	Search	483.00	0.0499
Detect	Check	506.00	0.010
Detect	Track	488.00	0.036
Discriminate	Search	587.50	<0.001
Discriminate	Check	593.00	<0.001
Discriminate	Track	587.50	<0.001
Search	Check	431.00	0.579
Search	Track	406.00	0.925
Check	Track	383.00	0.473

4.2 Effects of Workload Component Combinations and Scenario Complexity

In this section, we compared the differences between one workload component and its combinations with the other components. For example, the difference between the "detect" task and the combination of "detect" and "discriminate" tasks were examined in terms of error rate and response time. The detailed results are illustrated as follows.

The "Detect" Task and its Combinations

The results of Kruskal-Wallis test showed that under the medium ($p = 0.006$) and high ($p = 0.002$) scene complexities, the error rates between the "detect" task and its combinations with other tasks were significantly different. Under the high scenario complexity, the results of Mann-Whitney test showed that compared with the single "detect" task, adding the "search" task significantly increased the participant's error rate ($p = 0.019$). Adding the other tasks did not increase the error rate significantly.

In terms of response time, the results of ANOVA showed that the combinations of workload components had a significant effect ($p < 0.001$). The effects of scenario complexity ($p = 0.982$) and their interaction ($p = 0.792$) were not significant. Post hoc analysis revealed that compared with the "detect" task, adding the "search" ($p < 0.001$) or "check" ($p < 0.001$) tasks significantly increased the participant's response time.

The "Discriminate" Task and its Combinations

In terms of error rate, the results of Kruska-Wallis test did not show any difference between the "discriminate" task and its combinations with other tasks.

Compared with the single "discriminate" task, adding other tasks did not increase or decrease the participant's response time obviously.

The "Search" Task and its Combinations

The results of Kruskal-Wallis test revealed that the error rate between the "search" task and its combinations with other tasks were different under the medium ($p = 0.005$) and the high ($p = 0.017$) scenario complexity levels. Further analysis (Manny-Whitney test) showed adding the task of "detect" or "discriminate" significantly increased the participant's error rate.

As for response time, the effects of workload component combination and scenario complexity, and their interaction effect were significant. The response time was significantly increased when the "detect" ($p < 0.001$), "discriminate" ($p < 0.001$), or "track" ($p < 0.001$) task was added to the "search" task.

The "Check" Task and its Combinations

No matter under which level of scenario complexity, error rate did not exhibit difference between the "check" task and its combinations with other tasks, according to the results of Kruskal-Wallis test.

We only analyzed 14 participants' response data due to data missing. The ANOVA results showed that different combinations of workload components had a significant effect on response time ($p < 0.001$). The effect of scenario complexity ($p = 0.904$) and the interaction effect ($p = 0.206$) were not significant. By adding the "detect" ($p < 0.001$), "discriminate" ($p = 0.024$), or "track" ($p < 0.001$) task, the response time was significantly increased.

The "Track" Task and its Combinations

The results of Kruskal-Wallis test showed that no matter under which scenario complexity level, the error rates were not different between the "track" task and its combinations with other tasks.

In terms of response time, there were no significant effects of workload component combinations ($p = 0.347$), scenario complexity ($p = 0.446$), or their interaction ($p = 0.706$).

5   Discussion

According to the statistics of "mental workload" in the publication title from Ergo-Abs database, the mental workload research in maritime engineering seems not as active as those in driving and air-traffic control [11]. In this study, an experiment was conducted to investigate the effects of visual workload components on operator's mental workload inferred from task performance, and whether the effects would be different if the scenario complexity varied. Operator's mental workload was evaluated through two performance measures: error rate and response time.

Results showed that some combinations of workload components significantly increased the participant's mental workload (e.g. higher error rate, more response time). Compared with the single "detect" or "search" task, the combination of the two tasks led to higher workload. In another case, adding the "track" task on the basis of the "search" or "check" task obviously increased the workload. However, adding the "search" or "check" to the "track" task was not found to make a difference in error rate or response time. The possible reason is that the "track" task takes up a larger proportion of the visual resources than the "search" or "check" task and requires the participant's continuous attention. If operator performance is essential for the system effectiveness and safety, the workload component combinations that would degrade performance should be avoided during task design. If the combination cannot be avoided in real working environment, other channels (e.g. the auditory channel) should be used to release the workload.

The VACP rating scale provides standard values for workload components. However, scenario complexity was found to have interaction effects with workload component combinations. How much workload a workload component brings might be affected by the scenario complexity. In this case, the standard value given by the scale may not reflect the accurate workload that an operator would perceive. Further investigation is needed to make clear whether and how other factors interact with workload components to affect mental workload.

The study has several limitations. This study provided evidence that scenario complexity should be considered in the VACP component rating, but did not indicate how the scenario complexity should be quantified in the scale. Due to the limitation of the designed simulated system, the "comparison" task and the combination of "search" and "check" tasks were not included in the experiment. Moreover, the study focused on the single workload component and the combination of any two. It needs further investigation on whether and how the combinations of more than two components would influence task performance and mental workload.

6   Conclusion

This study reveals that workload components would lead to different workloads when they are combined, and they would interact with scenario complexity to influence operator's mental workload. Based on the experimental results, we provide some suggestions on task design and workload component scaling.

References

1. Wickens, C.D.: Multiple resources an: mental workload. Hum. Factors **50**, 449–455 (2008)
2. Parasuraman, R., Sheridan, T.B., Wickens, C.D.: Situation awareness, mental workload, and trust in automation: viable, empirically supported cognitive engineering constructs. J. Cognitive Eng. Decis. Making **2**, 140–160 (2008)
3. Embrey, D., Blackett, C., Marsden, P., Peachey, J.: Development of a Human Cognitive Workload Assessment Tool. Human Reliability Associates, Dalton (2006)
4. Aldrich, T.B., Szabo, S.M., Bierbaum, C.R.: The development and application of models to predict operator workload during system design. In: McMillan, G.R., Beevis, D., Salas, E., Strub, M.H., Sutton, R., Van Breda, L. (eds.) Applications of Human Performance Models to System Design, pp. 65–80. Springer, Boston (1989). https://doi.org/10.1007/978-1-4757-9244-7_5
5. McCracken, J.H., Aldrich, T.B.: Analyses of selected LHX mission functions: implications for operator workload and system automation goals (No. ASI479-024-84). Anacapa Sciences Inc., Fort Rucker AL (1984)
6. Vidulich, M.A., Tsang, P.S.: Mental workload and situation awareness. In: Salvendy, G. (ed.) Handbook of Human Factors and Ergonomics, vol. 4, pp. 243–273. (2012)
7. Wu, Y., Miwa, T., Uchida, M.: Using physiological signals to measure operator's mental workload in shipping – an engine room simulator study. J. Marine Eng. Technol. **16**, 61–69 (2017)
8. Galy, E., Cariou, M., Mélan, C.: What is the relationship between mental workload factors and cognitive load types? Int. J. Psychophysiol. **83**, 269–275 (2012)
9. Wickens, C.D., Hollands, J.G., Banbury, S., Parasuraman, R.: Engineering Psychology & Human Performance. Psychology Press, London (2015)
10. Mazur, L.M., Mosaly, P.R., Hoyle, L.M., Jones, E.L., Marks, L.B.: Subjective and objective quantification of physician's workload and performance during radiation therapy planning tasks. Pract. Radiat. Oncol. **3**, 171–177 (2013)
11. Young, M.S., Brookhuis, K.A., Wickens, C.D., Hancock, P.A.: State of science: mental workload in ergonomics. Ergonomics **58**, 1–17 (2015)