

Article

Workload Measurement Method for Manned Vehicles in Multitasking Environments

Chenyuan Yang ^{1,†}, Liping Pang ^{1,†}, Jie Zhang ^{2,*} and Xiaodong Cao ¹

¹ School of Aeronautic Science and Engineering, Beihang University, Beijing 100191, China; by2105512@buaa.edu.cn (C.Y.); pangliping@buaa.edu.cn (L.P.); caoxiaodong@buaa.edu.cn (X.C.)

² Department of Industrial Engineering, Tsinghua University, Beijing 100080, China

* Correspondence: zhangjie_thu@tsinghua.edu.cn; Tel.: +86-1881119962

† These authors contributed equally to this work.

Abstract: Workload (WL) measurement is a crucial foundation for human–machine collaboration, particularly in high-stress multitasking environments such as manned vehicle operations during emergencies, where operators often experience High Workload (HWL) levels, increasing the risk of human error. To address this challenge, this study introduces a novel WL measurement method that combines Task Demand Load (TDL) and Subject Load Capacity (SLC) to quantitatively assess operator workload. This method was validated through experiments with 45 subjects using the Environmental Control and Atmospheric Regeneration (ECAR) system. The statistical results showed that as the designed WL levels increased, the Average Workload (AWL), the NASA-TLX score, and the work time percentage increased significantly, while the task accuracy and the fixation duration decreased significantly. These results also revealed the impact of WL levels on human responses (such as subjective feeling, work performance, and eye movement). In addition, very strong correlations were found between AWL measurements and NASA-TLX scores ($r = 0.75$, $p < 0.01$), task accuracy ($r = -0.73$, $p < 0.01$), and work time percentage ($r = 0.97$, $p < 0.01$). Overall, these results proved the effectiveness of the proposed method for measuring WL. On this basis, this study defined WL thresholds by integrating task accuracy with AWL calculations, providing a framework for the dynamic management of task allocation between humans and machines to maintain operators within optimal WL ranges.



Citation: Yang, C.; Pang, L.; Zhang, J.; Cao, X. Workload Measurement Method for Manned Vehicles in Multitasking Environments. *Aerospace* **2024**, *11*, 406. <https://doi.org/10.3390/aerospace11050406>

Academic Editor: Paolo Tortora

Received: 14 March 2024

Revised: 3 May 2024

Accepted: 15 May 2024

Published: 16 May 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: workload; manned vehicle; task demand load; subject load capacity; NASA-TLX

1. Introduction

Accurately measuring the workload (WL) of operators in human–computer interaction systems is crucial for ensuring the safety of the system. Although the advancement of automation technology has reduced the physical burden on operators, at the same time, the amount of information output by the system is also rapidly increasing. This change is particularly prominent when performing complex multi-window tasks, such as operating manned vehicles in emergency situations, where operators need to monitor the status of multiple subsystems in real time and effectively perform multiple tasks under tight time pressure. This working environment can easily lead to operators bearing excessive WL, thereby increasing the operational risks of human–computer systems [1,2]. Therefore, this study aims to propose a new WL measurement method to better understand the WL of operators in complex environments, and based on this, optimize task allocation [3].

WL is a multidimensional concept, influenced by factors such as workload requirement, time constraint, operator ability, and behavioral performance [4]. Boff et al. [5] defined WL as the proportion of psychological resources required by the operator in the task. Brookhuis et al. [6] argued that WL is influenced by task characteristics and personal attributes, including age, experience, motivation, task execution strategies, physical

states, and emotional states [7]. These multiple influencing factors make measuring WL a difficult challenge. Now, the common methods for measuring WL can be divided into three categories: subjective measurements, performance measurements, and physiological measurements [8–11].

The subjective measurement methods require operators to self-assess the WL level when experienced in the current or recently completed tasks, including the NASA-TLX scale, Bedford scale, and the revised Cooper Harper scale [12–14]. The NASA-TLX scale developed by Hart et al. can comprehensively assess WL [12]. This scale mainly covers six dimensions: mental demand, physical demand, temporal demand, effort, own performance, and frustration level [15,16]. The scale was originally designed to meet the needs of the aerospace industry, and over time, its application scope has significantly expanded. Nowadays, the NASA-TLX scale is not limited to the aviation industry but is widely used in various fields such as transportation, nuclear power plants, healthcare, and human–computer interaction, becoming an important tool for assessing WL in these fields [17–19]. The subjective measurement methods are widely used due to their ease of management and efficiency [20,21]. However, the method requires the operator to temporarily interrupt the ongoing task during use. Therefore, it is difficult to assess the WL of an operator in real time using subjective measurement methods. In addition, the WL measurements obtained by the operator through the subjective measurement methods may differ from the actual WL experienced [22].

The physiological measurement methods provide an objective and real-time approach for assessing WL by analyzing physiological data from subjects during task execution. In previous studies, the correlations have been extensively examined between the WL level of subjects and their physiological indicators including electrooculography (EOG), electrocardiography (ECG), electroencephalography (EEG), respiration, skin conductance, blood pressure, and eye-tracking data [23,24]. Due to its sensitivity to transient fluctuations in WL, the physiological measurement methods have been adopted by many scholars to evaluate the WL level of operators in the work environment [25–27]. However, these methods require operators to wear physiological monitoring equipment, which not only increases the overall evaluation cost but may also cause interference with the performance of operators during task execution.

The performance is one of the main indicators for measuring WL, which can directly reflect the operational effectiveness of human–machine interaction systems [28,29]. Performance measurement methods are divided into primary task method and secondary task method. As the WL level increases, operators will consume more cognitive resources, leading to a decrease in performance [30].

Siegel and Wolf [31] proposed the Time Line Analysis and Prediction (TLAP) method to quantify time load by the ratio of required time to available time. Although the impact of task type on WL is not considered, the TLAP method provides valuable insights for subsequent research on WL measurement. For example, Liu et al. [31] proposed a pilot's WL prediction method based on timeline analysis. Park et al. [32] proposed the Adaptive Control of Thought-Rationale (ACT-R) model with cognitive architecture, which was experimentally validated through flight missions. However, both methods require the decomposition of specific human–machine systems and tasks into multiple modules. The standards for task decomposition have not yet been unified, which limits the application of these methods in other fields.

In sum, the current methods for measuring WL have their own application scopes and limitations. Firstly, the subjective measurement method is a post-event measurement method that cannot measure WL in real time, which restricts our timely intervention in task allocation of human–machine interaction systems. Secondly, although physiological measurement methods can measure WL in real time, these methods often cause unnecessary interference for operators and affect the effectiveness of task execution. Finally, the performance-based real-time WL measurement methods have received widespread attention but are mainly suitable for specific fields such as flight missions, and their universality

in other fields such as manned vehicles is limited. In addition, although there are some performance-based measurement methods for WL that could be used to monitor changes in operator's WL in real time, the effective rules for setting WL classification thresholds are still lacking. Wickens pointed out that the WL threshold could be objectively defined by an observed slowing in the rate of performance associated with increases in overall task demands [33]. But the current setting of WL threshold is mainly determined based on the subjective experiences of designers, which lead to the lack of commonality [34–36].

To solve the issued mentioned above, this study aims to establish a measurement method for WL to improve the problem of High Workload (HWL) among operators in multitasking environments of manned vehicles. And the WL assessment thresholds were further defined based on the decreasing tendency of performance accuracy associated with increased task demand. This study provides a theoretical and practical basis for the optimal design and adaptive adjustment of human–machine dynamic function allocation, which supports the improvement of operator efficiency and system safety.

2. WL Measurement Method

The operators in the manned vehicle are responsible for monitoring and operating computers and instruments. The differences between this multitasking system and others are that operators must access the human–machine interfaces of each subsystem through separate windows, and each subtask system demands the operator's complete attention. Multitasking systems with a high frequency of human–machine interaction are prone to put operators at HWL and even overload, thus creating safety risks. Therefore, the WL measurement method for operators in the manned vehicle was proposed. This method can help provide a basis for task allocation between humans and machines, and thus reduce the occurrence and duration of HWL states in operators.

2.1. Method Design

Given the limitations of existing methods [37], a new measurement method of WL was proposed based on the concepts of Task Demand Load (TDL) and Subject Load Capacity (SLC). The TDL refers to the workload imposed on operators by monitoring or operating systems [38,39], which is associated with the task attributes (such as task type and task difficulty). Besides the TDL, the workload capacity provided by an operator also varies with the task attributes, especially for multitasking human–machine interaction systems. Therefore, a new concept of SLC was defined in this study that refers to the ability of operators to complete tasks within limited available resources.

On the basis of existing methods for quantifying task complexity, this study proposes the quantization formulas for TDL, SLC, and Average Workload (AWL). The TDL and SLC were calculated considering two dimensions: the task complexity based on entropy theory, and the time factors including the planned completion time and the realized completion time. Based on the above definitions, the WL level of operator during tasks could be measured and classified in terms of thresholds related to performance accuracy.

Figure 1 illustrates the specific implementation process of the WL measurement method, as described below.

Step 1: Multitasking system analysis and task complexity estimation. The interface elements and operational flow were summarized via the system tasks analysis. On this basis, the interface–information interaction, the action control, and the interface control behavior were graphed. The types and numbers of nodes in the above three behavioral graphs were used as inputs of Equation (1), and the four sub-complexities were obtained including actions logic complexity, action size complexity, interface–information interaction complexity, and interface control behavior complexity. The specific methodology is described in Section 2.2.

Step 2: Calculation of AWL based on TDL and SLC. The TDL and SLC were calculated based on task complexity and time factors, and then were used to measure AWL combined

with other relevant factors including task frequency, task period, and actual service time of subjects. The specific methodology is described in Section 2.3.

Step 3: Classification of WL based on defined thresholds. The elevated WL can easily lead to operational errors, which may pose potential safety risks in human-machine systems. Therefore, the WL thresholds could be defined based on the task accuracy. Based on the established thresholds, the AWL measurement results could be classified into different levels. The specific methodology is described in Section 2.4.

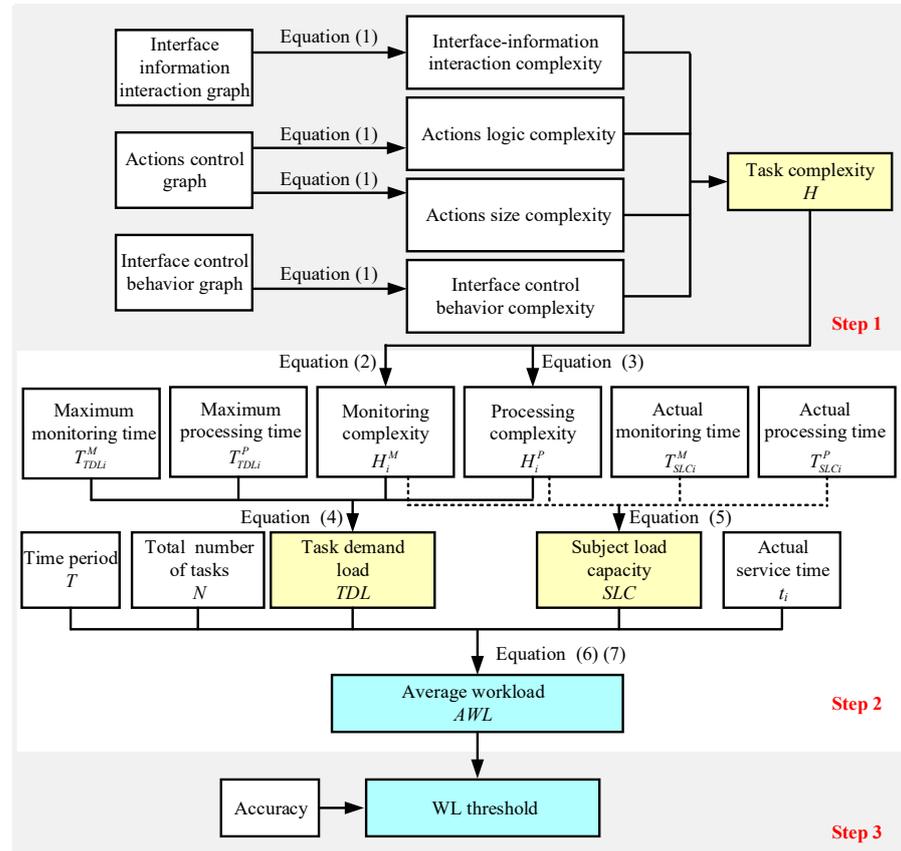


Figure 1. WL measurement method framework.

2.2. Task Complexity

The information theory has been proved feasible and effective in quantitative assessment of task complexity. Shannon introduced entropy as the most important measure into information theory. Entropy is a fundamental measure that includes information, decision-making, and uncertainty, and plays a core role in the field of information theory. To evaluate task complexity, Park et al. [40] and Davis et al. [41] adopted the entropy measurement method within the context of program control graphs. Mowshwitz [42] introduced two different entropy measures for graphical representations: first-order entropy for evaluating the logic of program control diagrams and second-order entropy for evaluating the complexity of program control diagrams. It is important to note that a higher entropy value signifies increased graph complexity. Equation (1) is the formal representation of the entropy measurement method [42]:

$$H = \sum_{i=1}^N p(A_i) \log_2 \frac{1}{p(A_i)}, \tag{1}$$

where H is the entropy value; N is the number of node types; $p(A_i)$ is the probability of the node belonging to class A_i .

Based on the entropy theory, the task complexity for multitasking systems was estimated from four different dimensions: interface–information interaction complexity, actions logic complexity, action size complexity, and interface control behavior complexity [41]. The interface–information interaction complexity refers to the intricacies of information interaction between operational behaviors and HMI, which could be quantified based on the interface–information interaction graph. The actions logic complexity and action size complexity, derived from the actions control graph, elucidate the logical sequence and the volume of requisite actions during task execution [43]. The interface control behavior complexity could be calculated based on the interface control behavior graph.

The operational requirements of the multitasking system can be divided into task monitoring and task processing. Accordingly, the task complexity in this study is divided into the monitoring complexity and the processing complexity. The monitoring complexity is determined using Equation (2) [43]. The processing complexity is determined by Equation (3) [43].

$$H^M = H_{IIC}, \quad (2)$$

$$H^P = \sqrt{H_{IIC}^2 + H_{ALC}^2 + H_{ASC}^2 + H_{ICBC}^2}, \quad (3)$$

where H^M is the monitoring complexity; H^P is the processing complexity; H_{IIC} is the interface–information interaction complexity; H_{ALC} is the actions logic complexity; H_{ASC} is the action size complexity; H_{ICBC} is the interface control behavior complexity.

2.3. AWL Measurement

The TDL describes the inherent attribute of the task, independent of the operator's ability. It is related to factors such as task complexity, task uncertainty, task frequency, and task performance requirements [44].

The existing scholarly literature has substantiated that the TDL can be quantified by assessing various task system parameters, with findings consistently highlighting the salient role of task complexity in influencing TDL relative to other determinants [40]. Furthermore, the TDL is closely intertwined with temporal constraints, commonly referred to as time pressure [45]. Consequently, the TDL in this study is quantified by considering the task complexity and the temporal requirements of the task, as follows:

$$TDL_i = \frac{H_i^M}{T_{TDLi}^M} + \frac{H_i^P}{T_{TDLi}^P}, \quad (4)$$

where TDL_i is the task demand load of task i ; H_i^M is the monitoring complexity of task i ; H_i^P is the processing complexity of task i ; T_{TDLi}^M is the maximum monitoring time of task i ; T_{TDLi}^P is maximum processing time of task i .

Besides the task-related factors, the WL is also influenced by the individual factors [6]. For example, the operators exhibit varied responses when participating in different task processes. Therefore, the SLC, defined in this study, is associated with the task parameters (such as the task complexity) and the operation proficiency (such as the actual operating time in Equation (5)).

$$SLC_i = \frac{H_i^M}{T_{SLCi}^M} + \frac{H_i^P}{T_{SLCi}^P}, \quad (5)$$

where SLC_i is the subject load capacity; H_i^M is the monitoring complexity of task i ; H_i^P is the processing complexity of task i ; T_{SLCi}^M is the actual monitoring time of subject for task i ; T_{SLCi}^P is the actual processing time of subject for task i .

The WL can be viewed as a resource pool. The available resources that operators can activate vary with the task complexity; meanwhile, the requirements for resources are different when performing various tasks. Therefore, this study compares SLC to activated available resources, while TDL is likened to occupied resources. And the WL is defined as

the ratio of occupied resources to activated available resources, with the value range of 0 to 1. Three cases of the instantaneous WL are considered in this study as follows:

1. When the subject is not actively engaged in task processing ($SLC_i = 0$), the instantaneous WL is defined as 0.
2. When SLC matches or exceeds TDL ($TDL_i \leq SLC_i$), the subject's instantaneous WL is calculated as the ratio between the two.
3. When SLC falls short of meeting TDL ($TDL_i > SLC_i$), the upper limit of WL is reached, designated as 1 in this study.

The mathematical expression for instantaneous WL is as follows:

$$WL_i = \begin{cases} 0, & SLC_i = 0 \\ \frac{TDL_i}{SLC_i}, & TDL_i \leq SLC_i, \\ 1, & TDL_i > SLC_i \end{cases} \quad (6)$$

where WL_i is the instantaneous WL at task i ; SLC_i is the subject workload capacity for task i ; TDL_i is the task demand load for task i .

The assessment of instantaneous WL only captures its fluctuations throughout the duration of task. To measure the WL of operator at a specific moment, the AWL is introduced as the accumulated WL per unit time, as described in Equation (7) [45]. The lower AWL in subjects implies a greater store of cognitive resources, thereby enhancing their capacity to effectively manage new tasks.

$$AWL = \frac{1}{T} \sum_{i=1}^N \int_0^{t_i} WL_i dt, \quad (7)$$

where t_i is the service time for task i ; N represents the total number of tasks; T is the duration of task time, $T = 600$ s.

2.4. WL Assessment Threshold

The multitasking systems in the manned vehicle mainly lead to HWL conditions and rarely cause operator underload, which is the main reason for many manned vehicle accidents. The human-machine dynamic functional allocation system closely links intelligent auxiliary support to the operator's WL level, providing automated support when the WL is too high [46]. The AWL measurements are able to assess changes in the real-time WL of the operator. However, in order to further clarify the intervention time for automation in human-machine dynamic function allocation systems, it is important to clarify the category in which the WL is located.

When task demand exceeds resource supply, operators would be in the state of overload, and their performance will sharply decline. The red line in Figure 2 shows the performance breakpoint. This red line divides the supply and demand space into two regions. The left region is the "remaining resources region", and the right region is the "overload region" [47,48]. Therefore, the setting of the red line is crucial for the evaluation of WL.

In order to ensure that operators have sufficient residual capacity to maintain a certain level of work performance in emergency situations, the WL assessment thresholds were defined in this study by determining the maximum allowable level of WL based on performance transition points. Based on these thresholds, the WL state was identified for better guidance of the high-performance allocation strategies of human-machine collaboration.

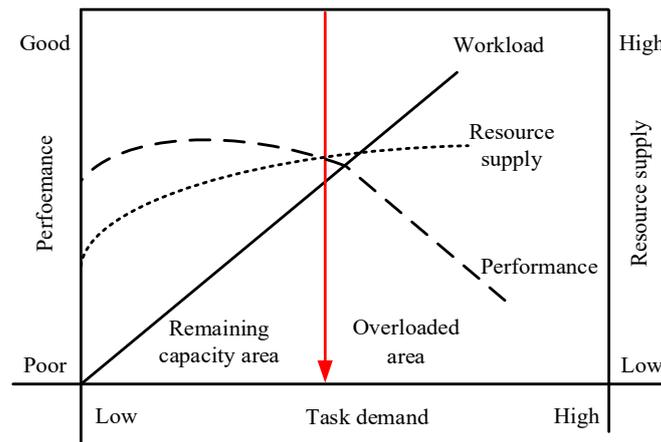


Figure 2. An interpretation of the supply–demand relationship associated with WL and performance, highlighting the redline of overload [48].

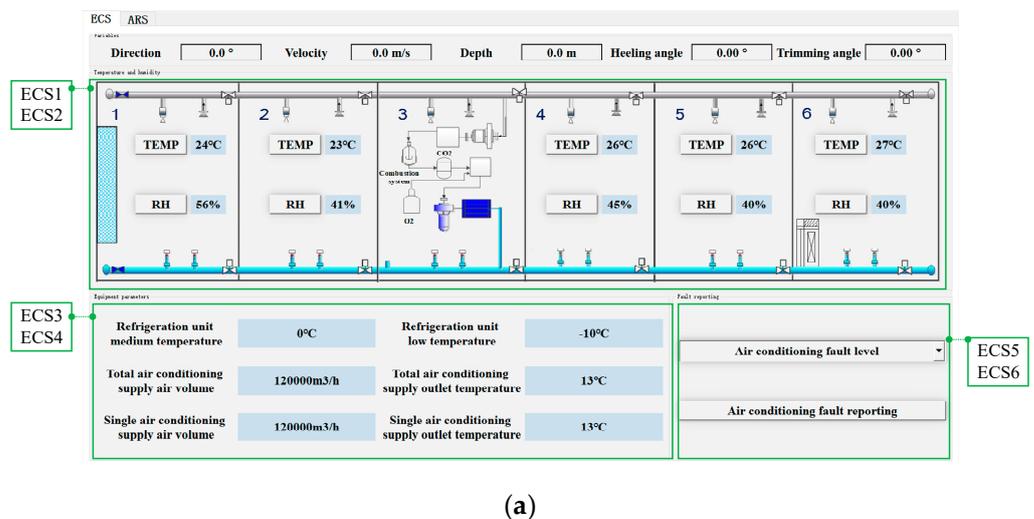
3. Experimental Methods

3.1. Subjects

Forty-five subjects between the ages of 20 and 27 years (23.03 ± 1.50 years) were recruited to participate in the ECAR system experiment. All subjects were physically healthy, right-handed, with normal or corrected vision, and with normal electrocardiograms. Prior to the formal experiment, an 8-h task training was conducted for each subject to ensure that they are fully familiar with the task operation of ECAR system under all conditions. All subjects had a good mental state before or during the experiment. All subjects completed the test without dropping out and received financial compensation after the experiment ended.

3.2. ECAR System Tasks

In the manned vehicle, the operator carries out multi-window tasks through the Environmental Control and Atmospheric Regeneration (ECAR) system. The ECAR system is a task platform developed in PyCharm Community Edition software (version 2020.2.4). It integrates the Environmental Control System (ECS) and the Atmospheric Regeneration System (ARS). The Human Machine Interfaces (HMIs) of the ECAR system are shown in Figure 3.



(a)

Figure 3. Cont.

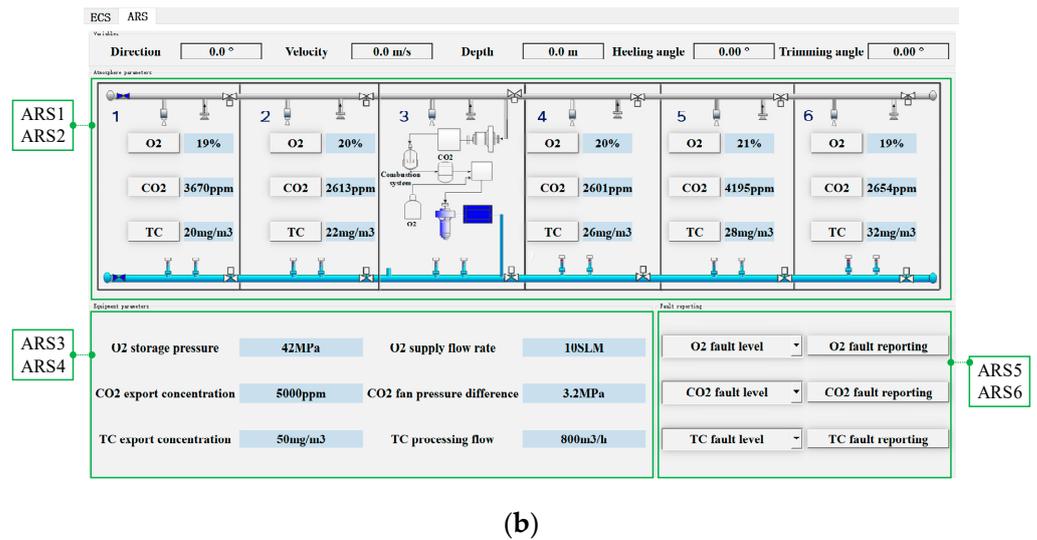


Figure 3. HMIs of the ECAR system. (a) ECS. (b) ARS.

Figure 3a shows the HMI of the ACS. The specific task operation process of the ECS is as follows: firstly, the operator observes the abnormal temperature and humidity parameters (ECS1), and clicks the “TEMP” or “RH” button (ECS2); secondly, the operator observes the abnormal equipment parameters (ECS3) and analyzes the cause of the failure (ECS4); finally, the operator selects the “Air conditioning fault level” (ECS5) and clicks the appropriate “Air conditioning fault reporting” button (ECS6).

Figure 3b shows the HMI of the ARS. The specific task operation process of the ARS is as follows: firstly, the operator observes the abnormal atmospheric parameters (ARS1), and clicks the “O2”, “CO2” or “TC” button (ARS2); secondly, the operator observes the abnormal equipment parameters (ARS3) and analyzes the cause of the failure (ARS4); finally, the operator selects the “Fault level” (ARS5) and clicks the appropriate “Fault reporting” button (ARS6).

3.3. Task Complexity of ECAR System

Taking the ECS as an example, the two types of complexity (monitoring complexity and processing complexity) and the sub complexity (interface–information interaction complexity, actions logic complexity, action size complexity, and interface control behavior complexity) were calculated.

The interface–information interaction complexity is equal to the second-order entropy of the interface–information interaction graph, as shown in Figure 4. In second-order entropy, if the amount of information flowing in and out is consistent, it is a type of node. The number of second-order entropy nodes in Figure 4 is 20. The calculated result is as follows:

$$H_{IIC} = \sum_{i=1}^{20} p(A_i) \log_2 \frac{1}{p(A_i)} = \frac{4}{41} \log_2 \frac{41}{4} + \frac{5}{41} \log_2 \frac{41}{5} + \frac{10}{41} \log_2 \frac{41}{10} + \frac{6}{41} \log_2 \frac{41}{6} + 16 \times \frac{1}{41} \log_2 \frac{41}{1} = 3.691, \quad (8)$$

Figure 5 shows the actions control graph. The actions logic complexity is equal to the first-order entropy of the actions control graph. In first-order entropy, when the elements and quantities of information flowing in and out are consistent, it is a type of node. The number of first-order entropy nodes in Figure 5 is 5. The specific formula is as follows:

$$H_{ALC} = \sum_{i=1}^5 p(A_i) \log_2 \frac{1}{p(A_i)} = \frac{2}{16} \log_2 \frac{16}{2} + \frac{3}{16} \log_2 \frac{16}{3} + \frac{9}{16} \log_2 \frac{16}{9} + 2 \times \frac{1}{16} \log_2 \frac{16}{1} = 1.795, \quad (9)$$

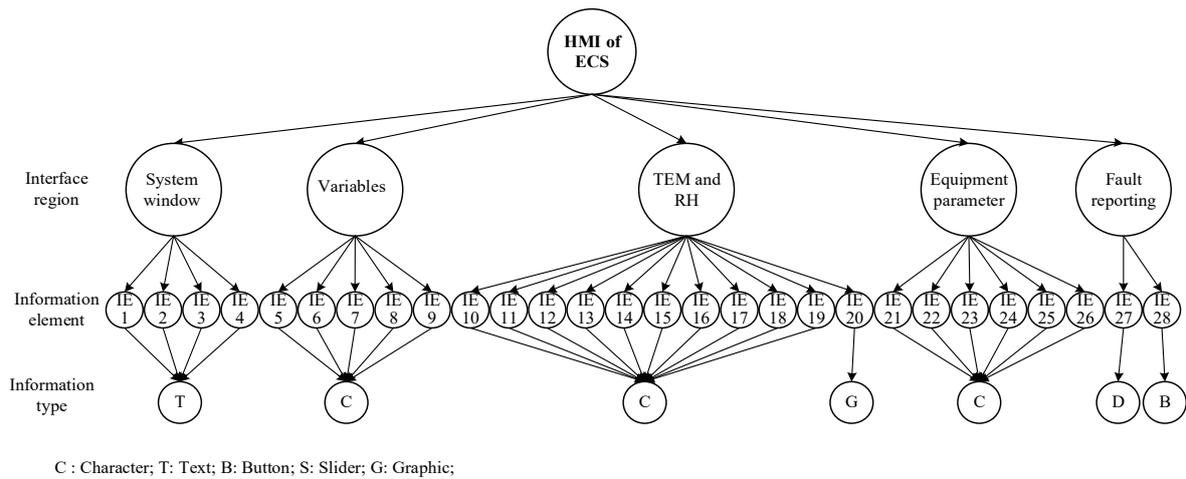


Figure 4. Interface-information interaction graph.

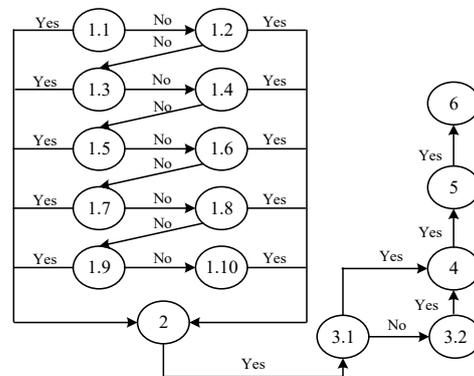


Figure 5. Actions control graph.

The action size complexity is equal to the second-order entropy of the actions control graph. The number of second-order entropy nodes is 16. The action size complexity value is as follows:

$$H_{ASC} = \sum_{i=1}^{16} p(A_i) \log_2 \frac{1}{p(A_i)} = 16 \times \frac{1}{16} \log_2 \frac{16}{1} = 4.000, \quad (10)$$

Similarly, the interface control behavior complexity is equal to the second-order entropy of the information control behavior graph. As shown in Figure 6, the number of second-order entropy nodes is 7. The interface control behavior complexity value is as follows:

$$H_{ICBC} = \sum_{i=1}^7 p(A_i) \log_2 \frac{1}{p(A_i)} = \frac{3}{20} \log_2 \frac{20}{3} + \frac{11}{20} \log_2 \frac{20}{11} + \frac{2}{20} \log_2 \frac{20}{2} + 4 \times \frac{1}{20} \log_2 \frac{20}{1} = 2.081, \quad (11)$$

According to Equations (2) and (3), the monitoring complexity and processing complexity are as follows:

$$H^M = H_{IIC} = 3.691, \quad (12)$$

$$H^P = \sqrt{H_{IIC}^2 + H_{ALC}^2 + H_{ASC}^2 + H_{ICBC}^2} = 6.097 \quad (13)$$

where H^M is the monitoring complexity; H^P is the processing complexity; H_{IIC} is the interface-information interaction complexity; H_{ALC} is the actions logic complexity; H_{ASC} is the action size complexity; H_{ICBC} is the interface control behavior complexity.

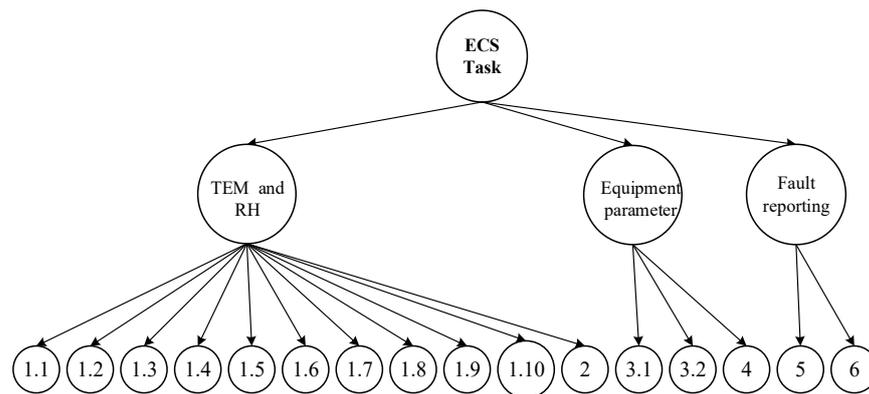


Figure 6. Interface control behavior graph.

The sub-complexity values (H_{IIC} , H_{ALC} , H_{ASC} , and H_{ICBC}) and the complexity type (H^M and H^P) of subtasks in ECAR system were summarized in Table 1.

Table 1. Summary of ECAR system task complexity results.

Task Type	Sub-Complexity Value				Complexity Type	
	H_{IIC}	H_{ALC}	H_{ASC}	H_{ICBC}	H^M	H^P
ECS	3.691	1.795	4.000	2.081	3.691	6.097
ARS	3.579	1.532	4.392	1.814	3.579	6.143

3.4. Experimental Design

In order to verify the effectiveness of the proposed WL measurement method, the experiments, adopting the within-subject design, were conducted with the three WL levels for the ECAR system tasks. The duration of each experimental test was set to 10 min in this study. The three WL levels were designed by setting different trigger numbers of tasks in fixed task duration of 10 min: LWL including eight tasks, MWL including 20 tasks, and HWL including 36 tasks. The dependent variables included the subjective NASA-TLX scale, work performance, and eye movement indicators.

3.5. Experimental Procedure

The experimental procedure is shown in Figure 7. Before the formal experiment began, the researchers carefully explained the experimental process to the subjects. Considering individual differences among the subjects, the researchers made appropriate adjustments to the seat height and distance from the operating computer based on their height. The computer screen used in the experiment was set to 14 inches in size and the screen brightness was controlled at 80%. Throughout the testing phase, the temperature in the laboratory was maintained between 21 °C and 25 °C, the relative humidity was maintained between 20% and 30%, and the working surface illumination was 500 lx.

During the experiments, the subjects were required to perform three groups of ECAR system tasks with different WL levels in the order designed by the Latin square method. The NASA-TLX scale was required to be filled out after completing the tasks at each WL level, and a 10-min break was reserved between adjacent WL levels. This study was approved by the Institutional Review Board (IRB) of Beihang University.

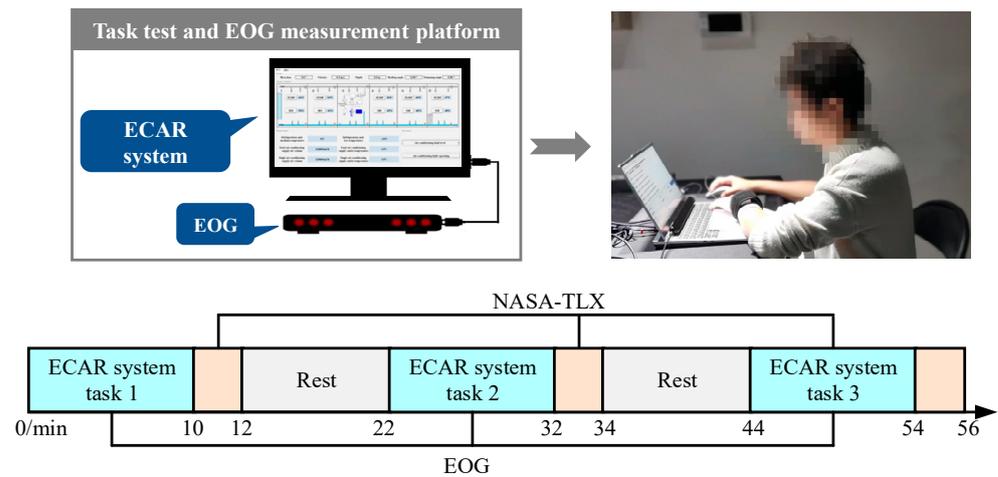


Figure 7. Flow chart of the formal experimental procedure.

3.6. Experimental Data

The experimental data consists of four types: subjective scale, work performance, eye movement, and AWL, as detailed in Table 2.

Table 2. Summary of parameters for this study.

Parameter		Unit
NASA-TLX scale		/
Work performance	Accuracy	%
	Work time percentage	%
Eye movement	Pupil diameter	px
	Fixation duration	ms
	Saccade amplitude	px
	Saccade velocity	px/ms
AWL		/

The NASA-TLX scale was adopted as the subjective tool for evaluating WL from six factors, namely mental demand, physical demand, temporal demand, effort, own performance, and frustration level. The use of the NASA-TLX scale is mainly divided into two parts. Firstly, the subjects rate each dimension based on their feelings and experiences during task execution; secondly, the subjects obtained the weight of each dimension by comparing the importance of the six dimensions [10]. Finally, the overall WL score is calculated by multiplying the weighted scores of each dimension by their scale scores, and then adding these products together. The higher the score is, the higher the WL is.

The actual operational data of ECAR tasks, including task execution accuracy, monitoring time, processing time, task accuracy, and task execution frequency, were automatically recorded by the computer system. The operator’s work time was calculated as the sum of the monitoring time and the processing time. The work time percentage is the ratio of the operator’s work time to the overall task duration (the task duration in this study was 10 min).

The visual attentional behavior of subjects was recorded using the aSee Pro desktop eye tracking device and aSee Studio software (version 0.2.29.3) in order to analyze its relationship with WL. ASee Pro utilizes telemetry-based gaze tracking technology based on human eye 3D models, without the need for fixed heads, and can accurately calculate gaze coordinates within a distance range of 55 cm to 80 cm. The pupil diameter, fixation duration, saccade range, and saccade velocity were used as indicators of eye movement. Among them, the pupil diameter is closely related to the changes of subject’s state and

the way of information processing and is a sensitive indicator for measuring WL [49]. The fixation duration refers to the duration length of the subject's fixation at each fixation point. The longer the fixation duration, the more difficult it is for the subject to process the information [50]. The saccade amplitude refers to the length of the subject's eye span from the end of the previous fixation behavior to the beginning of the next fixation behavior [51]. The more difficult the task, the smaller the saccade amplitude will be. The saccade velocity refers to the distance between the fixation points of the subjects per second [52]. It can indicate the efficiency of the subject's search for information in the current visual area, and the faster saccade velocity represents the more efficient the subject's search for information [52].

The AWL is quantitatively obtained based on the task complexity, maximum allowable time, actual work time, and task frequency. The detailed calculation process of task complexity is described in Section 3.2. The maximum allowable monitoring time for each subtask in the ECAR system is 5 s, and the maximum allowable processing time is 20 s. The actual monitoring time, the actual processing time, and the task frequency of each subtask in the ECAR system are recorded by the computer.

The repeated measures analysis of variance (ANOVA) was performed using the SPSS 25.0 to test the significance between all measurements. Given that the ANOVA necessitates adherence to normality and homogeneity of variance assumptions for all measures, the Shapiro-Wilk test was initially employed in this study to assess normality. Unfortunately, the examination revealed that none of the dependent variables met the criteria for normality. Subsequently, the one-way ANOVA was applied to assess homogeneity of variance, revealing that only NASA-TLX results demonstrated homogeneity. Considering the non-compliance with normality and homogeneity of variance, the Kruskal–Wallis test was adopted to examine the impact of WL levels on all measures. Additionally, the Spearman correlation analysis was performed using the open-source statistical software package R 3.6.1 to investigate the correlation between AWL measurement results and subjective NASA-TLX scales, work performance, and eye movement indicators. Statistical analysis was performed using $\alpha = 0.05$.

4. Results

4.1. WL Measurement Results

Table 3 presents the measurement results under different WL conditions. Figure 8 shows the measurements under three different WL conditions including the calculated AWL, NASA-TLX score, task accuracy, work time percentage, pupil diameter, fixation duration, saccade amplitude, and saccade velocity.

As shown in Figure 8a,b, the AWL measurements and NASA-TLX scores increased significantly with the increase of designed WL levels, indicating the sensitivity of the three designed WL levels and the effectiveness of the AWL measurement method.

As shown in Figure 8c,d, as the designed WL levels increased, the task accuracy decreases significantly, and meanwhile the work time percentage increased significantly. These results illustrated that an increase in the WL level caused subjects to spend more time processing tasks, while further contributing to subject errors.

Figure 8e–h depict the changes in eye movement of subjects when performing the tasks with different WL levels. The statistical results indicated that there was no significant effect of WL level on pupil diameter and saccade velocity. Moreover, the fixation duration was significantly higher at LWL than that at HWL, and the saccade amplitude under LWL was significantly lower than that under MWL and HWL. These results indicated that tasks with HWL level can lead to shorter fixation duration for subjects and force them to process more information in one fixation.

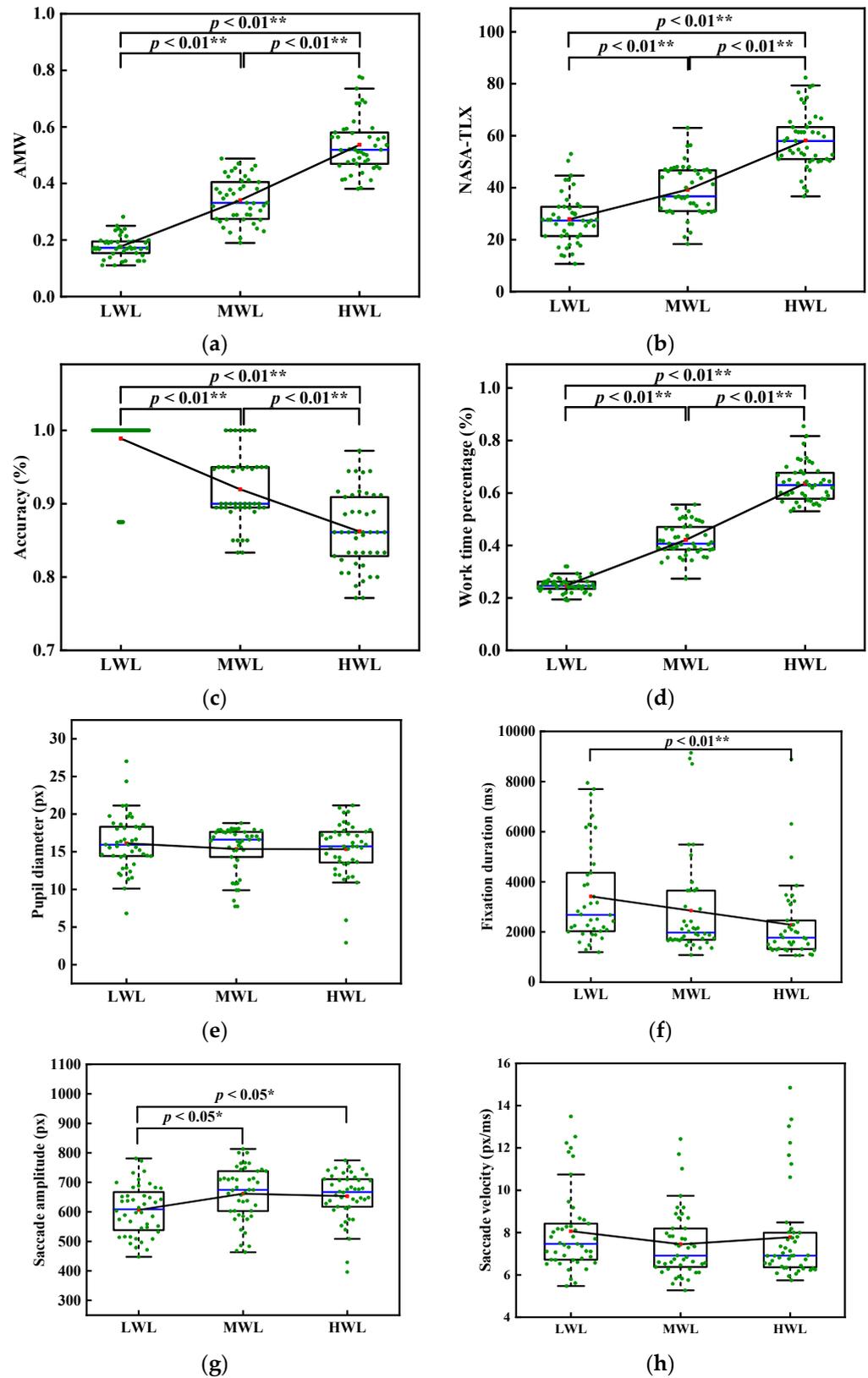


Figure 8. Measurement indicators under three different WL conditions. (a) AMW. (b) NASA-TLX. (c) Accuracy. (d) Work time percentage. (e) Pupil diameter. (f) Fixation duration. (g) Saccade amplitude. (h) Saccade velocity. In the box plot, green dots represent all data points, red dots represent the mean, and blue lines represent the median. The black line at the bottom of the box

indicates that 25% of the data is below this line, while the black line at the top of the box indicates that 75% of the data is below this line. The black whisker line at the bottom of the box extends to the minimum value of the data, while the black whisker line at the top of the box extends to the maximum value of the data (excluding outliers), ** ($p < 0.01$), * ($p < 0.05$).

Table 3. Measurement results under three different WL conditions.

Results	WL (Mean \pm SD)		
	LWL	MWL	HWL
AMW	0.177 \pm 0.038	0.341 \pm 0.077	0.537 \pm 0.098
NASA-TLX	27.852 \pm 9.449	39.237 \pm 9.579	58.193 \pm 10.732
Accuracy	0.989 \pm 0.036	0.920 \pm 0.047	0.862 \pm 0.050
Work time percentage	0.248 \pm 0.029	0.421 \pm 0.062	0.637 \pm 0.073
Pupil diameter	16.118 \pm 3.628	15.371 \pm 3.120	15.339 \pm 3.552
Fixation duration	3363.396 \pm 1883.851	2852.315 \pm 1999.529	2301.779 \pm 1525.657
Saccade amplitude	616.676 \pm 105.774	661.453 \pm 90.958	653.056 \pm 80.573
Saccade velocity	8.064 \pm 1.942	7.451 \pm 1.571	7.782 \pm 2.184

The above measurement results not only validated the effectiveness of the AWL measurement method proposed in this study, but also revealed the effects of changes in the WL level on the work performance and eye movement indexes.

4.2. WL Correlation Analysis

Figure 9 shows the correlation analysis results of all measurements during the processing of the ECAR system task. The absolute value of the correlation coefficient is regarded as extremely weak correlation between 0 and 0.1, weak correlation between 0.1 and 0.3, moderate correlation between 0.3 and 0.5, strong correlation between 0.5 and 0.7, and extremely strong correlation between 0.7 and 1.

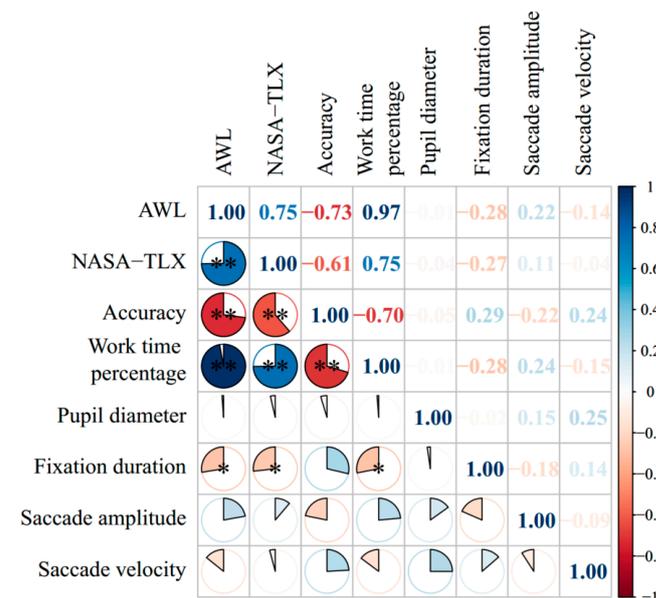


Figure 9. The correlation matrix heatmap between various WL measurement indicators during the ECAR system task. ** ($p < 0.01$), * ($p < 0.05$).

The correlation analysis results showed that there were significant and very strong positive correlations between the AWL measurements and the NASA-TLX scores ($r = 0.75$, $p < 0.01$) and work time percentage ($r = 0.97$, $p < 0.01$). A significant and very strong negative correlation was found between the AWL measurements and the task accuracy

($r = -0.73, p < 0.01$). And the significant and weak correlations were found between the AWL measurements and the fixation duration ($r = -0.28, p < 0.05$).

Based on the correlation between the AWL measurement results and the NASA-TLX scale, task accuracy, and working time percentage, it is proven that the AWL measurement results are effective in evaluating the operator's WL level. The larger the AWL measurement results, the higher the operator's WL level.

4.3. WL Assessment Threshold

The WL assessment thresholds were defined by integrating subjects' task accuracy with the calculated AWL measurements. In Figure 10, the solid black line represented the task accuracy, and the black dashed line denoted the fitted trend line for task accuracy. Similarly, the solid red line represented the AWL measurements calculated by the method proposed in this study, and the red dashed line indicated the fitted trend line for AWL measurements. Notably, as the designed WL level increased, the AWL measurements simultaneously increased, and the task execution accuracy progressively declined.

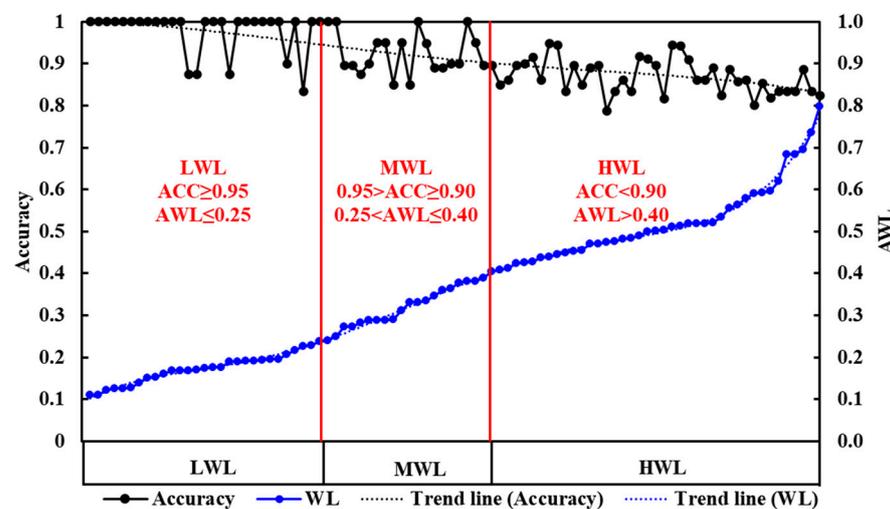


Figure 10. Determination of WL assessment thresholds.

When the trend line of task accuracy reached or exceeded 0.95, most subjects attained 100% task accuracy, signifying optimal performance of the human-machine system. At this point, the corresponding AWL measurement of 0.25 was designated as the LWL threshold. The performance of the human-machine system was deemed satisfactory when the trend line of task accuracy ranged between 0.90 and 1. However, the security of the human-machine system was threatened when the fitted trend line of task execution accuracy fell below 0.90. Consequently, the AWL value of 0.4, corresponding to the task execution accuracy of 0.9, was designated as the HWL threshold.

To summarize, based on the AWL measurement calculated using the method proposed in this study, the subject's WL state can be classified into different level. Specifically, subjects were regarded as being in the LWL state when the AWL measurement was below 0.25, in the MWL state when the AWL measurement was between 0.25 and 0.40, and in the HWL state when the AWL measurement surpassed 0.40.

5. Discussion

5.1. Comparison of WL Measurements

A measurement method of WL was proposed in this study based on TDL and SLC. To validate the effectiveness of proposed method, the experiments were conducted using the ECAR system task with three WL levels. The different WL levels were designed by setting different task departure frequencies. In the experiments, subjects were asked to fill out the NASA-TLX scale immediately after completing each ECAR system task. The results of the

NASA-TLX scale showed that the subjective WL increased significantly with the increase of the WL level of the ECAR system task, which justified the different WL levels induced by the task design [53].

The results of the subjects' work performance showed that the task accuracy decreased significantly with the increase of the designed WL level, and the work time percentage increased significantly with the increase of the designed WL level. These results are consistent with the three-stage model proposed by Meister [54].

In this study, the fixation duration decreased significantly as the designed WL level increased. However, de Greef et al. [49] found that fixation duration increased with the WL level. This inconsistency may be due to differences in study designs. Specifically, the WL levels in this study were designed by changing the task frequency, and high task frequency decreased the fixation duration of the operators. However, de Greef et al. increased the WL level by increasing the task difficulty. They found that high task difficulty causes difficulty in information processing for the operators, which increased the fixation duration. Besides, in this study, the saccade amplitude was found to be highest under MWL, which is consistent with the findings of Fan et al. [55].

AWL measurements were calculated for subjects at the three designed WL levels. The results showed that the AWL measurements increased significantly with the designed WL levels. In addition, the AWL measurements showed significant and very strong positive correlations with the NASA-TLX scores and the work time percentage. The AWL measurements showed significant and very strong negative correlations with task accuracy. These results suggest that the WL measurement method proposed in this study can effectively assess the WL level of operators, and the higher the AWL measurements, the higher the WL level of operators.

In conclusion, the above measures are effective in assessing the WL level of the subjects. However, the subjective scales do not enable real-time assessment, and the physiological measurements require high working conditions and are prone to disturbing the subjects [56,57]. In contrast, the WL measurement method proposed in this study can assess the operator's WL level in real time without interrupting task performance.

5.2. Real-Time WL Assessment

Real-time assessment of the operator's WL state enables the determination of whether the automation assistance functions are turned on or off during human–computer collaboration. Therefore, in order to get a better understanding of the operator's WL state, it is necessary to determine the WL threshold. The setting of the WL threshold determines the frequency of changes in the automation level. Higher WL thresholds lead to untimely automation interventions, while lower WL thresholds lead to more frequent automation interventions [58]. Wickens et al. [59] and Park et al. [60] have used the point at which task performance begins to decline as the WL threshold. This performance-guided WL threshold not only balances the operator's WL level, but also optimizes the performance of the human–computer interaction system. In this study, task execution accuracy was used to determine the LWL threshold (0.25) and HWL threshold (0.4). Based on the established WL thresholds, the AWL measurements were divided into three levels.

In order to validate the real-time WL measurement method proposed in this study and the established thresholds, Subject 1 underwent a 40-min ECAR system task again. The task was divided into three phases: the first phase was a 10-min LWL task; the second phase was a 20-min HWL task; and the third phase was a 10-min LWL task. During the experiment, the subjects' AWL measurements were recorded every 2 min. Figure 11 shows the AWL measurement results of Subject 1 during the completion of the 40-min ECAR system task. As shown in Figure 11, the subject was always in LWL during the first phase. However, during the second phase, the subject was in the HWL state for 80% of the time and in the MWL state for the rest of the time. This may be due to the fact that the work time percentage of Subject 1 was less than 50% during the time periods of 18–20 min and 26–28 min. During the third phase of the LWL task, subjects spent 80% of their time in the

LWL state and the rest of the time in the MWL state. This may be due to the fact that in the time period 32–34 min, the work time percentage of Subject 1 was higher than 30%.

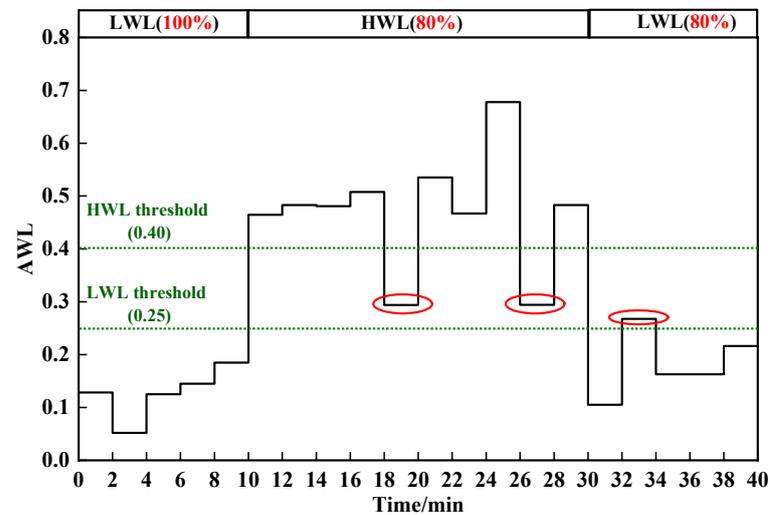


Figure 11. Subject 1's AWL measurements during completion of the 40-min ECAR system task. (Note: Red circles indicate incorrect classification of mental workload.)

The results of this study show that the proposed WL measurement method is not only capable of assessing the operator's WL level in real time. This method shows potential for application in human-machine dynamic function allocation systems. Based on the continuous monitoring of WL level, the optimal task assignment strategy could be autonomously triggered by the intelligent system to remain the operator at an appropriate WL state, which helps optimize the performance of the human-machine system.

5.3. Limitation

There are several limitations of this study, which are as follows: (1) Physiological measurements were limited to eye movement parameters. Future WL studies should incorporate multiple physiological indices (EEG, ECG, and EMG) for the comprehensive assessment of WL in subjects; (2) The ECAR tasks were conducted in a laboratory setting. Results from these simulated tasks could not fully mirror the dynamics of actual manned vehicle operational tasks. A more extensive investigation involving real-world manned vehicle operation tasks is recommended for future studies; (3) The subjects in this study were limited to healthy male college students, which may limit the generalizability of the findings to real manned vehicle operators. Future studies should aim to include a larger group of subjects, especially including real manned vehicle operators, to improve the applicability of the findings.

6. Conclusions

In the manned vehicle, the multitasking human-machine interaction system often leads to HWL for operators. The WL measurement method was proposed for multitasking processing based on TDL and SLC, in response to the limitations of traditional WL measurement methods in human-machine collaboration systems. To evaluate the effectiveness of this method, the ECAR system tasks with varying WL levels were designed, and 45 subjects were engaged in the experimental study. The main conclusions are summarized as follows:

1. In the measurement results of the AWL and subjective NASA-TLX scale, work time percentage increased significantly with the increase of WL level. The measurement results of accuracy and fixation duration decreased significantly with the increase of WL level.

2. There were very strong correlations existed between AWL measurements and NASA-TLX results ($r = 0.75, p < 0.01$), task execution accuracy ($r = -0.73, p < 0.01$), and work time percentage ($r = 0.97, p < 0.01$), respectively.
3. Thresholds for WL levels were established based on task execution accuracy. The thresholds for LWL and HWL were 0.25 and 0.40, respectively. Based on the established WL thresholds, this study evaluated the AWL measurements for three categories.

In conclusion, the results indicate that the AWL measurement method proposed in this study can effectively assess the WL of operators in real time, and can also automatically trigger the optimal task allocation strategy in human-machine dynamic function allocation system to keep operators in the appropriate WL state, thus optimizing the performance of the human-machine system.

Author Contributions: Conceptualization, C.Y., L.P. and J.Z.; methodology, C.Y.; software, C.Y.; validation, C.Y., L.P. and J.Z.; formal analysis, C.Y.; investigation, C.Y., L.P. and X.C.; resources, L.P. and J.Z.; data curation, C.Y.; writing—original draft preparation, C.Y.; writing—review and editing, L.P. and J.Z.; visualization, C.Y.; supervision, L.P. and J.Z.; project administration, L.P. and X.C.; funding acquisition, L.P. and X.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by 2022-JCJQ-JJ-0831, the National Natural Science Foundation of China (52008014) and the Fundamental Research Funds for the Central Universities (YWF-22-L-1009).

Data Availability Statement: Due to the limitations of the Ethics Review Committee, these data cannot be made public to protect the privacy and confidential information of the subjects. The data presented in this study are available upon request from the corresponding author.

Acknowledgments: The authors would like to thank all subjects involved in this study.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Hancock, P.A.; Desmond, P.A. *Stress, Workload, and Fatigue*; Lawrence Erlbaum Associates Publishers: Mahwah, NJ, USA, 2009.
2. Wickens, C.D. Mental Workload: Assessment, Prediction and Consequences. In *International Symposium on Human Mental Workload*; Springer International Publishing: Berlin/Heidelberg, Germany, 2017.
3. Wilson, G.F.; Russell, C.A. Real-Time Assessment of Mental Workload Using Psychophysiological Measures and Artificial Neural Networks. *Hum. Factors* **2003**, *45*, 635–644. [\[CrossRef\]](#)
4. Moray, N. *Mental Workload: Its Theory and Measurement*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2013.
5. Boff, K.R.; Kaufman, L.; Thomas, J.P. *Handbook of Perception and Human Performance*; Wiley: New York, NY, USA, 1986; Volume 1.
6. Brookhuis, K.A.; van Driel, C.J.; Hof, T.; van Arem, B.; Hoedemaeker, M. Driving with a Congestion Assistant; mental workload and acceptance. *Appl. Ergon.* **2009**, *40*, 1019–1025. [\[CrossRef\]](#)
7. Verwey, W.B. On-line driver workload estimation. Effects of road situation and age on secondary task measures. *Ergonomics* **2000**, *43*, 187–209. [\[CrossRef\]](#)
8. da Silva, F.P. Mental workload, task demand and driving performance: What relation. *Procedia-Soc. Behav. Sci.* **2014**, *162*, 310–319. [\[CrossRef\]](#)
9. Wang, P.; Fang, W.; Guo, B. A measure of mental workload during multitasking: Using performance-based Timed Petri Nets. *Int. J. Ind. Ergon.* **2020**, *75*, 102877. [\[CrossRef\]](#)
10. Rubio, S.; Díaz, E.; Martín, J.; Puente, J.M. Evaluation of subjective mental workload: A comparison of SWAT, NASA-TLX, and workload profile methods. *Appl. Psychol. Int. Rev.* **2004**, *531*, 61–86. [\[CrossRef\]](#)
11. Tao, D.; Tan, H.; Wang, H.; Zhang, X.; Qu, X.; Zhang, T. A Systematic Review of Physiological Measures of Mental Workload. *Int. J. Environ. Res. Public Health* **2019**, *16*, 2716. [\[CrossRef\]](#)
12. Hart, S.G.; Staveland, L.E. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. *Adv. Psychol.* **1988**, *52*, 139–183.
13. Roscoe, A.H.; Ellis, G.A. *A Subjective Rating Scale for Assessing Pilot Workload in Flight: A Decade of Practical Use*; Royal Aerospace Establishment: Bedford, UK, 1990.
14. Wierwille, W.W.; Casali, J.G. A Validated Rating Scale for Global Mental Workload Measurement Applications. *Proc. Hum. Factors Soc. Annu. Meet.* **1983**, *27*, 129–133. [\[CrossRef\]](#)
15. Zheng, B.; Jiang, X.; Tien, G.; Meneghetti, A.; Panton, O.N.; Atkins, M.S. Workload assessment of surgeons: Correlation between NASA TLX and blinks. *Surg. Endosc.* **2012**, *26*, 2746–2750. [\[CrossRef\]](#)
16. Astin, A.; Nussbaum, M.A. Interactive Effects of Physical and Mental Workload on Subjective Workload Assessment. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* **2002**, *46*, 1100–1104. [\[CrossRef\]](#)

17. Umair, M.; Sharafat, A.; Lee, D.; Seo, J. Impact of Virtual Reality-Based Design Review System on User's Performance and Cognitive Behavior for Building Design Review Tasks. *Appl. Sci.* **2022**, *12*, 7249. [[CrossRef](#)]
18. Hart, S.G. Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* **2006**, *50*, 904–908. [[CrossRef](#)]
19. Lin, C.J.; Yenn, T.; Jou, Y.; Hsieh, T.; Yang, C. Analyzing the staffing and workload in the main control room of the advanced nuclear power plant from the human information processing perspective. *Saf. Sci.* **2013**, *57*, 161–168. [[CrossRef](#)]
20. Paxion, J.; Galy, E.; Berthelon, C. Mental workload and driving. *Front. Psychol.* **2014**, *5*, 1344. [[CrossRef](#)] [[PubMed](#)]
21. Braarud, P.Ø. Investigating the validity of subjective workload rating (NASA TLX) and subjective situation awareness rating (SART) for cognitively complex human-machine work. *Int. J. Ind. Ergon.* **2021**, *86*, 103233. [[CrossRef](#)]
22. Hertzum, M.; Holmegaard, K.D. Perceived Time as a Measure of Mental Workload: Effects of Time Constraints and Task Success. *Int. J. Hum.-Comput. Interact.* **2013**, *29*, 26–39. [[CrossRef](#)]
23. Marquart, G.; Cabrall, C.D.; Winter, J.D. Review of eye-related measures of drivers' mental workload. *Procedia Manuf.* **2015**, *3*, 2854–2861. [[CrossRef](#)]
24. Charles, R.; Nixon, J. Measuring mental workload using physiological measures: A systematic review. *Appl. Ergon.* **2019**, *74*, 221–232. [[CrossRef](#)]
25. Zhang, J.; Yin, Z.; Wang, R. Design of an Adaptive Human-Machine System Based on Dynamical Pattern Recognition of Cognitive Task-Load. *Front. Neurosci.* **2017**, *11*, 129. [[CrossRef](#)]
26. Haarmann, A.; Boucsein, W.; Schaefer, F. Combining electrodermal responses and cardiovascular measures for probing adaptive automation during simulated flight. *Appl. Ergon.* **2009**, *40*, 1026–1040. [[CrossRef](#)] [[PubMed](#)]
27. Harriott, C.E.; Zhang, T.; Adams, J.A. Assessing physical workload for human-robot peer-based teams. *Int. J. Hum.-Comput. Stud.* **2013**, *71*, 821–837. [[CrossRef](#)]
28. Wei, Z.; Zhuang, D.; Wanyan, X.; Liu, C.; Zhuang, H. A model for discrimination and prediction of mental workload of aircraft cockpit display interface. *Chin. J. Aeronaut.* **2014**, *27*, 1070–1077. [[CrossRef](#)]
29. Tobaruela, G.; Schuster, W.; Majumdar, A.; Ochieng, W.Y.; Martinez, L.; Hendrickx, P. A method to estimate air traffic controller mental workload based on traffic clearances. *J. Air Transp. Manag.* **2014**, *39*, 59–71. [[CrossRef](#)]
30. Debusk, H.; Hill, C.M.; Chander, H.; Knight, A.C.; Babski-Reeves, K. Influence of military workload and footwear on static and dynamic balance performance. *Int. J. Ind. Ergon.* **2018**, *64*, 51–58. [[CrossRef](#)]
31. Liu, C.; Wanyan, X.; Xiao, X.; Zhao, J.; Duan, Y. Pilots' mental workload prediction based on timeline analysis. *Technol. Health Care Off. J. Eur. Soc. Eng. Med.* **2020**, *28*, 207–216. [[CrossRef](#)] [[PubMed](#)]
32. Park, S.; Jeong, S.; Myung, R. Modeling of multiple sources of workload and time pressure effect with ACT-R. *Int. J. Ind. Ergon.* **2017**, *63*, 37–48. [[CrossRef](#)]
33. Grier, R.; Wickens, C.; Kaber, D.; Strayer, D.; Boehm-Davis, D.; Trafton, J.G.; St. John, M. The Red-Line of Workload: Theory, Research, and Design. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* **2008**, *52*, 1204–1208. [[CrossRef](#)]
34. Rusnock, C.F.; Borghetti, B.J. Workload profiles: A continuous measure of mental workload. *Int. J. Ind. Ergon.* **2016**, *63*, 49–64. [[CrossRef](#)]
35. Mitchell, D.K. *Workload Analysis of the Crew of the Abrams V2 SEP: Phase I Baseline IMPRINT Model*; ARL: Adelphi, MD, USA, 2009.
36. Colombi, J.M.; Miller, M.E.; Schneider, M.F.; McGrogan, J.; Long, D.S.; Plaga, J. Predictive mental workload modeling for semiautonomous system design: Implications for systems of systems. *Syst. Eng.* **2012**, *15*, 448–460. [[CrossRef](#)]
37. Cain, B. *A Review of the Mental Workload Literature*; Defence Research and Development Toronto: Toronto, ON, Canada, 2007; Jul. Report No.: ADA474193.
38. Wei, Z.-G.; Macwan, A.P.; Wieringa, P.A. A Quantitative Measure for Degree of Automation and Its Relation to System Performance and Mental Load. *Hum. Factors* **1998**, *40*, 277–295. [[CrossRef](#)] [[PubMed](#)]
39. Heiligers, M.M.; Van Holten, T.; Mulder, M. Predicting pilot task demand load during final approach. *Int. J. Aviat. Psychol.* **2009**, *19*, 391–416. [[CrossRef](#)]
40. Park, J.; Jung, W.; Ha, J. Development of the step complexity measure for emergency operating procedures using entropy concepts. *Reliab. Eng. Syst. Saf.* **2001**, *71*, 115–130. [[CrossRef](#)]
41. Davis, J.S.; LeBlanc, R.J. A Study of the Applicability of Complexity Measures. *IEEE Trans. Softw. Eng.* **1988**, *14*, 1366–1372. [[CrossRef](#)]
42. Mowshowitz, A. Entropy and the complexity of graphs: I. an index of the relative complexity of a graph. *Bull. Math. Biophys.* **1968**, *30*, 175–204. [[CrossRef](#)]
43. Zheng, Y.; Lu, Y.; Wang, Z.; Huang, D.; Fu, S. Developing a Measurement for Task Complexity in Flight. *Aerosp. Med. Hum. Perform.* **2015**, *868*, 698–704. [[CrossRef](#)]
44. Bi, S.; Salvendy, G. Analytical modeling and experimental study of human workload in scheduling of advanced manufacturing systems. *Int. J. Hum. Factors Manuf.* **1994**, *4*, 205–234. [[CrossRef](#)]
45. Xie, B.; Salvendy, G. Prediction of Mental Workload in Single and Multiple Tasks Environments. *Int. J. Cogn. Ergon.* **2000**, *4*, 213–242. [[CrossRef](#)]
46. Schulte, A.; Donath, D.; Honecker, F. Human-System Interaction Analysis for Military Pilot Activity and Mental Workload Determination. *IEEE Int. Conf. Syst. Man Cybern.* **2015**, *9*, 1375–1380.

47. Wickens, C.D.; Gordon, S.E.; Liu, Y. *An Introduction to Human Factors Engineering*; Addison Wesley Longman Inc.: New York, NY, USA, 1998.
48. Young, M.S.; Brookhuis, K.A.; Wickens, C.D.; Hancock, P.A. State of science: Mental workload in ergonomics. *Ergonomics* **2015**, *58*, 1–17. [[CrossRef](#)]
49. de Greef, T.; Lafeber, H.; van Oostendorp, H.; Lindenberg, J. Eye movement as indicators of mental workload to trigger adaptive automation. In Proceedings of the 5th International Conference on Foundations of Augmented Cognition, Neuroergonomics and Operational Neuroscience: Held as Part of HCI International 2009 (FAC'09), San Diego, CA, USA, 19–24 July 2009.
50. Hooge, I.T.C.; Erkelens, C.J. Adjustment of fixation duration in visual search. *Vis. Res.* **1998**, *38*, 1295–IN4. [[CrossRef](#)] [[PubMed](#)]
51. Yuan, H. Research on the Measurement Method of Regulatory Workload Based on Human-Computer Interaction. Master's Thesis, Civil Aviation University of China, Tianjin, China, 2015. (In Chinese)
52. Zhu, G. Research on Eye Movement Characteristics during Simulated Approach Flight. Master's Thesis, Civil Aviation University of China, Tianjin, China, 2019. (In Chinese)
53. Zhang, J.; Pang, L.; Cao, X.; Wanyan, X.; Wang, X.; Liang, J.; Zhang, L. The effects of elevated carbon dioxide concentration and mental workload on task performance in an enclosed environmental chamber. *Build. Environ.* **2020**, *178*, 106938. [[CrossRef](#)]
54. Meister, D. *Behavioral Foundations of System Development*; Wiley: New York, NY, USA, 1976.
55. Fan, Y.; Liang, J.; Cao, X.; Pang, L.; Zhang, J. Effects of Noise Exposure and Mental Workload on Physiological Responses during Task Execution. *Int. J. Environ. Res. Public Health* **2022**, *19*, 12434. [[CrossRef](#)] [[PubMed](#)]
56. Sriranga, A.K.; Lu, Q.; Birrell, S.A. A Systematic Review of In-Vehicle Physiological Indices and Sensor Technology for Driver Mental Workload Monitoring. *Sensors* **2023**, *23*, 2214. [[CrossRef](#)] [[PubMed](#)]
57. Kramer, A.F. Physiological metrics of mental workload: A review of recent progress. *Mult.-Task Perform.* **1990**, *24*, 279–328.
58. IJtsma, M.; Borst, C.; van Paassen, M.M.; Mulder, M. Evaluation of a Decision-Based Invocation Strategy for Adaptive Support for Air Traffic Control. *IEEE Trans. Hum.-Mach. Syst.* **2022**, *52*, 1135–1146. [[CrossRef](#)]
59. Wickens, C.D.; Tsang, P. Workload. In *Handbook of Human-Systems Integration*; Durso, F., Ed.; American Psychological Association: Washington, DC, USA, 2014.
60. Parks, D.L.; Boucek, G.P. *Workload Prediction, Diagnosis, and Continuing Challenges*; Springer: Boston, MA, USA, 1989.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.