Taylor & Francis
Taylor & Francis Group

International Journal of Human–Computer Interaction

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/hihc20

Impact Analysis of Situation Awareness Based on a Multitasking Difficulty Quantitative Model

Xing Pan, Liuwang Sun, Song Ding & Dujun Zuo

To cite this article: Xing Pan, Liuwang Sun, Song Ding & Dujun Zuo (2024) Impact Analysis of Situation Awareness Based on a Multitasking Difficulty Quantitative Model, International Journal of Human–Computer Interaction, 40:17, 4758-4771, DOI: 10.1080/10447318.2023.2221601

To link to this article: https://doi.org/10.1080/10447318.2023.2221601

Published online: 25 Jun 2023.

Submit your article to this journal

Article views: 397

View related articles

View Crossmark data

Citing articles: 1 View citing articles

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=hihc20

INTERNATIONAL JOURNAL OF HUMAN–COMPUTER INTERACTION
2024, VOL. 40, NO. 17, 4758–4771
https://doi.org/10.1080/10447318.2023.2221601

Taylor & Francis
Taylor & Francis Group

Check for updates

Impact Analysis of Situation Awareness Based on a Multitasking Difficulty Quantitative Model

Xing Pan[a], Liuwang Sun[a], Song Ding[a], and Dujun Zuo[b]

[a]School of Reliability & Systems Engineering, Beihang University, Beijing, China; [b]Institute of System Engineering, China Academy of Engineering Physics, Mianyang, China

ABSTRACT

Situation awareness (SA) is a key indicator of operator task performance and behavioral safety in Human-Machine Interaction (HMI), and task difficulty is an important influencing factors in establishing and maintaining SA, especially in safety-critical scenarios such as flying and nuclear. However, the complexity and diversity of multitasking make it difficult to quantify task difficulty, which to a certain extent limits the research on cognitive research. To solve this problem, we proposed an approach to quantifying task difficulty in multitasking and analyzed the effect of task difficulty on different levels of SA through experiments in this paper. First, a task difficulty measure for a single meta-task was developed based on Shannon's information theory, including both discrete and continuous meta-tasks. Next, a three-dimensional attribute model of meta-tasks was proposed to measure the difficulty value added by concurrent tasks based on multiple resource theory. Finally, an experiment was conducted to measure SA at different levels based on SAGAT, and SA under different levels of task difficulty was analyzed. The results showed a strong negative correlation between SA and task difficulty. Specifically, compared to SA at the comprehension and prediction levels, SA at the perceptual level was more easily influenced by task difficulty. This study can provide some reference for the quantify of task difficulty in human factors experiments, the optimization of tasks in HMI, and the selection of operator attention allocation strategies.

KEYWORDS

Task difficulty; situation awareness; human-machine interaction; multitasking

1. Introduction

With the rapid development of technology, operators are required to make effective and immediate decisions in the face of increasingly complex operating systems (Lee et al., 2012; She & Li, 2017). A prerequisite factor for effective decision making and task performance is situation awareness (SA) (Endsley, 1995b), which is used in human factors research to explain the extent to which operators of complex human-machine systems are aware of what is happening in the system and the environment (Li et al., 2019). In safety-critical systems, incorrect establishment of SA could lead to human factor errors that may not only reduce task performance but even bring catastrophic consequences. Especially in the field of aviation, the level of SA of pilots is closely related to their operational safety. Endsley found that the causes of 88% of commercial aviation accidents caused by human errors have some connections with loss of SA (Endsley, 1995a). In several cases, including the accidents of AirAsia 8501 on December 28 2014 and Renaissance Air 235 on February 4 2015, SA errors are considered to be the most important contributing factor (Kharoufah et al., 2018). Some scholars even argue that SA is a direct cause of safety behaviors and human errors (Mohammadfam et al., 2021). Therefore, SA, as a potential safety indicator (Hauss & Eyferth, 2003), has become one of the current research hotspots in the fields of human-machine safety and ergonomics in recent years (Salmon & Stanton, 2013; Stanton et al., 2017; Wei et al., 2013).

Research on SA first requires a model explaining the cognitive mechanism of SA. Typical mechanistic explanations of SA include Endsley's three-level model of information processing, Bedny and Meister's perceptual action loop (Bedny & Meister, 1999) and Adams' event flow model (Adams et al., 1995). Endsley defines SA as "the perception of environmental elements and events in time or space, as well as the understanding of perceptual information and prediction of the future" (Endsley & Garland, 2000). Achieving SA involves a series of cognitive processes, including perception, attention and comprehension, and decision making. Considering the limited cognitive resources of operators, excessive competition for cognitive resources or distraction may lead to inadequate or even collapse of SA. For example, it is difficult for an operator who is manually tracking a target to notice the warning lights coming on. Therefore, cognitive overload is considered as a key reason for operators' incorrect SA establishment in complex task scenario (Plavsic et al., 2010), especially when the task scenario involves multiple types of subtasks. In this paper, we attempt to explore

CONTACT Xing Pan panxing@buaa.edu.cn School of Reliability & Systems Engineering, Beihang University, 37 Xueyuan Road, Haidian, Beijing 100191, P.R. China

© 2023 Taylor & Francis Group, LLC

the characteristics of operator SA in complex task situations from the perspective of cognitive resource competition.

Although task factors have been considered as an important category of SA influences (Endsley & Garland, 2000), few SA studies have considered them as major factors, probably because it is difficult for scholars to define precisely what a "task" is (Monsell, 2003), and some scholars define tasks broadly as "a collection of simple operations." Among the task factors affecting the operator's SA, task complexity is considered to be one of the most important ones. To understand the impact of task complexity on SA, it is necessary to define task complexity first. There have been three main perspectives among different models of task complexity in relevant qualitative studies: the structuralist perspective, the resource demand perspective, and the interaction perspective (Liu & Li, 2012). From the structuralist perspective, the definition of task complexity is obtained by analyzing the structure of the task, e.g., the number of the elements that make up the task and the function of the relationships between these elements. Task complexity models proposed by Wood (1986), Campbell (1988) and Ham et al. (2012) are based on this view. In the resource demand perspective, task complexity is considered as a result of resource demand, which refer to the perception and understanding of task attributes (Robinson, 2001), and some scholars have used resource demand to measure task complexity (Bedny et al., 2012; Sintchenko & Coiera, 2003). From the interaction perspective, task complexity is defined as the product of the interaction between the task and the task performer's characteristics (e.g., prior knowledge and experience). Several scholars (Chuanyan et al., 2020; Li et al., 2021) measured operators' SA at different levels of task complexity in experiments and found that operators' SA gradually decreased as task complexity increased. However, the lack of a section on measuring task complexity in these experimental studies leads us to think: what kind of tasks can be called high (low) complexity tasks? Although some progress has been made in developing methods to measure task complexity (Ham et al., 2011; Zheng et al., 2015), it is still difficult to compare the complexity between different task combinations in multitasking. Therefore, proposing a task complexity measure for multitasking has potential applications for us to effectively design HMI tasks or to match abilities between the operator and the task difficulty.

Multitasking can be described as the behavior of an operator processing multiple tasks simultaneously and is an amazing capability of our cognitive system (Himi et al., 2023). Multitasking has been a theme in experimental psychology and human factors science for decades (Hommel, 2020). Salvucci and Taatgen, working on the mechanics of human multitasking, proposed a theory of multitasking called threaded cognition (Dario & Niels, 2010). The core of threaded cognition is that multitasking behavior can be attributed to multiple threads of thought running simultaneously, which formalizes the notion of threads and how they execute, intertwine, and interfere during multitasking. In contrast, this paper attempts to formalize the operator's task behavior and the cognitive resource occupation it entails in order to explain the operator's cognitive overload and underload in multitasking. In this regard, this paper is in line with threaded cognition, both of which enhance the understanding of human multitasking more broadly. Unlike threaded cognition this paper provides a more abstract treatment of the information (cognitive resource) requirements of operators in HMI, incorporating them as difficulty factors in a complexity measure model for multitasking based on Shannon's information theory. This model is proposed following three basic assumptions: (1) complex tasks can be considered as a combination of simple tasks. (2) Even for the smallest task unit, there may be multiple attributes of cognitive resources demanded, or multiple threads may be invoked. (3) Each class of cognitive resources serves only one task at a time, which is the core assumption for the quantification of the value-added of difficulty proposed in the complexity measure model. Meanwhile, these three basic assumptions are the key assumptions that enable the task complexity measure model proposed in this paper to be embedded in the ACT-R model, which is a higher-level cognitive model (Ritter et al., 2019).

In summary, task complexity is a key factor in an operator's ability to establish and maintain SA in multitasking. Meanwhile, as task complexity is considered as a sub-concept or sub-element of task difficulty (Braarud & Kirwan, 2011), and the two are deemed interchangeable in certain tasks (Bell & Ruthven, 2004). This paper defines task difficulty as a measure of task complexity, which is a task characteristic that measures the information demand (attention resource demand) of operators during multitasking. The purpose of this paper is to propose a quantitative model of task difficulty in a multitasking to facilitate better exploration of the changing patterns of operators' SA at different levels of task difficulty.

This paper is structured as follows: Section 2 details the methodology of the paper, Section 3 describes the design of the SA measurement experiment and the analysis of the results, Section 4 discusses the results of the experiment, and Section 5 concludes the whole paper.

2. Methodology

2.1. Framework of the methodology

In this paper, multitasking is defined as the need for operators to complete at least two meta-tasks during the same period, and meta-tasks refer to the smallest task units obtained after the decomposition of a complex task (Zhao, 2021). The first step of quantifying the multitasking difficulty is to quantifying the difficulty of individual meta-tasks, which is based on Shannon's information theory. Then a three-dimensional attribute model of meta-task is developed based on multiple resource theory to measure the added difficulty caused by concurrent tasks in multitasking. Finally, this paper proposes a difficulty measure for multitasking in combination with the first two steps. Meanwhile, a SA measurement experiment is conducted to investigate the impact of task difficulty on SA at different levels. The methodological framework is shown in Figure 1.

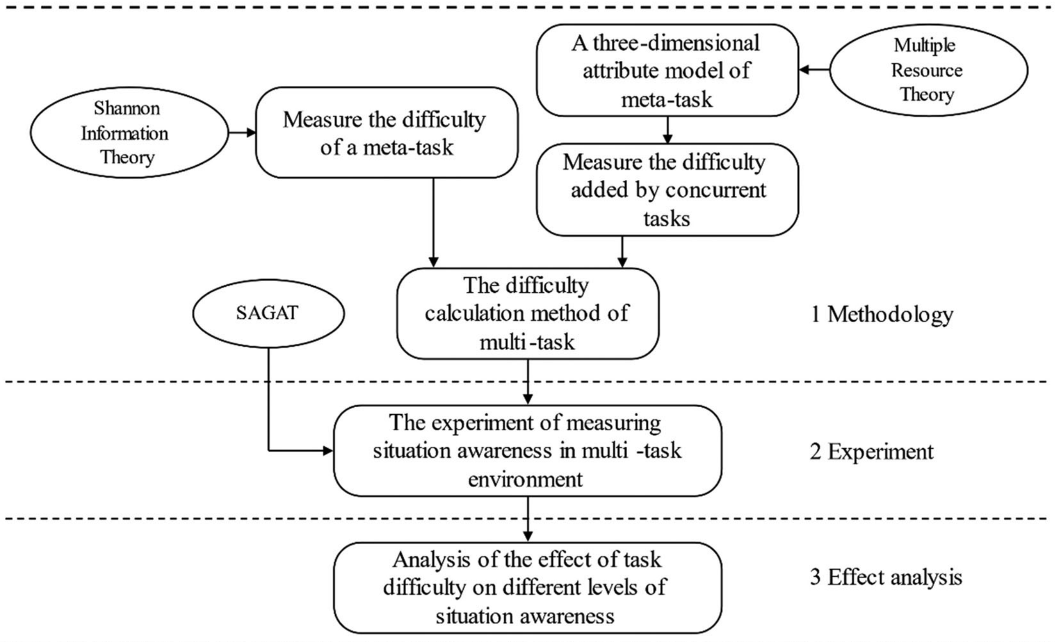

Figure 1. Framework of the method in this study.

2.2 A quantitative model of task difficulty in multitasking

Information-rich multitasking requires operators to allocate their attention to multiple meta-tasks, and when multiple meta-tasks compete for operators' limited attention resources, operators' performance, including SA, and the safety of the operation can be significantly affected due to information demand overload (Naderpour et al., 2014; Young et al., 2015). Task difficulty is a task characteristic that measures the information demand (attention resource demand) of operators during multitasking. Due to the resource competition among meta-tasks, the difficulty of multitasking is not the sum of the difficulty of all meta-tasks: the difficulty added by resource competition should also be considered.

2.2.1. A meta-task difficulty quantification method based on Shannon's information theory

Quantifying the difficulty of multitasking starts with measuring the difficulty of meta-tasks. In Human-Machine Interaction (HMI), the information requirement values of different meta-tasks determine the degree of difficulty for operators to perceive all the task information. In other words, the information requirement values of meta-tasks characterize their task difficulty. Therefore, this section measures the difficulty of meta-tasks by quantifying their information requirement values, which can be obtained based on certain characteristics associated with the meta-tasks.

Shannon's information theory precisely specifies human's capacity in certain sensory, perceptual, and perceptual-motor functions (Shannon, 1948), and Fitts extended this theory to human motor systems as a way to propose a difficulty index I_d which has been widely accepted as a way of quantifying the perceived difficulty of a task (Fitts, 1954). I_d is expressed as a logarithmic relationship between the task action amplitude A and the action tolerance range W_s to measure the average minimum information value required for a given task, while W_s is the tolerance range measured in inches and A is the specific action amplitude for a given task. For example, for a disc transfer task, where the operator needs to transfer a disc from one pillar to another, W_s represents the difference between the diameter of the pillar and the diameter of the center hole of the disc. and A represents the center-to-center distance between the two pillars. Fitts also conjectured that an operator's information processing capability I_p for any task action is nearly constant. As shown in Equation (2), there is a linear relationship between the difficulty index I_d and the action completion time FT of a task, with the slope a obtained from regression analysis of the experimental data and representing I_p. The usability of this conjecture and I_d can be proved in experiments.

$$I_d = -\log_2 \frac{W_s}{2A} = \log_2 \frac{2A}{W_s} \quad bits/response \tag{1}$$

$$FT = b + a \cdot I_d \tag{2}$$

I_d can be widely used in various action tasks (defined as Continuous Tasks (CTs) in this paper), but its application is also limited to action tasks (Fitts, 1954). In HMI, there are some tasks where both the amplitude and the tolerance range of an action are not suitable to be measured, especially those facing discrete response demands (defined as Discrete Tasks (DTs) in this paper). For example, a task requires the operator to monitor multiple lights that may come on at any time and respond to them within a short time, but the amplitude of the action and the allowable tolerance range cannot be accurately measured, and the information requirement values for the responses also cannot be characterized. For DTs, other task characteristics are needed to measure the information demand value. The solution lies in Hick/Hyman Law (Hyman, 1953), a development of

Shannon's information theory. It described a functional relationship between the stimulus information entropy Hs and the response time RT, as shown in Equation (3), where Hs is expressed as the logarithmic value of the number of stimuli j. There is a linear relationship between Hs and RT, and the slope a^* represents the ability of the operator to respond to the stimulus. As the operator's ability to respond to stimuli is fixed, the higher the Hs, the longer the RT. Based on this relationship, this paper proposes a second type of task difficulty index, I_d^*, as shown in Equation (4). I_d^* reflects the Hs of a task per unit of pending response time, and I_d^* has better applicability in difficulty quantification of DTs than I_d.

$$RT = b + a^* \cdot Hs = b + a^* \cdot \log_2 j \tag{3}$$

$$I_d^* = \frac{Hs}{\Delta T} = \frac{\log_2 j}{\Delta T} \ bits/response \tag{4}$$

I_d is well used in various CTs, while I_d^* can be used to quantify the difficulty of DTs. The difference between these two types of tasks is that they require different kinds of responses. CTs require operators to give continuous responses with specific action amplitude, and the amplitude and accuracy requirements are taken into account. By contrast, DTs require operators to respond to the task immediately within a given period of time, and the action amplitude and accuracy are often not considered. Despite the difference, I_d and I_d^* work under the same paradigm, which is expressed in Equation (5): the higher the difficulty level of responding to the task ID, the longer the response time T, under the condition that the information processing capability of the operator is fixed.

$$\begin{cases} T = \{FT, RT\} \\ ID = \{I_d, I_d^*\} \\ a' = \{a, a*\} \\ T = b + a' \cdot ID \end{cases} \tag{5}$$

Under this paradigm, I_d and I_d^* can be widely applied to task setting and difficulty quantification in HMI experiments, but there are still some problems. On the one hand, although the slopes a and a^* both represent the ability of the operator to process information, they cannot be considered numerically equivalent. In other words, I_d and I_d^* are unbalanced on the same numerical scale. On the other hand, according to the definition of I_d^*, the uncertainty of the stimulus will affect the quantitative value of I_d^* when the probability of occurrence of j different stimuli varies. To solve this problem, based on Equation (5), a correction method between different difficulty indices is proposed in this paper.

Assuming that a complex task $Task$ consists of m continuous tasks and $n - m$ discrete tasks, for the two types of meta-tasks shown in Equations (7) and (8), the correlation coefficient between the inverse of the task response rate $1/RR$ and ID is derived from a linear regression equation, and the correction factor μ is expressed as the ratio of the two coefficients. The response rate RR is defined as the ratio of the number of successful task responses to the number of tasks within the task time Δt under the specified response discrimination criterion, and i refers to the meta-task $Task_i$. The correction factor μ reflects the difference between the slope a and a^*, as shown in Equation (11), and in order to ensure the numerical consistency of the operator's ability to process information between the two types of difficulty indices, the task difficulty index for DTs is corrected into $\mu \cdot I_d^*$.

$$Task = \{Task_1, Task_2, \ldots, Task_m, Task_{m+1}, \ldots, Task_n\}, (m < n) \tag{6}$$

$$a = \frac{m \cdot \sum_{i=1}^{m} I_d(i) \cdot \frac{1}{RR(i)} - \sum_{i=1}^{m} I_d(i) \cdot \sum_{i=1}^{m} \frac{1}{RR(i)}}{m \cdot \sum_{i=1}^{m} I_d{}^2(i) - \left(\sum_{i=1}^{m} I_d(i)\right)^2} \tag{7}$$

$$a^* = \frac{(n-m) \cdot \sum_{i=m+1}^{n} I_d^*(i) \cdot \frac{1}{RR(i)} - \sum_{i=m+1}^{n} I_d^*(i) \cdot \sum_{i=m+1}^{n} \frac{1}{RR(i)}}{(n-m) \cdot \sum_{i=m+1}^{n} I_d^{*2}(i) - \left(\sum_{i=m+1}^{n} I_d^*(i)\right)^2} \tag{8}$$

$$\mu = \frac{a^*}{a} \tag{9}$$

$$RR(i) = I_d(i) \cdot a \ , (i = 1, 2, \ldots, m) \tag{10}$$

$$RR(i) = I_d^*(i) \cdot a^* = I_d^*(i) \cdot \mu \cdot a \ , (i = m+1, \ldots, n) \tag{11}$$

$$Id(i) = \begin{cases} \log_2 \dfrac{2A(i)}{W_s(i)}, \ for \ CT \\ \mu \cdot \dfrac{1}{\Delta T_i} \cdot \log_2 [j(i)], \ for \ DT \end{cases} \tag{12}$$

By combining the above methods, a formula that can be widely used to quantify the difficulty of meta-tasks in HMI experiments is derived, as shown in Equation (12). It can be applied to calculate Id of CTs and DTs respectively.

2.2.2. A three-dimensional attribute model of meta-task based on multiple resource theory

As discussed earlier, the difficulty of multitasking involves not only the difficulty of the meta-tasks but also the difficulty added by the resource competition among the meta-tasks. The key to measuring the added difficulty is to measure the degree of the resource competition among the meta-tasks. According to multiple resource theory, the competition is related to a variety of task factors, which are divided into three groups in this section: the way to respond to task requests, the type of resources required by the task, and the way of information exchange between the task and the operator. By using the three groups of factors as three basic sets of difficulty attributes of a meta-task, it is possible to determine the type of any meta-task in HMI and thus explore the resource competition between meta-tasks, as discussed below from these three perspectives.

1. Different types of responses to meta-task requirements: Discrete & Continuous

The two types of response to meta-task requirements have been discussed in detail in 2.2.1. Requirements for immediate response are defined as discrete requirements and requirements for continuous response are defined as continuous requirements. Some researchers (Wickens, 1976) have experimentally found that even though both monitoring tasks (monitoring a state change at a particular position on the screen) and tracking tasks (maintaining the stability at a particular position on the screen) occupy pools of visual resources, they are significantly different in the degree of interference with concurrent tasks and represent immediate response requirements and continuous response requirements, respectively. The above findings suggest that these two different kinds of response requirements have different degrees of impact on task difficulty in concurrent tasks, and also reveal a basic set of difficulty attributes: discrete versus continuous.

2. Different types of resources required for a meta-task: Visual & Auditory

In this paper, a resource pool is defined as a collection of sources of operator input to a meta-task. Specifying the resource pool types and which resource pool the meta-task specifically occupies is key to studying the type of meta-task. Some studies have found that visual interference affects operators more than auditory interference does when they drive a car through a curve (Donmez et al., 2006; Parkes & Coleman, 1990). Another study (Wickens et al., 2011) compared two kinds of concurrent tasks, i.e., visual-auditory and visual-visual, and found that operators performed better in the former kind of tasks. All of these findings indicate the existence of two resource pools, namely visual and auditory resource pools, and suggest that the more frequent the repeated occupation of the same resource pool by different meta-tasks, the greater their overall task difficulty. Therefore, depending on the resource attributes required by the meta-tasks, the second basic set of task difficulty attributes can be divided into visual and auditory attributes. However, for some special meta-tasks, visual and auditory resources are jointly demanded, and these meta-tasks are defined as visual-auditory Tasks.

3. Different ways of information exchange between the task and the operator: Input & Output

Task completion consists of two parts: operators' perception of information about the task environment and information feedback to the task environment, which are defined as operators' information input process and information output process respectively. For specific tasks, these two parts play either a primary or a secondary role in a cycle of completing the meta-task. In this paper, meta-tasks primarily involving information input process are defined as input tasks and meta-tasks primarily involving information output process as output tasks. Meanwhile, some experimental evidence showed that certain tasks with greater task difficulty (e.g., auditory monitoring) interfered less with another task (e.g., target tracking) than certain tasks with less task difficulty (e.g., maintaining constant force) (Wickens, 1976). As auditory monitoring and maintaining constant force are considered as an input task and an output task, respectively, in this paper, this finding illustrates the different influences of information input and output dominance on task difficulty. Further, we believe that the concept of System 1 and System 2 processing (Stanovich & West, 2000) is helpful here. System 1 processes are "fast, automatic, or unconscious" as input tasks, and System 2 processes "require access to a single, capacity-limited central working memory resource"(Evans, 2008) as output tasks. Therefore, this study considers input and out as the third basic set of difficulty attributes. However, for some special meta-tasks, both information input and output processes have a key influence on the completion of the task, and these meta-tasks are defined as input-output Tasks.

Integrating the above three groups of basic task difficulty attributes proposed according to multiple resource theory and based on Campbell's basic ideas about task classification methods (Campbell, 1988), this paper proposes a three-dimensional meta-task attribute model in cognitive process, as shown in Figure 2. The three groups of basic task difficulty attributes are embodied in three dimensions of space, thus dividing the cube into eight basic cells, each of which represents a basic meta-task attribute. With this model, any meta-task can be assigned to one or more cells in the cube according to its difficulty attributes in the three dimensions. Also, for any two meta-tasks in multitasking, the larger the number of overlapping cells between them, the stronger the resource competition between them and the greater the difficulty added. Based on this model, the difficulty added by multitasking can be quantified and calculated, offering the theoretical basis of the difficulty calculation method for multitasking in the next section.

2.2.3. A multitasking difficulty calculation method based on task profiles

The difficulty of multitasking depends on the difficulty of the meta-tasks and the difficulty added by multitasking. Concurrent tasks (i.e., the operator is asked to complete two meta-tasks at the same time) are the classic multitasking situation and the key influencing factor for multitasking performance. This paper proposes that the degree of resource competition between concurrent tasks is determined by the number of overlapping cells p between two meta-tasks: the larger the number, the stronger the competition, and the

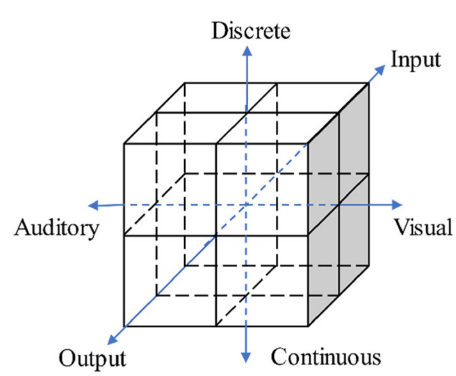

Figure 2. A three-dimensional attribute model of a meta-task.

greater the added task difficulty due to resource competition. In this view, an evaluation criterion of resource competition between concurrent tasks is obtained, as shown in Equation (13): the correlation coefficient c_{ij} between one meta-task $Task_i$ and another meta-task $Task_j$ is defined as the degree of resource competition between the two meta-tasks, and c_{ij} is positively correlated with the number of overlapping cells p_{ij} of two meta-tasks. If a complex task consists of n meta-tasks, the correlation coefficient between any two meta-tasks can be obtained from the correlation coefficient matrix $A_{n\times n}$. The matrix $A_{n\times n}$ represents the pressure of resource competition between meta-tasks in multitasking.

$$A_{n\times n} = \begin{bmatrix} 0 & c_{12} & \cdots & c_{1n} \\ c_{21} & 0 & \cdots & c_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ c_{n1} & c_{n2} & \cdots & 0 \end{bmatrix}, while\ c_{ij} = c_{ji}\ , c_{ij} \propto p_{ij} \tag{13}$$

In multitasking, each process from the beginning to the end of any meta-task is defined as one task event, and all task events of n types of meta-tasks in a given period Δt are integrated on the same task profile, which is shown in Figure 3. From the task profiles, we can understand whether these meta-tasks have time overlap at demand time and calculate the time overlap value between meta-tasks, which is also called task concurrency duration in this paper. t_{ij} is the concurrency duration of $Task_i$ and $Task_j$.

According to the difficulty quantification method of meta-tasks proposed in 2.2.1, the difficulty of n types of meta-tasks are $Id(1), ..., Id(n)$, respectively. In order to eliminate the influence caused by different dimensions of indicators within the task difficulty calculation formula, the meta-task difficulty is normalized to achieve comparability between different task difficulty values, and the process is as follows.

$$\overline{Id}(i) = \frac{Id(i)}{\max\{Id(1), Id(2), ..., Id(n)\}} \tag{14}$$

Resource competition between concurrent tasks increases the overall task difficulty. Suppose the added task difficulty caused by $Task_i$ and $Task_j$ in the complex task $Task$ is $\Delta Id(Task, i, j)$, and $\Delta Id(Task, i, j)$ is jointly determined by the difficulty of meta-tasks, the correlation coefficient c_{ij} between the meta-tasks, and the concurrency duration t_{ij} of the meta-tasks, which is defined as Equation (15). In Equation (15), $\overline{Id}(i)$ is the task difficulty of $Task_i$ after standardized processing, c_{ij} is the correlation coefficient between $Task_i$ and $Task_j$ and Δt is the length of a given period.

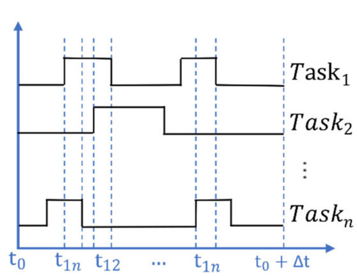

Figure 3. Task profiles in a multi-task model.

$$\Delta Id(Task, i, j) = \overline{Id}(i) \cdot \overline{Id}(j) \cdot c_{ij} \cdot t_{ij} \cdot \frac{1}{\Delta t} \tag{15}$$

In summary, the difficulty of multitasking $Id(Task)$ is determined by the task difficulty of all n meta-tasks and the difficulty added by all concurrent tasks, which is defined as Equation (16).

$$Id(Task) = \sum_{1\le i\le n} \overline{Id}(i) + \sum_{1\le i<j\le n} \Delta Id(Task, i, j) \tag{16}$$

2.3. A multilevel measure of SA based on SAGAT

Generally, existing SA measurement methods can be divided into two categories: inferential measurement methods and direct measurement methods (Miles & Strybel, 2017). Situation Awareness Global Assessment Technique (SAGAT) is one of the earliest and most widely used SA direct measurement methods, allowing immediate assessment of SA by asking operators about their perceptions of the current situation. When using SAGAT, simulations of representative tasks or scenarios are frozen at randomly selected times, and system displays are blanked while operator quickly answer questions about their current perceptions of the situation. The operators' answers are then compared with the actual situation to objectively measure their SA. SAGAT was found to be a highly sensitive, reliable, and predictive measure of SA that is useful across a wide variety of domains and experimental settings (Endsley, 2021; Salmon et al., 2006). In SAGAT, since it is not possible to ask the operators about all situations they face at the point of interruption, some SA questions are randomly selected, and this random sampling method is consistent and statistically valid (Endsley, 1988).

In HMI, the information exchange between the operator and the task environment is divided into two parts: information input and information output. The efficiency of information input is considered as the success rate of the operator's SA establishment, i.e., the ratio of output information to input information in the operator's cognitive process can measure the SA establishment. A randomly selected part of SA questions from all task situations are taken as the operator's information input, and the operator's response to the situation at the time of interruption is taken as the output information. As shown in Equation (17), SA is measured as the proportion of correct responses to SAGAT questions, where sa is the SA score and N is the total number of selected SA questions.

$$sa = \frac{1}{N}\sum_{j=1}^{N}\sigma(j)$$
$$whereby\ \sigma(x) = \begin{cases} 1 & \text{if the } x\text{th item is correctly answered} \\ 0 & \text{otherwise} \end{cases} \tag{17}$$

Overall, SAGAT adopts the perspective of information transfer accuracy in observing the HMI process and provides a more realistic picture of the operator's SA. As a global metric, SAGAT also allows the measurement of different levels of SA (i.e., level 1 (perception level), level 2 (understanding level), and level 3 (prediction level)) by different question designs. In the subsequent experiment, we used SAGAT to design questions for each of the three levels of SA and measured the operators' SA values at the three levels. Based on the quantitative model of multitasking difficulty proposed in the previous section, we explored the changes of the three levels of SA under different levels of task difficulty in multitasking.

3. Experiment

3.1. Subjects

This experiment had 18 participants, aged from 20 to 26 years old, ($M = 22.6$ and $SD = 2.6$), including 10 males and 8 females. All participants were students of Beihang University with knowledge of general flight operations and experience in operating aircraft simulators. On the day before the experiment, all subjects were informed of the experiment procedures and points to note. It was also assured that they had enough sleep the night before and were not tired on the day. All participants had normal hearing and normal or corrected-to-normal vision. Only right-handed participants were recruited for this study due to the operation of OPEN-MATB and to avoid the influence of handedness on the experiment.

3.2. Experimental equipment and materials

The experimental equipment consisted of a laptop computer with a 22-inch display, a mouse, and an x-box joystick. Subjects were asked to complete the set tasks on OPEN-MATB (Cegarra et al., 2020), a multi-attribute task software on the laptop computer, with the right hand using the mouse and the left hand controlling the joystick.

3.3. Experimental task

OPEN-MATB, derived from the Multi-Attribute Task Battery (MTAB) developed by NASA, provides a set of baseline tasks that can be used in a wide range of laboratory studies for human performance and workload assessment. Among the existing studies, many fields have leveraged MATB to simulate various attribute flight tasks faced by pilots in the air while exploring various theories (Feng et al., 2022; Mortazavi et al., 2019; Nixon & Charles, 2017), and studies have also made arguments for the persuasiveness of the MATB(C. D. Wickens et al., 2016). As shown in Figure 4, there were four meta-tasks provided by OPEN-MATB, namely, System Monitoring (SYSMON), Tracking (TRACK), Communications (COMM), and Resource Management (RESMAN).

The SYSMON required the subject to monitor the system status during the task, which consisted of two warning lights (F5 and F6) and four gauges (F1–F4). For most of the experiment, the warning lights remained normal (green light on and red light off) and the four gauges fluctuated within the normal range. When the status of the warning light changed or the gauges fell out of range, the subject was required to press the key corresponding to the failure point with the mouse.

The TRACK consists of two modes, automatic and manual. In the manual mode, it requires the subject to use the

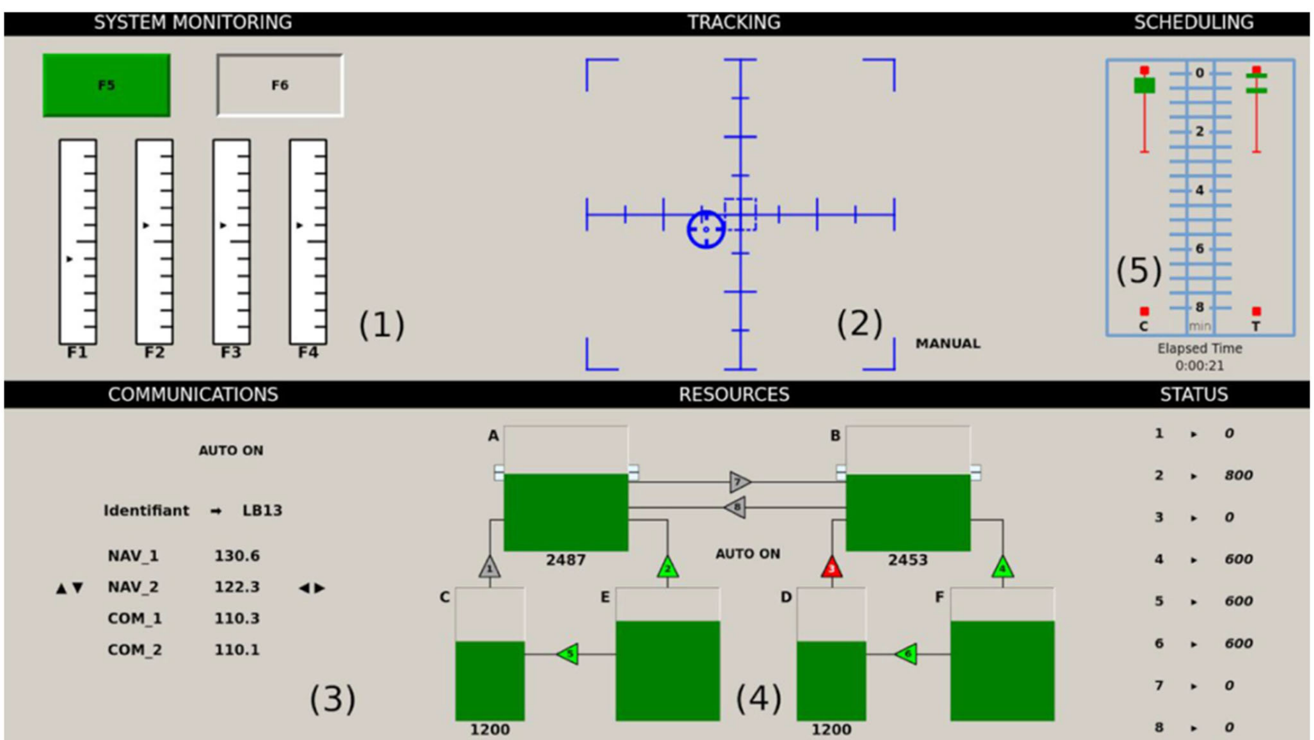

Figure 4. A Multi-attribute task battery.

joystick to maintain a circular icon in the center of the target area.

The COMM presents subject with auditory signals that prompt them to respond when they hear a specific call sign. Once the call sign is confirmed, the subject is asked to change the frequency of the radio.

The RESMAN simulated the in-flight fuel management task. Subject was asked to maintain optimal fuel levels (2400–2600 units) in two primary fuel tanks (labeled A and B). This was accomplished by transferring fuel from the four secondary tanks (labeled C-F) to the primary tanks. The subject can click on the valves (marked with the numbers 1–8) in order to transfer the fuel to the desired direction. Also, during the experiment, any valve can malfunction (change to red state), in which case the subject is required to keep the fuel level in the main tank by modifying the state of other valves.

3.4. Experimental design

3.4.1. Independent variable

The independent variable of the experiment is the difficulty of multitasking. The experiments were conducted by designing different task processes to achieve different difficulty levels, considering the effects of three main factors on task difficulty, including: numbers of concurrent tasks, task combinations and task concurrency duration. The different multitasking scenarios are implemented by writing Extensible Markup Language (XML).

3.4.2. Dependent variables

In order to investigate the effect of task difficulty on the subjects' SA at different levels, three dependent variables were counted: SAGAT score, task completion time and task response rate. SAGAT scores were obtained in the following way: the experiment was interrupted at a certain point, the subjects were asked to quickly answer questions about the current situation, and both the actual situation and the operator's responses were recorded and compared to determine the correctness of their responses. At each interruption point, three types of questions were designed to measure the three-level SA. SAGAT questions are documented in the Appendix of this paper. The time taken by the subject to complete the task was automatically recorded by the program and will be used as a criterion to determine whether the response to the task was successful. Task response rate referred to the ratio of the number of successful responses to the task to the total number of occurrences of the task in each experiment, which was collected for the task difficulty model.

3.4.3. Program design

The experiment uses a within-subject design, each subject was asked to complete three sets of tasks at different levels of difficulty, and was asked to complete a subjective questionnaire about the task difficulty after completion.

Six different tasks were designed in this experiment, each task was four minutes long, the first three tasks contained three different meta-tasks while the last three tasks contained four. For the three discrete meta-tasks SYSMON, COMM and RESMAN, the response demand time of each event is taken as the duration of this event; for the TRACK task, the time period of the manual mode is considered as the duration of the task. In different task designs, the occurrence frequency of each meta-task is kept the same, while the concurrent duration between meta-tasks is not the same. The event frequencies of different meta-tasks are shown in Table 1.

Meanwhile, in order to eliminate differences in the subjects' proficiency in completing the three sets of tasks, a Latin square was used to determine the tasks for each subject, thus avoiding the effects of experimental order. Each subject was asked to perform a six-minute pre-experiment before the formal experiment, during which the task completion time and the task response rate data for each meta-task were automatically recorded by the program.

3.5. Experimental data and pre-processing

3.5.1. Meta-task difficulty calculation

Based on the three-dimensional meta-task attribute model proposed in 2.2.2, the attributes of the four meta-tasks provided by the experimental platform were obtained (Table 2).

Based on the meta-task difficulty measurement method proposed in 2.2.1 and the task response data collected from multiple subjects through pre-experiments, the task difficulty of the four meta-tasks was calculated and shown in Table 3.

3.5.2. Complex task grouping design and its task difficulty calculation

Based on the evaluation criteria for the degree of resource competition between concurrent meta-tasks proposed in 2.2.3, the correlation coefficients between every two of the four types of meta-tasks provided by the experimental platform can be obtained, and the correlation coefficient matrix $A_{n\times n}$ was shown in Equation (18).

$$A_{4\times 4} = \begin{bmatrix} 0 & 2 & 3 & 2 \\ 2 & 0 & 1 & 2 \\ 3 & 1 & 0 & 2 \\ 2 & 2 & 2 & 0 \end{bmatrix} \tag{18}$$

Table 1. Frequency of events for different meta-tasks.

Mata-task	Events	Frequency
SYSMON	Green light off & Red light on & Gauges abnormal	Once in 15 s
TRACK	Manual mode	90 s
COMM	Call occurs	Once in 30 s
RESMAN	Valves failure & Valves repair	Once in 20 s

Table 2. Meta-task attributes.

Meta-task attribute	SYSMON	TRACK	COMM	RESMAN
Continuous & Discrete	Discrete	Continuous	Discrete	Con-Dis
Visual & Auditory	Visual	Visual	Auditory	Visual
Input & Output	Input	Input	In-Out	Output

Table 3. Meta-task difficulty measurement.

	SYSMON	TRACK	COMM	RESMAN
$j(i)$	12	—	4	16
t_i	10	—	15	20
$RR(i)$	0.729	0.889	0.944	0.791
$I_d(i)/I_d^*(i)$	0.358	0.270	0.133	0.2
$Id(i)$	0.451	0.270	0.167	0.252
$Id(i)$	1	0.599	0.370	0.558

Table 4. Task difficulty of different groups.

	*Task*1	*Task*2	*Task*3	*Task*4	*Task*5	*Task*6
$\sum_i Id(i)$	2.007	1.778	1.801	2.293	2.293	2.293
$\Delta Id(T,1,2)$	0.329	0.330	0	0.199	0.175	0.199
$\Delta Id(T,1,3)$	0	0.532	0.458	0.042	0.009	0.056
$\Delta Id(T,1,4)$	0.474	0	0.553	0.372	0.251	0.274
$\Delta Id(T,2,3)$	0	0.080	0	0	0.014	0.028
$\Delta Id(T,2,4)$	0.167	0	0	0.067	0.112	0.089
$\Delta Id(T,3,4)$	0	0	0.215	0.026	0	0
$\sum_{i,j} \Delta Id(T,i,j)$	0.971	0.941	1.226	0.707	0.560	0.647
$Id(T)$	2.978	2.719	3.027	2.999	2.853	2.939

Table 5. Different levels of SA in tasks with different difficulty levels.

	*Task*1	*Task*2	*Task*3	*Task*4	*Task*5	*Task*6
$I_d(T)$	2.978	2.719	3.027	2.999	2.853	2.939
Perception	—	—	—	0.589	0.839	0.732
Comprehension	—	—	—	0.786	0.857	0.714
Projection	—	—	—	0.857	0.786	0.857
SA	0.714	0.889	0.619	0.649	0.831	0.753

For the six groups of experimental tasks, the task difficulty was calculated based on the difficulty of the four types of meta-tasks used in the experiments, the correlation coefficients between the tasks, and the task concurrency duration, as shown in Table 4.

3.5.3. Measurement of three levels of SA

Based on the multilevel measurement method of SA in 2.3, the SAGAT scores for the three levels of SA (perception, comprehension, and prediction) were obtained, as shown in Table 5.

3.6. Results

Statistical analyses were performed using SPSS Statistics 26.0 ($\alpha = 0.05$ for all statistical tests). In this paper, a repeated-measures analysis of variance (ANOVA) was used to determine the main effect of task difficulty on the dependent variables. For within-subject variables, Mauchly's test was used to test the sphericity hypothesis and Pearson's correlation was used to calculate the level of correlation between the measures.

3.6.1. Correlation test between task difficulty measures and subjective scores

Based on the task difficulty quantitative model proposed in this paper, Section 3.5 completed the difficulty calculation for the six task groups. Meanwhile, the correlation test between task difficulty values and participants' subjective scores (through subjective questionnaires) was carried out in this section. Table 6 demonstrates the task difficulty values and subjective scores of the subjects for the different groups.

Table 6. Task difficulty values and subjective score for different groups.

	*Task*1	*Task*2	*Task*3	*Task*4	*Task*5	*Task*6
$Id(T)$	2.978	2.719	3.027	2.999	2.853	2.939
Subjective score	4.00	1.89	4.56	3.11	2.33	2.44

Table 7. Results of correlation analysis.

source	Pearson Correlation	Significance	Number of cases
Subjective score	0.819	0.046	6

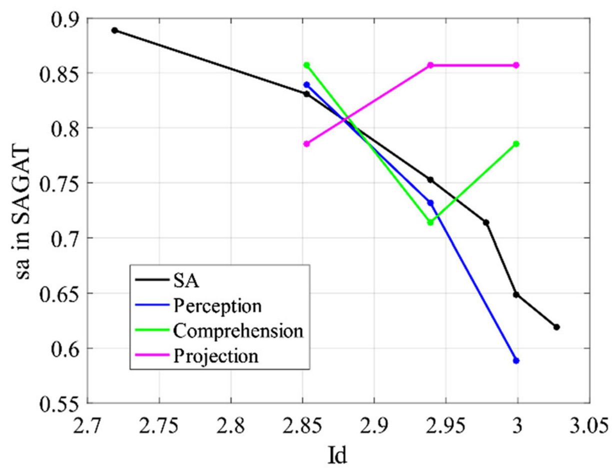

Figure 5. Correlation between three levels of SA and task difficulty.

Table 8. Mauchly's test of sphericity.

Source	Mauchly W	Approximate chi-square	Degree of freedom	Significance
Task difficulty	0.153	7.699	14	0.927

The results of the correlation analysis showed a positive correlation between task difficulty values and subjective scores. As shown in Table 7, at a significance level of 0.05, there was a significant correlation between the task difficulty value and the subjective scores ($sig = 0.046 < 0.05$).

3.6.2. Correlation analysis of task difficulty and SA scores

The Pearson correlation method was used to calculate the correlation level between task difficulty and SA scores. As shown in Figure 5, with the increase in task difficulty, the value of SA gradually decreased. The Pearson correlation coefficient between task difficulty and SA showed a strong negative correlation between the two ($r = -0.966$).

The results of Mauchly's test showed that SA scores at different levels of task difficulty meet the sphericity hypothesis ($p = 0.927 > 0.05$), as shown in Table 8. The main effect of task difficulty on SA scores was significant ($F = 9.176$, $\alpha = 0.05$, $p < 0.05$), as shown in Table 9. Therefore, it was concluded that the higher the task difficulty, the lower the SA level of the operator, and that there were significant and large differences between SA scores at different levels of task difficulty.

Table 9. Results of variance analysis of SA scores.

Source	Sum of squares of III class	Degree of freedom	Mean square	F	Significance
Task difficulty	0.379	5	0.076	9.176	0.001*

Table 10. Results of variance analysis of SA scores at different levels.

Source	Sum of squares of III class	Degree of freedom	Mean square	F	Significance
Perception	0.220	2	0.110	6.388	0.013
Comprehension	0.071	2	0.036	0.300	0.746
Projection	0.024	2	0.012	0.300	0.746

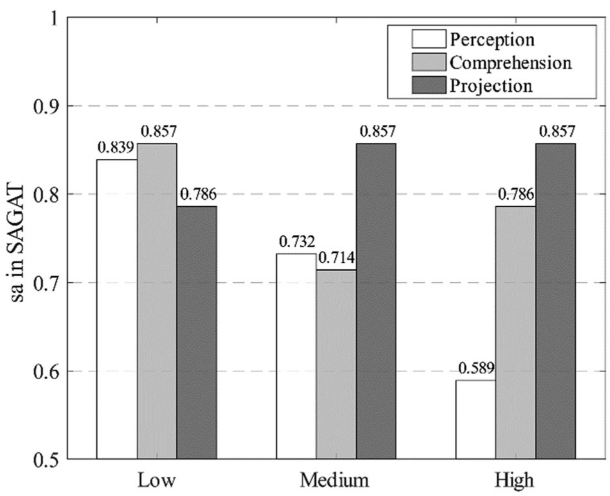

Figure 6. Three levels of SA results at different task difficulty levels.

3.6.3. Correlation analysis of task difficulty and three-level situation awareness

*Task*4, *Task*5, and *Task*6 were selected as three sets of experiment tasks with low, medium, and high difficulty respectively. The correlation analysis and ANOVA revealed the changes of SA at the perception, comprehension, and prediction levels when the task difficulty levels were different (Figure 6).

As shown in Table 10 and Figure 6, as task difficulty increased, a decreasing trend was found in perceptual-level SA scores, and results of ANOVA suggested a significant difference in perceptual-level SA scores when task difficulty changed ($F = 6.388, \ \alpha = 0.05, \ p = 0.013 < 0.05$). Post hoc comparisons indicated significantly higher perceptual-level SA scores for low task difficulty than those for medium task difficulty and high task difficulty, and perceptual-level SA scores for medium task difficulty were also significantly higher than those for high task difficulty. This result well supported the experimental control of task difficulty by selecting three sets of experimental tasks T_4, T_5, and T_6.

For SA at the comprehension level, there was a moderate negative correlation between task difficulty and SA scores ($r = -0.658$), which was significantly smaller than the correlation between task difficulty and SA scores at the perception level. For SA at the prediction level, there was no negative correlation between SA scores and task difficulty, and SA scores remained almost constant when the task difficulty was low, medium or high. Meanwhile, ANOVA showed that there was no significant difference between comprehension-level SA and prediction-level SA at different task difficulty levels ($p = 0.746 > 0.05$). Therefore, we concluded that task difficulty has a significant effect on SA at the perception level, a smaller effect on SA at the comprehension level, and almost no effect on SA at the prediction level.

4. Discussions

This paper proposes a task difficulty measurement method for multitasking in HMI, which provides a reference for quantifying of task-related factors such as task difficulty, complexity, and load in human factors experiments, and its implications are further discussed below.

1. Applicability analysis of the multitasking difficulty measurement method

Investigating the impact of various task (system) factors on people's cognitive processes through human factors experiments have been a popular research direction in the field of human factors reliability in recent years. In existing research, scholars have focused more on controlling task factors in experiments through qualitative comparison of task complexity, which has limitations when the experimental task composition is more complex, especially in multitasking situations. Such limitations are overcome by the multitasking model and its task difficulty calculation method proposed in this paper, which achieves task difficulty level classification and quantitative calculation in complex task situations.

The SA scores at different levels of task difficulty measured with SAGAT indicate a strong negative correlation between task difficulty and SA. This result is consistent with the findings of Lin, Heikoop, and Li et al (Heikoop et al., 2018; Lin et al., 2013; Lin & Lu, 2016), adding new quantitative evidence for the relationship between task difficulty and SA, as well as proving the rationality and validity of the model and calculation method.

The model and the method provide a new way to quantify task-related factors in human factors experiments. They allow the division of more task levels and help make the conclusions drawn from the inquiry in this direction more convincing.

2. Analysis of the effect of task difficulty on SA

The conceptual framework of SA has been extensively researched, and the relationships between the three levels of

SA are worth exploring in depth. The experimental data in this paper showed that the patterns of change in the establishment of the three levels of SA varied at different levels of task difficulty. Also, there was a strong correlation between the SA scores at the perceptual level and the task difficulty. Our interpretation of the experimental results is that the effect of different levels of task difficulty on SA is mainly reflected in the first level of SA, i.e., the perception level. By contrast, task difficulty has a much smaller effect on SA at the comprehension level and basically no effect on SA at the prediction level. In other words, the perceptual level of SA is most sensitive to changes in task difficulty and is more likely to lead to perceptual errors (ignoring or misperceiving information) when task difficulty reaches or exceeds the limits that the operator can handle, while the comprehension and prediction levels of SA appear to be more dependent on the operator's mental model and long term memory (Endsley, 1995a), and higher cognitive resource requirements (task difficulty) do not significantly affect this ability.

Moreover, the effect of operator experience level on SA cannot be ignored in the discussion of the results; for example, novice operators are more likely to make mistakes at lower levels of SA compared to experienced operators. This may alter the effect of task load on the mechanism of SA formation, just as it would alter the mechanism of SA formation by other factors (Gutzwiller & Clegg, 2013; Sohn & Doane, 2004). Although the subjects participating in this paper were trained on the operation of MATB, whether these results will vary with subject expertise remains an important question for future research. These results also explain the findings of a statistical analysis in the aviation safety report by Endsley et al., which suggests that about 72% of the SA-related errors occurred at the perception level, 22% at comprehension level and 6% at projection level(Mica R. Endsley, 1995a). This view is based on two ideas in this paper, (1) task overload is a significant contributor to SA errors (Endsley, 1995a) and (2) errors including perception, comprehension, and prediction are part of SA errors (Endsley, 1995b). Specifically, the perceptual aspect of SA is more sensitive to changes in task load and is more likely to lead to further SA errors when the task load reaches or exceeds the operator's tolerable limits.

The experimental findings in this paper also provide a new direction for subsequent research on three-level SA, namely exploring the relationship between the three levels of SA from the perspective of variable correlation quantitatively.

5. Conclusions

SA is critical to operator task performance and behavioral safety in HMI, and task difficulty is an important influencing factor of SA. In experimental investigations into human factors, the problem of how to set the difficulty level of experimental tasks is unavoidable. The traditional approach designs task difficulty levels by focusing on the difficulty contrast between different tasks, but this method has limitations in multitasking situations.

To address this problem, this paper proposes a multitasking difficulty measurement method and conducts experiments to measure SA under different levels of task difficulty. It presents a more general meta-task difficulty quantification method based on Shannon's information theory, a three-dimensional meta-task attribute model based on task taxonomy and multiple errors, and a way of measuring the resource competition between concurrent tasks, based on which a multitasking model and its task difficulty calculation method are put forward.

Through a SA assessment experiment, this research analyzed the correlation between task difficulty and SA scores as well as the correlation between the three levels of SA scores at different levels of task difficulty. The data analysis shows that there is a strong negative correlation between task difficulty and SA, and the perceptual level of SA is more easily affected by task difficulty than the comprehension and prediction levels. The experimental data and results of this research are consistent with those of other existing studies, which validates the rationality and usability of the proposed model and method. This research also provides a reference for task optimization, strategy selection, and operator attention allocation in HMI and offers new insights for subsequent research on multi-level SA.

This study also has certain limitations noticed during the experiment. One is that the meta-task three-dimensional attribute model proposed in this paper is presented from the perspective of resource requirements, which means that this model may not apply to other tasks beyond HMI. We believe, however, that the idea behind this model also works for tasks in other domains. The other limitation exists in the SAGAT questions pool designed in this paper, where some of the questions in the pool focus on interference information rather than task information. This is not consistent with most SA theories, and we will further discuss the impact of interference information in the SA process in our future work.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work is supported by the National Natural Science Foundation of China under Grant No. 72071011.

References

Adams, M. J., Tenney, Y. J., & Pew, R. W. (1995). Situation awareness and the cognitive management of complex systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *37*(1), 85–104. https://doi.org/10.1518/001872095779049462

Bedny, G., & Meister, D. (1999). Theory of activity and situation awareness. *International Journal of Cognitive Ergonomics*, *3*(1), 63–72. https://doi.org/10.1207/s15327566ijce0301_5

Bedny, G. Z., Karwowski, W., & Bedny, I. S. (2012). Complexity evaluation of computer-based tasks. *International Journal of Human-Computer Interaction*, *28*(4), 236–257. https://doi.org/10.1080/10447318.2011.581895

Bell, D. J., & Ruthven, I. (2004). *Searcher's assessments of task complexity for web searching* [Paper presentation]. European Conference on Information Retrieval, Berlin, Heidelberg.

Braarud, P. I., & Kirwan, B. (2011). *Task complexity: What challenges the crew and how do they cope*. Springer.

Campbell, D. J. (1988). Task complexity: A review and analysis. *The Academy of Management Review*, *13*(1), 40–52. https://doi.org/10.2307/258353

Cegarra, J., Valéry, B., Avril, E., Calmettes, C., & Navarro, J. (2020). OpenMATB: A Multi-Attribute Task Battery promoting task customization, software extensibility and experiment replicability. *Behavior Research Methods*, *52*(5), 1980–1990. https://doi.org/10.3758/s13428-020-01364-w

Chuanyan, F., Xiaoru, W., Shuang, L., Hao, C., & Damin, Z. (2020). Study on situation awareness under different mental workloads. *Journal of Northwestern Polytechnical University*, *38*(03), 610–618. https://doi.org/10.1051/jnwpu/20203830610

Dario, D. S., & Niels, A. T. (2010). *The multitasking mind*. Oxford University Press.

Donmez, B., Boyle, L. N., & Lee, J. D. (2006). The impact of distraction mitigation strategies on driving performance. *Human Factors*, *48*(4), 785–804. https://doi.org/10.1518/001872006779166415

Endsley, M., & Garland, D. (2000). *Situation awareness analysis and measurement*. Lawrence Erlbaum.

Endsley, M. R. (1988). *Situation awareness global assessment technique (SAGAT)* [Paper presentation]. Proceedings of the IEEE 1988 National Aerospace and Electronics Conference, Dayton, OH, USA. https://doi.org/10.1109/NAECON.1988.195097

Endsley, M. R. (1995a). A taxonomy of situation awareness errors. In R. Fuller, N. Johnston, & N. McDonald (Eds.), *Human factors in aviation operations*. Avebury Aviation, Ashgate Publishing Ltd.

Endsley, M. R. (1995b). Toward a theory of situation awareness in dynamic systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *37*(1), 32–64. https://doi.org/10.1518/001872095779049543

Endsley, M. R. (2021). A systematic review and meta-analysis of direct objective measures of situation awareness: A comparison of SAGAT and SPAM. *Human Factors*, *63*(1), 124–150. https://doi.org/10.1177/0018720819875376

Evans, J. S. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, *59*(1), 255–278. https://doi.org/10.1146/annurev.psych.59.103006.093629

Feng, C., Liu, S., Wanyan, X., Chen, H., Min, Y., & Ma, Y. (2022). EEG feature analysis related to situation awareness assessment and discrimination. *AEROSPACE*, *9*(10), 546. https://doi.org/10.3390/aerospace9100546

Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, *47*(6), 381–391. https://doi.org/10.1037/h0055392

Gutzwiller, R. S., & Clegg, B. A. (2013). The role of working memory in levels of situation awareness. *Journal of Cognitive Engineering and Decision Making*, *7*(2), 141–154. https://doi.org/10.1177/1555343412451749

Ham, D., Park, J., & Jung, W. (2011). Extension of TACOM to the complexity of tasks designed for abnormal situations in nuclear power plants. *Journal of Loss Prevention in the Process Industries*, *24*(5), 601–611. https://doi.org/10.1016/j.jlp.2011.04.006

Ham, D., Park, J., & Jung, W. (2012). Model-based identification and use of task complexity factors of human integrated systems. *Reliability Engineering & System Safety*, *100*(April), 33–47. https://doi.org/10.1016/j.ress.2011.12.019

Hauss, Y., & Eyferth, K. (2003). Securing future ATM-concepts' safety by measuring situation awareness in ATC. *Aerospace Science and Technology*, *7*(6), 417–427. https://doi.org/10.1016/S1270-9638(02)00011-1

Heikoop, D. D., de Winter, J. C., van Arem, B., & Stanton, N. A. (2018). Effects of mental demands on situation awareness during platooning: A driving simulator study. *Transportation Research Part F: Traffic Psychology and Behaviour*, *58*(October), 193–209. https://doi.org/10.1016/j.trf.2018.04.015

Himi, S. A., Volberg, G., Buhner, M., & Hilbert, S. (2023). Individual differences in everyday multitasking behavior and its relation to cognition and personality. *Psychological Research*, *87*(3), 655–685. https://doi.org/10.1007/s00426-022-01700-z

Hommel, B. (2020). Dual-task performance: Theoretical analysis and an event-coding account. *Journal of Cognition*, *3*(1), 29–29. https://doi.org/10.5334/joc.114

Hyman, R. (1953). Stimulus information as a determinant of reaction time. *Journal of Experimental Psychology*, *45*(3), 188–196. https://doi.org/10.1037/h0056940

Kharoufah, H., Murray, J., Baxter, G., & Wild, G. (2018). A review of human factors causations in commercial air transport accidents and incidents: From to 2000–2016. *Progress in Aerospace Sciences*, *99*(May), 1–13. https://doi.org/10.1016/j.paerosci.2018.03.002

Lee, S. W., Park, J., Kim, A. R., & Seong, P. H. (2012). Measuring situation awareness of operation teams in NPPs using a verbal protocol analysis - ScienceDirect. *Annals of Nuclear Energy*, *43*(43), 167–175. https://doi.org/10.1016/j.anucene.2011.12.005

Li, P., Wang, Y., Chen, J., Luo, Z., & Dai, L. (2021). An experimental study on the effects of task complexity and knowledge and experience level on SA, TSA and workload. *Nuclear Engineering and Design*, *376*(1), 111112. https://doi.org/10.1016/j.nucengdes.2021.111112

Li, P., Zhang, L., Dai, L., Zou, Y., & Li, X. (2019). An assessment method of operator's situation awareness reliability based on fuzzy logic-AHP. *Safety Science*, *119*(November), 330–343. https://doi.org/10.1016/j.ssci.2018.08.007

Lin, C. J., Hsieh, T.-L., & Lin, S.-F. (2013). Development of staffing evaluation principle for advanced main control room and the effect on situation awareness and mental workload. *Nuclear Engineering and Design*, *265*(December), 137–144. https://doi.org/10.1016/j.nucengdes.2013.08.056

Lin, L.-W., & Lu, M.-S. (2016). Empirical research on the relationship between Helicopter Pilots' mental workloads and situation awareness levels. *Journal of the American Helicopter Society*, *61*(3), 1–8. https://doi.org/10.4050/JAHS.61.032005

Liu, P., & Li, Z. (2012). Task complexity: A review and conceptualization framework. *International Journal of Industrial Ergonomics*, *42*(6), 553–568. https://doi.org/10.1016/j.ergon.2012.09.001

Miles, J. D., & Strybel, T. Z. (2017). Measuring situation awareness of student air traffic controllers with online probe queries: Are we asking the right questions?. *International Journal of Human-Computer Interaction*, *33*(1), 55–65. https://doi.org/10.1080/10447318.2016.1232231

Mohammadfam, I., Mahdinia, M., Sol Ta Nzadeh, A., Aliabadi, M. M., & Sol Ta Nian, A. R. (2021). A path analysis model of individual variables predicting safety behavior and human error: The mediating effect of situation awareness. *International Journal of Industrial Ergonomics*, *84*(3), 103144. https://doi.org/10.1016/j.ergon.2021.103144

Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences*, *7*(3), 134–140. https://doi.org/10.1016/s1364-6613(03)00028-7

Mortazavi, M. R., Raissi, K., & Mehne, S. H. H. (2019). A probabilistic approach to classification of human performance during interaction with a standard flight tasks simulator. *Journal of Ambient Intelligence and Humanized Computing*, *10*(8), 3211–3230. https://doi.org/10.1007/s12652-018-1038-2

Naderpour, M., Lu, J., & Zhang, G. (2014). A situation risk awareness approach for process systems safety. *Safety Science*, *64*(April), 173–189. https://doi.org/10.1016/j.ssci.2013.12.005

Nixon, J., & Charles, R. (2017). Understanding the human performance envelope using electrophysiological measures from wearable technology. *Cognition, Technology & Work*, *19*(4), 655–666. https://doi.org/10.1007/s10111-017-0431-5

Parkes, A., & Coleman, N. (1990). Route guidance systems: A comparison of methods of presenting directional information to the driver. *Contemporary Ergonomics*, *22*(2), 480–485. https://doi.org/10.1016/0003-6870(91)90330-k

Plavsic, M., Klinker, G., & Bubb, H. (2010). Situation awareness assessment in critical driving situations at intersections by task and human error analysis. *Human Factors and Ergonomics in Manufacturing & Service Industries*, *20*(3), 177–191. https://doi.org/10.1002/hfm.20173

Ritter, F. E., Tehranchi, F., & Oury, J. D. (2019). ACT-R: A cognitive architecture for modeling cognition. *Wiley Interdisciplinary Review-Cognitive Science*, *10*(3), e1488. https://doi.org/10.1002/wcs.1488

Robinson, P. (2001). Task complexity, task difficulty and task production: Exploring interactions in a componential framework. *Applied Linguistics*, *22*(1), 27–57. https://doi.org/10.1093/applin/22.1.27

Salmon, P., Stanton, N., Walker, G., & Green, D. (2006). Situation awareness measurement: A review of applicability for C4i environments. *Applied Ergonomics*, *37*(2), 225–238. https://doi.org/10.1016/j.apergo.2005.02.001

Salmon, P. M., & Stanton, N. A. (2013). Situation awareness and safety: Contribution or confusion? Situation awareness and safety editorial. *Safety Science*, *56*, 1–5. https://doi.org/10.1016/j.ssci.2012.10.011

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

She, M., & Li, Z. (2017). Design and evaluation of a team mutual awareness toolkit for digital interfaces of nuclear power plant context. *International Journal of Human–Computer Interaction*, *33*(9), 744–755. https://doi.org/10.1080/10447318.2017.1282188

Sintchenko, V., & Coiera, E. W. (2003). Which clinical decisions benefit from automation? A task complexity approach. *International Journal of Medical Informatics*, *70*(2–3), 309–316. https://doi.org/10.1016/S1386-5056(03)00040-6

Sohn, Y. W., & Doane, S. M. (2004). Memory processes of flight situation awareness: Interactive roles of working memory capacity, long-term working memory, and expertise. *Human Factors*, *46*(3), 461–475. https://doi.org/10.1518/hfes.46.3.461.50392

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *The Behavioral and Brain Sciences*, *23*(5), 645–665. https://doi.org/10.1017/S0140525X00003435

Stanton, N. A., Salmon, P. M., Walker, G. H., Hancock, P. A., & Salas, E. (2017). State-of-science: Situation awareness in individuals, teams and systems. *Ergonomics*, *60*(4), 449–466. https://doi.org/10.1080/00140139.2017.1278796

Wei, H., Zhuang, D., Wanyan, X., & Wang, Q. (2013). An experimental analysis of situation awareness for cockpit display interface evaluation based on flight simulation. *Chinese Journal of Aeronautics*, *26*(4), 884–889. https://doi.org/10.1016/j.cja.2013.04.053

Wickens, C., Prinet, J., Hutchins, S., Sarter, N., & Sebok, A. (2011). Auditory-visual redundancy in vehicle control interruptions: Two meta-analyses. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *55*(1), 1155–1159. https://doi.org/10.1177/1071181311551241

Wickens, C. D. (1976). The effects of divided attention on information processing in manual tracking. *Journal of Experimental Psychology. Human Perception and Performance*, *2*(1), 1–13. https://doi.org/10.1037//0096-1523.2.1.1

Wickens, C. D., Gutzwiller, R. S., Vieane, A., Clegg, B. A., Sebok, A., & Janes, J. (2016). Time sharing between robotics and process control: Validating a model of attention switching. *Human Factors*, *58*(2), 322–343. https://doi.org/10.1177/0018720815622761

Wood, R. E. (1986). Task complexity: Definition of the construct. *Organizational Behavior and Human Decision Processes*, *37*(1), 60–82. https://doi.org/10.1016/0749-5978(86)90044-0

Young, M. S., Brookhuis, K. A., Wickens, C. D., & Hancock, P. A. (2015). State of science: Mental workload in ergonomics. *Ergonomics*, *58*(1), 1–17. https://doi.org/10.1080/00140139.2014.956151

Zhao, Y. (2021). Modeling method of combat task based on attribute extraction. *Ship Electronic Engineering*, *41*(04), 74–78. https://doi.org/10.3969/j.issn.1672-9730.2021.04.018

Zheng, Y., Lu, Y., Wang, Z., Huang, D., & Fu, S. (2015). Developing a measurement for task complexity in flight. *Aerospace Medicine & Human Performance*, *86*(8), 698–704. https://doi.org/10.3357/AMHP.4189.2015

About the authors

Xing Pan received the PhD degree in systems engineering in 2005 from Beihang University, Beijing, China. He has been a Professor with the School of Reliability and Systems Engineering, Beihang University since 2020. His research interests include reliability engineering and system risk analysis.

Liuwang Sun received the BS degree in aircraft quality and reliability in 2021 from the School of Reliability and Systems Engineering, Beihang University, Beijing, China, where he is currently pursuing the MS degree in safety science and engineering. His research interests include human reliability and uncertainty analysis.

Song Ding received the BS degree in safety science and engineering in 2020 from School of China University of Geosciences, Beijing, China. He is currently pursuing the PhD degree in Beihang University, Beijing, China. His research interest includes risk assessment and human factors analysis.

Dujun Zuo received the MS degree in 2022 from the School of Reliability and Systems Engineering, Beihang University, Beijing, China. He is currently an engineer at Institute of System Engineering, China Academy of Engineering Physics, Mianyang, China. His research interests include risk assessment and uncertainty analysis.

Appendix

Appendix A. Questions pool of SAGAT.

No.	SAGAT question	SA level
1	Were F1-F4 all in the normal range during the time period just mentioned?	Perception level
2	Was the last light anomaly F5 or F6?	
3	Was Gauge F1 in an abnormal state at the moment before the experiment was interrupted?	
4	Did the recent gauges abnormal occur at the top or bottom side?	
5	What was the TRACK in automatic mode or manual mode before the experiment was interrupted?	
6	Did the operation mode of the TRACK change during the time period just now?	
7	Did you notice immediately when the TRACK switched to manual mode?	
8	Was the closest voice call to the end of the experiment from own channel or another channel?	
9	Has the COM2 band been selected during this time period?	
10	Were there any calls from other channels that occurred during the time period just now?	
11	What is the current fuel level of primary tank A and B respectively?	
12	What is the current fuel level of secondary tank C and D respectively?	
13	Which valve is currently red in color?	
14	What was the last faulty valve?	
15	Did valve 2 fail in the time period just now?	
16	Has the main tank been less than 2000 units in that time period?	
17	Does the red light come on at the same time when F1 or F2 is abnormal?	Comprehension level
18	Does the green light go off at the same time when F3 or F4 is abnormal?	
19	Are you aware of the fact that valve failures do not occur during the manual mode of the TRACK?	
20	Is the sensitivity of the two manual mode operations in the TRACK the same during the time period just now?	
21	Is operation sensitivity LOW, MEDIUM, or HIGH in the manual mode of the TRACK during this task?	
22	Do you know the meaning of the voice call when the object is NASA504?	
23	Is the timer synchronized with the call sign?	
24	Are you aware of a communication task occurring during manual mode for tracking tasks?	
25	What do you need to do when tank A and Tank B are below 2500?	
26	What do you need to do when the level of tank B is too low?	
27	Did you find that partial valve fail did not affect the fuel level of the primary tank?	
28	What do you need to do when tank A is below 2500 and tank B is above 2500?	
29	What happens when you click F1-F4?	Projection level
30	What happens to its state when you click F5 and F6?	
31	Is there any rule in the occurrence sequence of abnormal states of F5 and F6?	
32	Is there a pattern in the order in which F1-F4 appear abnormal?	
33	Do you notice a light anomaly before the TRACK mode changes?	
34	Does the center of the cursor shift the direction of the center box regularly in TRACK?	
35	Is the manual mode held for the same length of time in TRACK?	
36	Is there a pattern between the order of objects of call signs?	
37	Do you know what a call sign means when it asks you to change the station to the appropriate channel?	
38	What happens when a valve fails?	
39	What happens when all the valves are open?	
40	Is there a failure that cannot ensure that main tank A or B is not lowered?	