The Assessment of Aviation Situation Awareness Based on Objective Behavior Methods

Hao Zhou, Yanyu Lu[(✉)], and Shan Fu

School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University,
Shanghai 200240, People's Republic of China
`luyanyu@sjtu.edu.cn`

Abstract. Situation awareness assessment is one of the current focuses of aviation human factors research. In response to the current demand for real-time and interpretable situation awareness measurement, this study adopts a behaviour-based assessment approach and designs a simulated cockpit interface for the experiment. The response time is assessed as the performance indicator. The experimental periods can be divided into two categories: continuous and intermittent, depending on the way in which the tasks are performed. During the intermittent periods, the state differences in situation awareness are compared and analyzed from three main perspectives: First, the comparative analysis of situation awareness between different situation conditions. Second, the comparative analysis of situation awareness between Peripheral Visual Perception and Gaze Perception. Third, the hierarchical analysis of situation awareness of gaze perception data. It is found that there are significant differences in the states of situation awareness between some different conditions, which means that the criteria for assessing situation awareness should vary with situation conditions. This study paves the way for future analysis of real-time situation awareness.

Keywords: Aviation · Situation awareness assessment · Real-time · Intermittent period

1 Introduction

According to statistics, more than 70% of commercial aircraft flight accidents are caused by human errors. With the development of modern aircraft manufacturing technology, the number of accidents caused by human factors far exceeds those caused by mechanical failure, weather and other unforeseen circumstances. Aviation human factors have therefore long played an important part in the study of aviation safety.In the aviation industry, human factor researches cover a wide range of areas in the design,construction, operation, management and maintenance of products and systems in the aviation industry,with the goal of optimizing the relationship between operators, technology and the environment in order to eliminate or reduce errors in operations.

One important source of human errors is the decision error, and loss of situation awareness is an important cause of the decision error. Situation awareness (SA) has been

studied for a long time, and many models describing it exist to date. Of all these models, Endsley's three-level model is the most developed and has been used and studied by many other researches. In this model SA consists of three levels: perception,comprehension, and projection. Perception refers to the pilot's ability to perceive key task-related elements of the surrounding environment within a certain time and space. Comprehension is based on perception and requires pilots to be able to integrate and understand perceived information based on mission requirements and operational objectives. Projection is the highest level of SA which is based on perception and comprehension. It requires the pilot to make predictions for the near future based on his or her own judgment of the environment.

Measuring SA is an important means of understanding the pilot's cognitive state with respect to the current situation. By measuring SA, it is possible not only to obtain the reasons for the pilot's human errors, but even to predict how the pilot will perform. SA measurements can be divided into two main categories: subjective methods and objective methods. The main form of the subjective approaches is the post-test self-evaluation, where each subject fills in an evaluation scale of his or her overall SA after the experiment is completed. Previous work of this type mainly includes Situation Awareness Rating Technique (SART) [1], Situation Awareness Rating Scales Technique (SARS) [2], Cranfield Situation Awareness Scale (CSAS) [3], Crew Awareness Rating Scale (CARS) [4, 5] and so on. The advantages of post-exp self-evaluation are mainly its simplicity and convenience. For the fact that it is performed after the test, it is not invasive to the experiment process. However, the disadvantages of these methods are also obvious: One reason is that sometimes the subject cannot know whether he or she is in a good or bad state of SA. Therefore, the self-evaluation method measures more of the subject's confidence in his or her SA state [6]. Moreover, the method requires the subject to recall the entire experiment, which also depends to some extent on the level of memory. Another type of subjective form is the observer's evaluation, where an expert on the sidelines evaluates the SA of the subject by observing his or her performance during the experiment. The representative method is Situation awareness behavioral rating scale (SABARS) [7]. This method is non-invasive like the self-assessment, but it has been questioned whether it measures SA [8].

Objective methods can be further divided into two main categories: direct methods and indirect methods. The representative techniques of the former are Freeze-Probe Techniques and Real-Time Probe Techniques. The most famous of the Freeze-Probe Techniques is Situation Awareness Global Assessment Technique (SAGAT) proposed by Endsley [9], which requires a subject matter expert to develop a series of relevant questions examining the three levels of SA states before the experiment. The screen is frozen at random moments during the experiment and the subject is asked to recall the immediate situation and answer the relevant questions. Then the correctness of the answers is used as an assessment of current SA.A representative method of the real-time probe technique is the Situation Present Assessment Method (SPAM) [10], which, like SAGAT, gives a question probe during the experiment, but does not freeze the screen. The response time and the content of the answer are used as an assessment of SA. SPAM is somewhat less intrusive,but the degree of reduction cannot be estimated, and it is difficult to apply to complex and dynamic environments because SA questions in

SPAM need to be formulated in real time. In addition, although the probe approach is able to measure SA in real time states, its temporal resolution is poor due to the limitation of the number of probes [11].

Indirect methods can also be divided into two categories, one of which uses the subject's physiological data to assess the state of SA. The other method uses the subject's behavior performance. There have been many studies using physiological data to assess SA, such as EEG, ECG, EYE and EDA [12]. The correlation of the results between these physiological data and other methods of measurement has been validated to varying degrees, suggesting that there is some feasibility in using physiological data to assess SA, but the use of physiological data alone lacks significant persuasive power, which is due to the complexity of interpretation of physiological data. In order to obtain high temporal resolution measures of SA in a highly dynamic aviation environment in a non-invasive manner and give interpretable accounts, behavior measure is a feasible approach. By capturing subjects' eye-movement behaviors and manipulation behaviors in real time from high sampling rate sensors, real-time situation cognitive states of subjects can be analyzed.

In this study, a simulated experimental setting is designed by integrating several basic flight elements. The behavior responses of different subjects to different events are compared to analyze the SA according to two categories of behavioral performance indicators:eye-movement and manipulation. Experimental periods are first divided into two categories according to the tasks: continuous and intermittent. In the continuous periods, subjects are required to perform tasks that requires continuous operation, while in the intermittent periods, subjects take actions only in unusual situations. Intermittent periods are the main research periods of this study, and the behavioral performance indicator of SA used in this study is response time, which has been used as a measure of SA in other studies [13]. In this paper, response time is replaced by action response time (ART), which can further be subdivided into eye movement reaction time (ERT) and comprehension response time (CRT) in order to discuss SA in a hierarchical manner. The main research content of this paper are: First, factors that affect the overall SA, including workload changed by adjusting tracing tasks and different abnormal event designs. Second, the differences between using peripheral vision (PV) and central vision (CV) to perceive events. Third, further exploration of the evaluation indicators of different levels of SA for events perceived using CV.

2   Materials and Methods

2.1   Subjects

Eight volunteers (six males and two females) from Shanghai Jiao Tong University, ranging in age from 22 to 30 (mean 25), participated in this experiment. All subjects have normal or corrected normal vision and are right-handed. Prior to the start of the formal experiment, each subject underwent a 10-min training procedure.

2.2   Materials

Hardware.   The main experimental platform for this experiment is on an 18-in. monitor with a display resolution of 1920 × 1080 and a refresh rate of 60 Hz. Other equipment

used during the experiment contains a keyboard and a joystick. The subjects' eye movement is recorded by a portable head-mounted eye-tracking device (TOBII Glasses3, see Fig. 1). Eye-movement data is recorded at 100 Hz on an SD memory card, which is then transferred to a computer for further processing and analysis, mainly for mapping of gaze points in the panorama, data pre-processing and calculation of other eye-movement metrics. The eye-movement equipment will be calibrated before the start of the experiment.

Software. The layout of common elements in the flight task, such as alarm lights, instruments, fuel management, communication and target tracking, are simulated and simplified in the study and finally are integrated into a comprehensive multitasking experimental interface by using the built-in software (see Fig. 2). The software can also be used to set and execute the relevant experimental procedures, including recording the environmental variables and the operator's keystrokes in the experiment and saving them as a readable file with a frequency of 50 Hz. The software uses Microsoft Visual Studio as the development platform, C++ as the development language, and OpenGL as the basic development library.

Fig. 1. Experiment scene.

2.3 Tasks

Target Tracking. The target tracking task module is located on the top-center side of the interface (see Fig. 3, 3). The tracking target is a free-moving circle with a blue dot in its center. Two modes exist for this task: manual and automatic. In the manual mode, the subject is asked to manipulate the joystick to control the tracking box, and the task is to keep the target circle in the center of the tracking frame as much as possible. In the automatic mode, the tracking box will track the target circle automatically and the subject can focus more on other tasks.
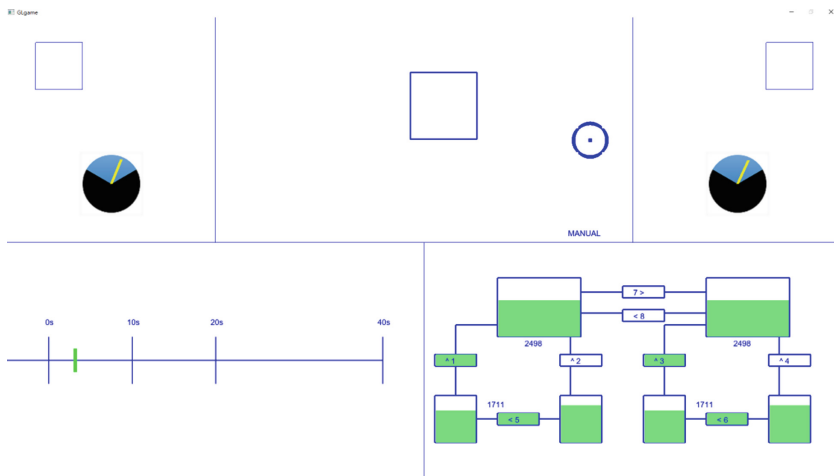
Fig. 2. Experiment software interface.

Abnormal Events Response. The abnormal event response task module is arranged on the left and right sides of the experimental interface (see Fig. 3, 1 and 2), which contains two parts, one is the alarm light event. The alarm light is usually colorless in the normal state,while in the alarm state, the light will turn red (see Fig. 4, left). The subject needs to press the corresponding button as a response, and then the abnormal event will be eliminated. Instruments are placed below the alarm lights, and each of them has a yellow pointer. The blue area indicates the normal area, while the black area is the abnormal area. The pointer oscillates in the normal zone during normal times and oscillates out of the normal zone when an abnormal event occurs (see Fig. 4, right). As with the alarm light event, the subject needs to notice the instrument abnormality event and press the corresponding button in time to eliminate the abnormal event.

Communication Response. The communication response task module is located at the bottom left of the interface (see Fig. 3, 4). In the event, response is similar to the abnormal event response. The subject has 5 s to press the corresponding button after the green bold line is slid to the '0s' position to indicate the response to the communication event. And then the event bar will be eliminated. The difference is that the event already appears in the schedule bar before reaching '0s'. By observing the position of the event on the schedule bar, the subjects can roughly determine how long it is before the communication event occurs. By observing the position of the event on the schedule bar, the subject can know the time before the event occurrence and make arrangements and plans in advance.

Fuel Management. The fuel management task module is located at the bottom right of the interface (see Fig. 3, 5).The module consists of 6 simulated fuel tanks with 8 simulated pumps. The upper two tanks are the main fuel tanks, which are consumed at a certain rate. The lower left tank is the supply tank, which feeds the main tank. The lower right-hand side is the reserve tank, which can supply both the main tank and the supply tank. The transfer of oil is controlled by eight pumps, the direction of which is marked by arrows. The flow rate of the pumps varies. At random moments some pumps will

fail (turn red) and the subject cannot operate the pump. After a certain time, the pump failure eliminates itself. The central task of the module is to maintain the oil volume of the left and right main tank at 2000 to 3000. In addition, the difference between the oil volume of the left and right main tank should not exceed 50.
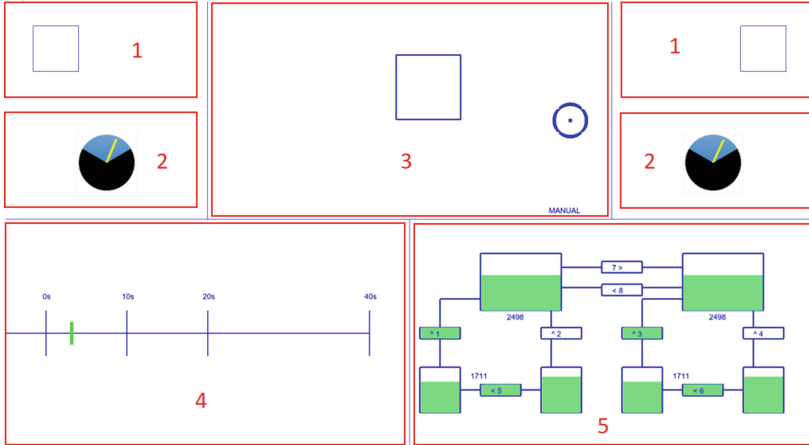

Fig. 3. Tasks area division.

Fig. 4. Alarm light and instrument events

2.4 Experiment Design

In this study, the experiment is divided into two types of time periods (see Fig. 5). The first type is the continuous period, in which subjects perform tasks that are continuous. In this experiment the communication task, the target tracking task and the fuel management task belong to this type. Although the communication task requires the subjects to respond after the bar reaches '0s', the subject has already looked ahead before. So the subject's normal responses are not caused by the '0s' events. The second type is the intermittent period, in which subjects complete tasks that do not require attention at all times, but only when an abnormal event occurs. In this experiment, the alarm light and instrument monitoring tasks belong to this type.

The experiments are arranged around the tracking module, with the tracking part alternating between automatic and manual modes, each lasting about 4 min. The number

of alarm light events and instrument events are set at the same level, but the moment of occurrence and sequence are randomized. The appearance of communication events is a Poisson process with a time constant of 20 s. The appearance of pump failures is also a Poisson process with a time constant of 15 s. The repair time meets a uniform distribution of 10 to 20 s. While performing the continuous tasks, Subjects are subjected to a number of intermittent tasks and they need to response to them in time.In the automatic mode, the continuous tasks include oil management and communication. In the manual mode, in addition to the two tasks above, the continuous tasks also include target tracking. In both two modes, the abnormal events response tasks are included in intermittent tasks. The experiment specifies the highest priority for the disposal of such tasks. The overall prioritization is arranged as follows: alarm light > instrument > communication > fuel management = target tracking.

This study focuses on the research of the SA states during intermittent periods. There are several issues that can be studied in these periods: First, factors that affect the overall SA, including workload and different abnormal event designs. Second, the differences between using peripheral vision and central vision to perceive events. Third, further exploration of the evaluation indicators of different levels of SA for events perceived using central vision.
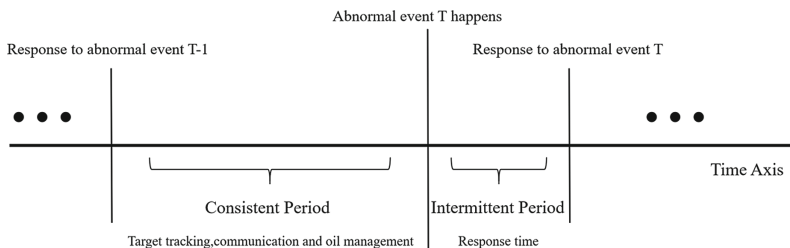

Fig. 5. Experimental periods division.

Effect of Subjects' Workload and Interface Design on Overall SA. There is a certain correlation between workload and SA. In addition, different interface designs also have effect on SA. As a result, the effects of subjects' workload is investigated through using the automatic or manual tracking mode, while the factor of interface design is studied through comparing alarm light and instrument abnormal event.

Central Vision vs Peripheral Vision. Through the eye-movement study, it is found that events during intermittent periods do not necessarily need to be perceived by subjects using central vision (CV) every time. Peripheral visual (PV) is common during such periods. Therefore, it is meaningful to make comparison between CV and PV. The percentage of total events perceived by PV and CV will be analyzed. Moreover, the action response time (ART) of subjects using PV and CV will also be compared.

Hierarchical Analysis of SA. According to the Endsley's SA model, SA is constructed by three levels:perception, comprehension and projection. To understand the SA state

more deeply, it is of great import importance to make hierarchical analysis of SA. Although it is hard to find the distinction between comprehension and projection, finding the boundary between perception and the other two is relatively simple. In this study, only those events perceived by CV can make further analysis because the hierarchical analysis needs the eye reaction time (ERT) which is the evaluation indicator of perception and the comprehension response time (CRT) which the evaluation indicator of comprehension and projection.

2.5 Data Processing

Response Time. In this study, response time is used as the evaluation indicator. To make deeper research on SA, the response time is subdivided here. Firstly, the key response time of the subjects to events is named as action response time (ART), which is the indicator of the overall SA. Then, to distinguish the SA1 (perception) and SA2/3 (comprehension and projection), ART will be further subdivided into eye reaction time (ERT) and comprehension response time (CRT). ERT, which refers to the time required from the occurrence to the point of eye movement fixation to the target AOI, only happens in those events perceived by central vision (CV). Similar to ERT, CRT which refers to ART minus ERT is also can only be extracted in CV events (see Fig. 6).

In the past researches on SA using the eye tracker, eye movement indicators such as fixation duration, fixation counts and saccade frequency are widely used. Eye reaction time (ERT) is rarely used because few researches use response time for SA hierarchical analysis. ERT is also an important indicator for SA1 (perception). CRT is another indicator which is seldom used while it can reflects the SA2/3 state (comprehension and projection) of the subjects from the side.
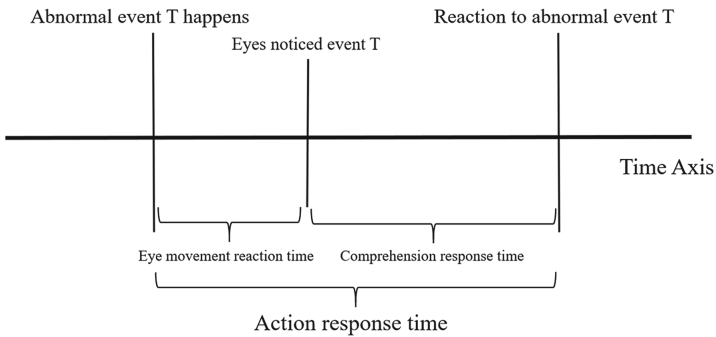

Fig. 6. Relationship between ART, ERT and CRT.

Statistical Methods. SPSS 25 is used as the main analysis tool in this study. Due to data volume limitations and the non-normality of some of the distributions, the statistical tests used in this study included the paired t-test, Wilcoxon signed-rank test, and Kruskal-Wallis H-test (K-W test). The normality test used is the Shapiro-Wilk method.

3   Results

The workload variation in this experiment comes from that whether the tracking mode is automatic or manual. The event type variables also contain two types: alarm light events and instrument events. Combining the workload and event types, four types of data are obtained, which are alarm light-manual tracking (Alt-Man), alarm light-automatic tracking (Alt-Aut), instrument-manual tracking (Ins-Man), and instrument-automatic tracking (Ins-Aut). The following content of the study results will be analyzed around the four types of data for comparison.

3.1   Effect of Subjects' Workload and Interface Design on ART

It is necessary to perform normality test on all subjects' data in the four types of data before comparing them.And then in each type, to test whether there is statistical variability in the distribution of all subjects' data, which determines whether subjects' data can be used as the analysis data. It is tested that all subjects did not satisfy the normality assumption in each type, and that the overall distribution test between subjects finds significant differences. This established that the statistical comparison method uses nonparametric methods. What's more, the data of all subjects cannot be directly compared as a whole to compare the differences between the two types of data.

To address the inter-subject differences, the subjects are divided into two categories overall as follows, one with longer overall ART (subject 3, 7, 8) and the other with shorter (subject 1, 2, 4, 5, 6):

1) The respective mean values of each subjects' data is calculated and sorted.
2) Data is divided into two groups with same amount of subjects according to the sorting order of mean values.
3) If subject differences still exist in one group, the group with differences will choose the data of the subject whose mean value is the closest to that of all data and transfer it to the other group.
4) Perform the above until there are no inter-subject differences within either group.
5) If the data of a subject cannot be put into any group to eliminate the differences, then this part of the data will be discarded.

It is found that in four conditions the subject can be divided into two groups with the same subjects and the final comparison can be seen below (see Table 1). After screening, the distribution differences are analyzed again within the respective groups, and it is found that there is no difference in the distribution of ART between the subjects in these two categories. Therefore, the ART under different conditions can be compared within the high ART group and the low ART group, respectively.

The Effect of The Workload with The Same Abnormal Event Type. The following figure compares the difference between the high and low ART group for different workloads respectively (see Fig. 7). It is shown that the change in workload in this experimental design does not cause significant differences in subjects' ART in the same type of event whether in the high or low ART group (Alt: LART: p = 0.290, HART: p = 0.485; Alt: LART: p = 0.693, HART: p = 0.187).

Table 1. ART comparison of the high and low ART group (unit: ms)

	High ART group ART mean (std)	Low ART group ART mean (std)
Alt-Man	1075 (316)	840 (160)
Alt-Aut	1236 (990)	875 (145)
Ins-Man	3159 (3792)	1165 (412)
Ins-Aut	1880 (1079)	1224 (880)

The Effect of Different Abnormal Event Types with The Same Workload. The following figures compare the ART variability of response times for the high and low reaction time groups for different events,respectively (see Fig. 7). Unlike the fixed event type, under the same workload, subjects in the high and low reaction time groups had significant ART variability between alarm light and instrument events. (Man: LART: $p = 0.001$, HART: $p = 0.000$; Aut: LART: $p = 0.006$, HART: $p = 0.000$), with significantly higher ART for instrument than alarm light.
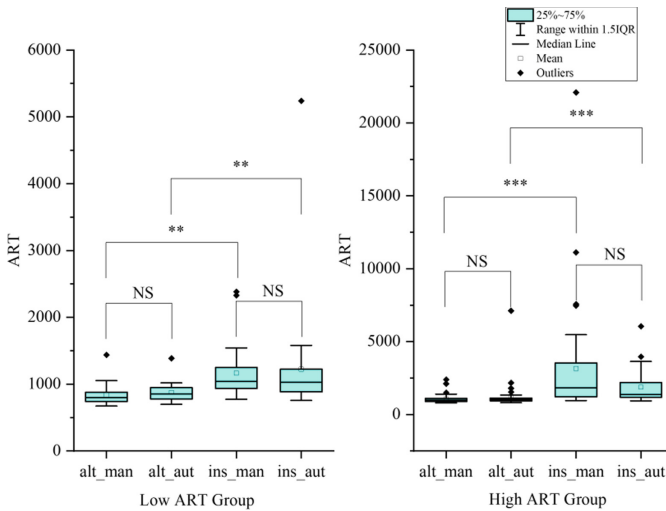

Fig. 7. ART comparison between different conditions in low and high ART group. (*: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$, NS: $p > 0.05$)

3.2 Central Vision vs Peripheral Vision

It is found that not all event responses are perceived using central vision (CV). This part of the study will first analyze whether there is variability in the proportion of events perceived using peripheral vision (PV) versus those perceived using CV, and

subsequently compare ART for events perceived using PV versus those perceived using CV.

Proportion of CV vs Proportion of PV. The number of PV response proportions and CV response proportions are the same for each subject, and the data under the same condition does not satisfy normality. So Wilcoxon's signed rank test for paired samples is performed directly. The results of the test are found that (see Table 2) when the occurrence is an alarm light event, the proportion of PVP used by the subjects is significantly higher than that of FP, regardless of the tracking pattern. While there is always no significant difference between the two proportions when an abnormal instrument event occurs.

Table 2. Matching test results of the proportion of PVP and FP under different conditions

	FP mean (Std)	PVP mean (Std)	p
Alt-Man	0.34 (0.15)	0.66 (0.15)	0.019
Alt-Aut	0.28 (0.22)	0.71 (0.22)	0.026
Ins-Man	0.56 (0.09)	0.44 (0.09)	0.102
Ins-Aut	0.53 (0.36)	0.46 (0.36)	0.815

Differences of ART between Events Perceived by CV and PV. Before formally comparing ART, it is still necessary to test for normality and subsequently for distributional differences in the data for each subject in each of the four categories of CV and PV separately. It is found that most of the subjects' data in the four categories does not meet the normality assumption, and then a non-parametric K-W test will be used next. And after doing the between-subjects difference test for each of the four categories of cases, it is found that there is no significant difference in the data between subjects under automatic tracking, while there is a significant difference between subjects under manual tracking. Using a similar method as in Sect. 3.1, the group of subjects is divided into two parts in the analysis of manual tracking according to the method in Sect. 3.1, the low ART group (subject 2, 4, 5, 6) and the high ART group (subject 1, 3, 7, 8) (comparison can be seen in Table 3). The data distribution between individual subjects within the same group is tested to be insignificantly different after regrouping.

Table 3. ART comparison of the high and low ART group (Unit: ms)

	High ART group ART mean (Std)	Low ART group ART mean (Std)
CV-Alt-Man	1300 (420)	913 (194)
PV-Alt-Man	1043 (270)	825 (99)
CV-Ins-Man	3450 (2695)	1434 (1087)
PV-Ins-Man	4347 (6808)	1331 (790)

After statistical analysis, it can be found that during automatic target tracking, all subjects have significantly lower ART using PV than that using CV, regardless of whether the occurring event is the alarm light or the instrument (see Fig. 8, Alt-Aut: p = 0.001; Ins-Aut: p = 0.002). In contrast, there are different statistical patterns between the low-ART and high-ART groups when manually tracking the target. There is no significant difference of ART between PV and CV in the low ART group regardless of the event that occurs (see Fig. 8, Alt-Man: p = 0.227; Ins-Man: p = 0.439), while there is a significant difference in the high ART group for both events that occurs (see Fig. 9, Alt-Man: p = 0.004; Ins-Man: p = 0.040).
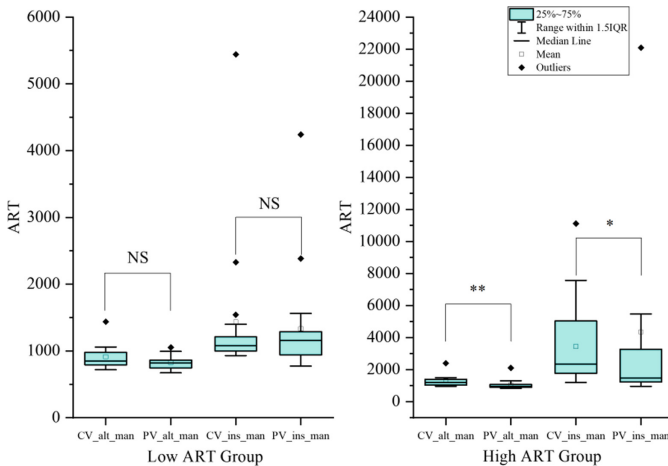

Fig. 8. ART comparison in low and high ART group between CV and PV under manual tracking.

3.3 Hierarchical Comparison of ART

For the experimental Intermittent periods in this study, not all data can be analyzed hierarchically, mainly because of the presence of PV. In this part of the results study, ART using CV will be compared across conditions and used as an indicator for different layers of SA.

Before hierarchical comparison, the ART of CV data is compared. The test for differences in distribution between subjects in the four categories is first done and found to be significantly different. So here too it is necessary to first divide the subjects into two groups with different ART (the classification method is the same as that in Sect. 3.1 and comparison can be seen in Table 4). The low ART group has subject 2, 4, 5, 6 and the high one has subject 1, 3, 7, 8. The test after regrouping finds that there are no significant differences between subjects within the groups. Using the non-parametric K-W test, the paired analysis of variance shows that in the high ART group, there is a significant difference in ART corresponding to different event types under fixed workload (see Fig. 10, Alt-Man/Alt-Aut, p = 1.000; Ins-Man/Ins-Aut, p = 0.380; Alt-Man/Ins-Man, p = 0.007; Alt-Aut/Ins-Aut, p = 0.000). While there is no significant difference in
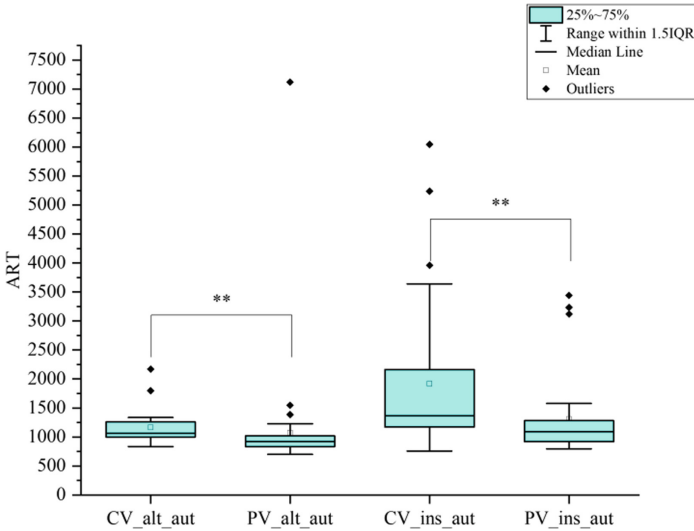
Fig. 9. ART comparison between CV and PV under automatic tracking.

ART corresponding to different workloads under fixed event types, which is the same as the results in Sect. 3.1. While slightly different from the previous results in the low ART group (see Fig. 10, Alt-Man/Alt-Aut, p = 0.882; Ins-Man/Ins-Aut, p = 1.000; Alt-Man/Ins-Man, p = 0.030; Alt-Aut/Ins-Aut, p = 0.345), only in the fixed workload type, there is a significant difference between the ART of alarm lights and instrument events when tracking manually, and pairwise comparative variability is not present for the remaining cases.

The statistical comparison between ERT and CRT likewise requires analyzing whether there are differences in the distribution of data for all subjects in different types of time periods, and after analysis it can be learned that there is no significant difference between ERT and CRT for all subjects under different conditions, so the data for all subjects are directly used as statistical data for comparison.

The following discussion is whether there is a significant difference between ERT and CRT at different periods. From the figure (see Fig. 11, Alt-Man/Alt-Aut, p = 0.381; Ins-Man/Ins-Aut, p = 0.575; Alt-Man/Ins-Man, p = 0.777; Alt-Aut/Ins-Aut, p = 1.000), it can be seen that in the comparison of pair types with fixed workload or fixed event type, there is no significant difference. And in terms of CRT's test (see Fig. 11, Alt-Man/Alt-Aut, p = 0.432; Ins-Man/Ins-Aut, p = 0.331; Alt-Man/Ins-Man, p = 0.762; Alt-Aut/Ins-Aut, p = 1.000), there is no significant difference in all types of periods overall.

Table 4. ART comparison of CV perceived events between the high and low ART group (Unit: ms)

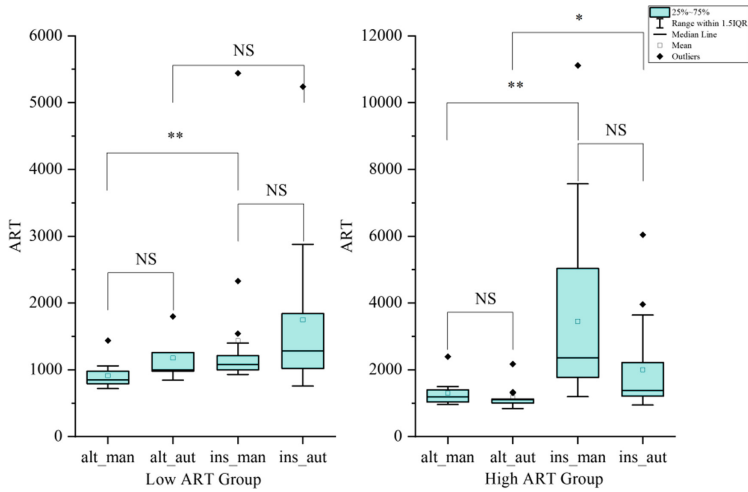|	High ART group ART mean (Std)	Low ART group ART mean (Std)
Alt-Man	1300 (420)	913 (193)
Alt-Aut	1162 (331)	1160 (379)
Ins-Man	3450 (2695)	1434 (1087)
Ins-Aut	1999 (1215)	1749 (1294)

Fig. 10. Comparison of ART of events perceived by CV between different conditions in low and high ART group.
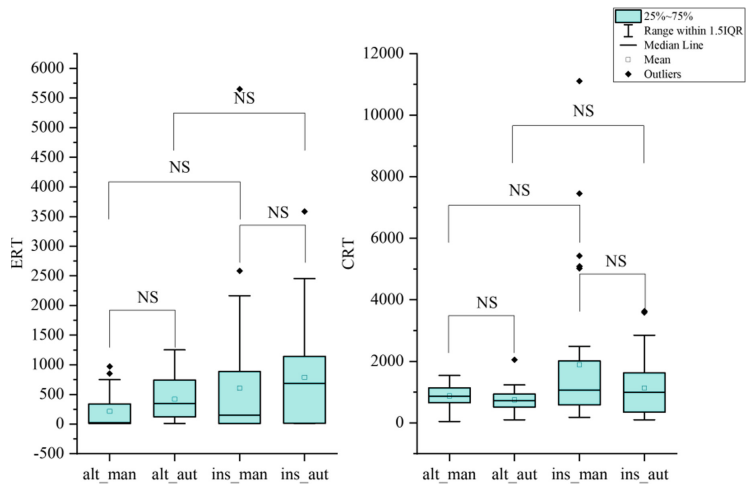

Fig. 11. Comparison of ERT and CRT of events perceived by CV between different conditions.

4   Discussion

4.1   Effect of Subjects' Workload and Interface Design on Overall SA

ART is used as an indicator to evaluate the overall SA. The longer the time, the lower the level of the overall SA state of the subjects. The analysis of the variability of the ART in the results corresponds to the division of the overall SA status into two groups: the high SA group and the low SA group.

The analysis of the results shows that under the condition of fixed event type, the workload changed in this experiment does not cause a significant difference in SA, regardless of the group of subjects with a high level of overall SA or a low level. According to the inverted U-shaped relationship curve between workload and performance [14], it indicates that the workload levels in this experiment are basically in the central part, so the changes in workload does not trigger significant differences in overall SA.

When the workload is fixed and the overall SA status is significantly different between the high SA and low SA groups under changing event types, the overall SA is better for the alarm light than the instrument. This variability mainly comes from the degree of significance of the appearance of the two events. The alarm light occurs from a white background color to a red alarm state that allows subjects to respond more quickly than the instrument's change in pointer position, and thus achieve a better SA state.

4.2   Central Vision vs Peripheral Vision

The results is shown that there is no significant difference in the proportion of subjects using CV versus PV when there is an instrument event. However, during the alarm light event, subjects use PV significantly more frequently than CV. The alarm light events require less perceptual and cognitive resources from subjects, so subjects tend to use less attention to response to the event. Whereas for the instrument event, the increased of perceptual and cognitive demands makes the subjects' choice between using CV and PV much less variable. As is in Sect. 4.1, ART is used here as an indicator of overall SA status to compare and analyze the overall SA state of CV and PV in different conditions.

According to the results it can be found that the inter-subject differences differ depending on whether the tracking is manual or automatic. In the automatic case, there are no differences in the inter-subject distribution, while in the manual case the differences are significant. This suggests that in the lower workload all subjects are likely to be in a more similar state of SA. While in the higher workload, subjects show significant differences in SA. It can be inferred that workload and SA show a non-linear negative relationship, and that the degree of non-linearity is determined by the subjects' individual difference.

In the automatic tracking condition, there are significant differences in the ART contrast between all subjects using CV and PV. In this case, the subjects are generally in a more relax state, and the events to which they response using PV indicates a better mastery of the surroundings (corresponding to a good SA), whereas the responses using CV are likely due to a sudden attention to the event after a long period of inattention, corresponding to a worse SA state.

In the manual situation, there is no significant difference in ART between CV and PV for the low ART group, whereas there is for the high ART group. This is mainly due to the fact that the low ART group (better SA state group) tends to be more of a personal random choice using the two types of perception types under higher load rather than a choice made due to being influenced by the current situation. So the SA states corresponding to the two types of perception modalities do not differ significantly in this category of subjects. In contrast, the SA state of the high ART group (worse SA state group) under high load resembles its SA state under low work load, so the use of PV is more likely to represent a better mastery of their environment, and the use of CV corresponded much to a sudden response after ignoring the event for a longer period of time.

4.3   Hierarchical Analysis of SA

It is possible to perform SA hierarchical analysis in the data where CV occurs. ERT is used as an indicator for SA1 analysis. CRT, which can be calculated by subtracting ERT from ART, is used as an indicator for understanding. The lower the reaction time, the better the corresponding SA state. Due to the hierarchical relationship of SA, a good SA2 presupposes a good SA1. Thus the CRT can not be directly used as an indicator of SA2. The final SA2 state is obtained by combining CRT on the basis of the evaluation of SA1.

In the analysis of the results, it can be found that overall SA differs significantly in the condition of fixed workload and changed event type, but does not differ significantly when assigned to either the perception or the comprehension layer. This suggests that the SA differences that occur between alarm light and instrument do not originate in the perceptual or comprehension layer alone. The source of variability is more complex.

5   Conclusion

This study use a behavior-based approach to assess SA in real time and interpretability in the current aviation environment. A simulated experimental interface for different flight elements is designed. The experimental periods are divided into continuous and intermittent periods according to the types of tasks performed. This study focuses on the SA state during intermittent periods and uses reaction time as the assessment indicator of SA.

The results revealed that the interface design has significant effect on the overall SA. The alert light with more obvious alarm signal is more possible for subjects to get good SA state than the instrument. The comparison between Peripheral vision and central vision show that people are more used to use PV to perform tasks with conspicuous appearance than CV and different workload causes different overall SA pattern. In general, those who use PV is more possible to have better SA under lighter workload. Finally, the hierarchical analysis of SA shows that the significant differences can be found in overall SA while not in individual level.

In conclusion, the results in the study demonstrated that eye-movement and behavioral performance indicators can be used to assess subjects' SA states in real time. The

SA criterion representing good or bad state in different types of conditions should also be different, and more data is still need to be obtained for further analysis to get a clear threshold for SA state segmentation.

References

1. Taylor, R.M.: Situational awareness rating technique (SART): the development of a tool for aircrew systems design. In: Proceedings AGARD AMP Symposium Situational Awareness Aerospace Operations, pp. 1–17 (1990)
2. Waag, W.L., Houck, M.R.: Tools for assessing situational awareness in an operational fighter environment. Aviat. Space Environ. A13–A19 (1994)
3. Denney, K.: Cranfield situation awareness scale: user's manual. Applications Psychology Unit College Aeronautics, Cranfield University, Cranfield, UK (1997)
4. McGuinness, B., Foy, L.: A subjective measure of SA: the crew awareness rating scale (CARS). In: Proceedings 1st human performance, Situation Awareness, automation conference, Savannah, GA, USA, pp. 286–291 (2000)
5. McGuinness, B., Ebbage, L.: Assessing Human Factors in Command and Control: Workload and Situational Awareness Metrics. BAE Systems Advanced Technology Centrer, Bristol, UK (2002)
6. Endsley, M.R.: Toward a theory of situation awareness in dynamic systems. Hum. Factors **37**(1), 32–64 (1995)
7. Matthews, M.D., Beal, S.A.: Assessing Situation Awareness in Field Training Exercises. US Military Aca, West Point, NY, USA (2002)
8. Nguyen, T., Lim, C.P., Nguyen, N.D., Gordon-Brown, L., Nahavandi, S.: A review of situation awareness assessment approaches in aviation environments. IEEE Syst. J. **13**(3), 3590–3603 (2019)
9. Endsley, M.R.: Measurement of situation awareness in dynamic systems. Hum. Factors **37**(1), 65–84 (1995)
10. Durso, F.T., Hackworth, C.A., Truitt, T.R., Crutchfield, J., Nikolic, D., Manning, C.A.: Situation awareness as a predictor of performance for en route air traffic controllers. Air Traffic Control Quart. **6**(1), 1–20 (1998)
11. Bacon, L.P., Strybel, T.Z.: Assessment of the validity and intrusiveness of online-probe questions for situation awareness in a simulated air-trafficmanagement task with student air-traffic controllers. Saf. Sci. **56**, 89–95 (2013)
12. Zhang, T., Yang, J., Liang, N., et al.: Physiological measurements of situation awareness: a systematic review. Hum. Factors (2020)
13. Endsley, M.R.: The divergence of objective and subjective situation awareness: a meta-analysis. J. Cogn. Eng. Decis. Making (2020)
14. Schnell, T., Cornwall, R., Walwanis, M., Grubb, J.: The quality of training effectiveness assessment (QTEA) tool applied to the naval aviation training context. In: Schmorrow, D.D., Estabrooke, I.V., Grootjen, M. (eds.) Foundations of Augmented Cognition. Neuroergonomics and Operational Neuroscience. LNCS (LNAI), vol. 5638, pp. 640–649. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-02812-0_73