DOT/FAA/AM-89/1

Office of Aviation Medicine
Washington, D.C. 20591

A Comparison of Detection Efficiency on an Air Traffic Control Monitoring Task with and without Computer Aiding

Richard I. Thackray, Ph.D. and
R. Mark Touchstone

Civil Aeromedical Institute
Federal Aviation Administration
Oklahoma City, OK 73125

January 1989

This document is available to the public through the National Technical Information Service, Springfield, Virginia 22161

U.S. Department of Transportation
Federal Aviation Administration

Technical Report Documentation Page

AD-A206 422

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
DOT/FAA/AM-89/ 1		

4. Title and Subtitle	5. Report Date
A COMPARISON OF DETECTION EFFICIENCY ON AN AIR TRAFFIC CONTROL MONITORING TASK WITH AND WITHOUT COMPUTER AIDING	JANUARY 1989
	6. Performing Organization Code
	8. Performing Organization Report No.
7. Author(s) Richard I. Thackray and R. Mark Touchstone	
9. Performing Organization Name and Address FAA Civil Aeromedical Institute P. O. Box 25082 Oklahoma City, OK 73125	10. Work Unit No. (TRAIS)
	11. Contract or Grant No.
	13. Type of Report and Period Covered
12. Sponsoring Agency Name and Address Office of Aviation Medicine Federal Aviation Administration 800 Independence Avenue, S.W. Washington, D.C. 20591	
	14. Sponsoring Agency Code

15. Supplementary Notes

This work was performed under tasks AM-C-86-PSY-106, AM-D-87-PSY-106 and AM-D-88-HRR-106.

16. Abstract

Future levels of air traffic control automation plan to incorporate computer-aiding features designed to alert the controller to upcoming problem situations by displaying information that will identify the situation and suggest possible solutions. Concerns have been expressed that reliance on such aids may lead to a reduced capacity to detect and respond to infrequent failures of the automation. The present study employed a simulated ATC monitoring task with a computer-aiding feature designed to detect possible aircraft conflict situations. The ability of subjects to detect occasional failures of the computer-aiding feature in detecting problem situations was compared with detection efficiency for these same situations when no computer aiding was provided. The hypothesis that alertness would be lower and detection less efficient with computer aiding than when no aiding was employed was not supported. Applications and limitations of the findings to the problem of complacency in automated systems are discussed.

17. Key Words	18. Distribution Statement
Air traffic control, attention, automation, complacency, computer aiding, monitoring, performance, vigilance	Document is available to the public through the National Technical Information Service, Springfield, Virginia 22161.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	10	

A COMPARISON OF DETECTION EFFICIENCY ON AN AIR TRAFFIC CONTROL MONITORING TASK WITH AND WITHOUT COMPUTER AIDING

The FAA is currently in the midst of a 20-year plan to modernize the nation's air traffic control (ATC) system. This plan, called the National Airspace System (NAS) Plan, includes a program consisting of a series of automated functions and enhancements to the present enroute ATC system that are collectively referred to as AERA (Automated En Route ATC). AERA will be implemented in stages, with each stage incorporating increasingly sophisticated computer aids (3). It is generally conceded that these aids will gradually change the role of the controller; he or she will increasingly become less of an active participant in control activities and more of a monitor of a largely computer-controlled system. If we make the reasonable assumption that future levels of automation will be extremely reliable with minimal need for routine controller intervention, it may become quite difficult to maintain controller involvement, alertness, and readiness to back up the system in response to unanticipated problems. At the same time, system planners emphasize that the controller cannot afford to become complacent; AERA cannot be relied upon to detect all possible problem situations and the controller must remain alert for such failures (1). The extent to which a controller performing in a relatively passive monitoring role will be able (or willing) to sustain a level of attention necessary to detect these failures if "misses" by the automation are quite infrequent and possibly quite subtle is not known. The present study was conducted to provide some initial information bearing on this question.

One aspect common to both the AERA 1 and 2 levels of ATC automation is automated detection and notification of possible problem situations, of which a potential aircraft-aircraft conflict situation would be one example (1). Our ATC monitoring task described in several previous publications (5,6) was modified to incorporate a computer-aiding feature that was programmed to detect most, but not all, possible conflict situations. Possible conflicts, when detected by the automation, were signalled to the subject through an advisory alert message presented in the lower left corner of the radar display. The ability of subjects to detect occasional misses by the computer-aiding feature was compared with detection efficiency for these same situations when no computer aiding was provided. It was hypothesized that subjects receiving computer aiding would become increasingly dependent on the aid, with a corresponding reduction in effort expended in searching for possible conflict situations. Thus, detection times and miss rates (i.e., detection efficiency) for situations that the computer aid failed to detect were predicted to exceed detection times and miss rates for those same situations when no computer aiding was provided. Further, it was anticipated that the reduction in detection efficiency resulting from computer aiding would be greater toward the end of a monitoring session than at the beginning. In addition to performance, subjective measures of strain, tiredness, attentiveness, irritation, boredom, effort, and difficulty were obtained.

MATERIALS AND METHODS

Subjects: Forty men and women, all paid university students, were randomly assigned to the two experimental conditions. Subjects ranged in age from 18 to 29 years, had 20/20 uncorrected vision, were nonsmokers, and had neither prior experience with the task used nor previous ATC training. None were currently taking any prescription medication on a regular basis.

Apparatus and Task Description: The basic experimental equipment consisted of a Digital Equipment Corporation (DEC) VS11 19-in (49-cm) graphics display, keyboard, and joystick, all of which were interfaced with a VAX 11/730 computer (DEC). The computer was used both to generate input to the display and to process subject responses. The VS11 was incorporated into a console designed to closely resemble an ATC radar unit. Two diagonal, nonintersecting flight paths were located on the display, along which aircraft targets could move in either direction. A given aircraft's location was displayed as a small rectangle on the flight path, and an adjacent alphanumeric data block identified the aircraft and gave its altitude and groundspeed. Aircraft were updated in position and any change in alphanumerics every 6 s. There were 8 aircraft on each flight path at all times; as one left the screen, another appeared. A typical target pattern as displayed in the subject's console is shown in Fig. 1.

FIGURE 1. SUBJECT'S CONSOLE WITH TYPICAL TARGET PATTERN DISPLAYED.

Two types of stimulus change (critical events) were employed. The duration of each type was 90 s; if a subject failed to detect a critical event within this 90-s period, the data block containing the change reverted to its previous state.

The first type of critical event was readily detectable and consisted of the appearance of three Xs in place of the three altitude numbers in a given data block. Subjects were told that this replacement of an altitude value by the three Xs signified that a malfunction had occurred in an aircraft's altitude transponder, resulting in a loss of altitude information. Upon detection of such an event, subjects were told to press a designated button on the console, move a joystick-controlled cursor over the data block containing the critical event, and to press another button located on the joystick control unit. This last response corrected the malfunction by replacing the three X's with the previous altitude value.

The second type of critical event was considerably more difficult to detect since it was not immediately apparent. This event was the occurrence of two aircraft at the same altitude on the same flight path. (Subjects assigned to the advisory alert condition received computer assistance in detecting this type of event.) Any time such an event was detected, whether by the automatic detection feature or by the subject, a designated console button was pressed. Subjects were then required to determine whether the two aircraft were moving towards each other, away from each other, or in the same direction. On the basis of this determination, subjects decided either that a conflict was likely or that no conflict was possible, and a console button appropriate to the decision made was pressed. Following a "conflict" decision, the cursor was positioned over one of the two conflicting aircraft, and the joystick control button was pressed. This caused the computer to assign a new altitude value to one of the two conflicting aircraft and display this value, along with the aircraft's identification in a box at the lower left of the screen. Subjects then verified that the computer-assigned altitude did not result in a conflict with some other aircraft on the flight path. If no new conflict was created, a keyboard entry was made that assigned the new altitude value to one of the two previously conflicting aircraft. (Although subjects were led to believe that a computer-assigned altitude might occasionally result in a conflict with some other aircraft, in actuality this never occurred.)

Whenever a "no conflict" response was made, no further action ensued since no change in altitude was required. Subjects were told that the altitude of one of the two nonconflicting aircraft would eventually change to some other value (this change always occurred 60 s after the no conflict response was made) and that they had to remember that they had responded to this particular pair of aircraft. If they failed to remember and reponded a second time, a memory error was recorded. (It should be noted here that the present study was concerned only with detection behavior (i.e., miss rates and detection times). The task behaviors described in this paragraph and in preceding paragraphs that deal with evaluation, implementation, or memory aspects of the task are included in this section only for the purpose of providing the reader with information concerning the total task performed by subjects. The interested reader is referred to a previous study (6) where data pertaining to these additional aspects of the task were analyzed and discussed in detail.)

Nine critical events occurred in each 30 min of task performance, with no more than one event present at any given time. Of these nine events, three were Xs, three were conflicting altitude changes, and three were nonconflicting changes. These events were arranged in a quasi-random order with the restriction that each of the three types of events had to occur at least once in both the first and second 15 min of each 30-min period. Subjects were given no information regarding the frequency of events or their order of occurrence. The times between events (interstimulus intervals) ranged from 126 to 302 s with a mean of 200 s.

The subject was observed from an adjacent room via closed-circuit TV. Indirect lighting was used in the subject's room, and the level of illumination at the display was 5.6 lux.

<u>Procedure:</u> Upon arriving at the laboratory, subjects were given general

information about the task that they would be performing, but no information about the computer aiding that some would receive. Following this, a 9-point subjective rating scale was administered dealing with present feelings of attentiveness, tiredness, strain, boredom, and irritation. Subjects then received information describing each kind of critical event along with the appropriate responses to each. For the condition in which no advisory alerts were provided (the NAA condition), it was emphasized that aircraft at the same altitude were more difficult to detect than the three Xs signifying a malfunction, and they would have to make constant comparisons of target altitudes in order to detect the former type of event as rapidly as possible. Subjects assigned to the condition in which advisory alerts were provided (the AA condition) were told that, because possible aircraft conflicts are often difficult to detect, future, more highly automated ATC systems will likely employ techniques to automatically alert the controller when such situations are detected. In the task that they would be performing, a computer-aiding feature would automatically alert them whenever it detected two aircraft at the same altitude on the same flight path. They were told, however, that this aid could not be relied upon to detect every situation that might constitute a possible conflict; some few would likely be missed. An important part of their task would be to act as a backup to the automation in order to detect those occasional instances in which aircraft at the same altitude were missed by the automatic detection feature. Aircraft at the same altitude, whether detected by the automation or by themselves, were to be responded to in the same way; in either case, they were to press a designated button on the console. In the former instance, this response was made as soon as they had located the two aircraft detected by the automation; in the latter case, the response was made whenever a subject detected two aircraft at the same altitude that were missed by the automation. Task instructions for both experimental conditions were interlaced with practice on each type of event or situation taken individually and a final combined session in which all of the different events were presented in random order using a less demanding (8-target) version of the task. On rare occasions, additional practice sessions were administered if the subject appeared to have difficulty with any of the procedures.

The same target display file was used in both the NAA and AA conditions. As just described, the only difference between the two conditions was the addition of an alerting feature in the latter condition. This feature was programmed so that it would fail to detect conflicting aircraft on two separate occasions. One of these occasions occurred approximately midway into the first half-hour of task performance and the other during the last 10-15 min of the 2-h session. For half of the subjects in the AA condition, the missed events were numbers 6 and 33; for the other half, events 7 and 34 were pre-selected to be missed. (All missed events were actual conflict situations.) The use of two different pairs of missed conflict situations reduced the likelihood that a particular pair selected might influence the results in some idiosyncratic way.

In order to add a greater element of realism to the task, a tape recording of background noises recorded in actual air traffic control radar rooms was played continuously during the 2-hour task session. Sound level of this noise at the subject's head location was 62 dBA. A previous study using an earlier version of this task showed no adverse effects on performance of this noise at a considerably higher (80 dBA) level (4).

At the completion of the 2-h task period, subjects were administered a second version of the rating scale that contained additional items relating to perceived task difficulty and the amount of effort required to maintain task concentration. This was followed by a thorough debriefing concerning the purposes of the experiment.

RESULTS

With a visual taskload of 16 targets to monitor, time to detect the presence of three Xs (designated as an altitude malfunction event) in a data block has been shown to take approximately 8-10 s, with no increase in this time over 2 h of monitoring (6). Consequently, detection time for this type of event serves as a useful estimate of general scanning activity during a monitoring session. Mean time to detect altitude malfunction events for subjects in the two experimental conditions is shown in Fig. 2.

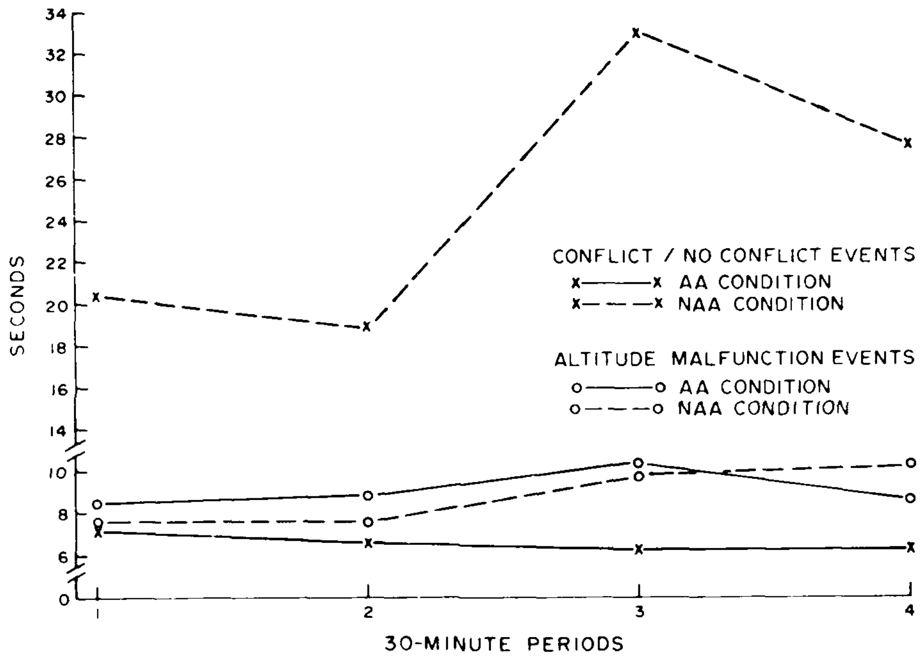

FIGURE 2. MEAN DETECTION TIMES ACROSS 30-MIN PERIODS FOR ALTITUDE MALFUNCTION EVENTS UNDER BOTH CONDITIONS AND FOR CONFLICT/NO CONFLICT EVENTS UNDER THE NONASSISTED (NAA) CONDITION. FOR PURPOSES OF COMPARISON, MEAN TIMES TO LOCATE AND ACKNOWLEDGE CONFLICT/NO CONFLICT EVENTS DETECTED BY THE AUTOMATION UNDER THE AA CONDITION ARE ALSO SHOWN IN THIS FIGURE.

A repeated measures analysis of variance (ANOVA) revealed no significant change across the 2-h session in detection time for malfunction events ($F(3/114)=1.84$, $p>.05$), no difference between conditions ($F(1/38)<1.00$) and no significant interaction ($F(3/114)<1.00$).

Also shown in Fig. 2 are mean times to detect or locate aircraft at the same altitude on the same flight path (designated conflict/no conflict events). Under the NAA condition, these data reflect the time required to detect the occurrence of these events and then to make the required button press response; for the AA condition, the data plotted are measures of the time required, following an advisory alert, to locate the two specified aircraft on the indicated flight path and make the required button press. It is evident from Fig. 2 that time required to locate and respond to specified aircraft following an advisory alert under the AA condition is relatively short (Mn=6.6 s) and shows no evidence of any change across the session. Time to detect aircraft at the same altitude without computer assistance (the NAA condition) shows a general increase over the 2-h monitoring session that is similar in magnitude to that found previously under comparable conditions (6). The irregular pattern of increase shown in this figure is probably without meaningful significance, since previous studies using the same basic task condition (i.e., without computer assistance) have typically shown a reasonably linear increase in detection time over a 2-h monitoring session (5,6). An ANOVA performed on the data of Fig. 2 yielded a significant difference between conditions ($F(1/38)=135.75$, $p<.001$), between periods ($F(3/114)=8.02$, $p<.001$), and a significant conditions by periods interaction ($F(1/114)=9.64$, $p<.001$).

Detection times for possible conflict situations missed by the automation (i.e., events for which no advisory alert message appeared at the time of target occurrence) for subjects assigned to the AA condition are shown in Table I. Also shown in this table are detection times for these same target events for subjects assigned to the NAA condition.

TABLE I. MEAN DETECTION TIMES (in s) FOR POSSIBLE CONFLICT SITUATIONS MISSED BY THE AUTOMATION DURING THE FIRST AND LAST 15 MIN OF THE SESSION. ALSO SHOWN ARE DETECTION TIMES FOR THE SAME SITUATIONS WHEN NO ADVISORY ASSISTANCE WAS PROVIDED. (VARIANCE IN PARENTHESES)

	Periods of the Session	
Condition	Beginning	End
AA	28.2(361.0)	26.0(272.2)
NAA	14.1(46.2)	23.9(269.0)

It is evident from Table I that time required for subjects to detect events missed by the automation under the AA condition exceeded detection time for these same events when no computer-aided detection was provided, with the greater numerical difference between conditions occurring at the beginning of the session. It is also apparent from this table that the two conditions were quite different in variance at the beginning of the session, but showed little difference at the end. Because this difference in variance at the beginning was found to be significant ($F(19/19)=7.79$, $p<.01$), and because the distribution of detection times for the initial AA missed event was extremely skewed in the direction of long detection times, a nonparametric (Mann-Whitney U) test was used to compare the two conditions. Separate comparisons revealed the difference between means of the two conditions to be significant ($p<.05$) for the initial missed event, but not for the final one ($p>.05$).

Of the 40 possible conflict events missed by the computer aid under the AA condition (20 subjects times 2 missed events per subject), 35 were detected by subjects in this condition. For the NAA condition, 36 of the same 40 events were detected. Thus, detection rates for these events were virtually the same under the two conditions; 87.5% under the former and 90% under the latter. Since all failures of subjects to detect the above events occurred during the last 30-min period in both conditions, it is evident that detection rates and patterns were essentially identical in both conditions.

Summary data obtained from the subjective rating scales administered at the beginning and end of the monitoring session are shown in Table II.

TABLE II. MEAN RATING SCALE VALUES FOR THE TWO EXPERIMENTAL CONDITIONS. ALSO SHOWN ARE COMBINED VALUES ACROSS CONDITIONS.*

		Measurement Periods			
Variable	Condition	Pre Task	Post Task	F**	p
Attentiveness	AA	6.70	4.90		
	NAA	6.95	5.55		
	Combined	6.82	5.22	21.86	<.001
Tiredness	AA	4.70	7.00		
	NAA	5.20	6.70		
	Combined	4.95	6.85	77.07	<.001
Strain	AA	3.15	4.85		
	NAA	3.60	4.50		
	Combined	3.37	4.67	18.09	<.001
Boredom	AA	3.20	6.25		
	NAA	3.05	5.90		
	Combined	3.12	6.07	65.00	<.001
Irritation	AA	1.30	2.55		
	NAA	1.45	2.00		
	Combined	1.37	2.27	7.76	<.01

*Higher numerical values indicate higher levels of the variable.
**Only the F values for the between periods effects are given, since none of the between conditions or interaction effects were significant (p>.05).

As indicated in the table, separate ANOVAs conducted on each of the rating scale variables yielded significant F values only for the pre to post change in ratings. None of the differences between groups were significant nor were any of the interactions (p>.05). For both conditions, self-reported attention decreased while tiredness, strain, boredom, and irritation all increased. Of these variables, those showing the greatest pre to post change during 2 h of monitoring were tiredness and boredom. The two variables not shown in the table (effort and difficulty) were measured only at the end of the session. Subjects were asked to recall and then to make comparisons of how much effort they felt was required and how difficult they experienced the task to be when they were performing it at the beginning and end of the session. ANOVAs performed on the ratings obtained for these two variables yielded no significant main effects or interactions. Performing the task under both conditions was experienced as moderately difficult, requiring only an average level of effort. Interestingly, subjects perceived no change in either difficulty or effort from beginning to end of the task period.

DISCUSSION

Detection times for the alphanumeric change used to indicate an altitude malfunction event showed no evidence of any increase over the 2-h session under either condition nor were any of these events missed. These findings are consistent with previous studies using this basic task (e.g.,5,6) and can be taken as a general indication that scanning activity remained relatively constant over the session.

Under the condition in which computer-aided detection was not available, the time to detect possible conflict situations increased significantly from an average of 19.8 s during the first hour to 30.4 s during the second. These findings likewise are in general agreement with previous studies under comparable conditions (5,6), and support our belief that the fatigue resulting from prolonged monitoring under high visual taskloads primarily impairs the ability to detect critical events that require considerable information processing in order to be "seen;" detection time for events that are directly apparent or perceivable to the observer (e.g., the altitude malfunction events of the present study) seem to show little or no evidence of impairment under these conditions.

Because the ability to detect possible conflict situations shows significant impairment during prolonged monitoring under high visual taskloads, the use of a computer aid to assist in detecting such situations clearly seems indicated. It was our initial hypothesis, however, that subjects would become overly reliant upon this aid, making them less efficient in detecting situations missed by the automation than if no computer aiding were provided at all. Contrary to this hypothesis, detection rates for missed situations when computer aiding was provided were found to be essentially the same as detection rates for the same situations when computer aiding was not available. While events missed by the automation were detected somewhat less rapidly when computer-aiding was employed than when such aiding was not provided, this difference between conditions was found to be significant only for the missed event that occurred during the first 30 min of task performance. Our expectation that the difference between conditions would be more pronounced at the end than at the beginning of the session clearly was not confirmed.

In attempting to account for the obtained findings, it may be significant to recall that subjects assigned to the computer-aided condition were told that an important part of their task would be to detect occasional failures of the automation, thereby acting as a systems backup. It has been our experience from previous studies that college students performing our simulated tasks generally take the tasks quite seriously, often appearing to view themselves as controllers dealing with real air traffic situations. It is certainly conceivable, therefore, that subjects in the computer-aided condition, perceiving their primary role as that of a backup to the computer in the event possible conflict situations were missed, were sufficiently motivated to sustain a level of attention at least as high as that of subjects in the condition without computer aiding. This possibility is supported by the fact that posttask ratings of effort and attentiveness for both conditions were equivalent. Of the remaining measures, none showed any evidence of differential change over the course of the session that would suggest differences between the conditions in motivation, attitude or feeling toward the task.

One of the frequently expressed concerns associated with increased automation is the concern that reliance on automated aids may lead to greater complacency, resulting in reduced alertness and lessened ability to detect and respond to infrequent failures of the automation (2,7). While the present study found no evidence that computer aiding contributed to either reduced alertness or reduced subject detection efficiency, it is evident that one must accept these findings with some degree of caution. Although the study suggests that complacency may not be a significant problem when computer-aided monitoring is confined to relatively short periods of time, the findings represent only the beginnings of research into this complex area. As Wiener (8) has noted, neither the mechanisms that influence the development of complacency nor its effects on performance are well understood. The important questions of whether future ATC systems, with failure rates that are likely to be extremely low, will induce unacceptable levels of complacency, or even stress, obviously cannot be answered in full by generalizing from the results of the present study. Although studies such as this are of value in helping to define those parameters that may or may not contribute to the development of complacency effects, definitive answers to the difficult questions posed above may well require lengthy field studies in which infrequent errors or failures are introduced while performing under real-life or highly realistic simulated conditions. Such research may become feasible when prototype equipment incorporating AERA automation is available.

REFERENCES

1. Ball CG, Dubofsky DF, Fernow JP, Gisch AH, Taber NJ, Weidner J F, White EB. AERA 2 operational description. Mitre Technical Report MTR-85W66 1986; The Mitre Corporation, McLean, Virginia.

2. Boehm-Davis DA, Renwick EC, Wiener EL, Harrison RL. Human factors of flight-deck automation: Report on a NASA-industry workshop. Ergonomics 1983; 26:953-961.

3. Lipps AW, Swedish WJ, Zimmerman BC. Operational and functional description of the AERA packages. 1983; FAA Office of Systems Management Technical Report No. DOT/FAA/ES-83/10.

4. Thackray RI. Some effects of noise on monitoring performance and physiological response. Academic Psych. Bull. 1982; 4:73-81.

5. Thackray RI, Touchstone RM. Complex monitoring performance and the coronary-prone type A behavior pattern. 1986; FAA Office of Aviation Medicine Report No. DOT/FAA/AM-86/4.

6. Thackray RI, Touchstone RM. An evaluation of the effects of high visual taskload on the separate behaviors involved in complex monitoring performance. 1988; FAA Office of Aviation Medicine Report No. DOT/FAA/AM-88/1.

7. Wiener EL, Curry RE. Flight-deck automation: Promises and problems. Ergonomics 1980; 23:995-1011.

8. Wiener EL. Complacency: Is the term useful for air safety? Proceedings of the 26th Corporate Aviation Safety Seminar, March 29-31, Denver, Colorado, 1981.

☆ U.S. Government Printing Office : 1989 - 617 000/88631