Naval Research Laboratory
Washington, DC 20375-5000

NRL/FR/5534—92-9375

Direct Manipulation and Intermittent Automation in Advanced Cockpits

JAMES A. BALLAS, CONSTANCE L. HEITMEYER, AND MANUEL A. PÉREZ

Human-Computer Interaction Lab
Information Technology Division

July 21, 1992

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	July 21, 1992	Final

4. TITLE AND SUBTITLE

Direct Manipulation and Intermittent Automation in Advanced Cockpits

5. FUNDING NUMBERS

PE - 62233N

6. AUTHOR(S)

James A. Ballas, Constance L. Heitmeyer, and Manuel A. Pérez

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Naval Research Laboratory
Washington, DC 20375-5000

8. PERFORMING ORGANIZATION REPORT NUMBER

NRL/FR/5534—92-9375

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Office of Naval Technology
Arlington, VA

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

Increasing use of automation in computer systems, such as advanced cockpits, present special challenges in the design of user interfaces. The challenge is particularly difficult when automation is intermittent because the interface must support smooth transitions from automated to manual mode. A theory of direct manipulation predicts that this interface style will smooth the transition. Interfaces were designed to test the prediction and to evaluate two aspects of direct manipulation: semantic distance and engagement. Empirical results supported the theoretical prediction and also showed that direct engagement can have some adverse effects on another concurrent manual task. Generalizations of our results to other complex systems are presented.

14. SUBJECT TERMS

Direct manipulation	Adaptive automation	Intelligent cockpit
Interface style	Intermittent automation	
Interface design	Aircraft interface	

15. NUMBER OF PAGES 52

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102

CONTENTS

INTRODUCTION 1

- HHN Theory of Direct Manipulation 1
- Related Research on Direct Manipulation 2
- Direct Manipulation in the Cockpit 3
- Experimental Hypothesis 3

EXPERIMENTAL DESIGN 4

- Subjects 4
- Experimental Tasks 4
- Interfaces for the Tactical Assessment Task 6
- Distance and Engagement in the Interfaces 9
- ANOVA Design 10
- Scenarios 10
- Training 12
- Accuracy Data 12
- Questionnaire and Workload Assessment 12

RESULTS 12

- Overview 12
- Interface Style and Automation Deficit in Response Time 13
- Tactical Assessment Speed and Accuracy 17
- Single, Dual, and Adaptive Automation Response Times 17
- Tracking Task Results 20
- Questionnaire Results 25
- Workload 26

DISCUSSION 26

- Relationships to Other Cockpit Studies 26
- Implications for Direct Manipulation Theory 29
- Generalizations to Other Domains 30
- Summary 31

ACKNOWLEDGMENTS 31

REFERENCES 31

APPENDIX A — Consent Form 35

APPENDIX B — Task Instructions 37

APPENDIX C — Questionnaire 43

APPENDIX D — TLX Instructions 49

DIRECT MANIPULATION AND INTERMITTENT AUTOMATION IN ADVANCED COCKPITS

INTRODUCTION

The rapid evolution of user interfaces is occurring not only in office systems but also in modern cockpits, which are computer-based and include advanced graphical displays (Wiener 1985). However, modern cockpits differ from traditional office systems in several fundamental ways. First, unlike office systems, they often include sophisticated automation, such as the ability to fly on automatic pilot. Moreover, unlike office applications, the cockpit application is dynamic and complex. The pilot must not only handle large quantities of real-time, often continuous, input data; he must also perform several demanding tasks concurrently, usually under severe timing constraints. Finally, unlike users of office systems who typically communicate via electronic mail, the pilot of a modern cockpit communicates in real-time via networked voice and data links. Given these differences, the cockpit interface presents many design challenges that the developers of office systems seldom encounter.

An important question in designing the user interface of modern cockpits is how to handle automation. Our research is part of a larger research program in adaptive automation, an automation philosophy of allocating tasks between the pilot and the computer system in an optimal manner (Parasuraman, Bahri, Deaton, Morrison, and Barnes 1990). In adaptive (i.e., intermittent) automation, the pilot performs a task only intermittently. Given a dual-task situation, a rise in the level of difficulty of one task could cause automation of the second task. Having the computer system take over the second task allows the pilot to focus his efforts on the increased difficulty task. Once the difficulty level of the first task returns to normal, the pilot resumes control of both tasks. Such an approach to automation is expected to result in better overall pilot/system performance (Parasuraman et al. 1990). Because the pilot only performs the first task intermittently, a challenging problem, and the problem that this paper addresses, is how to design an interface that supports a smooth transition from automated to manual mode.

This report presents the results of our empirical research on interface styles for adaptive automation. Our research is designed to test predictions from a theory of direct manipulation. A fundamental goal of the research is to determine whether a direct manipulation interface has performance benefits in adaptive automation; i.e., does direct manipulation lead to improved performance when a pilot must quickly resume a task that has been previously automated? A related goal is to separate and evaluate two aspects of direct manipulation identified by the theory, namely, distance and engagement. In this report, we introduce the direct manipulation theory, present our hypothesis about the effect of interface style in adaptive automation, describe the interfaces developed to test our hypothesis, and summarize the empirical results. We conclude with a discussion of the implications of our results.

HHN Theory of Direct Manipulation

Designing an interface for an adaptive system involves many issues and decisions, but little theoretical guidance or empirical information is available. There is general agreement on what the interface should accomplish. As a first priority, the interface should enable the pilot to maintain both situational awareness and system control (McDaniel 1988; Parasuraman et al. 1990). We define situational awareness as the extent to which the pilot has the knowledge needed to perform a specified task or tasks. Clearly, this knowledge depends upon the specific state of the aircraft and selected aspects of the aircraft environment. In adaptive automation, the pilot shifts from manually performing a task to monitoring its automated performance and then back to manual operation. In this situation, the key to assessing situational awareness is how well the pilot can resume a task that has been previously automated. We claim that a critical factor in achieving a smooth transition from automated to manual performance of a task is interface style.

Manuscript approved February 11, 1992.

Differences in interface style have been described metaphorically by Hutchins (1986). Direct manipulation interfaces behave according to a model world metaphor; the user interacts with an interface that represents the task domain, the domain objects, and the effect of user operations on those objects. Command language interfaces behave according to a conversational metaphor; the user and the computer have a conversation about the application domain. The interface acts as an intermediary between the user and the domain. Because the interface does not represent the task domain explicitly, the user is forced to maintain a mental model of the domain's state or make frequent queries about the state. Such requirements may place heavy cognitive burdens on the user. However, a command language interface can be very powerful if it is designed to cover most contingencies in a succinct manner.

Building upon these metaphors, Hutchins, Hollin, and Norman (HHN) have developed a theory of direct manipulation (Hutchins, Hollin, and Norman 1986). Although typically associated with desktop computer systems, direct manipulation is also being considered for large safety-critical systems, such as nuclear power plants (Beltracchi 1987; DeBor and Swezey 1989). HHN proposed models of the cognitive processes that users employ when interacting with a direct manipulation interface, concluding that two aspects of direct manipulation account for its performance advantages: low distance and direct engagement. According to HHN, the first aspect is the "information processing distance between the user's intentions and the facilities provided by the machine." Performance advantages come with less distance, because there is less cognitive effort needed to understand and manipulate the domain objects. HHN call such an interface *semantically direct* and claim that it can be achieved by "matching the level of description required by the interface language to the level at which the person thinks of the task."

Distance is of two types, semantic and articulatory. *Semantic* distance is the difference between the user's intentions and the meaning of the expressions available in the interface, both expressions that communicate the user's intentions to the computer and expressions whereby the computer system provides user feedback. For example, if the user wishes to delete all files whose names end in *text* and the computer system (e.g., the Macintosh) has no single expression for this purpose, then significant semantic distance exists between the user's intentions and the expressions available in the interface. *Articulatory* distance is the difference between the physical form of the expressions in the interface and the user's intentions. For example, when a UNIX user wants to display a file and to do so he must invoke a command named "cat," significant articulatory distance exists between the name of the UNIX command and the intended user operation. Our studies have focused on semantic distance. We have proposed follow-up studies to investigate issues concerned with articulatory distance.

The second aspect of direct manipulation is *engagement*, i.e., the involvement that comes when the user is able to interact directly with the application domain and the objects within it rather than interacting through an intermediary. The key to direct engagement is inter-referential I/O, which permits "an input expression to incorporate or make use of a previous output expression." For example, if a listing of filenames is displayed on the screen, one of these names can be selected and operated on without entering the name again. In Draper's (1986) view, the important aspect of inter-referential I/O is that the user and the computer system share a common communications medium. This takes the notion of inter-referential I/O beyond the UNIX concepts of channels and pipes. Jacob (1989) points out that while UNIX makes output usable as input, the medium of exchange is the unformatted (and invisible) text stream. In direct manipulation, the shared medium is usually a visual display that presents an explicit, often graphical, view of the task domain. Wolf and Rhyne (1987), in a survey and analysis of interfaces, concluded that all direct manipulation style interfaces share a visual representation of an object and a selection operator in the object's vicinity.

Related Research on Direct Manipulation

An early study comparing several interfaces (Whiteside, Jones, Levy, and Wixon 1985) concluded that usability depends more on specific interface design than interface style. Contrary to expectations, iconic interfaces were inferior to menu systems and command language interfaces for new and transfer users. More recent studies have generally shown advantages for direct manipulation over command language interfaces (Ziegler and Fahnrich 1988). For example, Karat (1987) found consistently faster times for several file management tasks in a direct manipulation interface that used pointing and dragging operations on iconic representations of files. However, Karat did find an advantage for the command language interface on one particular type of file management task. Thus, evaluations of interface styles need to be sensitive to task-specific effects. As cited by Kieras (1990), Elkerton and Palmiter suggest that the basic principle of direct manipulation lies in the replacement of complex cognitive operations with perceptual and motor activities. Thus, the advantage of direct manipulation may lie in tasks with complex cognitive operations that can be transformed into motor and perceptual operations.

Research on direct manipulation has been mostly on conventional applications, such as word processing and file management. A notable exception is a study by Benson and her colleagues (Benson, Govindaraj, Mitchell, and Krosner 1989) which compared a conventional interface to a direct manipulation interface for a parts manufacturing system. The conventional interface used menus, function keys, typed commands, displayed textual information, and paged displays. The direct manipulation interface used a mouse as the only input device and provided a continuous display of important information. The evaluation of these interfaces used performance measures relevant to manufacturing, such as cost, inventory levels and status, and late deliveries. Performance with direct manipulation was superior on three of five dependent measures.

All previous research on direct manipulation has not attempted to tease apart semantic distance and direct engagement, and determine which is important in user performance. Furthermore, previous research has evaluated applications designed for purposes other than evaluation of an interface style. The interfaces in our study were designed specifically to study direct manipulation by separating and evaluating the two aspects identified in the HHN theory.

Direct Manipulation in the Cockpit

Elsewhere, we have presented snapshot information about the types of interfaces that are currently used in cockpits (Ballas, Heitmeyer, and Pérez 1991) and an FAA report covers this in more detail (Federal Aviation Administration 1991). One point worth noting is that the effectiveness in the cockpit of a direct manipulation interface and its two aspects remains an open question. Some studies suggest that navigation displays should present a model world to the pilot. For example, Marshak, Kuperman, Ramsey, and Wilson (1987) found that moving-map displays in which the viewpoint is similar to what would actually be seen by looking outside the plane led to improved performance. However, Williams and Wickens (1991) found that simpler aspects of navigation are performed using verbal-analytic cognitive processes, not spatial processes. A spatial display may not support the verbal-analytic process as well as a textual display. Reising and Hartsock (1989) found that in warning/caution/advisory displays, a schematic of the cockpit showing the controls that were needed to handle an emergency did not improve performance. The important factor in improved performance was a checklist of the required procedures (which is closer to what a command language interface would offer). A test pilot pointed out that one of the best examples of an effective command language display in the modern cockpit is the "SHOOT" cue that appears in the HUD when the target is in weapons range (Maris 1990).

Ironically, in modern flight control systems, some trends have been away from direct manipulation. For example, fly-by-wire systems remove the pilot from direct control of wing surfaces. Bernotat (1981) and Zlotnik (1988) argue against this trend, suggesting that in such systems the pilot needs direct sensory feedback about the aircraft's performance. Such feedback is consistent with the notion of direct manipulation. Other trends in cockpit controls suggest a move toward direct manipulation, e.g., the incorporation of touchscreen displays. However, the incorporation of pointing devices into the flight deck needs to be carefully evaluated; e.g., what is the effect of the pilot's use of two pointing devices concurrently (a touchscreen and a joystick)?

Experimental Hypothesis

An issue in interface design for intermittent automation is *automation deficit*, the initial decrease in pilot performance that occurs when a task that has been previously automated is resumed. This deficit may reveal itself in several ways: slower human response, less accurate human response, subjective feelings of not being in control, subjective feelings of stress, etc. Some previous studies have shown an automation deficit for manual control tasks, while others have not (Parasuraman et al. 1990). In our research, we are interested in automation deficits in response time and the effect of interface style on automation deficit.

Our hypothesis is that direct manipulation interfaces lead to a reduction in automation deficit that is reflected in decreased response times right after automation ceases. The rationale underlying this hypothesis is that decreased semantic distance and improved direct engagement enhance a pilot's ability to monitor a task that is automated and then to quickly resume the task. Besides testing the general hypothesis, we evaluated the importance of each aspect of direct manipulation in minimizing automation deficit.

To test our hypothesis, we evaluated the effect of interface styles on a person's ability to resume a task quickly after a period of automation. Using different types of interfaces, we compared performance in the first few seconds of the manual mode to performance a minute later. To test our hypothesis and to achieve our goal of understanding the role of the two aspects of direct manipulation, we needed to solve three problems. First, we needed to develop interfaces that implemented different combinations of semantic distance and engagement. Solving this problem was particularly challenging, because there are no operational definitions of distance and engagement. Second, we needed a paradigm for assessing automation deficit. Third, we needed to ensure that this paradigm would allow us to separate the effects of the interface on automation deficit from the overall effects of the interface.

Solving these three problems required a lengthy period of interface development and analysis. Our solution to the first problem is covered in detail in the Experimental Design section. Basically, we manipulated semantic distance by developing interfaces that supported different user goals. We manipulated engagement by implementing two different communication mediums, one shared, the other split. In the shared case, both the subject's intentions and system feedback are expressed via the same visual medium. In the split case, the subject's intentions are expressed via one visual medium and the system feedback via another.

To solve the second problem, how to assess automation deficit, we needed alternating phases of a task, automated and manual, and performance measures in the initial part of the manual phase. To deal with the third problem, isolating the effects of interface style on automation deficit, we implemented similar task behavior during both the initial period of the manual phase (i.e., the first three responses) and later in the manual phase (i.e., the seventh through ninth responses) and then compared initial subject performance with later performance. Further, we set out to develop interfaces that would provide similar performance in single-task scenarios (without automation). In fact, we did not begin performance testing under intermittent automation until similar performance in the single-task scenarios was achieved.

EXPERIMENTAL DESIGN

Subjects

Twenty subjects (17 men and 3 women) were recruited from NRL personnel, with five randomly assigned to each of the four types of interfaces used in the tactical assessment task. All but two were right handed. Most were between 25 and 39 years old, were college graduates, reported themselves to be in good health, and had normal vision. All were screened for normal color vision. Two of the subjects were licensed pilots.

Experimental Tasks

The experiment required subjects to perform two tasks, a pursuit tracking task and a tactical assessment task. To establish a setting for adaptive automation, the difficulty of the tracking task alternated between moderate and high throughout the experiment. During the moderate difficulty phases of the tracking task, the subject performed both the tracking task and the tactical assessment task. Each time the difficulty of the tracking task rose to high, the tactical assessment task was automated, and the subject performed the tracking task only. The display screen used in the experiment was partitioned into two windows, one for the tracking task, the other for the tactical assessment task. Changes in the automation of the tactical assessment task were signaled in two modalities. A beep occurred at each change and a border was placed around the window when the task was performed manually. The color of this border matched the border of the tracking window so that the tasks would be integrated while both were in the manual mode. This approach was based upon Wickens and Andre (1990) who found that integration can be promoted by similar color coding. Thus, the subject had a consistent bordering cue indicating that the task within the window was to be performed manually.

The tracking task simulated air-to-air targeting of an enemy aircraft using a gun sight similar to the pipper and reticle on a typical head-up display. The target on the display was a graphical representation of an enemy aircraft. The target's driving function was the sum of nine nonharmonic sinusoids (.02, .03, .07, .13, .23, .41, .83, 1.51, and 3.07 Hz) with randomly determined starting phases. The amplitudes of these components were varied to produce two levels of tracking difficulty. The amplitudes for the "less difficult" tracking were flat up to a cutoff frequency of .07 Hz and reduced in amplitude 3 dB/octave above this frequency. The "difficult" function was flat up to a cutoff frequency of .23 Hz and reduced in amplitude 3 dB/octave above this frequency. The target position was updated every 83 ms and the control position was sampled at the same rate. The tracking control was a self-centering, displacement joystick. The control dynamics were a 25%/75% mixture of rate and acceleration. Performance measures included RMS amplitude calculated for each axis. In addition, a continuous record of the target and pipper position was recorded for later spectral analysis.

The second task, tactical assessment, is a critical task in a tactical aircraft and one that has become more challenging with the increased capabilities of modern aircraft. Our hypothesis was tested on four alternative interfaces for the tactical assessment task. The simulated tactical situation included three classes of targets—fighters, aircraft, and ground-based missiles—and contacts on the targets by sensor systems. The targets first were designated as possible threats using black color coding, but as they got closer to the *ownship* (the symbol for the aircraft the pilot was in), they were designated as neutral, hostile, or unknown, using blue, red, and amber color coding, respectively. The subjects were told that simulated sensor systems were assigning these designations.

Formal analysis (a partial GOMS analysis, Kieras 1988) led us to specify what aspects of the tactical assessment task would improve with a direct manipulation interface. It became apparent that the advantages of direct manipulation would only be seen if the tactical assessment task required the pilot to understand the status of targets in the tactical situation and act upon these targets. This meant that we needed to have responses reflective of a particular interpretation of the tactical situation. Furthermore, we had to have tactical situations and scenarios that were meaningful. This increased the realism of the simulation and enabled us to develop a direct manipulation interface that would present a view of the world in which meaningful actions were occurring. The use of meaningful tactical scenarios is also supported by Badre's (1982) study of representing tactical information. He evaluated the ability of experts and novices to encode and reconstruct structured and unstructured battlefield scenarios. He found that "there is a direct relationship between the level of coherence of a scenario and the capacity of the decision maker to encode it and represent it meaningfully" (p. 502). We currently require two types of decisions: confirmation and classification. The confirmation decision requires the pilot to recognize a color code for hostile or neutral and confirm this code. The classification decision requires the pilot to monitor the behavior of targets in the display and then, based on the target's behavior, to classify a target as hostile or neutral. These two types of decisions correspond to two levels within the aircrew decision model of situation awareness proposed by Endsley (1988). Level 1 of Situation Awareness (SA) in his model means perceiving that elements are present and perceiving the relevant properties of these elements such as color, speed and location. Level 2 SA means comprehending the significance of the elements and forming a holistic picture of the environment. Determining the hostile or neutral status would be an aspect of comprehending the significance of the elements and thus a behavior at Level 2 SA. Level 3 SA means making projections about the future course of the scenario. The experimental design did not require any responses that explicitly assessed Level 3 SA. However, we obtained some results that bear on awareness at this level of Endsley's model.

The subjects were required to perform two operations, confirm and classify. If the system designated a target as neutral or hostile (i.e., the target was colored blue or red), the subject had to *confirm* the designation by picking the target and then indicating the proper designation, i.e., neutral for blue targets and hostile for red targets. Thus, confirm decisions only required the subject to discriminate colors. If the system designated the target as unknown (i.e., the target was colored amber), the subject had to *classify* the target as hostile or neutral based on its behavior. Table 1 provides the rules for designating a target as hostile or neutral. The target class determines what target attribute the subject uses to determine the target's designation.

Table 1 — Rules for Tactical Assessment of Targets

Target Class	Hostile	Neutral
Fighter	Constant bearing	Bearing away
Airplane	Air speed ~ 800	Air speed ~ 300
Missile site	Within threat range	Outside threat range

To classify the amber targets, the subject needed to monitor heading for fighters, speed for aircraft, and projected lateral distance for ground missile threats. The responses were timed and analyzed to produce measures of accuracy and response time. The subject had a response interval of 10 seconds to make the assessment response. As recommended by Nunnally (1970), we substituted 9999 ms for the responses that were not completed within the deadline. We also performed confirmatory analyses using only responses that had been completed within the response interval. Generally, 95% of the responses were completed within the response interval.

Interfaces for the Tactical Assessment Task

To test our hypothesis, we designed and built four interfaces by using prototyping and iterative development. These four interfaces, which include a direct manipulation interface, a command language interface, and two hybrid interfaces, represent the four combinations of semantic distance and engagement shown in Fig. 1. The following paragraphs briefly describe each interface and discuss how each implements some combination of semantic distance and engagement. These interfaces were designed to be good representations of the four interface types and to support comparable performance in ordinary operation. We developed several versions of each interface during prototyping and collected performance data during the development of the interfaces.

Engagement \ Semantic Distance	Low	High
Direct	Graphical Display with Touchscreen (Direct Manipulation)	Tabular Display with Touchscreen
Indirect	Graphical Display with Keypad Input	Tabular Display with Keypad Input (Command Language)

Fig. 1 — Levels of engagement and semantic distance in the four interfaces for the tactical assessment task

The experimental software, which is based on an object-oriented design, is partitioned into user interface software and application software (see Fig. 2). The user interface software implements each of the five different interfaces, one for the tracking task and four for the tactical assessment task, as a subclass of an abstract interface class. The application software includes a simulation class shared by the different interfaces. The simulation class generates target information, controls the timing of the displayed events (e.g., target-detected), simulates user actions, and dispatches events to the interfaces. Each interface processes the events generated by the simulation class. Use of an object-oriented approach allows code to be shared across interfaces. For example, both the Tabular Display Interface and the Command Language Interface use the same code to display tabular information.

Building four interfaces that support equivalent performance required considerable prototyping and several iterations. Use of an object-oriented approach facilitated changes and extensions. In most cases, changes to the interfaces were achieved easily, since the code associated with each change was localized rather than distributed across the software. Extensions to each interface were produced by creating subclasses that provided the extended behavior. To maintain flexibility (e.g., allow the original code to be used) the original behavior was retained in the parent class.

The *direct manipulation interface* (Fig. 3(a)) has direct engagement and low semantic distance. It uses a shared communications medium: both the subject and the computer use the entire tactical assessment window to communicate. This interface simulates a radar display with continuously moving symbols representing the targets. The symbol used to represent a target is an intuitive graphical representation of the target class. Each target symbol is initially colored black but changes to red, blue, or amber once the system assigns the target a designation. A touchscreen overlays the display. The subject confirms or classifies a target by picking a target symbol on the display and selecting one of two strips, labeled HOSTILE and NEUTRAL, located on either side of the display. The subject accomplishes both the pick and the select by touching the appropriate part of the display screen. The words 'HOSTILE' and 'NEUTRAL' in the two side strips are colored red and blue, respectively. For classify decisions, the subject needs to observe the behavior of the graphical symbol that represents the target to determine the proper target designation. For confirm decisions, the subject needs to interpret the color of the target symbol. The use of a touchscreen to select targets in 2-D space not only enables the user and the computer to use the same communications medium, but is also consistent with Curry, Reising, and Zenyuh (1985) who found better performance in target designation with touch compared to a joystick and voice.

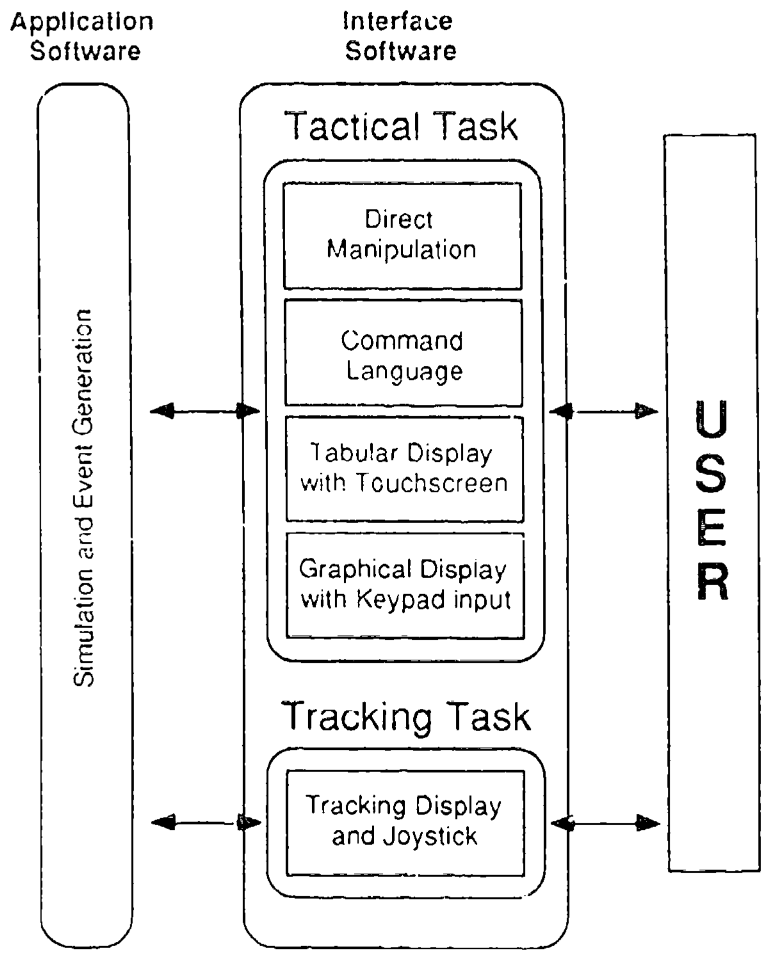
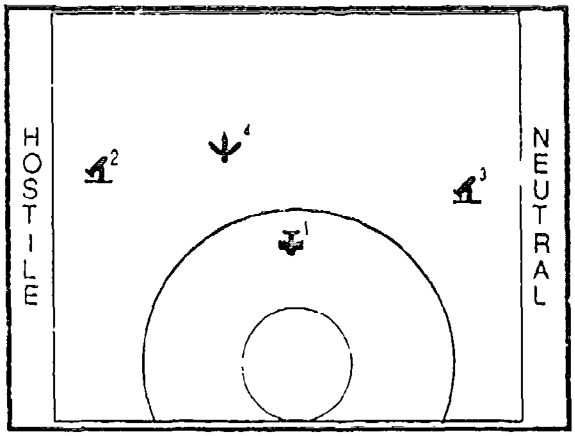
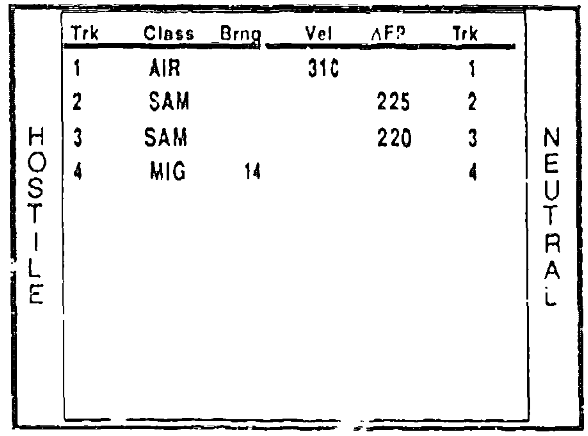


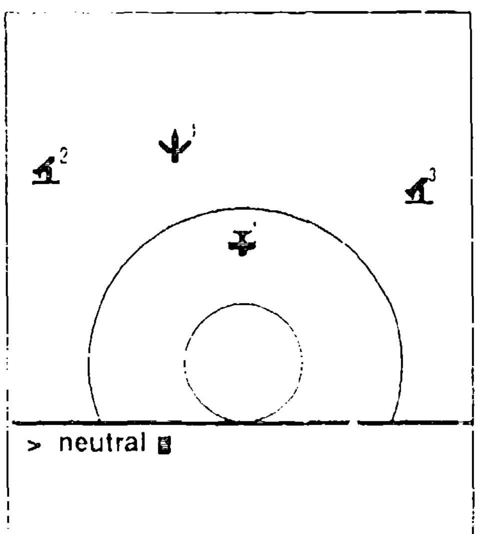
Fig. 2 — Object-oriented design of experimental software



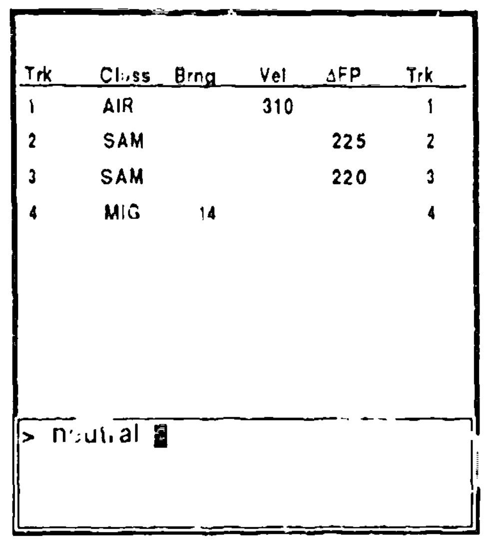
(a) Graphical Display with Touchscreen (Direct Manipulation)



(b) Tabular Display with Touchscreen



(c) Graphical Display with Keypad Input



(d) Tabular Display with Keypad Input (Command Language)

Fig. 3 — Four interfaces for the tactical assessment task, combining levels of semantic distance and engagement

The *command language interface* (Fig. 3(d)) has indirect engagement and high semantic distance. This interface uses a split visual medium: the tactical assessment window is partitioned into a top portion, which displays a table of target names and attributes, and a bottom portion, which is for subject input and error feedback. Each entry in the table describes a single target, providing the target's name (an integer), the target's class, and continuously updated data about the target. The name of the target class carries the system designation; initially black, it changes to red, blue, or amber once the system has assigned a designation. The table is decluttered: i.e., it only presents the critical attribute for the given target class. After the subject has completed a classify or a confirm operation on a target, the system removes the target entry from the table by scrolling the table. The subject uses a keypad to invoke a confirm or classify operation. For each operation, two sequential keypresses are required, one designating hostile or neutral, a second indicating the target number. For classify decisions, the subject needs to interpret the data in the table to determine the appropriate target designation. For confirm decisions, the subject needs to interpret the color of the word identifying the target class.

One important difference between the command language interface described above and the command language interfaces associated with more traditional office systems is that the table of target data is updated continuously. Such an approach is dictated in an aircraft context by the impact of external factors on the domain objects (i.e., the targets) and the real-time demands of the tactical domain. The approach makes less sense in an office system where, in most cases, changes to domain objects are made solely by the user and rapid response times are not as crucial.

The third interface (Fig. 3(c)), the *graphical/keypad interface*, combines the low semantic distance of the first interface with the less direct engagement of the second interface. Like the command language interface, this interface splits the tactical assessment window into two portions. The top portion contains the simulated radar display; the bottom portion is for subject input and error feedback. The subject uses the keypad to enter his classify and confirm decisions.

Finally, the fourth interface (Fig. 3(b)), the *tabular/pointer interface*, combines high semantic distance with direct selection of the tactical targets on the display using a touchscreen. The subject confirms or classifies a target by touching the appropriate table entry and touching either the HOSTILE or NEUTRAL strip at the sides of the display. This last interface is similar to a menu interface, except that the table items are updated dynamically. Scrolling in this interface occurs just after the subject completes entry of the confirm or classify decision and is thus associated with the completion of a user action.

Distance and Engagement in the Interfaces

Although the four interfaces intuitively represent different combinations of semantic distance and engagement, it is important to understand the theoretical rationale for the level of distance and engagement in each interface. Metaphorically, the direct manipulation interface represents a model world of the task domain, the command language interface a verbal description. A graphical representation more closely matches the way that a pilot thinks about the tactical situation. More importantly, these two interfaces support the user's goals differently. We distinguish two user goals: to remain aware of the current tactical configuration, and to perform the assigned task. The low-distance display was designed to support both goals. To support the first goal, the display continuously provided a graphical representation of the target's location and how the target was moving. To support the second goal, all relevant information about each target was encapsulated by this graphical representation.

The high-distance display was designed to support only the second goal, user performance of the assigned task. In developing the high-distance display, considerable effort was required to design a table that effectively supports the assigned task. For example, the target's spatial coordinates (x, y positions) were not provided because they are not relevant to the task and would have made the table harder to interpret. Moreover, the color code indicating the type of decision required was shown in the class column only, thus separating the system-assigned designation from the target attribute information. Finally, the columns were arranged to support efficient eye movements.

The levels of engagement can also be considered from several perspectives. We provide a pointing device (i.e., a touchscreen) for high engagement and a keypad for low engagement. The keypad uses a mode shift for two keys to preserve a common aspect of command language interfaces and to avoid introducing direct engagement with labeled keys for each action and object, a feature that Shneiderman associates with direct manipulation (Shneiderman 1982).

The theoretical difference between the levels of engagement in the interfaces is based upon the notion of a shared medium. In the direct engagement interfaces (the direct manipulation interface and the tabular/pointer interface), both the user and the computer system use a shared communications medium; that is, they both operate on the same objects. In the direct manipulation interface, the shared medium is the spatial display. The objects to be operated on are the target symbols and the strips labeled 'HOSTILE' and 'NEUTRAL.' In the tabular/pointer interface, the shared medium is the table, and the objects to be operated on are the table entries. In both direct engagement displays, the objects to be operated on and the strips share the same color code. Thus, for example, red in either the spatial display or the table of target attributes indicates that the subject should select the strip with the red wording.

In the indirect engagement interfaces (the command language interface and the graphical/keypad interface), the computer communicates to the user through one medium (i.e., section of the tactical display) and set of objects, while the user communicates to the computer through another medium (a keypad and another section of the tactical display) using a different set of objects. Thus, there is a separation of the user input and computer output.

ANOVA Design

Table 2 shows the ANOVA design, which was central to the testing of the hypothesis. The between-subject variable in this table is the type of interface. However, this variable was treated as a combination of two other variables, semantic distance and level of engagement as described in the previous section. Three within-subject variables are in the design. The first is the time interval after automation at which responses were required by the subject. This variable is the key to our assessment of automation deficit. We wanted to compare performance in the initial moments of resuming the task to performance later in the manual operation of the task. This variable was controlled by the scenarios so that responses were required as soon as the subject received the task from automation. Two more responses were required in quick succession. This pattern was repeated later in each manual period, and we compared performance right after automation to performance in the middle of a manual phase.

A second within-subject variable is the type of decision required: confirmation or classification. The confirmation decisions involved confirming the sensor classifications of hostile or neutral. The classification decisions involved the use of the rules outlined earlier. Both required the subject to select the target.

The third within-subject variable is the target type with a distinction being made between targets that have a critical parameter that is static vs targets that have a parameter that is dynamic. The key data for the fighters is dynamic in both graphical and tabular format since the heading information is either presented in the positional changes on the graphical display or the numerical changes in the tabular display. The data for the airplanes is a static velocity number in the tabular but dynamic positional changes in the graphical display. Data for the missiles is static in both types of display.

Scenarios

During the adaptive automation session, the scenarios produced 138 targets that had to be confirmed or classified within a 28-minute testing session, for an average event rate of one per 12 seconds. The overall duration of the session is similar to mission simulations in other research (e.g., Reising 1977). For experimental purposes, decisions were required when the target changed from contact status to presumed hostile, presumed neutral, or unknown. This change was signaled by switching the color of the target from black to red, blue, or amber. This change started the response time clock, and time to confirm or classify the target was taken. The scenario produced these changes at important points in the automated and manual phases of tactical assessment task. In particular, because of our hypothesis about the advantage of direct manipulation interfaces in resuming performance of the automated task, several decisions were required at the beginning of each manual phase. This meant that we were producing the transition from automated to manual operation at a period of high task demand. This demand was repeated later in the manual period to obtain comparison data.

After an initial period of manually performing the tactical assessment task for 3 minutes, the subject went through six cycles of automation to manual operation of this task. The duration of automation phases and the manual phases varied between 105 and 135 seconds. Coincident with the automation of the tactical assessment task, the tracking task increased in difficulty. It reverted back to the lower level when the automation of the tactical assessment task finished.

Table 2 — Experimental Design for ANOVAs

Between Subjects cells	Within	Subject	cells
Interface Style	Response after auto	Decision	Target
Direct Manipulation	First	Classify	AIR
	Seventh	Classify	AIR
	First	Confirm	AIR
	Seventh	Confirm	AIR
	First	Classify	MIG
	Seventh	Classify	MIG
	First	Confirm	MIG
	Seventh	Confirm	MIG
	First	Classify	SAM
	Seventh	Classify	SAM
	First	Confirm	SAM
	Seventh	Confirm	SAM
Command Language	First	Classify	AIR
	Seventh	Classify	AIR
	First	Confirm	AIR
	Seventh	Confirm	AIR
	First	Classify	MIG
	Seventh	Classify	MIG
	First	Confirm	MIG
	Seventh	Confirm	MIG
	First	Classify	SAM
	Seventh	Classify	SAM
	First	Confirm	SAM
	Seventh	Confirm	SAM
Graphical/Keypad	First	Classify	AIR
	Seventh	Classify	AIR
	First	Confirm	AIR
	Seventh	Confirm	AIR
	First	Classify	MIG
	Seventh	Classify	MIG
	First	Confirm	MIG
	Seventh	Confirm	MIG
	First	Classify	SAM
	Seventh	Classify	SAM
	First	Confirm	SAM
	Seventh	Confirm	SAM
Tabular/pointer	First	Classify	AIR
	Seventh	Classify	AIR
	First	Confirm	AIR
	Seventh	Confirm	AIR
	First	Classify	MIG
	Seventh	Classify	MIG
	First	Confirm	MIG
	Seventh	Confirm	MIG
	First	Classify	SAM
	Seventh	Classify	SAM
	First	Confirm	SAM
	Seventh	Confirm	SAM

The simulation was designed for an iteration loop of 12 Hz (83.3 ms). Tests indicated that this rate was achieved with all four of the interfaces. Within an iteration, the position of targets would be recalculated, responses from the joystick, keypad, and touchscreen would be retrieved, and the tracking display and tactical display would be updated.

Training

The subjects were trained on the two tasks separately with two 6-minute single-task sessions for tracking and two 10-minute single-task sessions for tactical assessment. The next day, they received training on the two tasks together in a 15.5 minute dual-task session before starting the *Adaptive Automation* session. The single- and dual-task training sessions were also used for data analysis of single-task and dual-task performance, except for the first 190 s of the dual-task session. The tactical scenarios were similar across the three types of sessions and tracking difficulty was varied in all three.

Accuracy Data

Twelve of the subjects were tested 4 months later to obtain better accuracy data. These subjects were retrained only for 3 minutes on both tasks before adaptive automation sequences began. The twelve subjects were selected on the basis of availability. Three additional subjects were tested but their data were not used because of a system crash in one instance and because the subjects neglected to follow instructions part way through the experiment in the other two instances. These three subjects were in three different interface groups and were replaced. Two changes were made in the experimental procedure for this retesting. First, a clearer touchscreen was obtained which would enable the users to read the tabular data easier. Second, each subject was tested with a unique scenario which conformed to the experimental design. This was done to eliminate the possibility that the two scenarios (one used with eight subjects, the other used for 12 subjects) in the initial testing contributed to the results.

Questionnaire and Workload Assessment

A questionnaire was prepared to obtain judgments about several aspects of the experiment, its tasks, and the interface the subject used for the tactical assessment task. It included questions about events that occurred during the automation of the tactical assessment task. These questions were used to make a post-experimental assessment of tactical situation awareness. A similar procedure was used by Kibbe (1988). An established alternative is to interrupt the scenario with probe questions (Endsley 1988). However, because we were interested in potential automation deficit effects and assessing these by comparing performance immediately after automation to performance after a period of manual operation, interrupting the scenarios was not feasible. The questionnaire also included rating scales on aspects of the tasks and the interfaces, and biographical information. This questionnaire was completed right after the last session involving periodic automation of the tactical assessment task. Workload was measured using the NASA TLX workload assessment technique (Hart and Staveland 1988). The subjects made their judgments after completing the last data collection session, rating each of the two tasks on the six TLX dimensions.

RESULTS

Overview

We found considerable support for our hypothesis: automation deficit was least with the direct manipulation interface and greatest in the interfaces that lacked one component of direct manipulation. Our notion of deficit was expanded to include not only the concept of automation deficit, but also a deficit associated with not performing the task for awhile. We also found some selected advantages of nondirect manipulation interfaces. In particular, the tabular display of information reduced automation deficit on simpler confirmation decisions. And we found that in the initial seconds of resuming the tactical assessment task, there was less disruption of tracking performance when the keypad rather than the touchscreen was used.

The results section includes analyses of three types of data: tactical assessment performance, tracking performance, and questionnaire and workload data. The analyses of tactical assessment performance focus on assessment of automation deficit and testing the experimental hypothesis, but include supplementary analyses to answer questions raised by key results. Analyses of tracking performance focus on comparisons between single, dual, and adaptive automation conditions, and transitional tracking performance. Analyses of questionnaire data focus on assessment of situation awareness and ratings of interface properties.

Interface Style and Automation Deficit in Response Time

We assessed automation deficit by comparing subject performance on the first decision after the tactical assessment task was resumed to performance on the seventh decision. The interaction of automation deficit and interface style was significant in the 12 subjects tested twice, $F(1,8) = 6.04, p < .04$ (Fig. 4). Similar results were found in the initial testing with the larger set of 20 subjects, although they were not significant, $F(1,16) = 1.05, p = .32$ (Fig. 5). Figures 4 and 5 show that with the direct manipulation interface, initial performance was almost as good as later performance. In other words, virtually no automation deficit was found with the direct manipulation interface. In contrast, automation deficit was clearly present in the two hybrid interfaces. The difference between the first and the later response is significant in both of these interfaces using a planned comparison test (crit. diff. = 803, $p < .05$). Later performance was improved significantly if either component of direct manipulation was present. This is shown by the reduction in response time for the later response in the two hybrid interfaces. The magnitude of the deficit is an increase in response time of 35-65% in these hybrid interfaces. If neither component of direct manipulation was present, as in the command language interface, both initial and later performance were poor. This result merits further scrutiny. As we note later, there was no significant effect of interface overall (i.e., on all responses). This is evident in a plot of the times for each of the responses which shows that the command language responses are rapid on the sixth response (Fig. 6). There is a general increase in the response times after the sixth response because the high event rate presented at the beginning of the manual period was repeated.

The elevated response times for the command language interface on both the first and seventh response could be due to two factors. First, a longer automation deficit effect might occur with the command language interface, and the heightened response times on the seventh response (about 60 seconds after the first response) could be due to the continued effect of automation deficit. Second, performance with the command language interface could be influenced more by high event rates compared to the other interfaces. Both the first and the seventh responses were made under comparable high event rate conditions. The data seem to support the second explanation. Note that in Fig. 6, the increase in response times associated with the high event rate after the sixth response is relatively greater for the command language interface, suggesting that this interface may still be deficient in handling events at a high rate. The analysis that follows provides further insight on this matter.

To further explore automation deficit, we analyzed responses made in the dual-task session under approximately the same conditions as the first and seventh response in the adaptive automation session. In dual task, the equivalent responses would be the first response after the transition from difficult to moderate tracking difficulty and the seventh response after this transition. The scenarios for the dual-task session were modifications of adaptive automation scenarios, so the changes in event rate were similar, and there was a higher event rate at the transition from high to moderate tracking difficulty. This event rate was repeated around the seventh target, just as in the adaptive automation. It should be noted that prior to the transition from high to moderate tracking difficulty, and the concurrent high tactical event rate, there was a "lull" in the tactical task. Our finding was that the response times for the first and the seventh response under dual-task conditions were not different except for the command language interface (Fig. 7). In Figs. 5 and 7, performance with the direct manipulation interface is as rapid initially as it is later in both adaptive automation and dual-task conditions. The initial deficit in response time in adaptive automation with the two hybrid interfaces is not seen in the dual-task conditions. The initial level for the command language interface is similar for adaptive automation and dual-task conditions, but improves in dual-task conditions. This suggests that the transition into the high event rates may have produced the initial poor performance in the command language interface. In addition, the intermittent automation did not support the improvement in the command language interface that occurred under dual-task conditions. These results suggest that the differences we have found between the initial and later responses are due to two types of deficit. One is produced by the complete automation of a task for a period of time (automation deficit), and the other produced by no active responses for a period of time (inactivity deficit). Both components of direct manipulation may be needed to offset these two effects. With one component missing, automation deficit will still have effects. With both components missing, both types of deficit may occur. Effects of event rate are also found under dual-task conditions. As shown in Fig. 8, the first response in dual-task conditions was rapid for the two hybrid interfaces, but the second and third responses, which are made in the midst of the high event rate, are longer for the two hybrid interfaces.

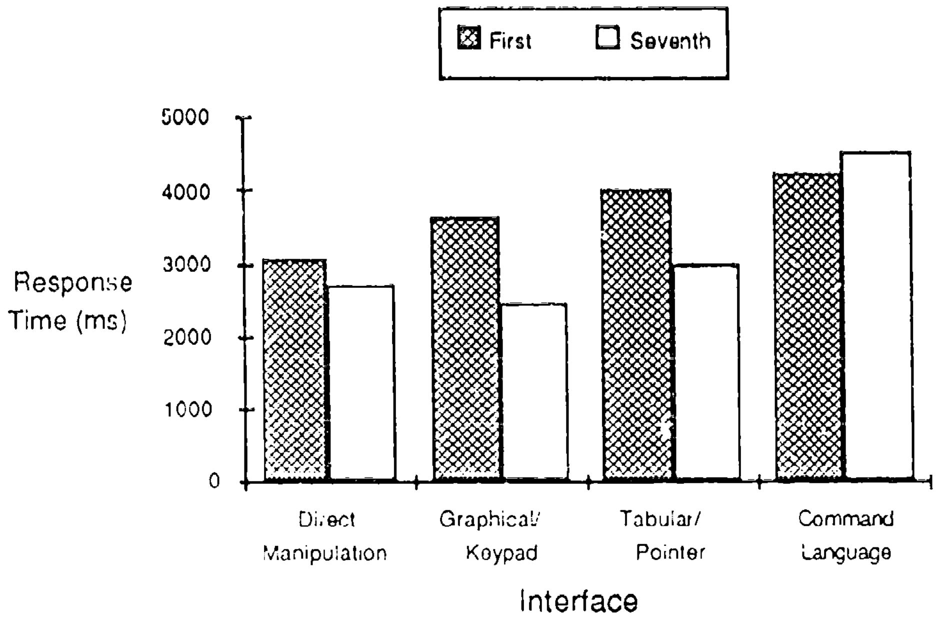


Fig. 4 -- Interface effect on automation deficit in response time in 12 subjects tested twice

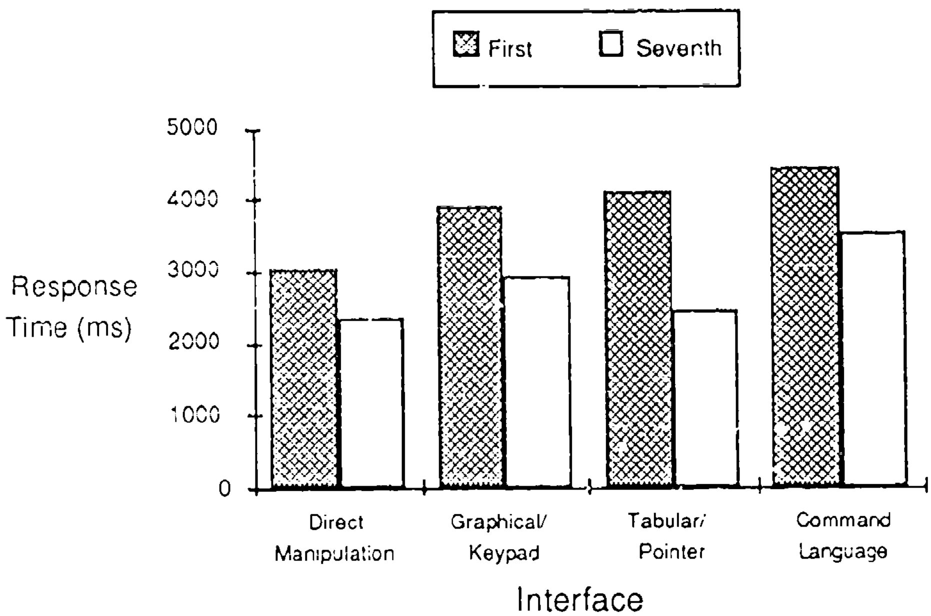


Fig. 5 — Interface effect on automation deficit in response time in initial group of 20 subjects

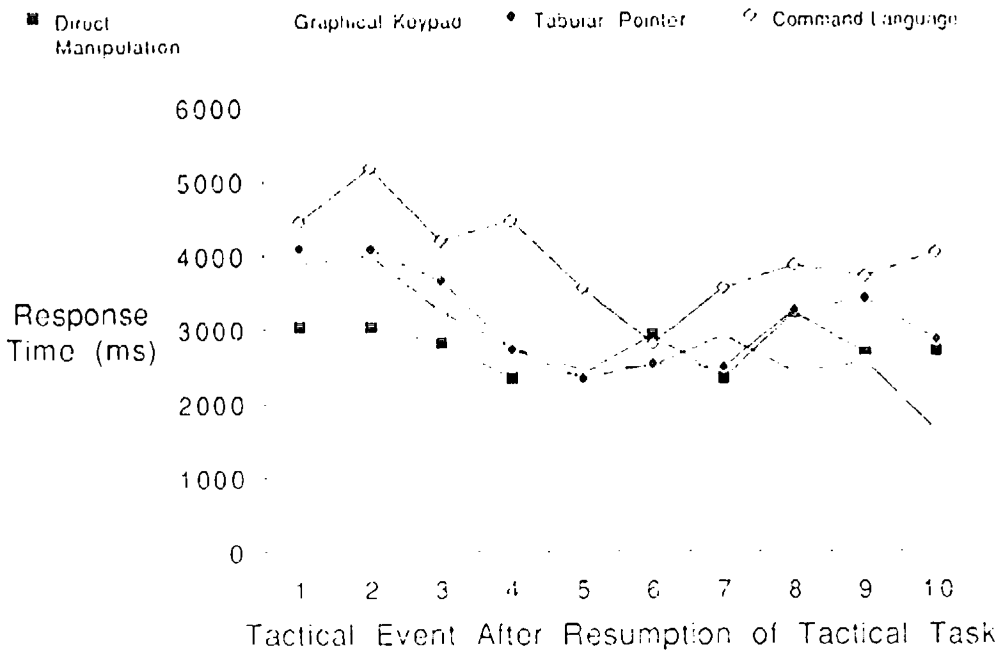


Fig. 6 — Response times for tactical events after manual resumption of the tactical assessment task

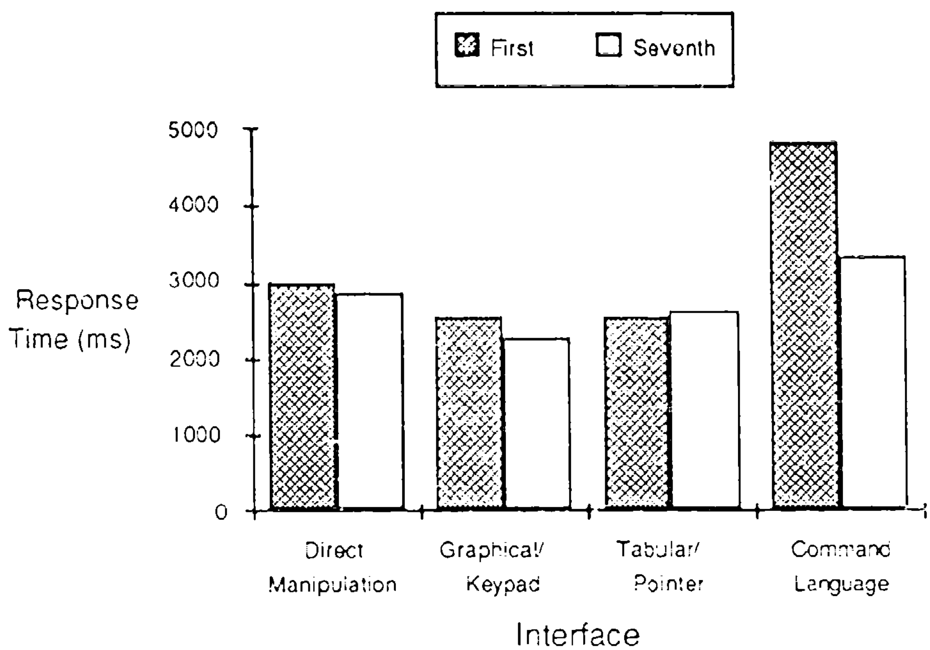


Fig. 7 — Response times in dual task with event rates and tracking conditions similar to the adaptive automation conditions for Figs. 4 and 5

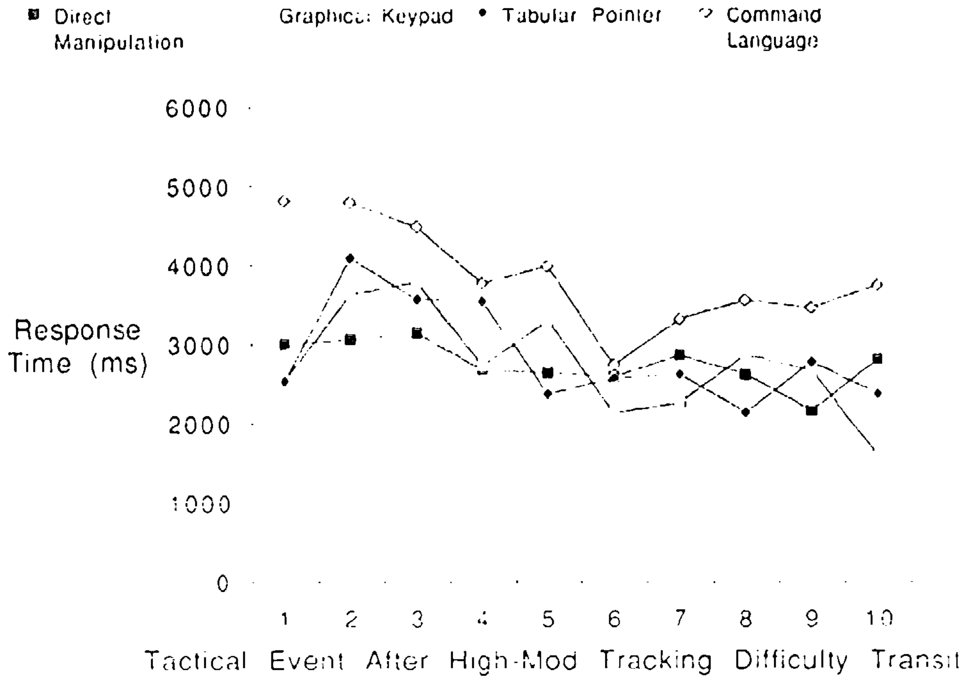


Fig. 8 — Response times for tactical events in dual-task conditions similar to the adaptive automation conditions for Fig. 6

We also found that automation deficit was related significantly to the interaction between the type of decision and the type of display, $F(1,16) = 7.89, p < .02$. On classification decisions, automation deficit was greater with the tabular displays. On confirmation decisions, the deficit was greater with the graphical displays. The interaction is best illustrated by calculating the difference between the first response and the seventh response (see Fig. 9). This pattern was also seen in the retesting four months later, although it was not as strong ($F(1,8) = 4.00, p = .08$).

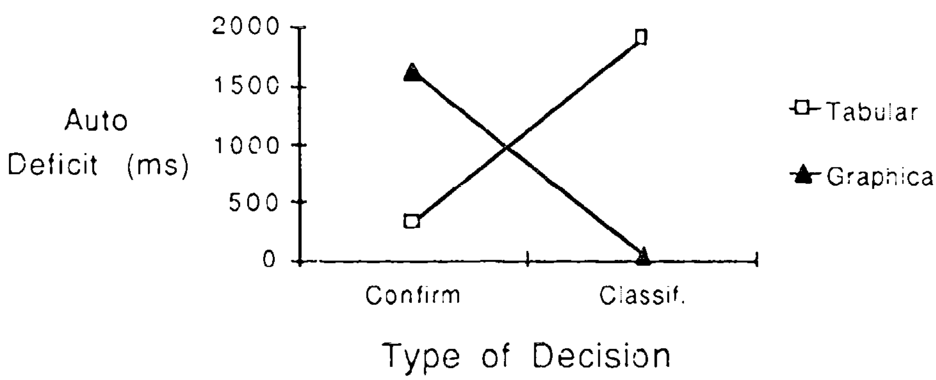


Fig. 9 — Automation deficit for different types of decisions with two types of tactical displays

Results for automation deficit were very similar when the subjects were retested. This occurred even though the touchscreen had been changed to improve the legibility of the tabular information and the retesting used a unique scenario for each subject. To analyze the effects of retesting, the ANOVA design in Table 1 was modified to include another within-subject variable called session, with two levels (original and the retesting). Both the original and the retested data for the 12 subjects was included in this analysis. The only significant effect involving session was an interaction between session and decision. Classification decisions were significantly slower in the retesting, $F(1,8) = 6.47, p < .05$. Confirmation decisions were similar in both sessions.

In our analyses of just the first and seventh response, other effects were also significant. Response time with the touchscreen was significantly quicker than with the keyboard, $F(1,16) = 4.85, p = .04$ in the group of 20 subjects. This also occurred in the group of 12 subjects tested twice, $F(1,8) = 7.13, p = .03$, and as well, responses were faster with the graphical display, $F(1,18) = 25.11, p < .001$. A significant main effect of automation deficit was also evident in both the 20 subjects, $F(1,16) = 10.26, p < .01$, and the subjects tested twice, $F(1,8) = 6.93, p < .05$. However, these main effects are secondary in importance to the interaction effect of automation deficit with interface style described above (e.g., on the basis of a main effect, one would expect quick responses with the touchscreen but this did not occur on the first response using the tabular pointer interface of Figs. 4 and 5). Two significant results were related to the type of target. There was a significant target by display type interaction ($F(2,32) = 6.70, p < .0037$, in the original data, and $F(2,16) = 12.31, p = .0006$ in the 12 subjects tested twice). Response times for the airplanes and fighters were faster with the graphical display than with the tabular display. The reverse occurred for the missile targets. This suggests that the graphical display was effective in portraying dynamic information, but not static positional information. There was also a significant interaction of target type, decision type, and automation deficit, but only in the initial testing. This effect could have been due to the specific scenarios used in the initial testing because it was not observed when we changed the scenarios for each subject in the retesting.

Tactical Assessment Speed and Accuracy

One of the reasons for retesting 12 of the subjects was to obtain data on response accuracy which was not available in the initial testing. In particular, we were interested in knowing if there were accuracy differences between the interfaces, whether there was any evidence for a speed-accuracy tradeoff, and whether there was an automation deficit in accuracy comparable to the deficit in response time. We found no significant differences either in response time or in accuracy between the four interfaces in the data obtained in retesting (Figs. 10 and 11). Thus, the four interfaces supported comparable speed and accuracy performance in "normal" operation. Accuracies for the first and the seventh responses are shown in Fig. 12. There is no evidence of a general speed-accuracy tradeoff in these data (the relationship between Figs. 10 and 11 is contrary to a speed-accuracy tradeoff), nor is there the consistent automation deficit effect that we found in the response times (Figs. 4 and 5). On the other hand, accuracy was related to the type of decision and the type of information that had to be interpreted. Accuracy for the confirmation decisions was 95% and for the classification decisions was 78%. Accuracy was lowest for classification decisions that depended upon monitoring whether a number was changing. This occurs when the subject monitors the numerical bearing of a fighter using the tabular display (Table 3).

Single, Dual, and Adaptive Automation Response Times

We were interested in comparing single-task, dual-task, and adaptive automation performance in the tactical assessment task to examine effects of learning, effects of tracking difficulty on tactical assessment, and effects of adaptive automation. Looking at the average response times per subject under single-, dual-, and adaptive automation conditions (averaging all responses, not just the first and the seventh response, and using data for 20 subjects), there were no significant differences between task conditions or between the interfaces unless the very first single task session is included. With this session in the analysis, there is a significant task effect, $F(4,64) = 8.76, p < .0001$, but still no significant effects of interface or an interaction effect between task and interface. The significant task effect represents an effect of learning for three of the interfaces. The learning showed up as an improvement in response times for all interfaces except the direct manipulation interface, which produced response times in the first single-task session that were as rapid as those in the second session (Fig. 13). The average response times show that the command language was on average slower than the others, but the difference was not significant . Furthermore, there were no significant effects as a result of changes in the difficulty of the tracking task during the dual-task session. Finally, the adaptive automation session did not result in differences in response times compared to dual-task times.

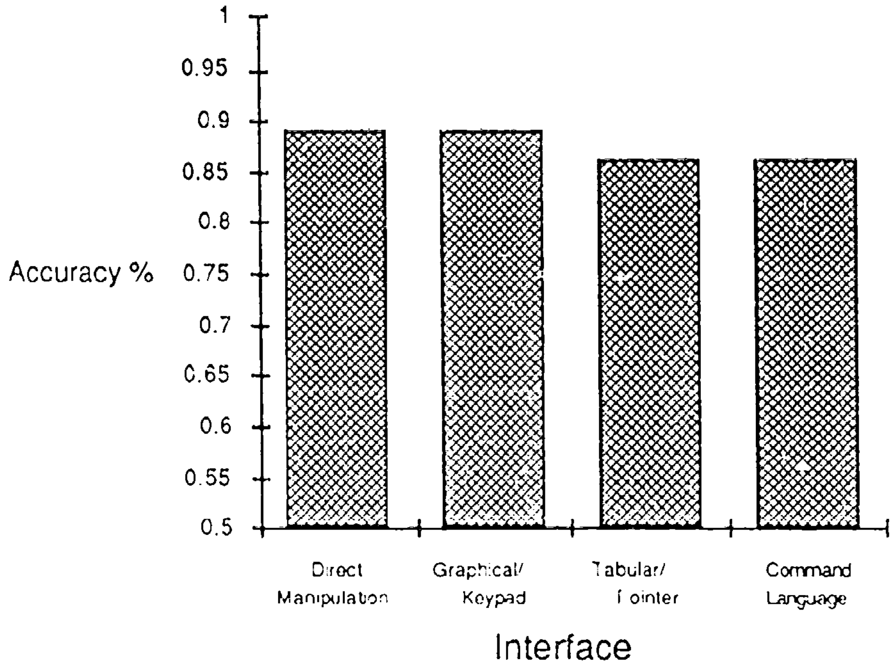


Fig. 10 — Accuracy data for the four interfaces obtained from 12 subjects who were tested a second time

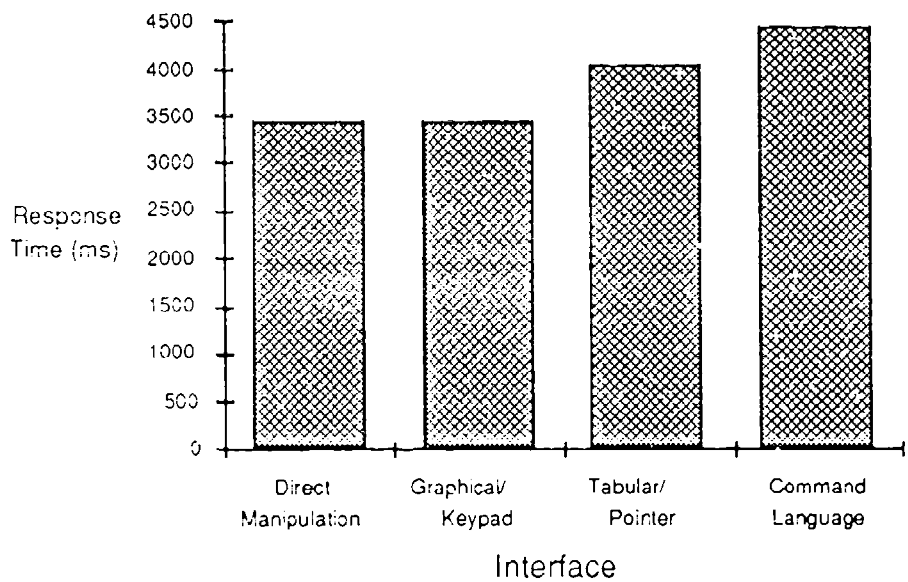


Fig. 11 — Response times for the second testing of 12 subjects when accuracy data were also obtained

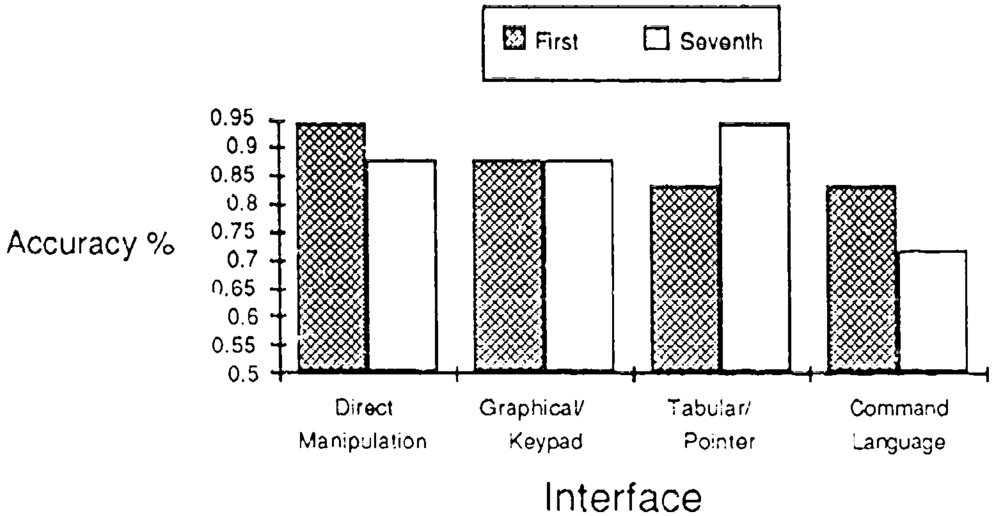


Fig. 12 — Accuracy on the first and seventh tactical events after resumption of the tactical task

Table 3 — Response Time and Accuracy on Targets for Tabular and Graphical Displays

Display	Target	Data type	Accuracy (mean)	Accuracy (stdev)	Response time (mean s)	Response time (stdev)
Graphical	AIR	Dynamic	.89	.32	3086	1958
	MIG	Dynamic	.90	.30	3076	2032
	SAM	Static	.88	.33	4092	2316
Tabular	AIR	Static	.96	.20	3831	1988
	MIG	Dynamic	.63	.49	5502	2926
	SAM	Static	.95	.22	3771	2071

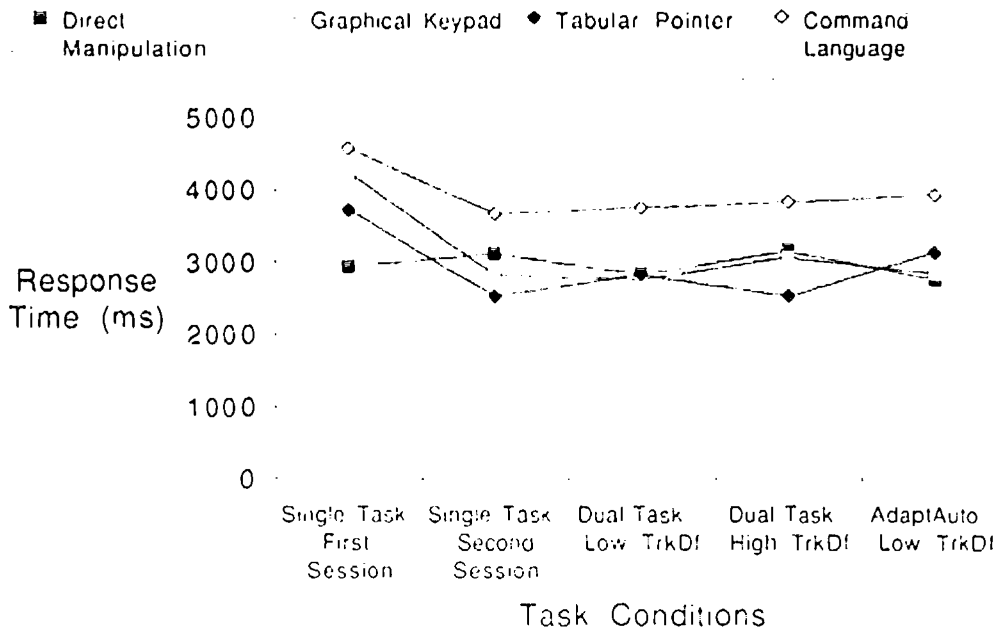


Fig. 13 - Average response times in single, dual task with two levels of tracking difficulty, and adaptive automation conditions with low tracking difficulty

A potential confound of the design is that it assesses automation deficit by comparing a particular response to a later response, and thus could be sensitive to learning effects. Several methods were used to offset this confound. First, two scenarios were used, one for 8 subjects, and the other for 12 subjects. Partial counter balancing of key events was used in these two scenarios to offset potential learning effects. Second, a regression analysis was used to specifically assess the learning effect in relation to the automation deficit effect. Recall that there were six manual periods. The cycle number was coded for the responses being analyzed as an amount of learning variable. Within each cycle, the first and the later response were coded as an automation deficit variable. In a stepwise regression, the effect of automation deficit correlated with response time more than the learning variable, accounting for 6% of the variance in the response times. Adding the learning variable to the regression accounted for an additional 2% of the variance. Thus, although response time was related to learning, the effect was minor and less than the effect of automation deficit. This regression analysis was also used to examine the effect of substituting a response time of 9999 for responses that were not completed within the 10 s response interval. Results indicated that the effect of automation deficit is greater when the substitution is *not* made because some of the substitutions are for the seventh response and would work against an automation deficit effect. Thus, we expect that the automation deficit effects of interface style that we have found would be heightened without the substitution.

Tracking Task Results

In the analyses of tracking performance, we were interested in single, dual, and adaptive automation comparisons, and cross-task effects of tactical interface style. Performance on the tracking task was analyzed using aggregate RMS (vector) error measures of accuracy as well as spectral comparisons between the driving functions and the produced tracking. RMS error and spectral analyses were available for three task conditions:

- single-task tracking;
- dual-task (tracking and tactical assessment); and
- adaptive automation (tracking alone during the difficult level and both tasks during the easier level of tracking difficulty).

An ANOVA of RMS error with interface, tracking difficulty, and task as independent variables showed significant effects of tracking difficulty ($F = 90.6$, $p < .0001$), task ($F = 68.96$, $p < .0001$), and the interaction of these two variables ($F = 38.7$, $p < .0001$). RMS error increased when the tracking difficulty was increased, and increased whenever both tasks were performed. Tracking under adaptive automation was equivalent to the appropriate single- or dual-task condition (Fig. 14).

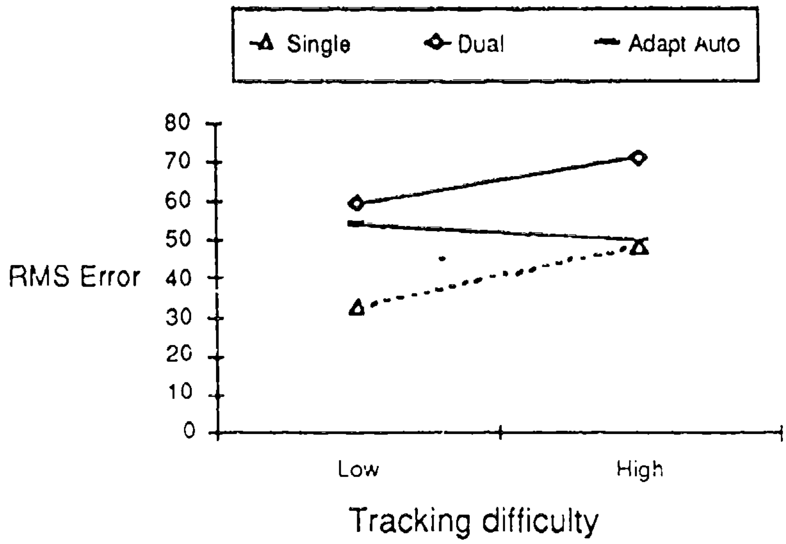


Fig. 14 — Tracking performance with changes in tracking difficulty under single-task, dual-task, and adaptive automation conditions

There were also effects related to interface style, including significant distance by engagement ($F = 4.77$, $p = .04$), and task by distance by engagement ($F = 6.50$, $p < .004$) interactions. Generally, subjects with the graphical/keypad interface had the best tracking performance and those with the graphical/touchscreen interface had the poorest performance. Much of this is due to individual differences, since this pattern was also evident in the single-task tracking condition (Fig. 16). Interestingly, the two hybrid interfaces (the graphical/keypad and the tabular/touchscreen) seemed to produce the least amount of decrement in dual-task performance, and subsequently the least amount of improvement from adaptive automation (Fig. 16). Those with the direct manipulation interface had as large a dual-task decrement as those with the command line interface, but showed a greater improvement with adaptive automation. These effects are clear when dual-task decrements and adaptive automation improvements are plotted (Fig. 17).

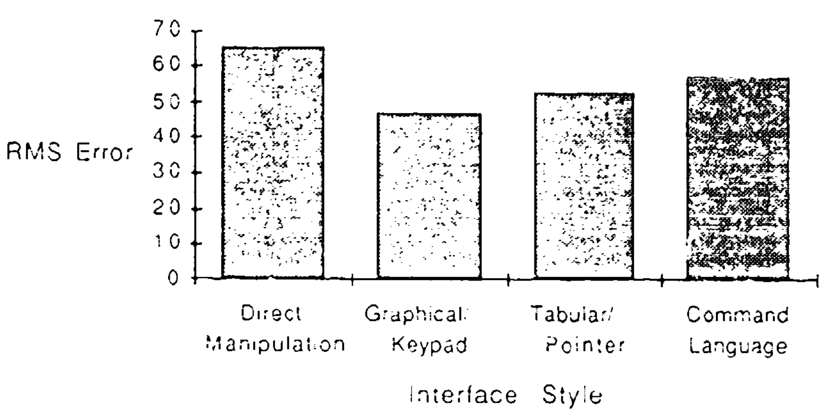


Fig. 15 — Tracking performance in all conditions by subjects assigned to the four types of tactical assessment interfaces

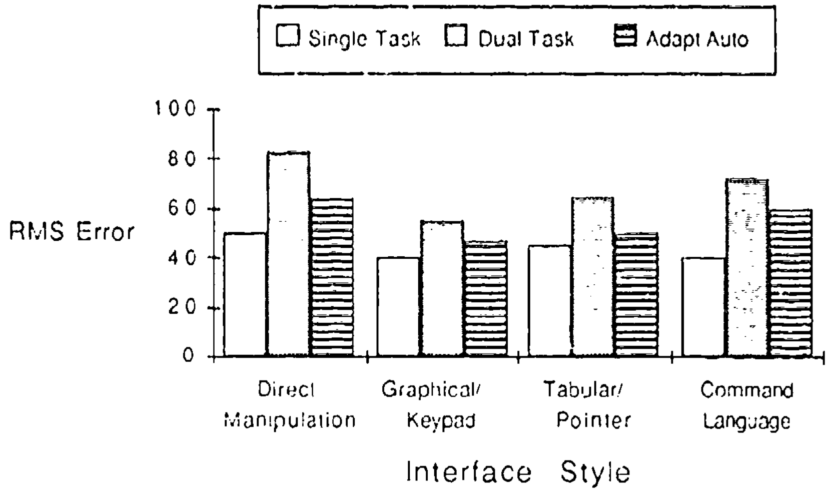


Fig. 16 — Tracking performance in each of the task conditions by subjects assigned to the four types of tactical assessment interfaces

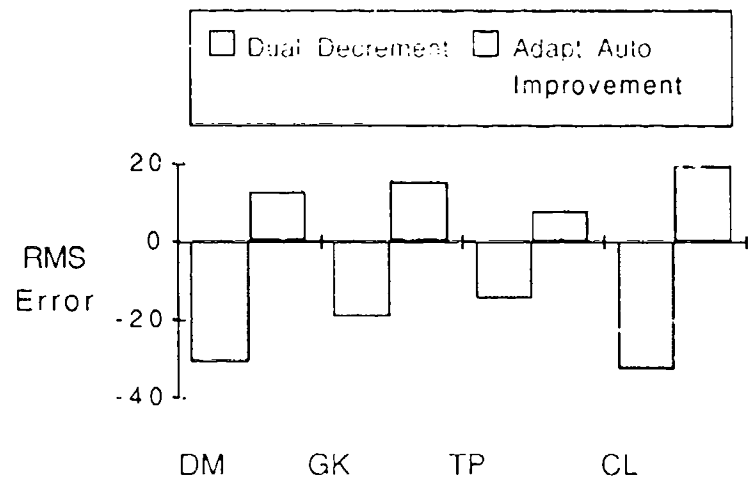


Fig. 17 — Dual-task decrement in tracking performance (single minus dual) and adaptive automation improvement (adapt auto minus dual) by interface style

Additional information about the cross task effect of tactical interface on tracking performance came when spectral comparisons were made between the driving function and the tracking performance. In particular, spectral analyses of tracking performance and the driving function were made for the initial part of each phase of low tracking difficulty. Under adaptive automation, this is when the subject must resume the tactical assessment task. For comparison, spectral analyses were made for a period later on in the low tracking difficulty phase. Care was taken so that the initial period and the comparison period had equivalent event rates in the tactical assessment task. Spectral analyses were based upon 256 point FFTs of a 21.3 s interval, sampled at .0833 Hz. From these FFTs, the spectral estimates at .05, .14, .23, .42, and .84 Hz were selected for detailed analysis since they closely matched the lower frequencies in the driving function. As noted earlier, the driving functions for the difficult and easy tracking levels were different above .97 Hz. The results reported here are ratios of tracking performance to driving function in the x axis. These data are equivalent to a measure of tracking gain. Interesting effects were found in comparing subjects who used the keypad versus those who used the touchscreen in the tactical assessment task.

With tracking alone, there were no large differences between subjects who were assigned to the keypad and those who used the touchscreen (Figs. 18(a)and 18(b)). Since the subjects had not yet used or even seen the interface for the tactical assessment task when the single-task tracking data were collected, any differences at this point would be due to individual differences in tracking performance. There is a difference at .23 Hz especially later in the period (Fig. 18(b)), but it is not consistently evident in later analyses as one would expect if it reflected a stable individual difference.

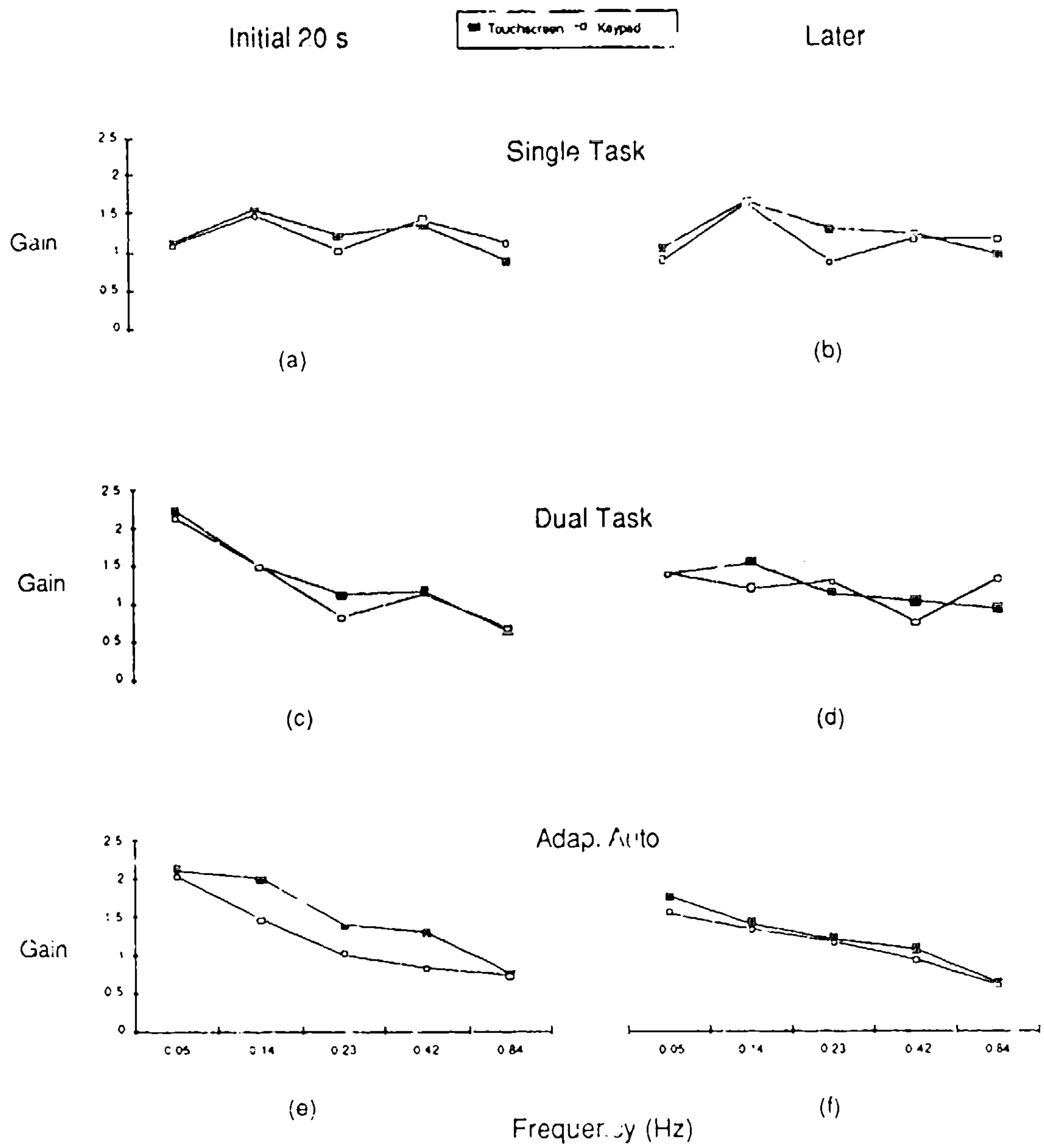


Fig. 18 — Tracking performance gain at driving function frequencies, in the initial 20 s after a change in the tracking difficulty, and later, for single, dual, and adaptive automation sessions

Under continuous dual task conditions, when the tactical assessment task was added, the gains at .05 Hz increased greatly, especially early in the phase (Fig. 18(c)). This reflects a shift in tracking toward very slow movements when the tactical assessment task is performed concurrently. This result is consistent with Wickens and Gopher (1977) who reported that power at low frequencies nearly doubled from single-task to dual-task conditions. In one analysis they did, this difference between single- and dual-task performance was reported for frequencies below .4 Hz. However, in their Fig. 6, the difference is apparent at about .66 Hz, and is reduced at about 1.5 Hz, results comparable to those here. It is not clear why the departure from the driving function at .05 Hz is so much greater earlier, compared to later (Fig. 18(d)). Note that the type of input device in the tactical assessment task did not produce any systematic effect.

With adaptive automation, the same type of disruption at .05 Hz occurs both early (Fig. 18(e)) and later (Fig. 18(f)), and again especially early in the phase. However, there is also a difference between those with the keypad and touchscreen. Those with the touchscreen show greater gain (i.e., effort) than those using the keypad at .14, .23, and .42 Hz early in the phase. Later in the phase, both are performing similarly. It could be that the subjects with the touchscreen were less adaptable to the changing demands of the tracking task, because the period just prior to this was when the tracking task required greater effort at .14, .23, and .42 Hz.

An alternative explanation to this result is that the increased tracking effort found with the touchscreen led to a reduction in tracking error (Wickens 1991). To assess this explanation, RMS errors were calculated for the initial and later periods in the single, dual, and adaptive automation conditions. These errors were calculated for the time periods that were used for the spectral analyses shown in Fig. 18. The results (Fig. 19) show no indication that the initial, greater tracking effort by those using the touchscreen produced a reduction in tracking error, and in fact, it appears that those using the touchscreen initially had higher tracking error than those using the keypad, in both dual and adaptive automation conditions. The results also show that tracking error rates were higher initially for both dual and adaptive automation conditions.

This finding suggests that touchscreen usage might conflict with a continuous tracking task. In the adaptive automation condition, initial control of this manual task may interfere with making required adjustments to the tracking task. However, the effect is transient. This result suggests that the touchscreen in the tactical assessment task induces a performance automation deficit in the tracking task. Those using the keypad for the tactical assessment task show better tracking in the initial phase of resuming the tactical assessment task. This occurs even though the subjects have been continually doing the tracking task. Thus, an hypothesis that input device in the intermittent tactical task produced transitory effects in the tracking task is viable.

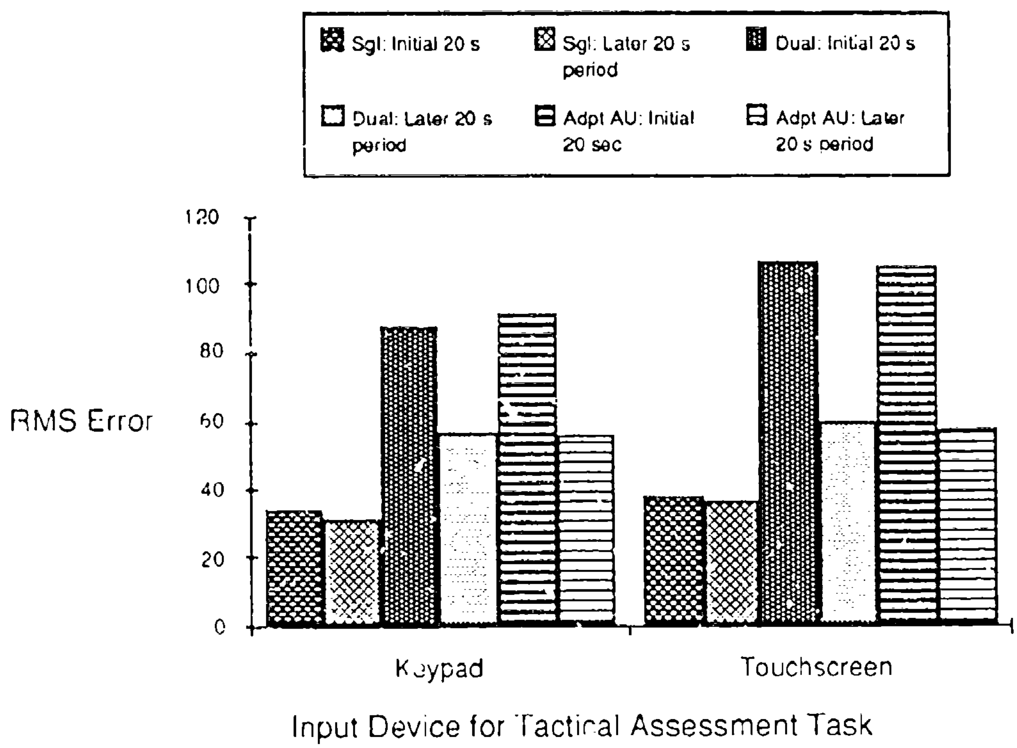


Fig. 19 — Tracking error for the conditions in Fig. 18

Questionnaire Results

Eleven questions probed for knowledge of events occurring in the tactical assessment task while it was automated. Each question had a correct answer. The questions asked about the number and type of tactical targets during automation, the accuracy and speed of the automation performance, and the duration of the automation periods. Two analyses were performed.

First, the number of correct responses from individuals who had used different interfaces was tallied. A two-tailed Fisher exact probability test was used for the individual questions because of the small sample sizes. Correct and incorrect responses were tallied for each question by the two interface variables, engagement and semantic distance, producing 2 X 2 contingency tables. None of the results were significant, indicating that neither semantic distance nor engagement was related to more correct responses on any of the specific questions.

Second, the total number of correct responses was calculated and used in a multiple ANOVA with engagement, semantic distance, and the interaction of these two as the independent variables None of the three was significant for the total number of correct responses. Additional analyses are planned for these questions to assess awareness of specific aspects of automated performance. For example, the scenarios were designed to make the automated system appear slower in successive automation periods, and we are interested in subjects who noticed this change.

Twenty-four rating scales were used to obtain subjective judgments about feelings of control, feelings of awareness, preferences for the interface, judgments of the difficulty in learning and performing the tasks and specific aspects of the tasks, and ability to anticipate the changes in automation. These ratings were analyzed in a multiple ANOVA with engagement, semantic distance, and the interaction of these two as the independent variables.

Significant results were found on five scales. The most interesting results were the ratings of ability to anticipate changes in automation and awareness of the tactical situation at the end of automation. Ability to anticipate the changes in automation was significantly different according to semantic distance. Those with the graphical display felt that they were able to anticipate the changes more often than those with the tabular interface. Furthermore, those with the graphical display felt that they were significantly more aware of the tactical situation at the end of automation. Debriefing confirmed that subjects with the graphical interface noticed the ebb and flow in activity during automation (i.e., activity picks up just before the task switches from automatic to manual), but those in the two tabular interfaces did not. This effect occurred despite the fact that tactical events were appearing in both types of display at the exact same time, that the number of items in both is always the same at any particular moment, and that the ebb and flow of activity is exactly the same in each type of display. Only two of the subjects in the tabular interfaces thought they could anticipate the changes, and one of these was referring to the ending of the manual task. Awareness of this type is based upon projections about the course of the scenario and is similar to the Level 3 SA in Endsley's (1988) model. This level of awareness was produced with the low semantic distance displays. Combined with the response time results, this result supports the conclusion that low distance displays are particularly important for maintenance of Levels 2 and 3 SA. However, Level 1 SA may not require a low distance display, and in fact, better SA for particular elements (e.g., color) may be produced with a display that enhances the separation of these elements. In our studies, this occurred with the command language interface.

Significant effects of minor importance were found on several other scales. The classification of airplane targets was rated as significantly easier by those using the tabular display and by those using the touchscreen. Those using the graphical/keypad interface felt that they had more control over the tracking task than those in the other three interfaces, and those with the graphical display felt that it was more difficult to classify fighters. Several of these ratings are consistent with the expected difference in the ability to interpret static and dynamic information.

The first 11 questions were intended to assess the person's knowledge of different aspects of the tactical situation during the automation periods. Answers on these questions were scored as correct or incorrect, and a χ^2 goodness-of-fit test was used to determine if the distribution of answers was significantly different from chance. Significance was obtained for four questions, $p < .05$. On three of these, the answers were significantly more correct than chance. These included judgments about the maximum number of targets simultaneously displayed during automation (Question 2, Appendix C, correct answer was "six"), whether all the targets were of the same type in one automation phase (Question 7, Appendix C, correct answer was "false"), and whether the automation and manual periods were different in duration (Question 8, Appendix C, correct answer was "about the same"). The answers to whether most of the amber tracks during automation were hostile or friendly (Question 9, Appendix C, correct answer was "hostile") were significantly different from chance, but were incorrect, probably reflecting a bias. There were some differences in accuracy related to interface style, but the small sample size in the cells of the χ^2 tables precluded statistical tests. The results suggest that correct responses on the type of targets during automation (Question 7, Appendix C) were related to graphical displays. Correct responses on whether the numbers of targets during automation were different from the numbers during manual phases were related to the graphical/keyboard interface. Finally, the incorrect responses on the disposition of amber targets during automation (Question 9, Appendix C) were mostly with indirect engagement (keyboard) interfaces. These results, combined with the result about ability to anticipate the changes in automation, suggest that the subjects were able to accurately report on some global characteristics of the tactical situation that existed during automation, but not on the details. Finding incorrect responses on Question 9 suggests questions for further research about whether judgments of automated behavior can be incorrect not only with respect to system reliability (e.g., Palmer and Degani 1991; Parasuraman, Bahri, Molloy and Singh 1991), but also with respect to specific types of tactical actions taken during automation.

Workload

The workload scores were used as the dependent variables in a multiple ANOVA with engagement, semantic distance, and the interaction of the two as between-subject variables, and task as a within-subject variable. None of the effects was significant, indicating that perceived workload with the four interfaces was not different, and perceived workload of tracking was not different for the four groups of subjects. Figure 20 shows the average ratings.

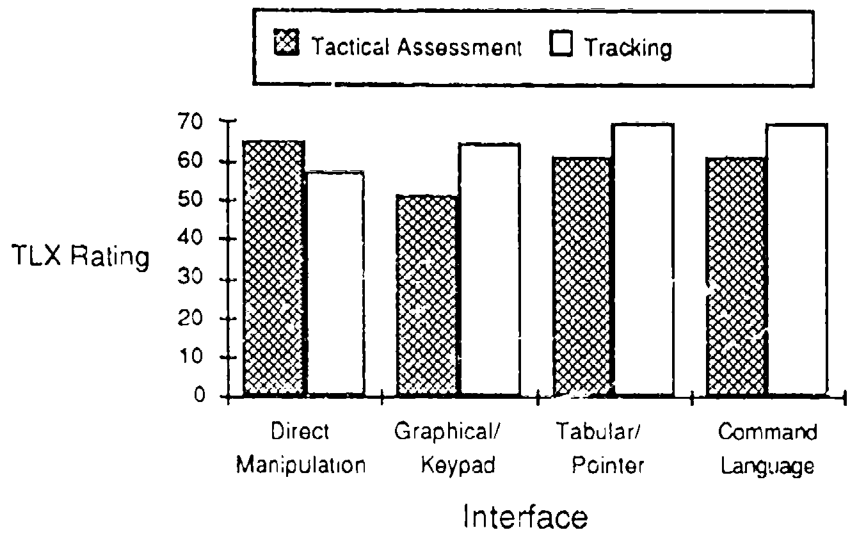


Fig. 20 — TLX ratings for the two tasks by interface for the tactical assessment task

DISCUSSION

Relationships to Other Cockpit Studies

A key departure of the present study from prior research on cockpits is our examination of the overall style of the interface rather than a narrow focus on selected aspects of displays and controls. Research on cockpit interfaces typically focuses on two areas: display formats and data entry techniques. In one of the few studies to examine the relationship between interface format and awareness of flight status, Steiner and Camacho (1989) presented flight status information in two display forms: alphanumeric and iconic. They also varied the amount of flight information, predicting that the best interface depends on how much information is presented. The display presentations were self-paced, and dependent measures included accuracy in answering questions about flight status and time to view the display. The researchers found that with small amounts of information, viewing time and errors were similar for the two formats. For larger amounts of information, the iconic format supported faster viewing times. However, the results have limited utility because a question-answer procedure was used rather than a flight simulation.

Reising and Hartsock (1989) evaluated three different designs of a warning/caution/advisory (w/c/a) display. All three designs described the status of several aircraft systems, including any malfunctions. In the first and second designs, only one aircraft system at a time was displayed and a complete system description was provided. In addition, a checklist of required responses was presented. In contrast, the third design presented an abbreviated description of the status of all aircraft systems simultaneously but no checklist. The only difference between the first and second designs was that the first included a pictorial layout of the switches the pilot used to respond to messages. Reising and Hartsock evaluated these displays with the subjects piloting a simulated training flight. Simulated emergencies were programmed into the flight to engage the w/c/a display. The results indicated that task completion was faster for the two designs that provided a complete description and checklist and slower for the abbreviated description without the checklist. The pictorial presentation of the switches did not improve performance.

Several studies have demonstrated the benefits of touch input. White and Beckett (1983) used a strike aircraft simulation to compare three forms of entering waypoint data into a navigation system. They compared the traditional mode of keyboard input to two alternatives: voice and touch-sensitive display. The last mode presented a touch-sensitive keypad on the display along with two data fields, several labeled buttons, and a directional representation of the four compass positions. They measured altitude variation during data entry, head-down time and data entry time. The direct voice input produced better altitude maintenance and less head-down time, but longer data entry time due to both a delay in the voice recognition system, and the tendency of the pilots to verify each entry before continuing. The benefits of the voice entry mode occurred because this mode enabled the display of verification data on the head-up display (HUD). The other two modes required this data on a head-down display. However, current technology enables keyed data to be displayed on the HUD rather than the head-down display. So the benefits of voice input found by White and Beckett can be obtained with other data entry techniques as long as the pilot does not have to go head-down. Similar results on data entry techniques were obtained by Smyth and Dominessy (1988) who, using a tactical assessment task, found that voice input was slower than both touch panel and switch entry. They combined these forms of data entry with three types of object selection: touch, eye gaze control of a cursor, and eye gaze alone without a cursor. They found that gaze control of a cursor produced faster selection than gaze alone. Touch panel selection was as rapid as gaze control of a cursor and more accurate. Reising and his colleagues (Curry, Reising, and Zenyuh 1985; Barthelemy, Reising, and Hartsock 1991) have examined target designation in both 2-D and 3-D space and found that touch and hand positioning produced better performance in 2-D and 3-D respectively compared to joystick and voice devices.

Besides the studies on interface format and interaction techniques, a variety of display and control parameters have been studied to determine the design of future glass cockpits. These include the gain function for cursor control (Rauch 1988), map magnification requirements (Allen 1988), moving map vs moving aircraft displays (Marshak, Kuperman, Ramsey, and Wilson 1987), and formatting of information on cockpit control/display units (Mann and Morrison 1986).

Several studies have addressed the design of flight control systems under different types of automation. Bernotat (1981) discusses trends in the automation of guidance and control systems, pointing out that in military aircraft the trend has been to achieve flight stabilization by having the computer handle the control dynamics. The pilot acts as a supervisory monitor, entering control values and monitoring performance. Bernotat discusses three types of function allocation that can be achieved. In the automatic mode, the pilot sets desired levels for altitude, speed, or heading and engages the autopilot to achieve the commanded settings. This capability exists in current aircraft, although there are problems with separation of displays and controls (Mitchell 1991). In a semi-automatic mode, the computer continues to maintain flight stabilization but the desired flight path is directed by the pilot. Control input is through an analog control such as a "natural feel stick" that provides kinesthetic and proprioceptive feedback. Finally, there is a back-up guidance mode. In the past, this would be a transition back to direct hydraulic control with the pilot taking over direct control of the flight surfaces. However, in modern aircraft which cannot be flown without computer-controlled stabilization, this refers to backup computers that function like the main system. The pilot's role would be similar with the backup system engaged.

Research in cockpit design like the above has produced essential information and insights about existing and emerging technology. However, past avionics advances have been limited by technology and change has been gradual. Within this context, a selective research strategy has been appropriate. However, with the development of reliable, readable CRT and liquid crystal displays, technology is less of a limiting factor in designing displays for the cockpit. Newer displays are programmable, and their development involves software developers and graphics specialists. In this atmosphere, greater variability and departure from existing practice is possible. Concurrent with rapid developments in display technology is the development of fly-by-wire control systems, which remove the mechanical linkage between the pilot's movements and control surface responses. Thus, the manner of the pilot's input control is programmable. This new flexibility in cockpit displays and pilot input allows an integrated approach to cockpit design, which considers the dialogue between the pilot and the computer system as a whole, i.e., the "look and feel" of the dialogue. Such an approach goes beyond a focus on control mode and display design—the focus of most cockpit interface research—to an analysis and evaluation of the complete interface. Our study was designed to take this latter approach.

Another important aspect of our research is the focus on automation. While the design of effective avionics interfaces is always challenging, the recent introduction of more complicated automation into the cockpit has added new dimensions to the challenge. With this automation, the pilot performs such tasks as programming, monitoring, and failure detection. Effective interface design for these types of tasks is especially challenging for the following reasons.

First, there is less experience with avionics interfaces for such tasks. Much of the research on cockpit interfaces has focused on the design of singular displays and controls. This research has been very effective in producing enhanced performance and safety. But the increasing complexity of modern aircraft has made it necessary to move away from singular display and control design to integrated cockpits, both in civilian and military aircraft. The peak of display complexity was reached in civilian aircraft with the Concorde and in military aircraft with the F-4. Since these designs, there has been a progressive reduction in the number of displays and an introduction of integrated and multimodal displays. Programmable control of these displays has also been introduced. But even with extensive development efforts, it is not always possible to anticipate how programmable systems will be used in actual service. For example, Wiener (1989) reports that pilots have learned how to program around the limitations of the computer to obtain results that cannot be programmed directly. Furthermore, although there has been a decrease in the number of discrete displays in modern aircraft, there has been an increase in the number of alerts (Veitengruber 1977). Ironically, the subsystem that has seen the greatest growth in alerts is the automatic flight control system (AFCS). According to Veitengruber, the number of alerts in this subsystem increased at about twice the rate of any other subsystem between 1965 and 1970. He also found that pilots were unanimous that any further increase in the number of alerts would be unacceptable. Irving and Irving (1990) point out that the automated flight management system is an additional subsystem overlaying the traditional subsystems in nonautomated aircraft, and thus has increased workload rather than reduced it. The basic problem may be that interfaces for automated systems are being modeled after the traditional aircraft interfaces.

Second, modern systems provide greater flexibility in display generation. Although the cockpit hardware places some limits on the flexibility of the software and thus the flexibility of the interface (Martz and Mueller 1989), advances in avionics and the incorporation of advanced software languages will increase this flexibility in the future. This flexibility can also increase the complexity of the system by providing multiple pages of information. In fact, the reduction of displays in newer aircraft has not meant a reduction in available information. As Rouse, Rouse, and Hammer (1982) point out, computer-generated displays may substitute the serial display of information for the parallel display of information. Although this provides more opportunity for creative solutions and integrated displays, it also means that it is less likely that there has been basic research that is relevant to evaluating the proposed solution. For example, much of the research on human-computer interaction has been performed on desktop business systems and applications and may have little relevance to an aerospace application. A complicated series of key commands to move through a database may be acceptable in a desktop application but is viewed by many pilots as inappropriate during the approach phase of a landing. An example of the difference is data entry procedures. In cockpits, procedures have been developed to verify the correct entry of data such as waypoints (e.g., Aarons 1988). Equivalent procedures are rarely considered in desktop systems.

Third, when automation is introduced, a pilot may move "in and out of the loop," with subsequent effects such as loss of situation awareness and need for performance "warm-up." For pilots, these effects are issues of great concern. Little is known about these effects, what conditions produce them, and how the interface might exacerbate or minimize them. Ironically, situations in which the pilot must assume control of an automated task may occur when the pilot is at a special disadvantage (Federal Aviation Administration 1990), i.e., when the situation is especially complex and automation is unable to handle an unforeseen scenario. In such situations, the pilot is typically involved in tasks other that those that have been automated, and is not aware of the situation in the tasks at which the automation will shortly fail. An example of this was the crash of a B-1A bomber (McDaniel 1988). The cause of the crash was instability produced by a mismatch between the center of gravity and the center of lift. This mismatch was produced because the fuel was not transferred as the wings were being moved forward. The transfer should have been done manually, but the pilot thought the transfer subsystem was under automatic control. However, the automatic stabilization system masked the degrading handling qualities of the plane until the situation was unrecoverable. Thus, the pilot was not only unaware that the automated fuel transfer was not occurring, but also that the plane's stability was degrading. Failure occurred when it was too late for the pilot to take corrective action.

Our results show that interface designs for automation must be based upon sensitive assessment of the transition period. Blocked designs that examine the aggregate effect of factors for an extended period of time may not show the improvements and deficits that accrue with different features of the interface. For example, Parasuraman, Bahri, Molloy, and Singh (1991) did not find any evidence of impaired performance in a manual period following automation when they examined average performance over 10-min blocks. The effects we have found, such as the advantage of the graphical display for transitions into a classification decision, the advantage of a tabular display for transitions into a confirmation decision, and the improved adaptivity of tracking when using a keypad, would not show up in a blocked design that did not carefully control the independent variables at the transition points from automated to manual operation. Unfortunately, this makes the experimental design extremely challenging because detailed temporal control of the events in the scenario is required, response sequences that are required must be carefully evaluated for confounding effects, and performance measures must be "windowed" into specific aspects of the data collection session. Paradigms for th s type of performance assessment are not well established.

Implications for Direct Manipulation Theory

Our research has implications for the theory of direct manipulation as well as for the design of interfaces for dynamic, multitask systems. The theoretical implications are based on both empirical results as well as observations we made during the course of developing the interfaces and conducting the experiment.

On the positive side, we found that the theoretical predictions that we made were generally supported. This result is noteworthy for several reasons. First, this research is a rare example of designing interfaces to test a theory explicitly. Previous studies of direct manipulation and command language interfaces have used interfaces for established applications which may not fairly represent the theoretical concepts. Second, our predictions concern a specific aspect of performance (automation deficit) in a complex, multitask situation. Either challenge—specificity of prediction or complexity of context—would put demands on a theory. Both were present in this research, which makes the successful predictions of the theory especially impressive.

However, we also found that the theory has limitations. First, the theory does not address interfaces that include a mixture of interface styles and that are probably the rule more than the exception in complex applications. The reason is that complex applications involve different types of tasks. A single interface style may not support all tasks in an optimal manner. In the HHN theory, a general interface for the application is assumed. This requires choosing a representation that is suitable for most tasks but may not be optimal for certain tasks. Thus, choosing a single interface style for a complex application may produce suboptimal performance on some aspects of the application.

This point is important because it is based not only on observation but on empirical results. In our data, we found evidence that the optimal display for reducing automation deficit depends upon the type of decision. In terms of theoretical predictions, the shortcoming of the HHN theory is that it (and we) did not make predictions about the simple decisions. In retrospect, it is evident that the theory would have to be modified to address decision complexity. It is likely that the confirmation decisions were best supported by the tabular display because the user did not need complete information about the object but simply needed to know the value of a single parameter. If the model world metaphor is implemented faithfully, then different representations for different decisions are not directly possible. Thus, an extension of the theory should be considered to support different levels of representation for different requirements.

Second, we found that the theory does not always help with detailed aspects of interface design. Our goal was to evaluate interfaces that had different levels of distance and engagement. The iterative design process we used forced many decisions about details of each of the four interfaces. Many of these decisions were based upon performance considerations and could not be based upon logical derivations from the tenets of the theory. Furthermore, the performance constraints were related to the specific application. For example, the relative placement of the two windows (horizontal or vertical) had an impact on how easy it was to use hands dedicated to the two tasks. This is a stimulus-response compatibility issue that the theory does not address. In essence, the theory is not performance based as are other formal models such as GOMS. Rather, its merit lies in explaining how aspects of the interface relate to cognitive complexity.

Finally, we found that distance and engagement are difficult terms to define operationally and to evaluate. Our experiment required interfaces that combined different levels of distance and engagement. In other words, these were design requirements for the interfaces. One of the problems is how to distinguish between distance and engagement. Our empirical results suggest that they are not independent, in that the degree of automation deficit in the command line interface was not a combination of the deficits in the two hybrid interfaces, which each lacked one aspect of direct manipulation. HHN themselves point out that engagement is only present when both semantic and articulatory directness are present.

The interfaces that we produced represented combinations of different levels of distance and engagement. What is not clear is how much distance and engagement were actually present. It is apparent that any interface that allows the person to perform a task successfully has bridged the distance of the gulfs of execution and evaluation as HHN discuss them. The command language interface we produced supported the user's goal of performing the task and, therefore, reduced semantic distance to a greater degree than would an interface that would not support this goal. And yet, it did not provide a view of the model world as a pilot would normally think of it, so considerable distance still remained. Better precision about the degree of distance and engagement in an interface would be helpful.

Generalizations to Other Domains

Based upon our findings, we expect that intermittent operation of comple 'asks will be more effective with direct manipulation interfaces in a variety of dynamic, real-time systems. Although our results were found in a cockpit application, extension to other systems is appropriate, particularly systems in which the operator is intermittently moving from one task to another. To envision potential generalizations, it is helpful to characterize our application in abstract terms. The dual-task application we tested included 1) a continuous task with simple perceptual demands, rigorous manual demands, and minor cognitive complexity; and 2) an intermittent task with varying cognitive and perceptual complexity and minimal manual demands. The cognitive complexity of the intermittent task was manipulated by changing the interfaces and by changing the decisions. The results were interpretable at an abstract level: increases in the cognitive complexity of an interface adversely affect the resumption of its use after a period of automation. This principle certainly holds for systems that include the two types of tasks. The principle would probably hold for systems that have greater complexity on the continuous task. In fact, the effects of interface would probably be greater. The key to appropriate generalization is that relatively little cognitive interaction existed between the two tasks. There was some manual interaction as noted below.

Generalization may not be warranted if the system includes multiple tasks that use similar cognitive processes. In a multitask application, there may be different forms of expressions to the various tasks; the interaction of these expressions is an important issue. Direct engagement in particular may introduce incompatibilities. We found that tracking performance was adversely affected in the initial seconds of resuming pointing with the touchscreen. The cause was an incompatibility between the two forms of manual manipulation. The important issue is whether direct manipulation interfaces to different tasks could compete. According to Wickens (Wickens and Liu 1988), the answer is yes. In his resource theory, competition for attentional resources occurs whenever information to the user is in similar modalities or is in a similar code (e.g., spatial or verbal). Competition also occurs whenever responses are similar. Thus, two direct manipulation interfaces, which both have spatial graphical displays and which both require pointing devices, could produce competition for attentional resources. Thus, the generalization of our results to other multiple task systems should be made with consideration given to possible competition between aspects of the direct manipulation interface.

Summary

The study reported here was an experimental test of a theory of direct manipulation applied to simulated cockpit interfaces operated under intermittent automation. The hypothesis was that a direct manipulation interface would produce less automation deficit in resuming a task that had been automated for awhile, compared to other interfaces that did not implement direct manipulation fully. Two components of direct manipulation were examined systematically: semantic distance and engagement. The experiment used a dual-task paradigm with the subjects constantly performing a tracking task and intermittently performing a tactical assessment task, using different interfaces. Results supported the hypothesis and provided additional insight into the specific conditions in which direct manipulation leads to improved performance. Results also showed some advantages of nondirect manipulation interfaces.

ACKNOWLEDGMENTS

We acknowledge Rob Jacob for contributing to the initial idea for the hypothesis and Rob Carter and Diane Damos for contributions to the design. Many individuals reviewed the interfaces and contributed to their development. We thank the subjects for their participation.

REFERENCES

Aarons, R. N. (1988). Waypoint insertion errors. *Business and Commercial Aviation.* June, 78-85.

Allen, C. P. (1988). Variable magnification considerations for airborne, moving map displays. *Proceedings of the Human Factors Society 32nd Annual Meeting.* Santa Monica, CA: Human Factors Society.

Badre, A. N. (1982). Selecting and representing information structures for visual presentation. *IEEE Transactions on Systems, Man, and Cybernetics,* 12(4), 495-504.

Ballas, J. A., Heitmeyer, C. L., and Pérez, M. A. (1991). Interface styles for the intelligent cockpit: Factors influencing automation deficit. *Proceedings of the AIAA Computers in Aerospace 8 Conference.* Baltimore, MD, Oct. 22-24. AIAA, Washington, D.C.

Barthelemy, K. K., Reising, J. M., and Hartsock, D. C. (1991). Target designation in a perspective view, 3-D map using a joystick, hand tracker, or voice. *Proceedings of the Human Factors Society 35th Annual Meeting.* Santa Monica, CA: Human Factors Society, 97-101.

Beltracchi, L. (1987). A direct manipulation interface for heat engines based upon the Rankine cycle. *IEEE Transactions on Systems, Man and Cybernetics,* 17(3), 478-487.

Benson, C. R., Govindaraj, T., Mitchell, C. M., and Krosner, S. P. (1989). Effectiveness of direct manipulation interaction in the supervisory control of FMS parts movement. *Proc. IEEE International Conference on Systems, Man, and Cybernetics,* Nov. 14-17, 1989, Cambridge, MA. Piscataway, NJ: IEEE, 947-952.

Bernotat, R. K. (1981). Man and computer in future on-board guidance and control systems of aircraft. In B. Shackel (Ed.) *Man-computer interaction: Human factors aspects of computers and people.* Rockville, MD: Sijthoff and Noordhoff.

Curry, D. G., Reising, J. M., and Zenyuh, J. P. (1985). Target designation by speech, joystick, and touch control: format design implications, in *Proceedings of the National Aerospace and Electronics Conference.* Dayton, Ohio. Piscataway, NJ: IEEE, 923-928.

DeBor, J. and Swezey, R. (1989). *Man-machine interface issues in nuclear power plants.* U. S. Nuclear Regulatory Commission, NUREG/CR-5348.

Draper, S. W. (1986). Display managers as the basis for user-machine communication. In D. A. Norman and S. W. Draper (Eds.) *User centered system design.* Hillsdale, NJ: Erlbaum Associates, 339-352.

Endsley, M. R. (1988). Situation awareness global assessment technique (SAGAT). *Proceedings of the IEEE 1989 Aerospace and Electronics Conference NAECON 1988.* (Cat. N. 88CH2596-5). Piscataway, NJ: IEEE, 789-795.

Federal Aviation Administration. (1991). *Report of Study on Airlines' Anticipated Near Future Cockpit Control and Display Capabilities and Plans for Data Link Communication.* Report No. DOT/FAA/CT-TN91/7 Washington, DC: U. S. Department of Transportation.

Federal Aviation Administration. (1990). *The National Plan for Aviation Human Factors* (Draft), Vol. 1.Washington, DC: U. S. Department of Transportation.

Hart S. G. and Staveland L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In P. A. Hancock and N. Meshkati (Eds.) *Human Mental Workload.* North-Holland: Elsevier Science Publishers, 139-183.

Hutchins, E. (1986). *Metaphors for interface design.* Paper presented at NATO-sponsored workshop on Multimodal Dialogues Including Voice. Vanaco, Corsica, France.

Hutchins, E., Hollan, J. D., and Norman, D. A. (1986). Direct manipulation interfaces. In D. A. Norman and S. W. Draper (Eds.) *User centered system design.* Hillsdale, NJ: Erlbaum Associates.

Irving, S. and Irving, J. (1990). Workload and human-computer interaction issues in the automated cockpit/air transport environment. *CHI '90 Workshop on Computer-Human Interaction in Aerospace Systems*, Seattle, WA, April 1-2, 1990.

Jacob, R. J. K. (1989). Direct manipulation in the intelligent interface. In P. A. Hancock and M. H. Chignell (Eds.) *Intelligent interfaces: Theory, research and design.* North-Holland: Elsevier Science Publishers.

Karat, J. (1987). Evaluating user interface complexity. *Proceedings of the Human Factors Society 31st Annual Meeting.* Santa Monica, CA: Human Factors Society, 566-570.

Kibbe, M. P. (1988). Information transfer from intelligent EW displays. *Proceedings of the Human Factors Society 32nd Annual Meeting.* Santa Monica, CA: Human Factors Society, 107-110.

Kieras, D. E. (1988). Towards a practical GOMS model methodology for user interface design. In M. Helender (Ed.) *Handbook of Human-Computer Interaction.* North-Holland: Elsevier Science Publishers.

Kieras, D. E. (1990). An overview of human-computer interaction. *Journal of the Washington Academy of Sciences*, 80(2), 39 - 70.

Maris, J. (1990). Personal communication.

Marshak, W. P., Kuperman, G., Ramsey, E. G., and Wilson D. (1987). Situation awareness in map displays. *Proceedings of the Human Factors Society 31st Annual Meeting*, Santa Monica, CA: Human Factors Society.

Martz, S. and Mueller, E. (1989). Expert system-conventional processing interface. *Proceedings of the IEEE 1989 Aerospace and Electronics Conference NAECON 1989.* Piscataway, NJ: IEEE, 1020-1026.

Mann, T. L. and Morrison, J. G. (1986). Effects of display density and format type on control display unit format design. *Proceedings of the IEEE /AIAA 7th Digital Avionics Systems Conference* (Cat. no. 86CH2359-8) Piscataway, NJ: IEEE, 330-337.

McDaniel, J. (1988). Rules for fighter cockpit automation. *Proceedings of the IEEE 1988 National Aerospace and Electronics Conference: NAECON 1988.* (Cat. no. 88CH2596-5) Piscataway, NJ: IEEE, 831-838.

Mitchell, C. M. (1991). Human-computer interaction in the glass cockpit: Issues, problems and possibilities. *CHI '91 Workshop on Advances in Computer-Human Interaction in Complex Systems*. New Orleans, LA, April 28-29, 1991.

Nunnally, J. C. (1970). *Introduction to psychological measurement*. NY: McGraw-Hill.

Palmer, E. and Degani, A. (1991). Electronic checklists: Evaluation of two levels of automation. *Proceedings of the Sixth International Symposium on Aviation Psychology*, Columbus, Ohio, The Department of Aviation, The Ohio State University, 178-183.

Parasuraman, R., Bahri, T., Molloy, and Singh, I. (1991). Effects of shifts in the level of automation on operator performance. *Proceedings of the Sixth International Symposium on Aviation Psychology*, Columbus, Ohio, The Department of Aviation, The Ohio State University, 102-107.

Parasuraman, R., Bahri, T., Deaton, J. E., Morrison, J. G., and Barnes, M. (1990). *Theory and design of adaptive automation in aviation systems* (progress report). Washington, D.C.: Cognitive Science Laboratory, The Catholic University of America.

Rauch, S. (1988). Determination of a gain-function relating control force to cursor velocity. *Proceedings of the Human Factors Society 32nd Annual Meeting*. Santa Monica, CA: Human Factors Society.

Reising, J. M. and Hartsock, D. C. (1989). Advanced warning/caution/advisory displays for fighter aircraft. *Proceedings of the Human Factors Society 33rd Annual Meeting*. Santa Monica, CA: Human Factors Society, 66-70.

Reising, J. M. (1977). Multifunction keyboard configurations for single-seat, air-to-ground fighter cockpits. *Proceedings of the Human Factors Society 21st Annual Meeting*. Santa Monica, CA: Human Factors Society, 363-366.

Rouse, S. H., Rouse, W. B and Hammer, J. M. (1982). Design and evaluation of an onboard computer-based information system for aircraft. *IEEE Transactions on Systems, Man, and Cybernetics*, **12**(4), 451-463.

Shneiderman, B. (1982). The future of interactive systems and the emergence of direct manipulation. *Behavior and Information Technology*, **1**, 237-256.

Smyth, C. C. and Dominessy, M. K. (1988). Comparing oculometer and head-fixed reticle with voice or switch for tactical display interaction. *Proceedings of the Human Factors Society 32nd Annual Meeting*. Santa Monica, CA: Human Factors Society.

Steiner, B. A. and Camacho, M. J. (1989). Situation awareness: icons vs alphanumerics. *Proceedings of the Human Factors Society 33rd Annual Meeting*. Santa Monica, CA: Human Factors Society.

Veitengruber, J. E.(1977). Design criteria for aircraft warning, caution, and advisory systems. *Journal of Aircraft*, **15**(2), 574-581.

White, R. G. and Beckett, P. (1983). Increased aircraft survivability using direct voice input. In *AGARD-CP-347*, Paris, Advisory Group on Aerospace Research and Development.

Whiteside, J., Jones, S., Levy, P. S., and Wixon, D. (1985). User performance with command, menu, and conic interfaces. In *Proc. CHI '85 Human Factors in Computing Systems Conference*. ACM Press, 185-91.

Wickens, C. D. (1991). Personal communication.

Wickens, C. D. and Andre, A. D. (1990). Proximity compatibility and information display: Effects of color, space, and objectness on information integration. *Human Factors*, 32(1), 61-78.

Wickens, C. D. and Gopher, D. (1977). Control theory measures of tracking as indices of attention allocation strategies. *Human Factors*, **19**(4), 349-366.

Wickens, C. D. and Liu, Y. (1988). Codes and modalities in multiple resources: A success and a qualification. *Human Factors*, 30(5), 599-616.

Wiener, E. L. (1989). *Human Factors of Advanced Technology ("Glass Cockpit") Transport Aircraft.* Tech. Rep. NASA CR 177528. Washington, DC: NASA.

Wiener, E. L. (1985). Beyond the sterile cockpit. *Human Factors*, 27, 75-90.

Williams, H. P. and Wickens, C. D. (1991). Electronic map interpretation in a dual-task context. *Proceedings of the Human Factors Society 35th Annual Meeting*. Santa Monica, CA: Human Factors Society, 92-96.

Wolf, C. G. and Rhyne, J. R. (1987). A taxonomic approach to understanding direct manipulation. *Proceedings of the Human Factors Society 31st Annual Meeting*. Santa Monica, CA: Human Factors Society.

Ziegler, J. E. and Fahnrich, K. P. (1988). Direct manipulation. In M. Helender (Ed.) *Handbook of Human-Computer Interaction*. North-Holland: Elsevier Science Publishers.

Zlotnik, M. A. (1988). Applying electro-optical display technology to fighter aircraft-flying with feeling again. *Proceedings of the IEEE 1988 National Aerospace and Electronics Conference: NAECON 1988.* (Cat. No. 88CH2596-5), Piscataway, NJ: IEEE, 191-197.

Appendix A
CONSENT FORM

Project: Intelligent cockpit interface

The purpose of this experiment is to develop an understanding of how interfaces for automated systems ought to be designed. The subjects will perform tasks that are similar to those performed in an aircraft cockpit. Two tasks will be done:

1. a tracking task in which a joystick is moved to keep a sight on a moving target, and
2. a tactical assessment task in which decisions about tactical threats are made and entered into the simulated cockpit computer system.

Both tasks will involve simple hand and arm movements. A questionnaire will be given at the end of the experiment, which includes questions about the experiment and other information which might be related to how well people can perform the experimental tasks.

The benefits of this research include advancement of our knowledge about computer interface design and interfaces for automated systems.

All data collected will be kept confidential and will not be recorded with personal identification information. Published reports of the research will not include any data on the performance of specific people.

There are no known risks or discomforts in this experiment. The experimental sessions will be held in a comfortable environment. In the event that a subject has unexpected discomfort or has a complaint, he or she should contact Jim Ballas, room 203, building 16, phone 404-7988 or 767-2774.

Participation in the experiment is voluntary and may be terminated at any time for any reason.

As a voluntary participant, I have read the above description of the research project. Anything I did not understand was explained to my satisfaction. I agree to participate in this research.

______________________________ (Participant) __________ (date)

______________________________ (Witness) __________ (date)

______________________________ (Investigator) __________ (date)

Appendix B
TASK INSTRUCTIONS

Introduction and tracking instructions

You will soon be starting the experiment involving two tasks: tracking and tactical assessment. The tracking task is to keep a "gunsight" on a moving target and the tactical assessment task is to make tactical decisions about potential threats and targets. We will be doing the experiment in phases. You will be doing the tracking task first, then the tactical assessment task, and finally both under different conditions. Today you will be doing the two tasks separately. At the next session, we will combine the two tasks. We are interested in how well you do the tasks alone and together, and will be measuring how accurately and how quickly you perform the tasks.

The first task you will do is a tracking task. In the bottom right of the screen, you will see a small image of an airplane moving. You will use the joystick to move the "gunsight" and try to keep it on the plane. The software is programmed to act somewhat like an airplane, so it will take some practice to use it. You will receive practice on this task before we start the full experiment.

Periodically, the tracking task will become more difficult. You will notice this because the plane "target" will start to move around more quickly. When this happens, you will have to devote most of your attention to this task. Now you will do this task alone for about 15 minutes.

Instructions: Graphical Keypad

The second task is a tactical decision task. To do this task, you will have to interpret information in the right window about **fighters**, **airplanes**, and **missile** sites. Each of these can be **hostile** or **neutral** depending on what they are doing.

The fighters are symbolized by swept back wings and they are hostile if they are heading toward your location in the center of the radar range lines. If they are flying away from you, they are neutral.

The airplanes are the fatter plane symbols with the square wings and are hostile bombers if they are flying fast. If they are flying slowly, they are commercial airline planes.

The missile sites are hostile if they are within the horizontal range of the outer radar line. This means you will eventually fly close enough for them to hit you.

When the items come on the screen, they are colored black because the automatic sensors do not have enough information to classify them as hostile or neutral. After a while, the color will go to red, blue or amber. If the color is red or blue, then the computer system has been able to classify them as hostile (red) or neutral (blue). However, you must confirm the computer. Using the keypad, you enter 5 for neutral or 6 for hostile and then the number of the item. You must enter either 5 or 6 first. If you make a mistake on the 5 or 6 you can clear it with the * key on the keypad. The computer is never wrong about its selections, so this task is very easy for you.

When the item goes to amber, you must decide whether it is hostile or neutral. You do this by using the three rules above:

Fighters:
heading toward your location = hostile.
flying away from you = neutral.

Airplanes:
flying fast = hostile bombers.
flying slow = neutral commercial airline planes.

Missile sites
within outer range on x axis = hostile
outside of outer range on x axis = neutral

Once you have decided, you use the keypad to enter 5 for neutral or 6 for hostile and then the number of the item. You must enter either 5 or 6 first. If you make a mistake on the 5 or 6, you can clear it with the * key on the keypad.

In order to do this decision task as effectively as possible, you should watch each item when it is black to determine whether it is hostile or neutral. Then if it goes to amber, you will be ready to make your decision.

To ensure that you understand the tactical assessment rules, would you please rephrase them in your own words to the experimenter:

Rule for fighters -

Rule for airplanes -

Rule for missiles -

Instructions: Command Language

The second task is a tactical decision task. To do this task, you will have to interpret information in the right window about **fighters**, **airplanes**, and **missile** sites. Each of these can be **hostile** or **neutral** depending on what they are doing.

The fighters are abbreviated MIG and they are hostile if their bearing in the first column is *not* changing. This means that they are heading toward your location. If the bearing is changing, they are flying away from you and are neutral.

The airplanes are abbreviated AIR and are hostile bombers if their velocity in the second column is around 800. If their velocity is around 300, they are commercial airline planes.

The missile sites are abbreviated SAM and are hostile if their distance from your flight path is 150 or less in the third column. This means you will eventually fly close enough for them to hit you. If this distance is greater than 150, they do not pose a threat and are neutral.

When the abbreviations come on the screen, they are colored black because the automatic sensors do not have enough information to classify them as hostile or neutral. After a while, the color of the abbreviation will go to red, blue or amber. If the color is red or blue, then the computer system has been able to classify them as hostile (red) or neutral (blue). However, you must confirm the computer. Using the keypad, you enter 5 for neutral or 6 for hostile and then the number of the item. You must enter either 5 or 6 first. If you make a mistake on the 5 or 6 you can clear it with the * key on the keypad. The computer is never wrong about its selections, so this task is very easy for you.

When the item goes to amber, you must decide whether it is hostile or neutral. You do this by using the three rules above:

Fighters:
bearing constant = hostile.
bearing changing = neutral.

Airplanes:
velocity about 800 = hostile bombers.
velocity about 300 = neutral commercial airline planes.

Missile sites
within 150 of flight path = hostile
greater than 150 of flight path = neutral

Once you have decided, you use the keypad to enter 5 for neutral or 6 for hostile and then the number of the item. You must enter either 5 or 6 first. If you make a mistake on the 5 or 6 you can clear it with the * key on the keypad.

In order to do this decision task as effectively as possible, you should watch each item when it is blue to determine whether it is hostile or neutral. Then if it goes to amber, you will be ready to make your decision.

To ensure that you understand the tactical assessment rules, would you please rephrase them in your own words to the experimenter:

Rule for fighters -

Rule for airplanes -

Rule for missiles -

Instructions: Direct manipulation

The second task is a tactical decision task. To do this task, you will have to interpret information in the right window about **fighters**, **airplanes**, and **missile** sites. Each of these can be **hostile** or **neutral** depending on what they are doing.

The fighters are symbolized by swept back wings and they are hostile if they are heading toward your location in the center of the radar range lines. If they are flying away from you, they are neutral.

The airplanes are the fatter plane symbol with the square wings and are hostile bombers if they are flying fast. If they are flying slowly, they are commercial airline planes.

The missile sites are hostile if they are within the horizontal range of the outer radar line. This means you will eventually fly close enough for them to hit you.

When the items come on the screen, they are colored black because the automatic sensors do not have enough information to classify them as hostile or neutral. After a while, the color will go to red, blue or amber. If the color is red or blue, then the computer system has been able to classify them as hostile (red) or neutral (blue). However, you must confirm the computer. You simply select the item by touching it and then touch the appropriate panel on the side. The computer is never wrong about its selections, so this task is very easy for you.

When the item goes to amber, you must decide whether it is hostile or neutral. You do this by using the three rules above:

Fighters:
heading toward your location = hostile.
flying away from you = neutral.

Airplanes:
flying fast = hostile bombers.
flying slow = neutral commercial airline planes.

Missile sites
within outer range on x axis = hostile
outside of outer range on x axis = neutral

Once you have decided, you select the item by touching it and select the identification by touching one of the side panels.

In order to do this decision task as effectively as possible, you should watch each item when it is black to determine whether it is hostile or neutral. Then if it goes to amber, you will be ready to make your decision.

To ensure that you understand the tactical assessment rules, would you please rephrase them in your own words to the experimenter:

Rule for fighters -

Rule for airplanes -

Rule for missiles -

Dual-task Instructions

In this part of the experiment you will do both tasks. When you are doing both tasks, you will need to scan back and forth. Each task is too difficult to do with peripheral vision. You should use a strategy of moving your eyes back and forth between the two task windows. Once you have something to respond to in the tactical window, respond as fast as possible with an accurate response. Then get back to the tracking window.

The color of the "gunsight" will tell you if you are *tracking satisfactorily*. If the "gunsight" goes to yellow, you are not doing the task well enough, and you should devote more attention to the tracking. The criterion for this signal is based upon how well you did this task alone.

Instructions for Adaptive Automation Session

The purpose of this part of the experiment is to examine some of the effects of automation on human performance of tasks in the cockpit. To do this, we will have you working with a system that will periodically have automation introduced.

The automation will take over the tactical assessment task. The software is programmed to take over this task when the tracking task becomes more difficult. You will be doing the tracking task **all** of the time, and **intermittently** doing the tactical assessment task.

When the computer is doing the tactical task, you should periodically check it to keep up to date. This will enable you to resume this task effectively. At the end of the experiment we will be asking questions about what was happening in the tactical situation while the computer was performing this task.

Two signals will keep you informed about automation of the tactical assessment task:

1. A beep signals a *change* in the automation of the tactical task. A low pitched beep sounds when this task is automated, and a high pitched beep sounds when you must resume the task.

2. The color of the border around the tactical window *always* indicates if you should be doing it. When the border is green, you should be doing this task. When it is black, the computer is doing the task.

In summary:

1. Do the tracking all the time: do it better if the gunsight goes yellow.

2. Drop the tactical assessment when you hear a beep, start it when you hear the next beep. Check the border if you are unsure.

3. Use the rules to figure out the status of every track while it is black so you can handle the amber items when the color changes. If the item goes to red or blue, simply confirm this.

4. Scan back and forth between the windows to do both tasks. Try to respond to the color changes in the tactical window as soon as they occur.

5. Check the tactical window periodically while the computer is doing it to keep up to date so that you will be prepared to resume the task.

Appendix C
QUESTIONNAIRE

Subject ID: ___________ Date:______________

1. There were _____ tracks during automation than during manual phases.
a. fewer
b. more
c. about the same

2. The largest number of items simultaneously displayed during automation was
a. two
b. three
c. four
d. five
e. six
f. seven

3. The first track you handled in each manual phase was a type (fighter, air, missile) that
a. that had occurred first in the preceding automation phase.
b. had not occurred in the preceding automation phase
c. showed up as amber in the preceding automation phase

4. Automation made one mistake in each phase in classifying the tracks.
a. True
b. False

5. From one phase to the next, automation became _______ in responding after the tracks changed from black to red/blue/amber.
a. slower
b. faster

6. Automation was slower in handling the amber tracks.
a. True
b. False

7. In one automation phase, all the tracks were the same type (fighter, air, missile).
a. True
b. False

8. The automation periods were ______ than the manual phases.
a. shorter
b. longer
c. about the same

9. Most of the amber tracks during the automation turned out to be
a. neutral
b. hostile

10. The first event in each automation phase was always a ______
a. fighter
b. missile
c. air

11. How many amber tracks occurred in each automation phase?
a. one
b. two
c. three
d. four

12. Do you feel that you had control over the tracking task?

___1___	___2___	___3___	___4___	___5___	___6___	___7___
strongly agree						strongly disagree

13. Do you feel that you had control over the tactical assessment task when you took over after automation?

___1___	___2___	___3___	___4___	___5___	___6___	___7___
strongly agree						strongly disagree

14. Do you feel that you had control over the tactical assessment task after you had been doing it for a few minutes?

___1___	___2___	___3___	___4___	___5___	___6___	___7___
strongly agree						strongly disagree

15. Do you feel that you had control over the tactical assessment task while it was automated?

___1___	___2___	___3___	___4___	___5___	___6___	___7___
strongly agree			neutral			strongly disagree

16. How much did you like the interface for the tactical assessment task?

___1___	___2___	___3___	___4___	___5___	___6___	___7___
very much			neutral			very little

17. How well did the interface match the way you would naturally think about the tactical assessment task?

___1___ ___2___ ___3___ ___4___ ___5___ ___6___ ___7___

very compatible — neutral — very incompatible

18. How directly did the interface enable you to perform the tactical assessment task?

___1___ ___2___ ___3___ ___4___ ___5___ ___6___ ___7___

very direct — neutral — not very direct

19. How slow or fast were you able to select items?

___1___ ___2___ ___3___ ___4___ ___5___ ___6___ ___7___

very slow — neutral — very fast

20. How slow or fast were you able to decide if an item was hostile or neutral?

___1___ ___2___ ___3___ ___4___ ___5___ ___6___ ___7___

very slow — neutral — very fast

21. How slow or fast were you able to tell the computer if an item was hostile or neutral?

___1___ ___2___ ___3___ ___4___ ___5___ ___6___ ___7___

very slow — neutral — very fast

22. How easy or difficult was it to *learn* to do the tracking task?

___1___ ___2___ ___3___ ___4___ ___5___ ___6___ ___7___

very easy — neutral — very difficult

23. How easy or difficult was it to *learn* to do the tactical assessment task?

___1___ ___2___ ___3___ ___4___ ___5___ ___6___ ___7___

very easy — neutral — very difficult

24. How easy or difficult was it to classify fighters?

___1___ ___2___ ___3___ ___4___ ___5___ ___6___ ___7___

very easy — neutral — very difficult

25. How easy or difficult was it to classify airplanes?

___1___ ___2___ ___3___ ___4___ ___5___ ___6___ ___7___

very easy — neutral — very difficult

26. How easy or difficult was it to classify missiles?

1	2	3	4	5	6	7
very easy			neutral			very difficult

27. How aware were you of the number of targets the automated system was handling?

1	2	3	4	5	6	7
very aware			neutral			not very aware

28. How aware were you of the types of targets the automated system was handling?

1	2	3	4	5	6	7
very aware			neutral			not very aware

29. How aware were you of the classifications the automated system was making?

1	2	3	4	5	6	7
very aware			neutral			not very aware

30. How aware were you of the occurrence of amber items while automation was on?

1	2	3	4	5	6	7
very aware			neutral			not very aware

31. How aware were you of the tactical situation at the beginning of automation?

1	2	3	4	5	6	7
very aware			neutral			not very aware

32. How aware were you of the tactical situation at the end of automation?

1	2	3	4	5	6	7
very aware			neutral			not very aware

33. How easy or difficult was it to do the tactical assessment task immediately following the automation period?

1	2	3	4	5	6	7
very easy			neutral			very difficult

34. How easy or difficult was it to do the tactical assessment task after you had been doing it for a few minutes?

___1___ ___2___ ___3___ ___4___ ___5___ ___6___ ___7___

very easy — neutral — very difficult

35. Were you able to anticipate the changes in automation?

___1___ ___2___ ___3___ ___4___ ___5___ ___6___ ___7___

never — sometimes — always

36. Please check the appropriate category for your profession:

___engineering ___computer science ___psychology
___business ___pilot (___fulltime ___part time)
___other: ______________________________

37. If you are a pilot, please list the types of aircraft you have flown and the approximate number of hours in each:

38. Please check the appropriate category for your age:

___under 20 ___20-24 ___25-39 ___40-44 ___45-49 ___50-54
___55-60 ___60-64 ___65-69

39. Please check the appropriate category for your education level:

___less than high school ___high school ___some college
___college ___some graduate ___masters
___doctorate

40. How would you rate your health today?

___1___ ___2___ ___3___ ___4___ ___5___ ___6___ ___7___

very good — neutral — not very good

41. Please check the appropriate category for the amount of sleep you have received within the past 24 hrs

___under 2 hrs ___2-4 hrs ___5-6 hrs ___7-8 hrs ___9+ hrs

42. Please check how long it has been since you last ate?

___under 1 hr ___1-2 hrs ___3-4 hrs ___5-6 hrs ___6+ hrs

43. Please indicate the following personal characteristics:
___Female ___Male
___Left handed ___Right handed
___Corrective lenses used during experiment (vision is corrected to: ______)
___Corrective lenses not required

44. Do you have any disabilities which may have had an effect on your performance in this experiment? If so please briefly describe them?

45. What is your opinion of the tactical assessment display? How well does it provide information about the tactical situation? Any suggestions for changes? On what types of tasks would the display and control be especially useful?

46. We would appreciate any comments you could make about the experiment from your perspective:

Appendix D
TLX INSTRUCTIONS

TLX rating

We are not only interested in assessing your performance but also the experiences you had during the different task conditions. Right now we are going to use a technique to examine the workload you experienced. Because workload may be caused by many different factors, we would like you to evaluate several of them individually, using six scales. Please read the following descriptions of the six scales carefully. If you have a question about any of the scales in the table, please ask me about it. Then evaluate each task by putting an "X" on each of the six scales at the point which matches your experience. Each line has two endpoint descriptors that describe the scale. Note that the scale goes from "good" on the left to "bad" on the right. Please consider your responses carefully for each of the two tasks.

TLX weights

The rating scales are extremely helpful but their utility suffers from the tendency people have to interpret them in different ways. For example, some people feel that mental or temporal demands are the essential aspects of workload regardless of the effort they expended on a given task or the level of performance they achieved. Others will have very different feelings. The evaluation you are about to perform is a technique to assess the relative importance **to you** of the six scales you used to rate the tasks. The procedure is simple: You will be presented with a series of pairs of rating scale titles and asked to choose which of the items was more important to your experience of workload in the task. Each pair of scale titles will appear on a separate card.

Circle the scale title that represents the more important contributor to workload for the specific task.

Please consider your choices carefully and make them consistent with how you used the rating scales for the task. Don't think that there is any correct pattern; we are only interested in your opinions.

If you have any questions, please ask them now. Otherwise, start whenever you are ready.