

The effects of dialogue-based task allocation on system performance in a computer-aided air traffic control task

CHRISTINE A. RIEGER and JOEL S. GREENSTEIN
Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061

The present study investigated dialogue-based human-computer task allocation. Five levels of allocation mode, four command strategies and a control condition, were examined in a high-task-load situation using a Latin square design. Five male participants acted as controllers in a simplified representation of an air traffic control system, sharing their task with a computer controller. Performance measures were metered on-line, and subjective ratings were recorded between trials. There were significant differences among the groups in the percentage of planes landed by the human-computer system, the total number of commands executed, and the number of allocation commands used. Other variables suggested for further study include comparisons of input mode, dialogue-initiation style, and degraded performance levels of the computer controller.

The tremendous impact of computer technology has advanced the operation of many systems. As systems become increasingly automated, the role of the human operator will change. One goal may be to modify this role to increase productivity and yet enhance interest and motivation to work. Similarly, a basic tenet of human-factors engineering is to allocate tasks in a way that achieves an optimum human-machine balance.

There are a number of problems associated with the introduction of computerized automation. First, economic considerations resulting in a strategy to computerize everything possible may result in underutilizing human resources (Tomeski & Lazarus, 1975). Second, Danaher (1980) suggests that the introduction of automation to air traffic control systems and the associated operator unburdening can result in boredom, inattentiveness, and the development of unsound work habits. Third, Wiener and Curry (1980) report a perceptible loss of skill proficiency in pilots aided by flight-deck automation. For these reasons, the central task allocation problem is to achieve a balance that assigns functions of acceptable difficulty to the human, as Chapanis (1965) suggests, while taking advantage of the computer resources available. Specifying that tasks be shared between the human and computer components of a system may be an operative solution to this problem.

An on-line shared task allocation schema has been examined in a number of studies (Chu & Rouse, 1979; Greenstein & Rouse, 1982; Rouse, 1977, 1981; cf. Rieger & Greenstein, 1982, for a recent review). Com-

This research was supported by the Office of Naval Research under ONR Contract N00014-81-K-0143 and Work Unit SRO-101. The effort was supported by the Engineering Psychology Programs, Office of Naval Research, under the technical direction of John J. O'Hare. Reproduction in whole or part is permitted for any purpose of the United States government.

puter aiding was introduced in an attempt to facilitate system performance by coordinating multiple tasks. A multiple task was defined as a set of discrete, but related, tasks, such as monitoring a number of related instruments in an industrial plant or guiding a number of aircraft in an air traffic control system.

Because there are multiple-task situations in which the human and computer are mutually capable, Rouse (1977, 1981) suggests that each decision-making party should have responsibility for all tasks that can be adequately performed. Instead of a static allocation plan, one that is unchanging over time, a situation-dependent or dynamic allocation plan, in which tasks are allocated on the basis of temporal availability, is advocated. Unlike a static allocation plan, a dynamic one specifies that both the human and computer can perform the same types of tasks in parallel.

To summarize the literature, a dynamic approach to task allocation is deemed advantageous for a number of reasons (Greenstein & Rouse, 1982; Rouse, 1977, 1981). First, the human retains skill proficiency and knowledge of the overall system operation; this retention is less likely with the static assignment of a subset of tasks. Second, dynamic task allocation may result in a more fault-tolerant system in the event of computer failure. Such an event cannot be treated so flexibly if certain tasks are allocated exclusively to either the human or the computer. Third, by implementing a computer aid that is adaptive, or sensitive to the human's task load, a less variable and more optimal level of human workload may be achieved.

In situations in which a computer aid is introduced, there must be a means for both the human and the computer to transmit their respective plans of action. Greenstein and Revesman (1981) examined modes of human-computer communication in task allocation: The

first, implicit or model-based communication, is implemented by supplying the computer with a predictive model of the human's performance for the task environment. With this predictive model, the computer can predict the human's course of action and can act to complement the human by avoiding conflicting or redundant actions. Thus, computer assistance is activated without an explicit request or the execution of any commands. The second mode of communication, explicit or dialogue-based communication, completely avoids conflict with the human's actions because the computer is expressly informed of the human's plan of action.

The results indicate that the costs associated with model-based and dialogue-based communication should be traded off against each other when selecting a communication mode for particular applications (Greenstein & Revesman, 1981). For example, there is a significant increase in human workload when dialogue-based communication is employed for task allocation, but this mode results in more accurate communication and less redundant actions when compared to model-based communication.

The purpose of the study reported here was to investigate dialogue-based human-computer task allocation. In general, a task-sharing approach specified by dynamic task allocation has been advocated as a means of partitioning work between the human and the computer. Given that there are cases in which a dialogue-based communication mode is desirable for situations involving dynamic task allocation, it is necessary to design human-computer dialogues to convey information regarding the status of tasks. Since there is a significant time cost associated with the increased human workload incurred by dialogue-based communication, a research goal might then be to develop guidelines for the design of efficient task allocation dialogues. The goals of this research are to develop operative modes of task allocation appropriate for high-task-load situations and to provide implications for the design of effective human-computer task allocation dialogues. This preliminary study serves as a basis for choosing variables to be studied in a sequential program of research.

METHOD

Subjects

Five male students at Virginia Polytechnic Institute and State University received \$3.50/h for their participation. The subjects were screened for corrected 20/20 vision using a Bausch & Lomb Ortho-Rater and were given a simple test of typing proficiency. Extensive computer experience was not necessary, but pilots and air traffic controllers were ineligible to participate.

Experimental Task

The software system GENIE (a generic environment for interactive experiments) was designed to provide a task environment for behavioral research of human-computer interaction (Lindquist, Fainter, Guy, Hakkinen, & Maynard, Note 1). GENIE software was used to examine the effects of the design of the task allocation dialogue on human-computer multiple-

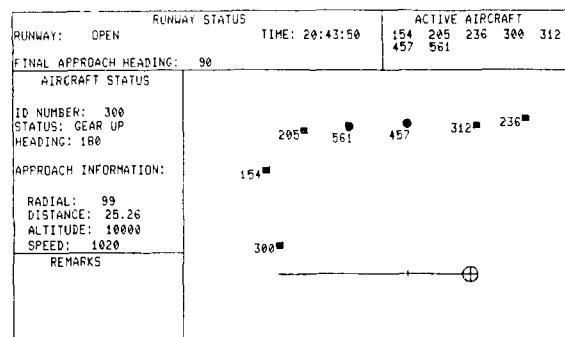


Figure 1. Primary display of the air traffic control task environment.

task performance. The task environment consisted of a simplified representation of a computer-aided air traffic control system, in which the subject acted as air traffic controller. The subject shared the task with the computer by delegating responsibility for landing some of the aircraft to the computer aid.

The primary display monitor showed runway and aircraft status information, identification numbers of the active aircraft, and a radar-like graphic display of the aircraft and runway, as shown in Figure 1. A secondary display monitor was reserved for the command input area, error messages, and other feedback. The subject was required to detect the entrance of an aircraft in the upper right corner of the control area and was then required to direct the planes to the airport runway in a specific landing pattern by issuing a series of three to four "TURN" commands. If a plane's identification number was displayed in red, the subject detected an emergency situation and directed the other aircraft in an alternate landing pattern by issuing two to three "TURN" commands. In the treatment conditions, the subject issued "TAKE" commands to delegate responsibility for landing certain planes to the computer controller. After a plane was delegated to the computer controller, it was displayed as a round blip instead of a square one.

Independent Variables

Originally, two independent variables, allocation mode, and task load, were to be investigated. During pretesting, however, it was found that subjects did not invoke the aid provided by the computer controller unless the task-load level was so high that it was extremely difficult to perform the multiple tasks unaided. Therefore, only one task-load level was investigated. In the high-task-load level, aircraft entered the control area at a rate of one every 55 sec, with a total of 21 aircraft in 20 min, of which 4 were aircraft in emergency situations.

"Allocation mode" refers to the manner in which tasks were allocated to the computer. Through implementation of different TAKE commands, the following treatment conditions and a control condition were examined: (1) a control condition in which the subject performed all tasks unaided to establish a baseline measure of performance, (2) ID mode, in which planes were allocated by their respective identification numbers [e.g., "TAKE (aircraft) 200"] and one to four aircraft could be allocated with each command, (3) NEXT mode, in which the next specified number of planes to enter the control area were allocated [e.g., "TAKE NEXT 3 (aircraft)"] and one to four aircraft could be allocated with each command, (4) SUBTASK mode, in which all planes currently in an emergency situation or planes inadvertently directed off the display were allocated [e.g., "TAKE EMERG (aircraft)" or "TAKE DISPLAY" (display error aircraft)] and any plane that currently met the given condition could be allocated, and (5) QUAD mode, in which all planes currently in a given display quadrant were allocated [e.g., "TAKE (display) QUAD 2"] and one to two quadrants could be allocated with each command.

Experimental Design

The four levels of allocation mode and the control condition yielded five treatment conditions. The subjects received all five treatments in a within-subjects Latin square design, which was chosen to partition intersubject variability and control for order effects.

Dependent Variables

A number of measures of system performance were recorded using metering software embedded within GENIE. Table 1 describes the performance measures collected. A 7-point subjective rating scale questionnaire was administered to evaluate perceived workload on dimensions of time-constraint loading, mental effort loading, and psychological stress loading. These three dimensions were adapted from Sheridan and Simpson (Note 2). Ratings for the TAKE commands on dimensions of the complexity of planning strategy required, ease of use, effectiveness, and overall ratings of satisfaction were also collected. A sample item is shown in Table 2.

Apparatus

The hardware configuration consisted of a DEC GIGI terminal (Model VK-100) and two Barco Model GD33 color monitors interfaced to a DEC VAX-11/780 host computer. The primary display was the monitor located on the right side, with a secondary display on the left. A circular protractor was affixed to the GIGI terminal to aid subjects in the control task, and a training manual was available for reference.

Procedures

Two 3-h sessions were conducted over 2 days. Paced by a tape recording, the subject read a brief training manual on Day 1. This manual explained information about the task, the computer terminal and displays, and the command syntax. Excluding the four experimental allocation (TAKE) commands, only three other commands were needed to guide the aircraft. The subjects then performed a 15-min practice session for each of the five conditions.

Table 1
Descriptions of Performance Measures Collected

Measure	Description
System-Land	Percentage of all planes entering control area that were landed by the system.
H-Land	Percentage of all planes entering control area that were landed by the human.
C-Land	Percentage of all planes entering control area that were landed by the computer.
H-Plane	Percentage of all planes entering control area that the human controlled.
Errors	Total number of human errors, including syntax errors and action errors.
N-Command	Total number of correct commands.
P-Command	Percentage of correct commands.
Take	Total number of TAKE commands.

Table 2
Sample Item on Rating Scale

1	2	3	4	5	6	7	No Spare Time
Much Spare Time		Some Spare Time					

Note—Time constraint loading refers to the number of interruptions that occur during the planning, execution, or monitoring of tasks; it also refers to the amount of spare time experienced.

After demonstrating the ability to land two aircraft in a training trial, the subject returned on Day 2 to participate in five counterbalanced 20-min trials. After each trial, the subject completed a rating scale questionnaire, and after three trials, he was given a 10-min break. At the end of the second session, subjects completed the questionnaire by making overall ratings.

RESULTS

Four conditions of task allocation mode and a control condition were considered.

System-Land refers to the percentage of all planes entering the control area that were landed by the human-computer system. The mean of System-Land was 32.76%, with a standard deviation of 10.50. A one-way analysis of variance revealed significant differences among the five groups [$F(4,16) = 3.07$, $p < .05$]. A post hoc l.s.d. test indicated that the ID, NEXT, and SUBTASK conditions resulted in greater mean performance than the control condition. The group means and standard deviations for System-Land are shown in Table 3.

A one-way analysis of variance suggested that H-Land, the percentage of all planes entering the control area that were landed by the human, did not differ significantly among the five groups [$F(4,16) = .72$, $p > .05$]. Similarly, C-Land, the percentage of all planes entering the control area that were landed by the computer, did not differ significantly among the four treatment groups [$F(3,12) = .72$, $p > .05$].

H-Plane refers to the percentage of planes the human chose to control in the treatment conditions. A one-way analysis of variance revealed no significant differences in H-Plane among the treatment groups [$F(3,12) = 1.05$, $p > .05$].

Human errors in the control task included syntax errors in the commands, inadvertently directing planes out of the control area, and missed approaches, which refers to guiding a plane to the final approach, but missing the airport runway. A one-way analysis of variance revealed no significant differences among the five groups in the number of errors committed [$F(4,16) = 1.29$, $p > .05$].

With regard to the total number of correct commands, N-Command, a one-way analysis of variance revealed significant differences among the five groups [$F(4,16) = 4.04$, $p < .05$]. The mean of N-Command was 60.04 commands with a standard deviation of 9.54. A post hoc l.s.d. test indicated that significantly more commands were executed in the control condition than in the ID, NEXT, and QUAD conditions. Also, there were significantly more commands executed in the SUBTASK condition than in the QUAD condition. The group means and standard deviations for N-Command are shown in Table 3. Regarding P-Command, the percentage of correct commands, there were no significant differences among the five groups [$F(4,16) = .59$, $p > .05$].

Regarding Take, the number of TAKE commands executed, there were significant differences among the four treatment groups [$F(3,12) = 5.57$, $p < .05$].

Table 3
Descriptive Statistics by Mode for Reliable Performance Measures

Mode	Performance Measure					
	System-Land		N-Command		Take	
	Mean	SD	Mean	SD	Mean	SD
Control	22.86	9.76	68.20	5.26		
ID	37.14	14.04	57.80	6.98	7.40	2.70
NEXT	35.24	5.43	55.60	4.88	6.40	2.30
SUBTASK	39.05	6.21	64.60	5.94	5.20	1.10
QUAD	29.52	9.16	54.00	15.00	2.80	.84

Note—Subjects in the control condition did not use the TAKE command.

Table 4
Correlations of Performance Measures With Subjective Measures

Correlated Measures	r
H-Plane	
Time Constraint	.64†
Mental Effort	.59**
Psychological Stress	.66†
Preference of Conditions	-.60**
Ease in Using Take Command††	.49*
N-Command	
Time Constraint	.51**
Mental Effort	.57**
Psychological Stress	.58**
Preference of Conditions	-.41*
C-Land	
Preference of Conditions	.50*
Take	
Preference of Conditions	.59*

Note—Unless otherwise noted, df = 24 for all entries.

*p < .05. **p < .01. †p < .001. ††df = 19.

A post hoc *t*-s.d. test indicated that a significantly lower number of TAKE commands were used in the QUAD condition than in either the NEXT or the ID condition. The group means and standard deviations of Take are shown in Table 3.

Pearson product-moment correlations were used to correlate the performance measures with the subjective measures. As shown in Table 4, a number of correlations were significant.

DISCUSSION

The results of the F tests regarding System-Land can be considered graphically, as shown in Figure 2, by comparing the mean percentages of planes landed by the human and the computer controllers. This representation demonstrates that the introduction of the computer controller does not improve the performance of the human controller per se but contributes additively to the overall system performance. It is notable that the QUAD allocation mode did not contribute to significantly greater system performance than the control

condition did. This phenomenon may have occurred because use of the QUAD mode allocated groups of planes with one command, leaving the human heavily loaded with other groups of planes. The other allocation modes may have allowed the human to use a different strategy of alternating the responsibility of controlling planes with the computer controller.

The results of the F tests of N-Command and Take may be considered jointly. Although invoking aid from the computer controller required typing TAKE commands, the human was still able to type fewer commands overall, as shown in Figure 3. Although the treatment conditions lessened the human's burden of typing commands to varying degrees, there were no differences in System-Land among these four conditions. Even though the QUAD mode allowed the human to type the fewest commands, it did not result in improved system performance. Again, this finding may have occurred because the nature of the QUAD mode may have caused the human to be intermittently burdened with typing commands while somewhat idle at other times. It is reasonable, however, to conjecture that the sustained burden of typing commands might be considered a factor contributing to relatively depressed System-Land performance in the control condition.

Some of the significant correlations appear to corroborate these conjectures. The ratings of the time-constraint, mental effort, and psychological stress dimensions of perceived workload were all positively correlated with H-Plane and N-Command. Thus it appears that the ratings of perceived workload were related to the number of planes for which the human was responsible and to the degree of command input the task demands required. H-Plane was also correlated with the ease-of-use ratings for the TAKE commands; it appears that the syntax difficulty associated with

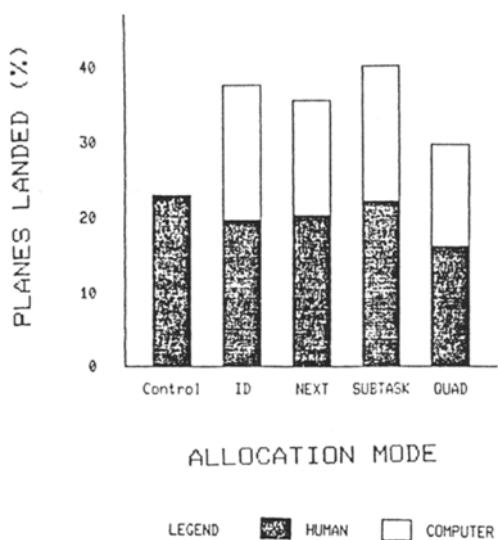


Figure 2. Mean percentage of planes landed by controller.

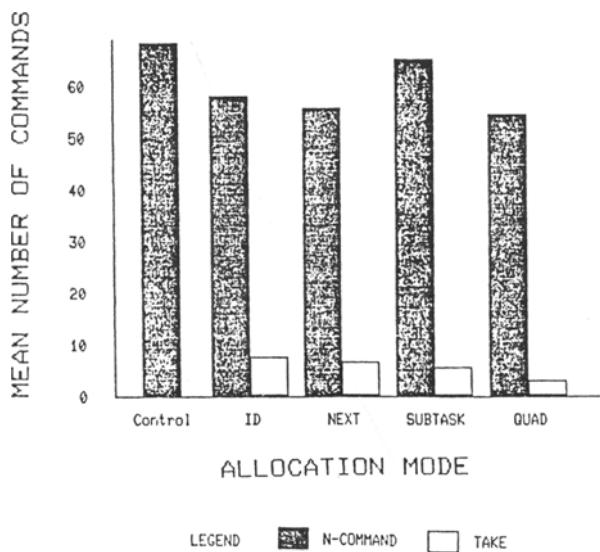


Figure 3. Means of N-Command and Take by allocation mode.

the various task allocation modes was related to the number of planes the human controlled.

The overall preference ratings for the five conditions were related to a number of performance measures. It appears that greater preference ratings were associated with lower values of H-Plane, the percentage of planes the human controlled. Also, greater preference ratings were associated with lower values of N-Command, the total number of commands executed. Greater preference ratings were also associated with greater values of Take, the number of TAKE commands executed (execution of a TAKE command allowed the human to reduce the task demands).

The results provide implications for the design of effective task allocation dialogues, which implies minimizing both the likelihood of errors and the time spent engaging in dialogue. Because perceived workload ratings were correlated with N-Command, the human-computer dialogue in general may not have been an optimal one, since it caused the human to be taxed with typing command input. The task allocation dialogue, in particular, did not seem to adversely affect the level of error, since there were no differences among the five groups in the percentage of correct commands or the total number of errors. Performance differences among the four task allocation modes may be attributed to strategy differences rather than the dialogue design.

This preliminary research may serve as a basis for investigating other variables. First, a comparison of input modes, such as keyboard, key-pad function keys, touch panel, and voice input, might be conducted. To investigate input mode, embedded metering software must accurately record the time engaged in dialogue. Second, dialogue styles, including computer-initiated, human-

initiated, and mixed-initiative dialogue styles, might be compared. Examples of computer-initiated dialogues include form filling and menu selection; examples of human-initiated dialogues include command languages and formal query languages. Under a mixed-initiative dialogue style, the computer is programmed with a schema of the task. The user leads the interaction, but the computer takes initiative when the user overlooks some aspect of the task or requests assistance. A comparison of these styles may indicate which is appropriate for situations involving stress and high workload. Finally, the performance of the computer aid might be degraded to various levels to obtain more generalizable results. By conducting a carefully planned sequential program of research, one may gain knowledge about both human-computer task allocation and dialogue design.

REFERENCE NOTES

1. Lindquist, T. E., Fainter, R. G., Guy, S. R., Hakkinen, M. T., & Maynard, J. F. *GENIE: A dialogue-rich task environment for empirical studies of human-computer interactions* (Draft Technical Report). Blacksburg: Virginia Polytechnic Institute and State University, Department of Computer Science, August 1982.

2. Sheridan, T. B., & Simpson, R. W. *Toward the definition and measurement of the mental workload of transport pilots* (FTL Report No. R79-4). Cambridge, Mass: M.I.T. Flight Transportation Laboratory, 1979.

REFERENCES

- CHAPANIS, A. On the allocation of functions between men and machines. *Occupational Psychology*, 1965, **39**, 1-11.
- CHU, Y. Y., & ROUSE, W. B. Adaptive allocation of decision-making responsibility between human and computer in multi-task situations. *IEEE Transactions on Systems, Man, and Cybernetics*, 1979, **SMC-9**, 769-778.
- DANAHER, J. W. Human error in ATC system operations. *Human Factors*, 1980, **22**, 535-545.
- GREENSTEIN, J. S., & REVESMAN, M. E. A Monte-Carlo simulation investigating means of human-computer communication for dynamic task allocation. In *Proceedings of the IEEE 1981 International Conference on Cybernetics and Society*. Atlanta: Institute of Electrical and Electronics Engineers, 1981.
- GREENSTEIN, J. S., & ROUSE, W. B. A model of human decisionmaking in multiple process monitoring situations. *IEEE Transactions on Systems, Man, and Cybernetics*, 1982, **SMC-12**, 182-193.
- RIEGER, C. A., & GREENSTEIN, J. S. The allocation of tasks between the human and computer in automated systems. In *Proceedings of the IEEE 1982 International Conference on Cybernetics and Society*. Seattle: Institute of Electrical and Electronics Engineers, 1982.
- ROUSE, W. B. Human-computer interaction in multi-task situations. *IEEE Transactions on Systems, Man, and Cybernetics*, 1977, **7**, 384-392.
- ROUSE, W. B. Human-computer interaction in the control of dynamic systems. *ACM Computing Surveys*, 1981, **13**, 71-99.
- TOMESKI, E. A., & LAZARUS, H. *People-oriented computer systems*. New York: Van Nostrand Reinhold, 1975.
- WIENER, E. L., & CURRY, R. E. Flight-deck automation: Promises and problems. *Ergonomics*, 1980, **23**, 995-1011.