Characteristics of Fluent Skills in a Complex, Dynamic Problem-Solving Task

Myeong-Ho Sohn, Scott A. Douglass, Mon-Chu Chen and John R. Anderson

Human Factors: The Journal of the Human Factors and Ergonomics Society 2005 47: 742
DOI: 10.1518/001872005775570943

Characteristics of Fluent Skills in a Complex, Dynamic Problem-Solving Task

Myeong-Ho Sohn, George Washington University, Washington, D.C., and **Scott A. Douglass, Mon-Chu Chen,** and **John R. Anderson,** Carnegie Mellon University, Pittsburgh, Pennsylvania

We examined critical characteristics of fluent cognitive skills, using the Georgia Tech Aegis Simulation Program, a tactical decision-making computer game that simulates tasks of an anti-air-warfare coordinator. To characterize learning, we adopted the unit-task analysis framework, in which a task is decomposed into several unit tasks that are further decomposed into functional-level subtasks. Our results showed that learning at a global level could be decomposed into learning smaller component tasks. Further, most learning was associated with a reduction in cognitive processes, in which people make inferences from the currently available information. Eye-movement data also revealed that the time spent on task-irrelevant regions of the display decreased more than did the time spent on task-relevant regions. In sum, although fluency in dynamic, complex problem solving was achieved by attaining efficiency in perceptual, motor, and cognitive processes, the magnitude of the gains depended on the preexisting fluency of the component skills. These results imply that a training program should decompose a task into its component skills and emphasize those components with which trainees have relatively little prior experience. Actual or potential applications of this research include learning and training of complex tasks as well as evaluation of performance on those tasks.

INTRODUCTION

In the field of cognitive psychology, a number of theories have emerged that describe skill acquisition in laboratory tasks quite successfully. However, it is not clear how these theories would generalize to the solution of complex real-world tasks. In this paper, we extend unit-task analysis (Card, Moran, & Newell, 1983) of skill acquisition to a synthetic task that has some of the complexity of real-world tasks. We picked the Georgia Tech Aegis Simulation Program (GT-ASP; Hodge et al., 1995), which is a tactical decision-making computer game that simulates tasks of anti-air-warfare coordinators (AAWCs) on board U.S. Navy cruisers and destroyers. A participant assumes the role of an AAWC, who monitors a radar screen for unknown aircraft, requests and collects information, and updates the identity of the aircraft. If an unknown aircraft turns out to be hostile, the AAWC keeps monitoring its trajectory and issues a military action in accordance with the rules of engagement. Although vastly simplified from the actual AAWC workstation, GT-ASP is a reasonably realistic cognitive task with medium fidelity to the system that is currently used in the U.S. Navy. Using this task, Hodge (1998) collected a large data set (the Georgia Tech data set). This data set, which tracked fairly long-term practice of the task, allowed us to examine the characteristics of skill learning in an ideal situation. In this paper, we will describe an analysis of the Georgia Tech data set and a new eye-tracking experiment specifically designed to investigate the nature of learning in a complex, dynamic task.

The aim of this study is to identify characteristics of fluent cognitive skill in a task such as GT-ASP and to examine whether these characteristics can be understood within a single

Address correspondence to Myeong-Ho Sohn, 2125 G St. NW, 313A Building GG, Department of Psychology, George Washington University, Washington, DC 20054; mhsohn@gwu.edu. **HUMAN FACTORS,** Vol. 47, No. 4, Winter 2005, pp. 742–752.

framework. We are especially interested in showing that fluency in a complex skill exhibits three characteristics that have been identified in the skill acquisition literature. First, learning a complex skill can be reduced to learning its component skills – the *decomposition hypothesis.* This hypothesis is consistent with one of the primary assumptions about part-task training in skill acquisition (Fisk, Ackerman, & Schneider, 1987): that fluency with component skills can facilitate learning of the whole task. Hodge (1998) demonstrated that people who learned the GT-ASP task by practicing only some components of the whole task in a piecemeal fashion could perform the entire task as fluently as those who learned the task as a whole. This result implies that part-task training can be more beneficial than whole-task training because part-task training requires less time and effort. In a different task domain, Anderson, Conrad, and Corbett (1989) decomposed learning of LISP programming skills into about 500 production rules and showed that the execution time for each production sped up with learning. Moreover, this speedup was identified all the way down to the eye-movement level. Further, in the context of whole-task learning, Lee and Anderson (2001) demonstrated that learning the Kanfer-Ackerman Air Traffic Controller (Ackerman, 1988) task can be explained by the learning of its component tasks.

Second, different component skills may have different degrees of prior learning and therefore show different amounts of learning with practice – the *practice-relativity hypothesis.* Many complex tasks have components with predominantly cognitive demands as well as components with predominantly motor demands. When people are learning a novel task, it is often the case that the cognitive components, which reflect unique aspects of the task, are less practiced than the motor components, which typically involve well-practiced skills such as key pressing. According to MacKay (1982), learning takes place at multiple layers of a task hierarchy, and the learning of each layer is encapsulated in the sense that the efficiency or the amount of learning of each layer is independent from the amount of learning at other layers. If prior learning at a certain layer is already quite sufficient, then additional practice does not add much. Thus the practice-relativity hypothesis suggests that how much can be learned from each component may differ depending on task demands.

Third, part of the practice effect reflects a reduction in the amount of task-irrelevant information gathered by the learner – the *attentional focus hypothesis.* The ability to streamline the information flow in a dynamic display is critical to staying focused on the task. Reducing the time spent on task-irrelevant information results in increased cognitive processing efficiency. The attentional focus hypothesis has been explored by tracking eye movements during problem solving with a display that provides both task-relevant and task-irrelevant information (Haider & Frensch, 1999; Lee & Anderson, 2001). Lee and Anderson (2001) showed that people were sensitive to which regions provided the task-relevant information and, moreover, that eye fixation times on the task-irrelevant regions were reduced dramatically with practice.

We report our data analysis in light of the three hypotheses just discussed and relate them to the power-law learning curve (Newell & Rosenbloom, 1981), which represents one of several efforts to provide a quantified description of the speedup in task performance with practice. A power function relates the time (T) to perform a task to the amount of practice by the function $T = A + BN^{-c}$, in which A is an asymptote, B is a scale factor reflecting the amount of time by which performance is reduced over N number of practice units, and c is the exponent that specifies the rate of the speedup with practice. The task performance time decreases from an initial time of $A + B$ to A in the limit. Each hypothesis mentioned earlier makes a unique claim regarding the three parameters of the best fitting power functions for the unit tasks that make up the entire task as well as components of each unit task. The decomposition hypothesis suggests that one single exponent, which reflects the learning rate, should be able to fit performance at different levels of task decomposition. We will test this hypothesis by comparing a single-exponent model with alternative models. The practice-relativity hypothesis suggests that although the scale factors (how much can be learned) follow the same learning rate, they should be greater for cognitively loaded subtasks than for subtasks with heavy motor demands. The attentional focus hypothesis suggests that the differential amount of learning should also be identified at the level of eye movements. That is, scale factors associated with the time spent on task-irrelevant regions should be greater than those associated with the time spent on task-relevant regions.

We will describe the GT-ASP task in detail, focusing on the critical aspects of interest. Then we will provide our analyses of the Georgia Tech data in light of the decomposition and the practice-relativity hypotheses. Then we will report an eye-tracking experiment with GT-ASP conducted at Carnegie Mellon University (CMU).

TASK ANALYSIS OF GT-ASP

The GT-ASP task has been widely used by various research groups, and different task analyses are available (Cannon-Bowers & Salas, 1998). Many of these analyses focus on the effectiveness of part-task training and types of feedback (Kirlik, Fisk, Walker, & Rothrock, 1998), effective design of a learning environment (Johnston, Poirier, & Smith-Jentsch, 1998), various aspects for task analysis and modeling (Zachary, Ryder, & Hicinbothom, 1998), and so forth. Because we are interested in hierarchical organization of the GT-ASP task, we adopted the unit-task analysis of Card et al. (1983), which suggests that a task can be decomposed into increasingly specific subtasks, all the way down to the keystroke level of elementary cognitive and perceptual-motor subtasks. Although it is not always the case with all cognitive tasks, many complex, dynamic, real-world tasks can be readily decomposed this way. This task analysis includes three levels: a unit-task level, a functional level, and a keystroke level. At the unit-task level, the main task is decomposed into a set of independent unit tasks that are repeatedly executed. At the functional level, a unit task is further decomposed into smaller, functional-level subtasks. The keystroke level is the most detailed level of analysis and consists of elementary motor and perceptual subtasks (e.g., pressing keys, finding and encoding information from the environment, retrieving information from long-term memory).

In the full version of the GT-ASP task (see Figure 1 for the task screen), a participant is responsible for monitoring the air traffic that appears on the radar screen by carrying out five unit tasks. First, the participant is supposed to identify as many unknown aircraft as possible: the *identification* task. Identification specifies primary intent (hostile, friendly, etc.) and air type (helicopter, strike, commercial airliner, etc.). Second, the participant gathers information about the unknown aircraft: the *information search* task. Some information is relayed through other simulated officers and combat air patrols (CAPs). Third, when the participant needs information from a CAP, the participant can send a CAP close to an unknown aircraft by changing altitude, speed, and/or course: the *CAP control* task. Fourth, the participant issues military action commands (e.g., warnings, assignment or engagement of a missile) against a hostile aircraft: the *military action* task. Fifth, the participant is allowed to control the display, such as changing its radius and center and to add supplementary range rings: the *display control* task. Of these five unit tasks, identification and CAP control are further analyzed at the functional level because these two tasks have well-defined functional-level subtasks.

Identification

The identification task starts with a mouse hook of the target aircraft and is executed in the order of *search, initiate, classify,* and *save* subtasks. Figure 2a shows the correct key presses for each subtask. A participant first executes the search subtask to gather information about the aircraft of current interest (i.e., the currently hooked aircraft) by requesting information from other virtual officers, reading profiles from the character readout box, or picking up requested information from the message box. Because many types of information in the GT-ASP task are probabilistic rather than deterministic and partial rather than complete, the participant has to collect various kinds of information and make inferences to determine the intent of the aircraft. The initiate and classify subtasks demand fairly mechanistic key presses. If a participant is confident with the information at hand, he or she executes a sequence of key presses, including several function keys, as shown in Figure 2. The save subtask requires only one keystroke to finalize and confirm the updating. Although this seems quite simple, participants often forget to perform this last keystroke.

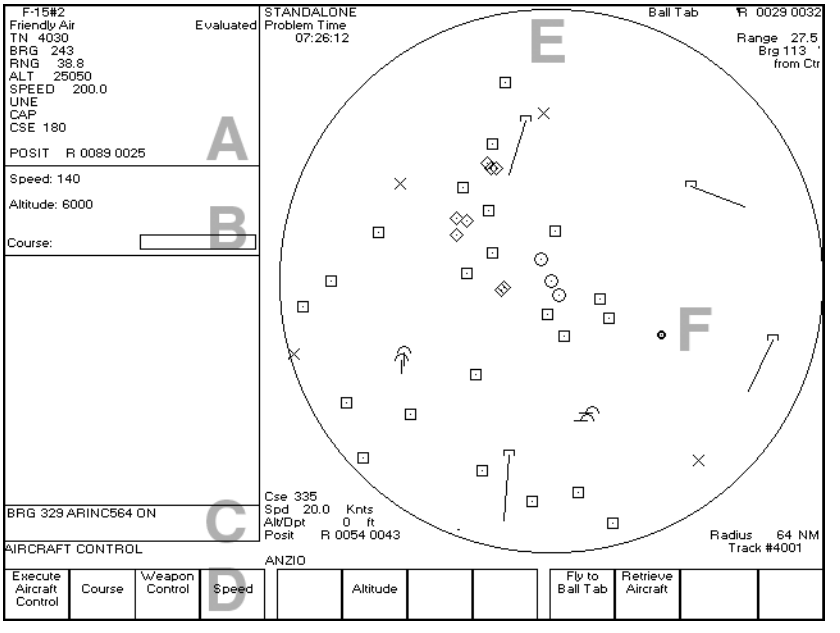

Figure 1. An example of the GT-ASP display. The character readout box (Region A) provides available information of the currently hooked aircraft (e.g., speed, altitude, bearing, course). The character type-in box (Region B) is used when the participant wants to change speed, altitude, and course of a friendly fighter jet, combat air patrol (CAP). The track-reporting area (Region C) displays information requested by the participant, which is also presented in an auditory channel. The menu panel (Region D) shows the currently available function keys (F1–F12 on the computer keyboard) and their functions. The menu panel is divided into three banks of four function keys just as the function keys are arranged on a computer keyboard. Because the GT-ASP menu has a hierarchical structure, the same key can map to several different functions depending on the current task. The scope (Region E) shows all the air tracks and surface tracks. Open rectangles represent unknown aircraft. The half circles represent friendly CAPs. The vector emanating from these two types of aircraft indicates the flight course (the direction of the vector) and speed (the length of the vector). The other symbols represent surface tracks. Within the radarscope, the ball tab (Region F) indicates where the current mouse is located.

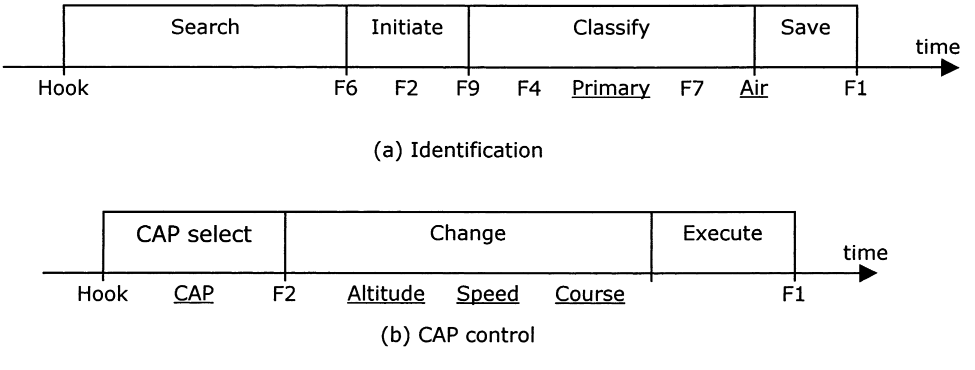

Figure 2. Decomposition of unit tasks into functional subtasks for (a) the identification task and (b) the CAP control task. In each panel, the boxes above the time line represent functional subtasks constituting the unit task. Below the time line are the relevant key presses. For the underlined key press activities, the locations and exact number of key presses may vary.

The motivation for this breakdown of the identification unit task is that these subtasks not only represent distinctive segmentation but are also differently loaded with cognitive and motor demands. During the search subtask, a participant has to make inferences about the kind of information to request as well as the relationship between the currently available information and the identity of the aircraft. Also, during the save subtask, a participant has to explicitly remember to perform the last keystroke. These subtasks are more cognitively loaded. In contrast, during the initiate and classify subtasks, the participant simply performs a sequence of key presses. Therefore, most of the learning with these two subtasks is related to gaining familiarity with the key locations and promptness of key presses. We can test the decomposition hypothesis by examining whether the time to perform each subtask decreases with practice following the same learning rate. We can also test the practice-relativity hypothesis by examining whether the amount of learning differs between the cognitive and motor subtasks.

CAP Control

The most dependable source of information in GT-ASP is visual inspection by a CAP pilot. For this purpose, a participant has to explicitly designate which CAP should do this job and to specify its speed, altitude, and course to send the CAP to fly close to the target. The CAP control task has three subtasks, as shown in Figure 2b. During the *CAP selection* subtask, the participant decides which CAP should do visual inspection. This subtask represents processing demands, such as searching for CAPs in the visually cluttered display and selecting one that is relatively close to the target aircraft. During the *change* subtask, the participant specifies the speed, altitude, and course of the CAP. The *execute* subtask, which requires only one keystroke to complete CAP control, is often neglected, similar to the save subtask of the identification task.

These three subtasks represent different functionalities. During the CAP selection subtask, a participant has to make a decision by considering the current locations of the CAPs and the target airplane. During the change subtask, the participant determines the appropriate set of control parameters and enters these parameters. The execute subtask can be demanding because the structure of GT-ASP makes it easy to neglect actually executing the changes made to the CAP. We treat the CAP selection and the execute subtasks as cognitively loaded and the change subtask as having mainly motor processing demands, similar to the identification unit task.

ANALYSIS OF GEORGIA TECH DATA

In the whole-practice condition of the Georgia Tech experimentation, 48 people practiced 18 30-min scenarios over 10 days. First there was a demonstration by an experimenter. Then participants performed 2 short scenarios with over-the-shoulder coaching (e.g., key press sequences) before they started to perform main scenarios. The first 15 scenarios were followed by immediate feedback on the accuracy of identification and missing opportunities. The last 3 scenarios were performed without feedback. All the scenarios were different from one another (e.g., different numbers of unidentified aircraft). Consequently, task difficulty also varied from scenario to scenario. To reduce between-scenario variability, we grouped 3 consecutive scenarios into a practice period and analyzed the data accordingly.

Unit-Task Analysis

From the five unit tasks, we excluded military action from analysis because the frequency of this task was relatively low and varied a lot depending on the specific scenario. The performance time sped up with practice by factors ranging from 1.5 to 2.5, indicating that improvement at the unit-task level is a major source of improvement in the GT-ASP task.

Table 1 presents the results of power-law fitting to the unit tasks. The power-function fitting was done by estimating nine free parameters (scale factors and asymptotes for four unit tasks and one common exponent for all of them). The goodness of fit was determined by χ^2 deviation, which is the ratio of the actual deviation (sum of the squared deviations of the predicted means from the actual means) to the estimated error in the means (squared standard errors obtained from the participant-by-practice interaction for each curve). The asymptotes, the scale factors, and the exponent were estimated such that the

TABLE 1: Exponent-Constrained Power-Law Fitting of Georgia Tech Data Set

	Power Functions	R^2	*df*	χ^2
	Unit-Task Level			
Identification	$T = 6.53 + 20.23\ N^{-0.56}$	.98	15	22.33
CAP control	$T = 3.40 + 11.19\ N^{-0.56}$	.99		
Information search	$T = 1.60 + 2.15\ N^{-0.56}$	.97		
Display control	$T = 0.46 + 4.82\ N^{-0.56}$	.99		
	Functional Level: Identification			
Search	$T = 1.15 + 2.48\ N^{-0.56}$	.99	16	16.67
Initiate	$T = 0.45 + 1.13\ N^{-0.56}$			
Update	$T = 0.33 + 1.43\ N^{-0.56}$			
Save	$T = 0.29 + 0.92\ N^{-0.56}$			
	Functional-Task Level: CAP Control			
CAP selection	$T = 0.13 + 1.88\ N^{-0.56}$	.98	12	11.48
Change	$T = 0.84 + 1.45\ N^{-0.56}$			
Execute	$T = 0.21 + 2.05\ N^{-0.56}$			

total χ^2 deviation was minimized. The fit of this constrained model, $\chi^2(15) = 22.33$, $R^2 = .997$, is quite good relative to the 15 degrees of freedom, which is the total number of observations (24) minus the number of estimated parameters (9).

Functional-Level Subtasks

We chose to analyze the mean time per keystroke instead of the total interval time because each interval requires a different number of keystrokes, which makes it difficult to attribute the differences in the asymptotes and scale factors to the task demands of intervals. The results of fitting the exponent-constrained power-law models are given in Table 1.

At the functional level, we examined the decomposition and practice-relativity hypotheses. The decomposition hypothesis suggests that learning of functional subtasks also followed the learning rate of the unit task. To test this, we used the same exponent (0.56) from the unit-task power-law fittings and estimated the asymptotes and scale factors for functional subtasks. When considering both the identification and the CAP control tasks, the model did not significantly deviate from the data, $\chi^2(28) = 28.15$, $R^2 = .99$, suggesting a common learning rate for different levels of learning and therefore supporting the decomposition hypothesis.

The practice-relativity hypothesis claims that there should be more learning for cognitive subtasks than for motor subtasks because prior learning may be quite high for motor subtasks but relatively low for cognitive subtasks. Note that the power-function fitting in Table 1 assumes a single exponent but different scale factors and asymptotes. We decided to perform statistical tests of the differences in scale factors, which reflect the amounts to be learned. To do this, we fit power functions to each participant's mean times per keystroke for each subtask of the identification and the CAP control unit tasks and estimated the asymptote and the scale factor for the participant. In doing this, we used the 0.56 exponent. The individual scale factors were analyzed in terms of the processing load associated with a functional subtask (cognitive and motor) and the type of unit task (identification and CAP control). For the identification unit task, the search and the save subtasks are cognitively demanding, whereas the initiate and the update subtasks have mainly motor demands. For the CAP control task, the CAP selection and the execute subtasks are cognitively demanding, whereas the change subtask has mainly motor demands.

The mean scale factors are presented in Figure 3a. The scale factor was greater for the CAP control unit task (2.42 s) than for the identification unit task (1.09 s), $F(1, 47) = 94.11$, $MSE = .91$, $p < .0001$. Also, the scale factor was greater for cognitively loaded subtasks (2.34 s) than for

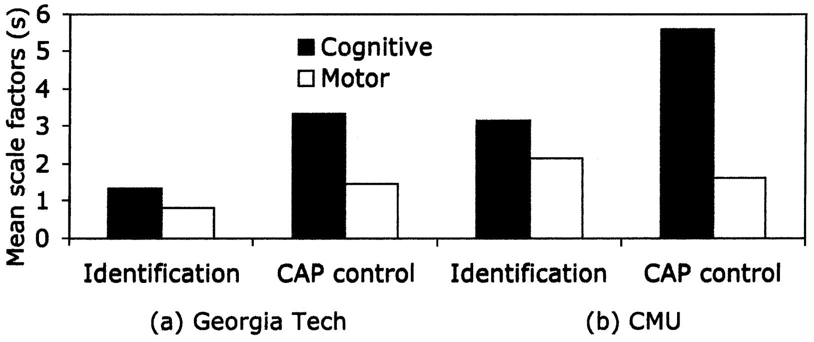

Figure 3. Mean scale factors for the cognitive subtasks and the motor subtasks of the identification and the CAP control tasks; (a) Georgia Tech data and (b) CMU data.

the motor subtasks (1.16 s), $F(1, 47) = 81.86$, $MSE = .84$, $p < .0001$. Also, the interaction was significant, $F(1, 47) = 30.71$, $MSE = .72$, $p < .0001$. For both unit tasks, the scale factor was greater for the cognitive than for the motor subtasks, but this trend was greater for CAP control, $t(47) = 9.65$, $p < .0001$, than for identification, $t(47) = 3.12$, $p < .01$. The analysis of the scale factors clearly shows that there is more to learn in cognitive skills than in motor skills, supporting the practice-relativity hypothesis.

Summary

Our analysis of the Georgia Tech data shows that speed of performance improves as a power function of practice, which concurs with many other studies. As Lee and Anderson (2001) did, we showed that the overall learning curve could be decomposed into learning curves for components. Moreover, these subcurves share the same power exponent. We have also gone beyond the decomposition hypothesis by further elaborating the practice-relativity hypothesis. In the following section, on the CMU experiment, we will report the test of the attentional focus hypothesis along with tests of the decomposition and the practice-relativity hypotheses.

CMU EXPERIMENT

The CMU experiment had two purposes: (a) to replicate our tests of the decomposition and the practice-relativity hypotheses and (b) to test the attentional focus hypothesis by examining eye movements. One important characteristic of skill learning is that people learn to distinguish task-relevant information from task-irrelevant information (Haider & Frensch, 1999). Lee and Anderson (2001) showed that when participants were learning a complex problem-solving task, much of the inefficiency could be attributed to suboptimal scanning of the screen. They looked at participants' eye movements and examined fixation time on several regions of the screen as a function of whether the information available in the region was relevant to performing the task. They found that as participants' performance became more fluent, the proportion of fixations on the irrelevant regions decreased.

Similarly, in the GT-ASP task, the participant does not have to pay attention to every part of the screen. The relative importance of a region depends on the current unit task. For example, before a unit task is selected, a participant may actively search the radarscope to find aircraft to work on. Once the task is initiated, gathering information about the selected airplane is important, as is keeping track of the target in the radarscope. Therefore we categorized the radarscope and the information boxes as *on-task regions.* The rest of the areas on which eyes fixate are called *off-task regions.* The attentional focus hypothesis predicts that learning in GT-ASP should be reflected in the pattern of attention distribution, as assessed through eye fixations in the current study. People should learn to pay more attention to on-task regions with practice and to pay less attention to the off-task regions. Therefore, the scale factor associated with the off-task regions should be greater than the scale factor associated with the on-task regions.

Although we closely replicated the Georgia

Tech experiment, we made a few changes. Because it is hard to control for the complexity of military actions across scenarios, we decided to drop the military action unit task. As this simplification makes the task easier than the original GT-ASP task, it allows us to better examine the practice effect on the unit tasks that we are interested in. Also, we used only four scenarios and repeated these across practice days to keep the difficulty consistent. Participants were unaware of this repetition because the configuration of the planes was rotated for each practice day.

Method

Participants. Twelve undergraduate students, graduate students, and staff members of Carnegie Mellon University were paid for 5 days of participation.

Task and equipment. The GT-ASP task was performed with a Windows-operated personal computer. The GT-ASP system we adopted was developed by Hodge et al. (1995) and modified by CHI Systems (Zachary et al., 1998). The nature of the modification by CHI Systems involved adaptation of the original source code, which was run in the DOS mode, into a Windows-based operating system, without changing the actual course of events in scenarios. We used an ETL-500 video-based, head-mounted eye-tracking system with a magnetic-based head tracker from ISCAN®, Inc. The software for collecting and analyzing eye data consisted of the Eye Point-of-Regard Analysis Lab (EPAL; Douglass, 1998) software suite, which was developed in our lab to facilitate the development of eye-tracking experiments and their analyses.

Procedure. Among the modified scenarios by CHI Systems, we chose four scenarios with no military action within the first 15 min. The experiment was conducted individually and lasted for 5 days. On the 1st day, a participant studied the briefing materials with the guidance of an experimenter. Then the experimenter demonstrated the GT-ASP task for 30 min, using another scenario with no military action, and this scenario was not played in the actual experiment. There was no separate practice on key press sequences. From the 2nd day on, participants performed four 15-min scenarios each day. These scenarios were repeated across 4 days in a counterbalanced order for each participant. The configuration of a scenario was rotated by 90° for each repetition. Performing the GT-ASP task requires relatively good mastery of the briefing materials, which convey an extensive amount of information, including the purpose of the task, useful strategies, types of aircraft and their characteristics, and relationships between specific information and the intents and types of aircraft. To encourage participants to study the material, we conducted two brief quizzes on the first 2 days of actual task performance, which is different from the Georgia Tech experiment. Participants were told that for actual task performance, the experimenter would provide any information that they would need from the briefing packet. Before every scenario, the quality of eye tracking was recalibrated. After each scenario, participants were given feedback on their timing and accuracy. Each day, the experiment lasted up to 2 hr.

Results and Discussion

Unit-task analysis. The time to perform unit tasks sped up with practice by factors ranging from 1.9 to 2.6. The extent of improvement is similar to that of the Georgia Tech data. Table 2 presents the results of power-law fitting to the unit tasks with the single-exponent constraint. As in the Georgia Tech data analysis, the asymptotes, the scale factors, and the overall exponent were estimated so that the sum of χ^2 deviation for each unit task would be minimized. The overall learning rate (exponent = 1.28) in the CMU experiment was much higher than in the Georgia Tech experiment (0.56). Presumably, this is because in the CMU experiment there was no key press practice of unit tasks, which was a part of the Georgia Tech experiment. The overall goodness of fit of the single-exponent model is quite good, $\chi^2(7) = 0.37$, $R^2 = .999$.

Functional-level subtasks. The results of power-law fitting are presented in Table 2. The mean latencies of both identification and CAP control unit tasks were fit well with the single-exponent model, $\chi^2(28) = 6.50$, $R^2 = .98$. This single-exponent model inherited the exponent from the unit-task analysis described previously. The successful fitting of the single-exponent model supports the decomposition hypothesis. Using the exponent, we estimated the best fitting individual power function for each participant

TABLE 2: Exponent-Constrained Power-Law Fitting of CMU Experiment

	Power Functions	R^2	*df*	χ^2
	Unit-Task Level			
Identification	$T = 10.34 + 17.20\ N^{-1.28}$	.99	7	0.37
CAP control	$T = 8.62 + 23.50\ N^{-1.28}$	.99		
Information search	$T = 1.67 + 2.33\ N^{-1.28}$	.99		
Display Control	$T = 1.91 + 5.19\ N^{-1.28}$	.99		
	Functional Level: Identification			
Search	$T = 1.52 + 2.35\ N^{-1.28}$	.99	8	1.58
Initiate	$T = 0.85 + 1.19\ N^{-1.28}$			
Update	$T = 0.87 + 0.95\ N^{-1.28}$			
Save	$T = 0.56 + 0.81\ N^{-1.28}$			
	Functional-Task Level: CAP Control			
CAP selection	$T = 1.73 + 2.74\ N^{-1.28}$	.99	6	0.40
Change	$T = 1.82 + 1.63\ N^{-1.28}$			
Execute	$T = 0.83 + 2.56\ N^{-1.28}$			
	Eye Movement: Identification			
Cognitive: on task	$T = 1.14 + 1.23\ N^{-1.28}$	.99	8	0.58
Cognitive: off task	$T = 1.58 + 2.91\ N^{-1.28}$			
Motor: on task	$T = 0.07 + 0.10\ N^{-1.28}$			
Motor: off task	$T = 0.78 + 0.94\ N^{-1.28}$			
	Eye Movement: CAP Control			
Cognitive: on task	$T = 0.45 + 3.43\ N^{-1.28}$	.99	8	1.46
Cognitive: off task	$T = 1.14 + 3.86\ N^{-1.28}$			
Motor: on task	$T = 0.82 + 0.65\ N^{-1.28}$			
Motor: off task	$T = 0.97 + 0.88\ N^{-1.28}$			

and obtained individual asymptotes and scale factors. The mean scale factors are presented in Figure 3b. The scale factor was greater for subtasks with cognitive demands (4.37 s) than for those with motor demands (1.88 s), $F(1, 11) = 9.74$, $MSE = 7.65$, $p < .01$. Also, the interaction was significant, $F(1, 11) = 7.21$, $MSE = 3.58$, $p < .05$. For both unit tasks, the scale factor was greater for the cognitive than for the motor subtasks, but this trend was greater for CAP control, $t(11) = 3.96$, $p < .05$, than for identification, $t(11) = 1.03$, $p < .05$. Consistent with the Georgia Tech data, this analysis confirms that there is more to learn in cognitive skills than in motor skills, supporting the practice-relativity hypothesis.

Eye movements. In addition to the regions containing the information boxes, menu panels, and radarscope, discussed previously, we added two regions: middle of nowhere and off screen. The middle-of-nowhere region consisted of areas on the screen that were not classified as meaningful. Eye movements could be directed to off-screen areas when participants looked at the keyboard to press the appropriate keys. We then categorized these five regions as either on-task regions (radarscope and information boxes) or off-task regions (middle of nowhere, off screen, and menu panel). These are the regions the participant does not necessarily have to pay attention to while performing unit tasks.

The attentional focus hypothesis claims that one factor in skilled performance is decreased attention to task-irrelevant information. During performance of the GT-ASP task, this should be revealed by changes in eye fixation times reflecting the functionality of the fixated regions. Moreover, the eye fixation times should change with how far an individual is in the identification unit task. For example, the on-task regions in the GT-ASP task are critical for information gathering, which is the main function of the search subtask, but these regions are not important during the subsequent motor subtasks. The previous analyses of functional-level subtasks showed that the identification and CAP control tasks consist of intervals tapping different processing demands. Therefore, we report eye movement data in terms of processing demands. The results of power-law fitting are given in Table 2. We used the same exponent from the unit-task analysis and estimated the asymptotes and the scale factors. The goodness of fit was good, $\chi^2(16) = 2.04$, $R^2 = .99$, indicating that the common learning rate for the unit tasks can apply even to the eye movement data.

As with the latency data, we estimated the best fitting individual power function for each participant and obtained individual asymptotes and scale factors with the exponent mentioned previously. The mean scale factors of fixation time from both identification and CAP control are presented in Figure 4 as a function of processing demands and relevancy of regions. According to the practice-relativity hypothesis, more learning should take place during cognitive subtasks than during motor subtasks. Consistent with this prediction, the scale factor was greater for the cognitive interval (2.82 s) than for the motor interval (0.73 s), $F(1, 11) = 25.68$, $MSE = 4.06$, $p < .01$. According to the attentional focus hypothesis, increased learning should be reflected in fewer eye fixations to the off-task region because participants should learn not to pay attention to the region that does not afford task-relevant information. Consistent with this prediction, the scale factor for the off-task region was greater (2.21 s) than for the on-task region (1.34 s), $F(1, 11) = 49.06$, $MSE = .37$, $p < .0001$. Also, the relevance effect was greater for the cognitive subtasks (1.11 s) than for motor intervals (0.63 s), $t(11) = 2.77$, $p < .05$.

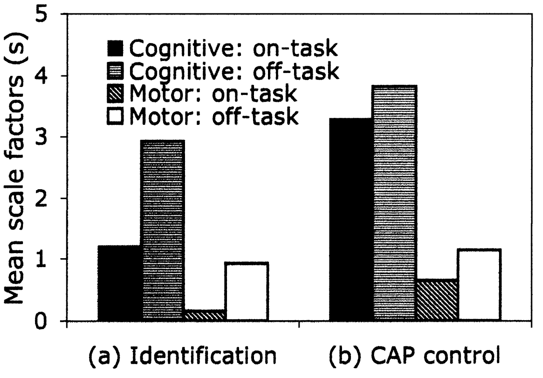

Figure 4. Mean scale factors of fixation time for the cognitive subtasks and the motor subtasks, (a) during the identification task and (b) during the CAP control task.

Summary

The CMU data, like the Georgia Tech data, supported the decomposition and the practice-relativity hypotheses. The same exponent that explains the unit-task performance with learning could also explain the changes in eye fixations with learning, further supporting the decomposition hypothesis. Moreover, the pattern of eye movement data showed that learning at the eye-movement level may be quite sensitive to the task relevance of the information available from the region where the participants were investing their attentional resources. Although this hypothesis has gained support from several studies (Haider & Frensch, 1999; Lee & Anderson, 2001), we further showed that the task relevance of information can dynamically change depending on the task currently being performed, and our eye-fixation data were sensitive to this.

GENERAL DISCUSSION

In the current study, we adopted the unit-task analysis paradigm (Card et al., 1983) to examine three hypotheses of skilled performance in a dynamic, complex problem-solving situation. The evidence for the decomposition hypothesis is that the same learning rate can be applied to different levels of task organization, suggesting that component-level fluency is truly critical for achieving overall fluency. Further, the practice-relativity hypothesis suggests that the contributions of each component to overall learning may not be the same. Specifically, in the current study, the cognitive components explained more of the learning in general, as compared with the motor components, because the level of prior learning of the cognitive component was lower when a new task was introduced.

The current study has implications for the design of part-task training sessions. There are a number of reasons to consider part-task training. For instance, it is easier and less overwhelming to administer practice in a piecemeal manner (Hodge et al., 1995; Wightman & Lintern, 1985). In addition, part-task training on component skills can lead to automatic processing, especially through consistent mapping between a specific stimulus and a response (Schneider & Shiffrin, 1977). This component-level fluency has been hypothesized to facilitate the learning of higher levels of organization (Fisk et al., 1987). The practice-relativity hypothesis suggests that component skills should be selected for part practice on the basis of their task demands. In particular, part tasks that emphasize less practiced cognitive components would produce greater benefit than part tasks that emphasize motor components.

The attentional focus hypothesis of the current study suggests that consistent mapping is crucial when arranging the source of information. In a dynamic display environment such as the one used in the current study, the same spatial location of a display may provide different types of information, depending on the context. The attentional focus hypothesis suggests that part-task training will be more effective if the spatial location of each type of information is consistent, which will eventually facilitate perceptual learning of information-seeking behavior.

ACKNOWLEDGMENTS

This research is supported by Grant N00014-99-1-0097 from the Office of Naval Research and is performed in collaboration with the Training Systems Division of the Naval Air Warfare Center. We thank Gwen Campbell, Craig Haimson, and Frank Lee for their helpful comments on earlier drafts.

REFERENCES

Ackerman, P. L. (1988). Determinants of individual differences during skill acquisition: Cognitive abilities and information processing. *Journal of Experimental Psychology: General, 117,* 288–318.

Anderson, J. R., Conrad, F. G., & Corbett, A. T. (1989). Skill acquisition and the LISP tutor. *Cognitive Science, 13,* 467–505.

Cannon-Bowers, J. A., & Salas, E. (Eds.). (1998). *Decision making under stress: Implications for individual and team training.* Washington, DC: American Psychological Association.

Card, S. K., Moran, T. P., & Newell, A. (1983). *The psychology of human-computer interaction.* Hillsdale, NJ: Erlbaum.

Douglass, S. A. (1998). EPAL: *Data collection and analysis software for eye-tracking experiments.* Pittsburgh, PA: Carnegie Mellon University.

Fisk, A. D., Ackerman, P. L., & Schneider, W. (1987). Automatic and controlled processing theory and its application to human factors problems. In P. A. Hancock (Ed.), *Human factors psychology* (pp. 159–197). New York: North Holland.

Haider, H., & Frensch, P. A. (1999). Eye movement during skill acquisition: More evidence for the information-reduction hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 172–190.

Hodge, K. A. (1998). *Training for decision-making in a complex dynamic task environment: An empirical investigation focusing on part-task training and feedback.* Unpublished dissertation, Georgia Institute of Technology, Atlanta.

Hodge, K. A., Rothrock, L., Kirlik, A. C., Walker, N., Fisk, A. D., Phipps, D. A., et al. (1995). *Trainings for tactical decision making under stress: Towards automatization of component skills* (HAPL-9501). Atlanta: Georgia Institute of Technology, School of Psychology, Human Attention and Performance Laboratory.

Johnston, J. H., Poirier, J., & Smith-Jentsch, K. A. (1998). Decision making under stress: Creating a research methodology. In J. A. Cannon-Bowers & E. Salas (Eds.), *Decision making under stress: Implications for individual and team training* (pp. 39–60). Washington, DC: American Psychological Association.

Kirlik, A., Fisk, A. D., Walker, N., & Rothrock, L. (1998). Feedback augmentation and part-task practice in training dynamic decision making skills. In J. A. Cannon-Bowers & E. Salas (Eds.), *Decision making under stress: Implications for individual and team training* (pp. 91–114). Washington, DC: American Psychological Association.

Lee, F. J., & Anderson, J. R. (2001). Does learning of a complex task have to be complex? A study in learning decomposition. *Cognitive Psychology, 42,* 267–316.

MacKay, D. G. (1982). The problems of flexibility, fluency, and speed-accuracy trade-off in skilled behavior. *Psychological Review, 89,* 483–506.

Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 1–55). Hillsdale, NJ: Erlbaum.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search and attention. *Psychological Review, 84,* 1–66.

Wightman, D. C., & Lintern, G. (1985). Part-task training for tracking and manual control. *Human Factors, 27,* 267–283.

Zachary, W. W., Ryder, J. M., & Hicinbothom, J. H. (1998). Cognitive task analysis and modeling of decision making in complex environments. In J. A. Cannon-Bowers & E. Salas (Eds.), *Decision making under stress: Implications for individual and team training* (pp. 315–344). Washington, DC: American Psychological Association.

Myeong-Ho Sohn is an assistant professor in the Department of Psychology at George Washington University. He received his Ph.D. in cognitive psychology at Pennsylvania State University in 1998.

Scott A. Douglass is a Ph.D. student in the Department of Psychology at Carnegie Mellon University. He received his B.A. in psychology at San Diego State University in 1989.

Mon-Chu Chen is a graduate student at the Human-Computer Interaction Institute, School of Computer Science, Carnegie Mellon University, where he received his M.S. in human-computer interaction in 2001.

John R. Anderson is the R. K. Mellon University Professor of Psychology and Computer Science at Carnegie Mellon University. He received his Ph.D. in psychology at Stanford University in 1972.

Date received: August 13, 2003
Date accepted: January 25, 2005