Visual Scanning and Pilot Expertise: The Role of Attentional Flexibility and Mental Model Development

ANDREW H. BELLENKES, M.A., CHRISTOPHER D. WICKENS, Ph.D., and ARTHUR F. KRAMER, Ph.D.

BELLENKES AH, WICKENS CD, KRAMER AF. *Visual scanning and pilot expertise: the role of attentional flexibility and mental model development.* **Aviat Space Environ Med 1997; 68:569–79.**

In order to examine differences in flying expertise, 12 novice and 12 expert pilots flew a 7-segment simulation pattern under specific attentional constraints while cockpit instrument visual scan was recorded. Flight segments involved various combinations of maneuvering of heading, altitude and airspeed. Expert pilots performed better than novices on vertical and longitudinal, but not lateral control. They accomplished their superior vertical tracking by allocating more control resources to the vertical control. Analyses of scanning strategies revealed that experts: a) had shorter dwells and more frequent visits to most instruments; b) adapted their visiting strategy more flexibly in response to changing task demands; c) demonstrated a better mental model of cross-coupling and predictive relations between and within axes; and d) showed more frequent checking of axes whose values remained constant. The data is discussed in terms of their implications in pilot cockpit scan training program development.

THIS STUDY ADDRESSES how novice pilots differ from experienced pilots in attention control, as the latter is measured by visual scanning and control response velocity. If such differences can be carefully identified and reliably measured, then training strategies can be formulated to accelerate the development of effective attention control for the novice pilot and thus, ideally, accelerate the development of expertise. In order to provide background for the experiment reported here, we review literature bearing on the intersections of four areas; attention control, visual scanning, expertise, and aviation, progressing toward an identification of the specific context for the experiment that we describe.

Attention control: Attention control and attentional flexibility have been identified as critical components of human performance in the operation of high risk dynamic systems (15), like flying an aircraft (18), monitoring a complex industrial process plant (26), supervising aircraft in air traffic control (34,35), or driving a ground vehicle (15,27). Numerous aircraft accidents, for example, have been attributed to "neglect" of the monitoring of altitude or other attributes of aircraft orientation (37,44), a type of behavior that might be characterized by "rigid" (7), in contrast to "flexible" attention allocation.

In particular, the research of Gopher and his colleagues has concluded that measures of the speed of attention switching, assessed in a dichotic listening task, predict both the success (13,18), and failure (i.e., accidents; 22) of vehicular control both in the air and on the ground. Gopher's research has also revealed that training people to flexibly allocate their attention between different tasks provides positive transfer to overall dual-task performance, when such flexibility is not explicitly required, but is implicitly imposed by changes in the difficulty of one or the other of two time shared tasks (19); i.e., high difficulty tasks demand more attention.

Expressed within an information processing framework, attention control can be broken down into its perception (what channels to select), and response (what actions to perform) components. For visual tasks, in a naturalistic environment, perceptual selection is closely associated with, and typically measured by, visual scanning behavior. For tasks involving continuous actions (e.g., the tracking tasks found in many vehicular control environments), gradations of attention can be measured by the open loop gain or power of the response (38,43), a measure that closely correlates with mean control velocity. Hence, in a dual-axis tracking task, emphasis on one axis over the other will produce a reduction in error, and an increase in mean control velocity on the favored axis (12).

Aviation and attention control: Researchers have examined both control behavior and visual scanning in the cockpit, or in other simulated aviation environments. With regard to control behavior, a handful of researchers have modeled the allocation of attention toward favored axes of flight path control, in terms of increases in gain, or reductions of noise variance associated with task-relevant variance in control position (3,5).

Considerably more research has examined the role of visual scanning as an index of attention allocation in the

From the University of Illinois at Urbana-Champaign, Institute of Aviation, Aviation Research Laboratory, #1 Airport Road, Savoy, IL.

This manuscript was received for review in June 1996. It was revised in October 1996 and accepted for publication in November 1996.

Address reprint requests to Andrew H. Bellenkes, who is a Commander in the U.S. Navy and presently working on his doctorate in human factors at the University of Illinois at Urbana-Champaign, Aviation Research Lab, #1 Airport Road, Savoy, IL 61874.

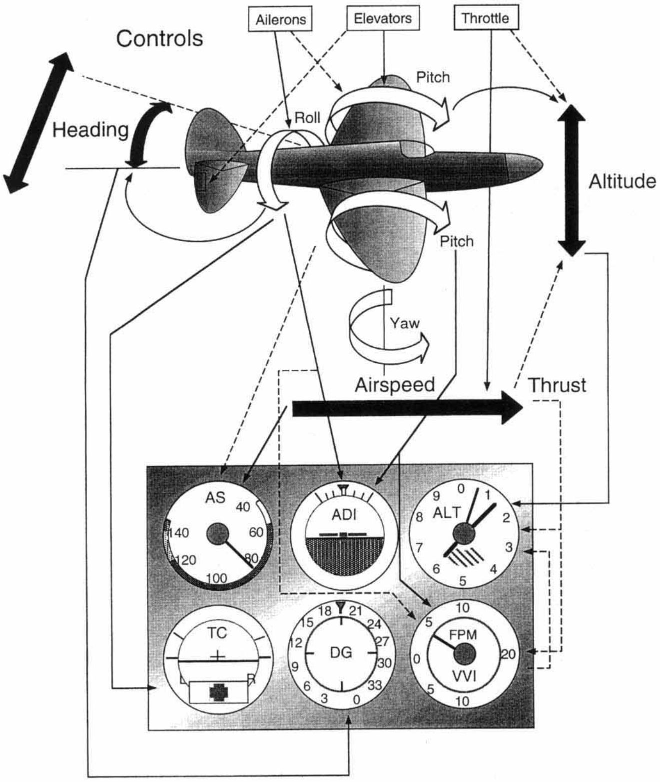


Fig. 1. Schematic representation of the causal influences in flight dynamics which underlie variation in flight instrument readings. The solid thin lines represent sources of influence within a primary flight axis (lateral, vertical, longitudinal). The dashed lines represent cross-coupling influence between axes.

cockpit, an association which is easy to make for three reasons: a) because of the clearly defined, and spatially separated flight instruments that invite feasible scanning measurement; b) because differences in the dynamic bandwidth of information revealed by instrument fluctuation make instrument scanning amenable to supervisory sampling models of optimal scan patterns (4,25,31) and c) because of the clearly evident consequences of the breakdowns in attention control—here the failure to attend to critical instruments, such as the altimeter at low flight levels.

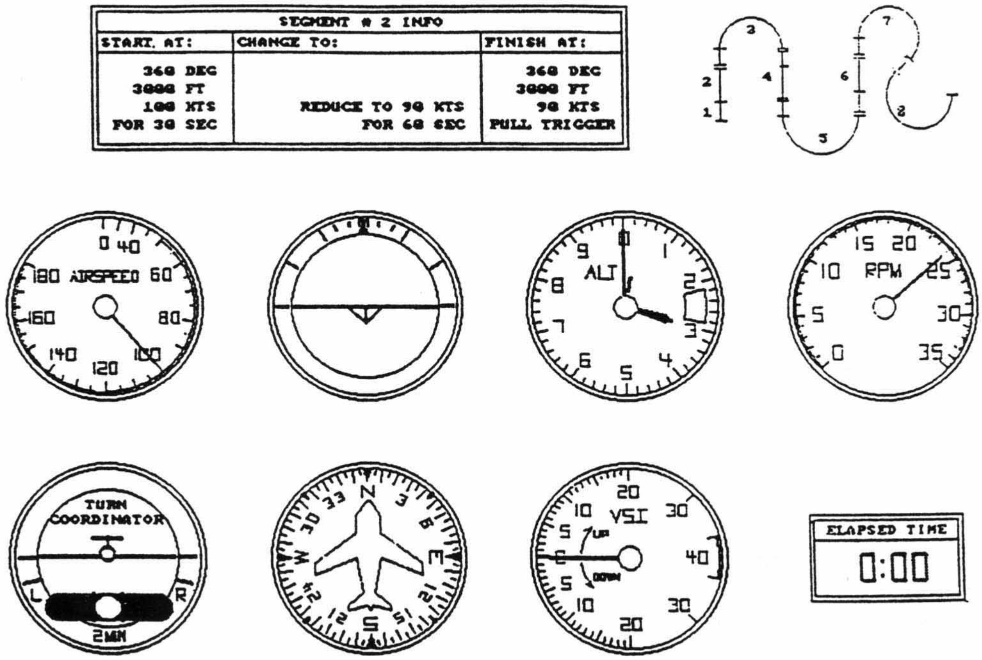
Visual scanning may be assumed to be driven by a mental model of the process whose elements are being displayed (25). The expert pilot's mental model of flight dynamics, which drives the scan across the instrument panel, is complex, reflecting the complexity of the dynamics themselves. These dynamics are shown schematically in **Fig. 1,** which depicts the three primary flight controls, the six degrees of freedom of aircraft movement (3 of rotation, 3 of translation; although the lateral translation axis is controlled via heading change), and the six critical flight instruments (in visual flight) which are the destination of the visual scan.

As any pilot well knows, three features make the flight dynamics particularly challenging. First, attention is limited and therefore to some extent the pilot must trade off the allocation of resources between the three primary tasks or axes of control (longitudinal, lateral, vertical). The appropriate allocation of resources to axes that require positive control (because they are changing) while not altogether neglecting those that must be monitored (so they do not diverge from target values) requires a high skill of attentional flexibility (15). Second, all three axes are somewhat sluggish, defining second and (in the case of lateral deviations), third order systems. This imposes a need for consulting predictive indicators; for example, the vertical speed indicator is predictive of altitude changes. Third, dynamics are interactive in complex ways, as indicated by the thin arrows in Fig. 1. For example, increases in bank causes a pitch downward leading to a loss of altitude and, as a consequence, a gain in airspeed.

All of the features described above should be reflected by the scan of the well-trained pilot. Analysis of this scan behavior can provide information both regarding where the pilot looks, and also how long the pilot fixates or dwells on a particular instrument from which information is being extracted.

The seminal work of Fitts, Jones, and Milton, (11) and Jones, Milton, and Fitts (21) laid the groundwork for subsequent aircraft instrument scanning research. Among several other findings, these investigators observed that the most important (most visited) instrument was the attitude directional indicator (ADI). The cause of such importance can be attributable to four characteristics: a) the ADI has the highest bandwidth (31); b) it is important because changes in pitch or roll are necessary to carry out changes in lateral and vertical deviation; c) it is important because excessive pitch or roll can lead to an aircraft stall; and d) the ADI is the only instrument that combines two attributes of information in a single object (the vertical displacement, and angular orientation of the artificial horizon). Fitts and his colleagues also established that dwells on the ADI tend to be longer (and more variable) than on other instruments (see also 21). Further studies of pilot instrument scanning by Carbonnell et al. (4), Hameluck (20), Harris and Christhilf (21), and Tole et al. (36) have all served to refine a model of instrument scanning based upon the pilot's assumed mental model of the underlying flight dynamics in a way that is consistent with the representation in Fig. 1. That is, more important instruments are fixated more often, and the correlation between axes (and therefore instruments) dictates certain patterns of contingencies or linkages between scans to pairs of instruments (see also ref. 33 for similar analyses of scanning in an air traffic control setting).

In spite of the substantial number of studies examining scanning and expertise in other skilled domains such as driving (27), radiology (24), or athletic performance (1), only a few investigations have hinted at scanning differences related to pilot expertise. Fitts et al. (11) noted that more experienced pilots tended to make shorter dwells. DeMaio and his colleagues (8,9) examined pilots' ability to note deviant readings of statistically presented indicators, and found that experts could do so more rapidly than novices as if they could either process more information from a single glance, or could sequentially direct their scan path more rapidly to the deviant indicator. Tole et al. (36) measured instrument scanning behavior of low and high time pilots at lower and higher levels of workload. They found that experts tended to fixate



Vertical Axis	Lateral Axis	Longitudinal Axis
AI (pitch)	AI (roll)	Airspeed
VSI	Heading Indicator	VSI
Altimeter	Turn coordinator	
Airspeed		
Elevator control	Aileron control	Throttle control

Fig. 2. Cockpit instrumentation layout as displayed to subjects. The location of the display approximates actual cockpit viewing angle and distance. A description of the instruments and their corresponding control actions is listed.

the ADI more frequently and with shorter dwells than did novices at low workload. Increases in workload brought about an increase in ADI fixation for novices (at the expense of other instruments), while experts did not adjust their scan strategy as much, when workload was increased. Finally, Spady and Harris (32) noted that low time pilots tended to have a more predictable scan pattern across instruments than did high time pilots, as if the former were driven by a more rigid open loop schema, whereas the expert's scanning was more flexible, dictated by the dynamically changing state of the aircraft.

A limitation of many of the studies reported above is that they were based upon a very small sample size; hence statistically reliable indices of scanning behavior were difficult to infer. An experiment which provided the foundation for that described here was carried out by Kramer et al. (23) and used a larger sample size. Two groups of pilots, students and flight instructors, flew a series of simulated maneuvers using the instrument flight simulator depicted in **Fig. 2.** The maneuvers varied in the extent to which they required lateral, vertical and longitudinal (i.e., airspeed) change, and scan patterns were measured. A large number of measures were extracted, and generally revealed that students tended to dwell longer on all instruments, and in particular, visited the ADI more frequently than did the flight instructors, while the latter maintained a more evenly distributed scan pattern across all instruments. However, analysis of the performance data revealed that differences between the two groups were attenuated because many of the "expert" flight instructors did not substantially differ from the novices in their flight path tracking performance, in part because many in the former group were not "current" in their flying proficiency. Another difficulty in interpreting the data resulted because there was considerable variation between pilots in terms of when they initiated each maneuver (turn, climb, etc.). This variance turned out to make precise scoring of the maneuvers somewhat difficult to achieve.

The current study: The purpose of the current study is to build upon the simulation paradigm developed by Kramer et al., in order to examine how differences in flight expertise are reflected in attentional strategies, as the latter are revealed by differences in scanning and control power. Because the study of attentional flexibility requires experimental manipulation of the demand or priority of different tasks and information sources, we chose to implement such a manipulation in three ways: a) natural variations in the bandwidth of the different instruments provide an implicit manipulation of instrument importance (31); b) because the mission segments varied in terms of whether the parameters on each axis of the three axes were to be maintained at a constant value (easier) or varied along a commanded trajectory (more difficult), a second implicit manipulation of priority was imposed and varying axes (along which a change is requested) were assumed to demand more attention; c) across blocks, we explicitly varied the priority of lateral vs vertical control (16,28). This variable turned out to have little influence on performance or scan measures, and will not be discussed further (see ref. 42 for details).

We assumed that expertise would lead to better performance, in terms of smaller deviations from the flight path. In addition, we predicted that expertise would re-

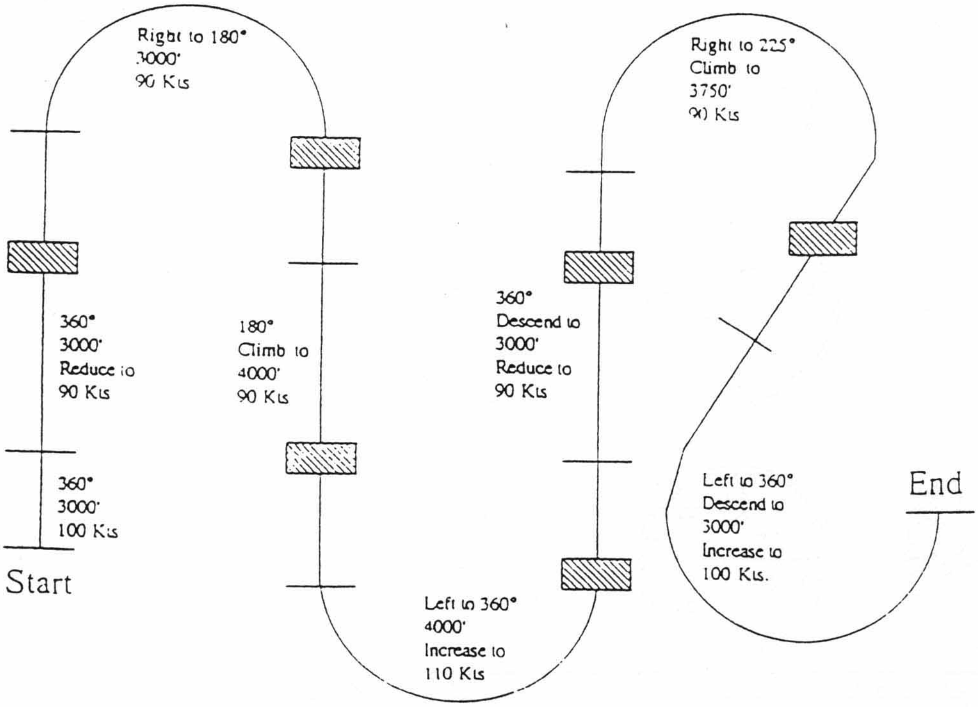


Fig. 3. Mission profile for simulated 7-segment flight (23).

veal greater responsiveness to demand changes, as reflected by the flexibility of scanning and/or control measures, although there was no apparent basis for predicting that expertise differences would be manifest more in one measure than the other. Our review also suggested that expertise would lead to the development of a more refined mental model that would reveal greater sensitivity to cross-coupling between axes, and to predictive elements or leading indicator instruments within an axis. The more refined mental model should provide a basis for information-guided scanning, reflected in attention flexibility. Finally, we assumed that greater expertise would lead to greater efficiency in information extraction, and hence, to shorter dwell durations, particularly on those instruments of greater complexity. With greater efficiency of extracting information from any single source, there should be more reserve capacity or free time available for the expert to sample other sources.

METHODS

Participants

Participants consisted of 24 pilots (20 male and 4 female; age range 18–26 yr) involved in the private pilot flight curriculum at the University of Illinois Institute of Aviation. As a function of pre-test questionnaire data obtained from each participant, the pilots were assigned to one of two experimental groups based on level of expertise—12 student pilots with a mean of 1 h instrument flight time (novices) and 12 flight instructors with a mean of 80 h instrument flight time (experts).

Equipment and Data Collection Requirements

Flight simulator: The simulation employed software designed specifically for the study. It was designed to recreate the flight dynamics of a Beach Sport (Sundowner) aircraft; using first-order equations and tuned to cruise flight only. The aircraft instrument panel and all appropriate data were generated and displayed using a Gateway computer (Model 4DX2-66V, Gateway 2000, N. Sioux City, SD) with a SVGA graphics card (Fig. 2). A Mitsubishi color 19-in monitor (Model HL6905ATK) displayed the instrument panel, standard primary instruments, and instructions. The display subtended a visual angle of 29° horizontally and 22.5° vertically with a viewing distance of 66 cm. The instruments were 6° in diameter with a minimum separation of 2.2°. All control inputs were made via a right sidearm-mounted joystick. Roll and pitch were controlled by lateral and fore-aft stick movements. A button atop the stick controlled aircraft power and there was a trim knob at the base of the stick. Stick inputs were sampled at 5 Hz.

Head-mounted eye/head tracker: Eye scan measures were made using an ASL series 4000 head-mounted eye tracker (Model 4100H, Applied Sciences Laboratory, Waltham, MA). The sampling and output rate of the tracking camera was 60 Hz.

Procedures

Prior to commencing the study, all subjects completed a questionnaire regarding their flight experience (hours, aircraft type flown, currency, etc.). Each participant was then given a standardized briefing regarding the flight task and maneuvers to be completed. Cockpit panel instrumentation (displayed on the CRT in front of the participant) was described (Fig. 2). Following the brief, the eye tracking apparatus was placed on the participant's head, and eye location calibrated. Once done, the experimental session began.

The mission profile used in the current study was identical to that used by Kramer et al. (23) and is illustrated in **Fig. 3.** Each flight was broken down into 14 segments.

All odd-numbered segments involved straight and level lead-in legs holding the altitude, heading, and airspeed from the finish of the previous leg. Every even-numbered segment (a maneuver segment) required at least one change to the speed, heading or altitude during the maneuver. The change(s) for each segment, were stipulated in an instruction box displayed on the instrument panel and could be accomplished within 60 s in the first five maneuver segments, and within 75 s in the last two. As shown in Fig. 3, maneuver segments 2, 4, and 6 required only one axis to be changed. Segments 8, 10, and 12 required two axes to be changed, and segment 14 required changes on all three axes.

At the beginning of each segment, the experimenter provided each participant with a 5-s verbal countdown prior to the start of maneuver. If executed correctly and within the allotted time, the end of each maneuver placed the participant at the lead-in leg of the next maneuver. If a maneuver was not completed within the allotted time, the program reset the pilot to the lead-in leg for the next maneuver. The time required to complete the full simulated mission (all eight segments) was approximately 13 min.

Each pilot flew four consecutive missions in a single session totaling approximately 1.5 h. The first mission was a practice session which was used to acquaint the pilot with the mission profile and use of the eye tracker. In order to ensure optimal performance throughout practice sessions, the experimenter provided each pilot with feedback as necessary (i.e., "watch your heading" or "watch your altitude"). The subsequent three data collection runs varied in the emphasis placed on lateral vs vertical flight path control, but this manipulation will not be discussed in the current report. Visual scan and tracking performance data were recorded, and stored for later analysis.

Analysis framework: Our analyses are based upon the idea that the pilot allocates attention to tasks at a global level (altitude, lateral and longitudinal control), and adopts specific scanning and control activities to accomplish those tasks. Our approach suggests that evidence for attention to lateral control is provided by fixations on the following instruments (Fig. 2); a) the Attitude-Directional Indicator (ADI) which presents pitch and roll information, the leading indicators for control of both the vertical and lateral axes; b) the Directional Gyro (DG); and c) the Turn Coordinator (TC), and any active manipulations of control on the X axis control (ailerons). Attention to the vertical axis is evidenced by fixations on the ADI, the Altimeter (Alt), the Vertical Velocity Indicator (VVI) which indicates the rate of ascent or descent, and active manipulations of control on the Y axis (elevators). Attention to the longitudinal control is evidenced by fixations on the Airspeed Indicator (ASI) and, to a lesser extent, on the Altimeter and the VVI, and by throttle control.

We examined the data for two general kinds of effects: a) tracking performance on the three primary axes; and b) strategic behavior, which can be partitioned into that related to scanning (visit frequency and dwell duration) and to manual control behavior.

Performance data were digitized and analyses carried out for each of the seven flight segments to examine group differences and data trends. The effects of pilot expertise were assessed by measuring three categories of dependent variables: a) performance measures—RMS error for airspeed, altitude, and heading; b) scan measures—the number of visits and mean dwell time on each of the 6 instruments; and c) elevator and aileron control velocity. A 2 × 3 × 7 (expertise × emphasis level × segment) repeated mixed design analysis of variance was conducted on each dependent variable to ascertain the experimental effects of the three independent variables.

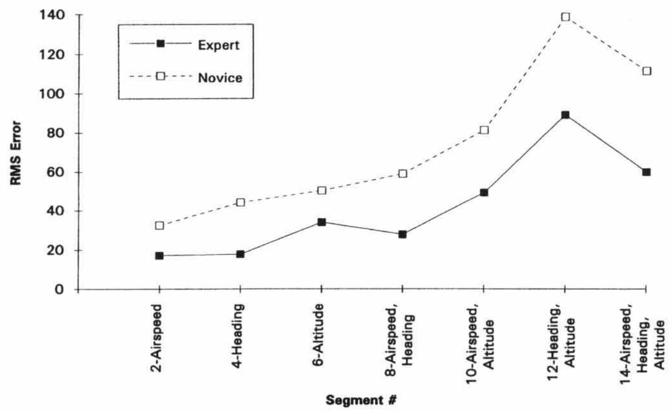


Fig. 4. Vertical axis (altitude) control performance.

RESULTS

Flight Path Tracking Performance

RMS heading error (lateral axis): Participants were required to maintain heading to within ±5° of the criterion heading value (as per Fig. 3). The data reveals a highly significant effect of segment ($F(6,132) = 54.58$; $p < 0.001$). Both novices and experts showed similar performance patterns in that RMS error was highest during those segments (2,6,10,12,14) where a change in heading was required. The data indicated a lack of a significant expertise effect ($F(1,22) = 1.06$; $p < 0.10$).

RMS altitude (vertical axis): Participants were required to maintain altitude to within ±50 ft of the criterion altitude (as per Fig. 3). **Fig. 4** illustrates altitude tracking performance of novices and experts over segments. The ANOVA revealed a highly significant effect of both segment ($F(6,132) = 43.80$; $p < 0.001$) and expertise ($F(1,22) = 26.01$; $p < 0.001$). There was a general tendency for performance to suffer as the flight progressed from segments 2 through 12. RMS error for both groups improved slightly in segment 14, and experts performed better than novices. The data also suggest a significant segment × expertise effect ($F(6,132) = 2.44$; $p < 0.03$) as can be seen in the figure. The advantage of experts over novices was greatest in the two most difficult segments (12 and 14), a difficulty resulting from the pilot's need to simultaneously change values on the two higher order, interacting axes—lateral and vertical.

RMS airspeed error (longitudinal axis): The pilot's ability to maintain airspeed to within ±5 knots of criterion airspeed (as per Fig. 3) is illustrated in **Fig. 5.** As with the other performance measures, there was a highly significant effect of segment ($F(6,132) = 24.2$; $p < 0.001$). As

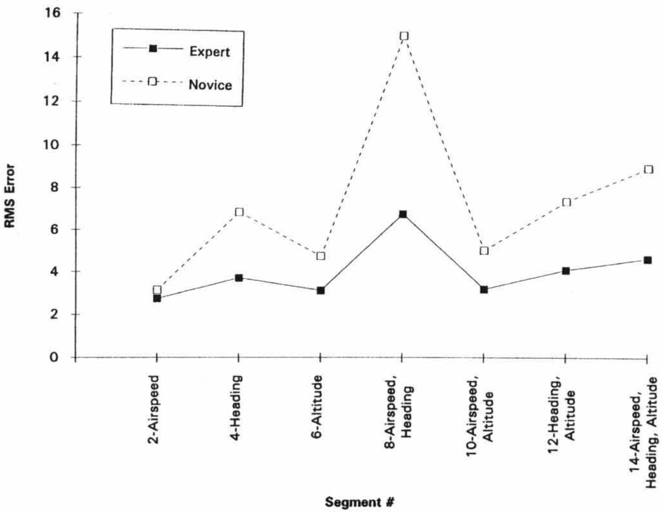


Fig. 5. Longitudinal axis (airspeed) control performance.

heading changes were made (segments 4, 8, 12, 14), RMS tended to increase, relative to the segments when no heading changes occurred (2,6, and 10). There was likewise a highly significant difference between expertise levels ($F(1,22) = 19.54$; $p < 0.002$) with novices having a consistently higher RMS error. Novices particularly showed significantly greater performance decrements than experts during those segments requiring heading changes (4,8,12,14), thereby leading to a significant segment × expertise interaction ($F(6,132) = 5.98$; $p < 0.001$).

Control input velocity: The absolute control input velocities for both the aileron (horizontal) and elevator (vertical) axes were analyzed across segments for each expertise group. The analysis of aileron velocity revealed a significant effect of segment ($F(6,132) = 5.79$; $p < 0.03$), and suggests that there was higher velocity lateral control on those segments during which heading was changing.

Elevator (Y-axis) control velocity data similarly revealed a significant segment effect ($F(6,132) = 4.14$; $p < 0.008$). There was lower elevator control velocity for both groups on those segments where only airspeed or altitude were changing (2 and 6). Elevator control velocity increased for both groups in segments where pilots were required to change more than a single flight parameter (8,10,12,14). The data suggested a marginally significant expertise effect ($F(1,22) = 2.96$; $p < 0.1$) wherein experts input higher vertical velocities than did novices.

Visual Scan Performance

In order to assess the cockpit visual scan pattern, an analysis was made of the number of visits and duration of dwells (in seconds) on each of the cockpit instruments. In the following, we examine these processing effects on the three different control axes in turn, by focusing on the scanning data for the flight instruments:

Lateral (Heading)—Directional Gyro (DG), Turn Coordinator (TC), and Attitude Directional Indicator (ADI).
Vertical (Altitude)—Altimeter (ALT), Vertical Velocity Indicator (VVI), and ADI.
Longitudinal (Airspeed)—Airspeed Indicator (AI) and VVI.

The results for each instrument are described below.

Directional Gyro (DG)—The Primary Heading Indicator

Number of visits: The number of visits to the DG is illustrated in **Fig. 6.** Experts visited the DG significantly more often than did novices ($F(1,22) = 10.2$; $p < 0.004$). Analysis of the data also showed a significant effect of segment ($F(6,132) = 13.41$; $p < 0.001$) as well as a significant segment × expertise interaction ($F(6,132) = 4.64$; $p < 0.005$). Experts made the greatest number of visits to the DG during those segments in which altitude was changing (6,10,12,14) while novices visited most often during those segments in which heading was changing (4,8,12,14). This effect for novices was particularly pronounced during segments 12 and 14 where both heading and altitude were changing.

Dwell durations: There were significant effects of segment ($F(6,132) = 3.18$; $p < 0.006$) and expertise ($F(1,22) = 13.22$; $p < 0.002$). The novices had consistently longer dwell times than did the experts; the latter dwelling an average of approximately 100 ms shorter on the DG.

Attitude Directional Indicator (ADI)

Number of visits: The data reveal a highly significant segment main effect ($F(6,132) = 46.55$; $p < 0.001$) for the number of visits to the ADI. The greatest number of visits (for both groups of pilots) occurred during segments 12 and 14, the two segments during which both altitude and heading were changing. Unlike most of the other instruments, there was no significant expertise effect on ADI visits ($F(1,22) < 1$).

Dwell durations: The data in **Fig. 7** clearly show that novices consistently spent significantly longer dwelling on the ADI than experts ($F(1,22) = 14.55$; $p < 0.001$), a difference of approximately 400 ms. There was also a significant effect of segment ($F(6,132) = 3.05$; $p < 0.01$), with the longest dwells on those two segments (12 and 14) during which heading was changed.

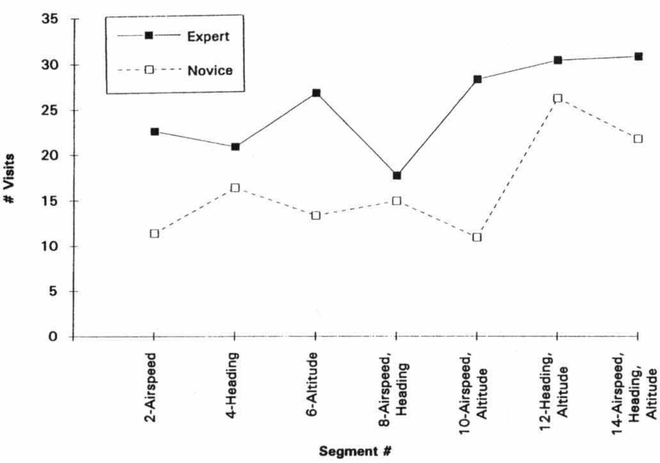


Fig. 6. The average number of visits to the directional gyro.

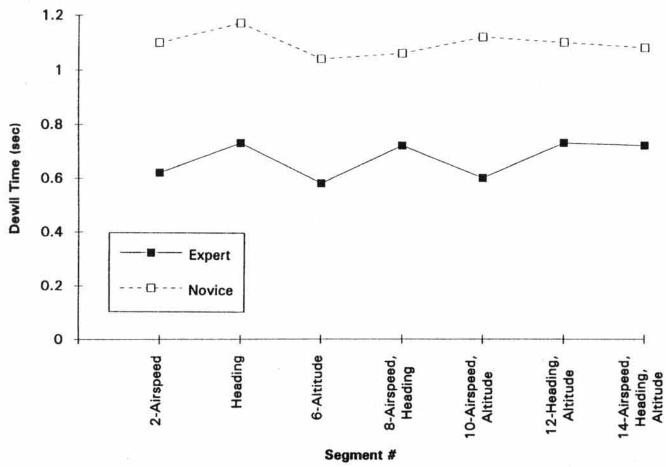


Fig. 7. The average dwell durations on the attitude directional indicator.

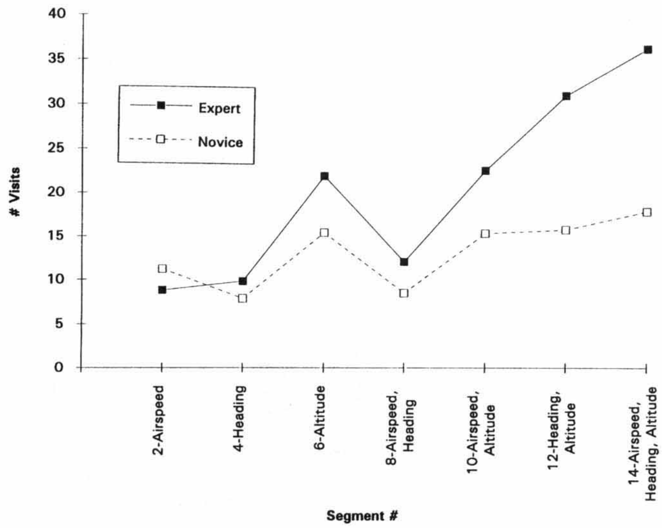


Fig. 9. The average number of visits to the vertical velocity indicator.

Turn Coordinator (TC)

Number of visits: The data reveal a significant effect of segment ($F(6,132) = 17.11$; $p < 0.001$) on TC visits which mimicked that observed on visits to the DG; that is, visits to the TC were more frequent on turning segments (4,8,12,14). The data also revealed a lack of a significant effects of pilot expertise.

Dwell durations: As with the number of visits to the TC, the dwell durations on the TC were longest during those segments during which changes of heading were required ($F(6,132) = 66.67$; $p < 0.001$). The dwell data revealed a significant expertise effect wherein novices consistently dwelled longer than did experts ($F(1,22) = 11.51$; $p < 0.002$).

Altimeter (Alt)

Number of visits: As can be seen in **Fig. 8,** expert pilots visited the altimeter significantly more frequently than did the novices ($F(1,22) = 13.85$; $p < 0.001$). The data also revealed significant effects of segment ($F(6,132) = 7.44$; $p < 0.001$), and a significant expertise × segment interaction ($F(6,132) = 3.33$; $p < 0.004$). The latter suggested that the expert's tendency to visit the altimeter more frequently was greatest on those segments (2, 4, and 8) during which altitude was not changing.

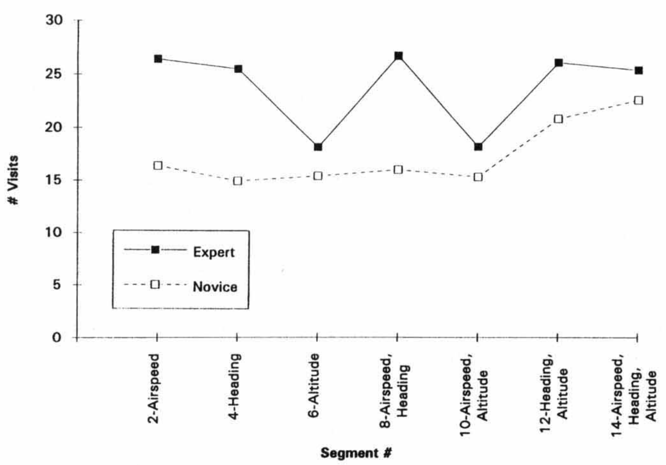


Fig. 8. The average number of visits to the altimeter.

Dwell durations: There was a significant effect of segment on dwell duration for both groups ($F(6,132) = 6.65$; $p < 0.001$). Pilots dwelled the longest on the altimeter during those segments where heading was changing, either alone or with one other parameter (4,8,12). Novices tended to dwell longer on the altimeter than did experts ($F(1,22) = 4.27$; $p < 0.05$).

Vertical Velocity Indicator (VVI)

Number of visits: **Fig. 9** reveals a marginally significant effect of expertise ($F(1,22) = 2.48$; $p < 0.1$) with experts in general visiting the VVI more often than novices. The data also indicate a significant effect of segment ($F(6,132) = 44.02$; $p < 0.001$) wherein pilots most often visited the VVI during those segments when altitude was changing. A significant segment × expertise interaction ($F(1,22) = 11.53$; $p < 0.001$) indicated that novices maintained a fairly constant number of visits during the last three segments (altitude changes in each), whereas experts showed a disproportionately higher number of visits during the two segments in which both altitude and heading were simultaneously changing.

Dwell duration: As with all of the other instruments, novices dwelled significantly longer on the VVI than did the experts ($F(1,22) = 7.63$; $p < 0.01$). The data also show that there is a significant effect of segment ($F(6,132) = 20.54$; $p < 0.001$) whereby pilots dwelled longest on the VVI during those segments where altitude was changing.

Airspeed Indicator (ASI)

Number of visits: **Fig. 10** shows that as with most other instruments, experts visited the ASI significantly more frequently than did novices ($F(1,22) = 9.27$; $p < 0.01$). There was also a significant effect of segment ($F(6,132) = 9.39$; $p < 0.001$) with the highest number of visits occurring during those segments where altitude was changing, either alone or in conjunction with other pa-

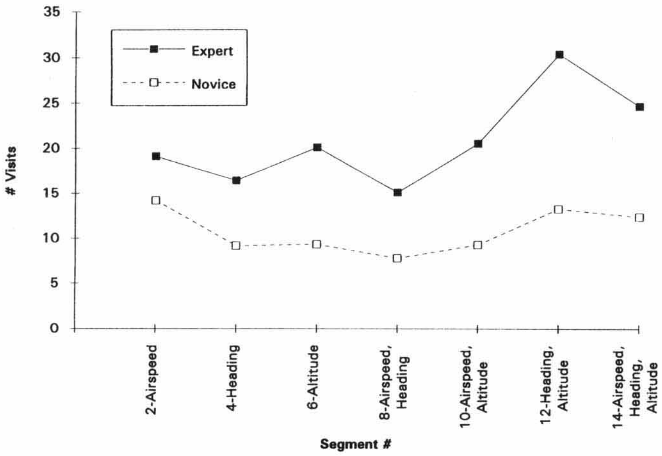


Fig. 10. The average number of visits to the airspeed indicator.

rameters (6,10,12,14). A significant 2-way segment × expertise interaction ($F(6,132) = 3.12$; $p < 0.01$) revealed that the more frequent pilot sampling of the ASI was particularly evident in the two most difficult segments, when heading and altitude were both changing, an effect similar to that seen with the VVI.

Dwell durations: The experimental variables exerted no significant effects on the duration of dwells on the airspeed indicator.

DISCUSSION

The data from the experiment are complex and multifaceted. Nevertheless, some important trends emerged that illuminate the relationship between attention and flight expertise.

Performance Analysis

With regard to the performance data, the results revealed that novices tracked the lateral axis as well as did experts, both groups showing greater error when changes were made on that axis. However, novices suffered in tracking accuracy relative to experts in both vertical (Fig. 4) and longitudinal (airspeed) control, and were particularly vulnerable in certain segments; for vertical control these were the two most difficult dual axis segments (12 and 14), and for longitudinal control these were on segments when heading changes were made.

These performance results may be well explained in the context of resource theory (6,30,40,43). Pilots in both groups supplied sufficient resources to the lateral axis in order to maintain heading parameters within the same acceptable bounds. However, the greater resource demands of this task for the less skilled novices left fewer resources available for them to allocate to the axes of apparently lower priority: the vertical (altitude) and longitudinal (airspeed) axes. Hence, error along these axes showed a clear novice cost. For the vertical axis, this cost was amplified still further during the most difficult, resource demanding segments (12 and 14), in which both lateral and vertical maneuvers were required concurrently. For airspeed tracking, the novice cost was born most severely on those segments during which heading was being changed.

For airspeed tracking, this pattern of effects is an intriguing one which will have parallels in our analysis of scanning data below; novices do as well as experts, or come closer to expert performance on controlling a variable that is changing (i.e., the changing airspeed segments 2, 8, 10, and 14). Where novices seem to fall particularly short is on maintaining, or stabilizing, a variable that is to be held constant. We say then that experts do a better job of "minding the store," when the "action" (i.e., a changing parameter) is elsewhere. We will see this effect revealed in the scan measures.

Why the lateral tracking task is treated as "primary" by the novices, and not allowed to suffer (relative to the experts), is not entirely clear. One plausible hypothesis is that, in normal flight, lateral tracking of the deviation from a flight path is an example of a third-order tracking task.* This is the task of matching the curved flight path with the curved ideal on the map, depicted in Fig. 2 and 3. In contrast, tracking of altitude is second order, and tracking of airspeed is first order. It is well established that higher order tracking tasks are more difficult (17,39), and hence, may intrinsically demand a greater allocation of resources (i.e., be treated as the "primary task" in a dual-task situation).

Control Analysis

The measure of control velocity was used to infer the amount of control effort allocated to an axis (43). Here the pattern of results are consistent with the overall performance effects. First, the amount of lateral control effort was equivalent across the two groups, as was heading error toward which this effort was directed. Second, control effort increased on turn segments, and so did error. Thus, we can infer that experts apparently achieved their more accurate vertical tracking, in part, by allocating more control effort to the vertical axis. However, the added expertise benefit observed in the difficult combined lateral and vertical segments (segments 12 and 14) cannot apparently be attributed to differences in control effort, since this particular interaction did not appear in the vertical control velocity data. The expert benefit for those two segments must instead relate to differences in scanning, to which we now turn. These scanning data provide further insights into differences in strategy between novices and experts.

Scanning Analysis

The most noticeable trend in the scanning data is the observation that experts universally tended to visit instruments more frequently, while novices tended to dwell for a longer time on each, a pattern of behavior that has been observed previously in scanning data (11,23,34), and also in data on the frequency of aviator task switching (29). More interesting still is how the pattern of novice-expert differences in instrument visits informs us as to the strategic differences between the two groups in their control of the three axes.

* The order of a tracking task is formally defined by the number of time integrations between control input and system response. Higher order dynamics are more sluggish, and hence, require greater prediction and mental workload to control.

While both experts and novices attained the same pattern of lateral error and control, the patterns of fixations on the most important (and most visited) of the two lateral instruments, the compass or DG was significantly different between the two groups. Experts not only visited the DG more frequently (Fig. 6), but they also tended to do so particularly and specifically on those segments during which heading was *not* maneuvered. Hence, this is another direct example of better "minding the store."

While this particular attention allocation strategy did not directly benefit the experts in their lateral tracking accuracy, it was presumably an adaptive one that enabled more resources to be freed for dealing with information on the vertical and longitudinal axes and contributed to the expert's improved performance on these axes. To understand this improvement, we consider the scanning data from the viewpoint of three alternative characteristics that have in the past been used to characterize "expertise": automaticity, mental model development, and attentional flexibility.

Automaticity: It is reasonable to assume that experts are simply better pilots than novices, and can extract the necessary information more efficiently. Accordingly, they should show an across-the-board improvement in all aspects of flying, but particularly those that are most difficult. In this regard the most directly supporting evidence is provided by the shorter dwell time of the experts. Information is extracted more efficiently by experts from nearly all instruments, and particularly from the most information-rich instrument, the ADI (Fig. 7). The more rapid extraction of information from the ADI leaves experts with considerably more free time or residual capacity to sample other instruments on the display.

However, a general improvement in information extraction efficiency does not fully account for all aspects of novice-expert differences revealed by the data, since the improvement in performance was, as we noted, selectively targeted to the vertical and longitudinal axes, but not to the lateral axis. Had experts simply become more automated at all aspects of flight, we would have anticipated equal improvements in tracking performance on all three axes. Other aspects of expertise then appear to be revealed in the mental model, and in attentional flexibility.

Mental model development: The pattern of scanning on the vertical and longitudinal axis instruments suggests that there are more subtle differences that reflect greater expert understanding of the contingencies within and between axes (i.e., cross-coupling). This we define as characterizing a more refined mental model of the flight dynamics and it is manifest in three aspects of the scanning data.

First, experts visit the VVI more frequently than novices (Fig. 9), but do so particularly on segments when altitude is changed, in a way that suggests that experts are better tuned to this leading indicator of altitude change. The more frequent expert visits are amplified still further on those two segments (12 and 14) when altitude and heading are both changed, indicating that experts are more aware of the pronounced cross-coupling between these two axes, whereby a bank will lead to a loss of altitude, and that loss can be predicted by an early glance at the VVI (see Fig. 1).

Second, the more frequent visits to the altimeter by the experts (Fig. 8) were particularly enhanced when one of the other parameters (heading or airspeed) was changing; a change that would produce cross-coupling effects on altitude. A more refined mental model that incorporated that cross-coupling should trigger the more frequent visits.

Third, the more frequent expert visits to the airspeed indicator (Fig. 10) were also amplified on those segments in which altitude was changed, again reflecting the greater knowledge of the cross-coupling between the vertical and longitudinal axes, an increased awareness that altitude changes would lead to unwanted airspeed changes.

Finally, we can draw further evidence regarding the role of mental models by noting that more frequent expert visits were not observed to the ADI, that instrument whose change is not directly caused by variations reflected in changes in other instruments. That is, while the ADI is information rich and of high priority, its behavior is fairly straightforward, and most of its variance is directly coupled to control stick activity. Hence, a sophisticated model of cross-coupling is not needed to understand its variance. Instead, both groups equally visited the ADI more frequently than any other instrument (11,21), and did so particularly on the two segments (12 and 14) when both the pitch and roll dimensions of the ADI were conveying information about changing axes (Fig. 7).

Attentional flexibility: As we have noted, the above results also reflect a pattern of scanning whereby experts are more likely than novices to visit an instrument that reflects information on the non-changing axis in a maneuver. We have referred to this pattern as that of "minding the store." By itself, this pattern says little about attentional flexibility, since such a pattern could be revealed if the experts simply maintained a rigid scan pattern across all segments, and failed to diminish their visit frequency when an axis was not changed. We will argue below, however, that this interpretation is unlikely.

The "minding the store" pattern can also be interpreted as reflecting the fact that experts simply had more spare time available away from the ADI, given their greater efficiency in extracting information from that complex instrument. Indeed an estimate of the total time spent looking at the ADI (mean dwell × # visits), reveals that the experts had approximately 24 additional seconds (than novices) during segments 2–10, and 32 additional seconds during segments 12 and 14 to scan other instruments (the latter two segments took about 15 s longer to complete). Hence, one could argue that the experts had more time to sample all other instruments, those depicting a changing axis and those depicting a constant one.

However, as we noted, the differential allocation of this spare scanning capacity (away from the ADI) was not consistent across instruments and segments for the experts, but varied in a way that suggested experts to be more flexible from segment to segment in how they allocated their attention. This flexibility was revealed in three ways. First, the amount of variance, across segments in instrument visits was much greater in experts than novices for three of the instruments. For the altimeter, the experts showed 60% more variance, while for the airspeed indicator and VVI expert variance was 4 times and 6 times that of novices, respectively. These figures indicate greater expert modulation of fixations, as the relative demands on the three axes were changed.

Secondly, as we have noted above, the particular pattern of differential scanning is one that reveals expert scanning to be more driven by a mental model of the cross-coupling of elements, a model that dictates differential viewing, when the axis demands change. Here, for example, we saw the more frequent expert visits on the VVI on those segments in which altitude and heading were simultaneously changing, but more frequent visits to the altimeter on those segments during which heading and airspeed were changing. Novices, in contrast, did not flexibly modulate their scan strategy in this manner.

Finally, a Markov analysis of sequential scanning behavior reflecting the consistency of sequential scans within a segment was carried out. The analysis revealed that the experts demonstrated a less homogeneous scan pattern, particularly on those two most difficult segments, 12 and 14, which required the greatest degree of expertise ($F(6,132) = 2.82$; $p < 0.04$).

In this regard the data are consistent with the general pattern of effects that Gopher and his colleagues have reported: some aspects of expertise lie in greater attentional flexibility (14,18), as well as in the possession of greater reserve capacity (6).

CONCLUSIONS

In conclusion, our eye movement and control velocity data revealed a fairly clear picture of the deployment of perceptual and response resources in simulated flight. Pilots tended to treat their lateral axis as primary, more aggressive (higher velocity) control action reflected more attention to an axis of flight (which partly accounts for why experts were better at altitude control), and more frequent fixations on an instrument were induced by increases in the importance of the instrument as defined implicitly by greater information conveyance about a manuevered axis (rather than a constant one).

More subtle variations of fixation, across instruments and maneuvers, differed between novice and expert pilots in a way that revealed experts to have more automatized skill in extracting information, and a more refined mental model of cross-coupling and leading indicators. This refinement in turn supported more flexible allocation of visual attention across different maneuvers and hence, supported overall better performance (15). Experts also differed from novices in terms of a fixation pattern that guarded against unwanted "tunneling"; they "mind the store" by checking primary indicators of the health of axes that are not explicitly changed.

Such findings have important implications for how novices might be better trained if, indeed, one shortcut to expertise involves the targeted training of expert strategies (2,10). In particular, we believe that better understanding of cross-coupling, of the need to check unchanging axes, and possibly the introduction of enhanced targeted practice on ADI information extraction, might provide cornerstones for more efficient, theory-based, pilot training.

ACKNOWLEDGMENTS

The authors wish to acknowledge the programming skills of Roger Marsh who developed the software for the simulation and eye-movement recording. Thanks also go to Julianne Fox and David Merwin who assisted with scenario development, data collection, and analysis. Drs. Gavan Lintern and Stephanie Doane also contributed to the scientific thinking that provided the foundation for this work. This research was sponsored by a grant from the Office of Naval Research N00014-93-1-0253, and was carried out in collaboration with a parallel program of research at the Navy Aerospace Medical Research Laboratory, Pensacola, FL. Dr. Harold Hawkins was the scientific monitor.

DISCLAIMER

Opinions, interpretations, conclusions, and recommendations are those of the authors and are not necessarily endorsed by the United States Department of the Navy or the United States Department of Defense.

REFERENCES

1. Abernethy B. Visual search in sport and ergonomics. Its relationship to selective attention and performer expertise. Hum Performance 1988; 1:205–35.
2. Acta Psychologica: Special Issue—The learning strategies program: an examination of the strategies in skill acquisition. Donehin E, Fabiani M, Sanders A, eds. Amsterdam, The Netherlands: North Holland, 1989, Vol. 2.
3. Baron S, Levison WH. Display analysis with the optimal control model of the human operator. Hum Factors 1977; 19:437–57.
4. Carbonnell JR, Ward JL, Senders JW. A queuing model of visual sampling: experimental validation. IEEE Trans on Man-Machine Sys 1968; MMS-9:82–7.
5. Curry RE, Kleinman DL, Hoffman WC. A design procedure for control/display systems. Hum Factors 1977; 19:421–36.
6. Damos D. Residual attention as a predictor of pilot performance. Hum Factors 1978; 20:435–40.
7. David JF. The human factor problem in the Canadian forces aviation. In: Shanahan F, Anton DJ, Green RA, Leger A, eds. Aircraft accidents: trends in aerospace medical investigation techniques. Neuilly Sur Seine, France: Advisory Group for Aerospace Research and Development. 1991. AGARD-CP-532, 4-1/4-9.
8. DeMaio J, Parkinson SR, Crosby JV. A reaction time analysis of instrument scanning. Hum Factors 1978; 20:467–71.
9. DeMaio J, Parkinson S, Leshowitz B, et al. Visual scanning: comparisons between student and instructor pilots. Williams AFB, AZ. USAF Human Resources Laboratory. 1976. Technical Report 76-10, AD-A023634.
10. Fabiani M, Buckley J, Gratton G, et al. The training of complex task performance. Acta Psychol 1989; 71:259–99.
11. Fitts PM, Jones RE, Milton JL. Eye movements of aircraft pilots during instrument landing approaches. Aeronautical Engin Rev 1950; 9:24–9.
12. Fracker ML, Wickens CD. Resources, confusions, and compatibility in dual-axis tracking: displays, controls, and dynamics. J Exp Psychol [Hum Percept] 1989; 15:80–96.
13. Gopher D. A selective attention test as predictor of success in flight training. Hum Factors 1982; 24:174–83.
14. Gopher D. Cognition at your finger tips—a cognitive approach to the design of data entry devices. In: Salvendy G, ed. Human computer interaction. The Netherlands: Elsevier Science Publishing, 1992.
15. Gopher D. The skill of attention control: acquisition and execution of attention strategies. In: Meyer DE, Kornblum S, eds. Attention and performance XIV: synergies in experimental psychology, artificial intelligence, and cognitive neuroscience—a silver jubilee. Cambridge, MA: MIT Press. 1993:299–322.
16. Gopher D, Brickner M, Navon D. Different difficulty manipulations interact differently with task emphasis: evidence for multiple resources. J Exp Psychol [Hum Percept] 1982; 8:146–57.
17. Gopher D, Navon D. How is performance limited: testing the notion of central capacity. Acta Psychol 1980; 46:161–80.
18. Gopher D, Weil M, Bareket T. Transfer of skill from a computer game trainer to flight. Hum Factors 1994; 36:387–405.
19. Gopher D, Weil M, Siegel D. Practice under changing priorities: an approach to the training of complex skills. Acta Psychol 1989; 71:147–79.

20. Hameluck DE. Mental models, mental workload, and instrument scanning in flight [dissertation]. York, Ontario: York University; 1990.
21. Harris RL, Christhilf DM. What do pilots see in displays? Proceedings of the 24th Annual Meeting of the Human Factors Society. Santa Monica, CA: Human Factors Society, 1980:22–6.
21. Jones RE, Milton JL, Fitts PM. Eye fixations of aircraft pilots IV: frequency, duration, and sequence of fixations during routine instrument flight. Wright-Patterson AFB, OH, 1950. USAF Technical Report No. 5795.
22. Kahneman D, Ben-Ishai R, Lotan M. Relation of a test of attention to road accidents. J Appl Psych 1973; 58:113–5.
23. Kramer A, Tham M, Konrad C, et al. Instrument scan and pilot expertise. Proceedings of the 38th Annual Meeting of the Human Factors Society. Santa Monica, CA: Human Factors and Ergonomics Society, 1994:36–40.
24. Kundel HL, La Follette PS. Visual search patterns and experience with radiological images. Radiology 1972; 103:523–8.
25. Moray N. Monitoring behavior and supervisory control. In: Boff KR, Kaufman L, Thomas JP, eds. Handbook of perception and performance, Vol. II (pp. 40-1/40-51). New York: Wiley & Sons, 1986.
26. Moray N, Rotenberg I. Fault management in process control: eye movements and action. Ergonomics 1989; 32:1319–42.
27. Mourant RR, Rockwell TH. Strategies of visual search by novice and experienced drivers. Hum Factors 1972; 14:325–36.
28. Navon D, Gopher D, Chillag N, Spitz G. On separability of and interference between tracking dimensions in dual-axis tracking. Israel Institute of Technology Research Centre for Work Safety and Human Engineering Report (HEIS-82-9). Haifa, Israel: Technion, 1982.
29. Raby M, Wickens CD. Strategic workload management and decision biases in aviation. Intl J Aviat Psych 1994; 4:211–40.
30. Schneider W, Detweiler M. The role of practice in dual-task performance: toward workload modeling in a connectionist/control architecture. Hum Factors 1988; 30:539–66.
31. Senders JW. The human operator as a monitor and controller of multidegree of freedom systems. IEEE Trans on Hum Factors Electron, HFE-5, 2–6.
32. Spady A, Harris R. Summary of NASA Langley's pilot scan behavior research. Proceedings of the Society of Automotive Engineers Conference. Warrendale, PA: Society of Automotive Engineers, 1983.
33. Stark L, Ellis S. Statistical dependencies in visual scanning. Hum Factors 1986; 32:421–38.
34. Stein ES. Air traffic control visual scanning (DOT/FAA/CT-TN 92/16). Atlantic City International Airport, NJ: U.S. Department of Transportation, Federal Aviation Administration, 1992.
35. Stager P, Hameluck D. Ergonomics in air traffic control. Ergonomics 1990; 33:493–9.
36. Tole JR, Stephens AT, Harris RL, Ephrath AR. Visual scanning behavior and mental workload in aircraft pilots. Aviat Space Environ Med 1982; 53:54–61.
37. Vanderbeek RD. Human factors causes and management strategies in U.S. Air Force F-16 mishaps 1984–present. In: Shanahan F, Anton DJ, Green R, Leger A, eds. Aircraft accidents: trends in aerospace medical investigation techniques. Neuilly Sur Seine, France: Advisory Group for Aerospace Research and Development. AGARD-CP-532, 1992:22-1/22-5.
38. Wickens CD. The effects of divided attention on information processing in tracking. J Exp Psychol [Hum Percept] 1976; 2:1–13.
39. Wickens CD. The effects of control dynamics on performance. In: Boff K, Kaufman L, Thomas J, eds. Handbook of perception and performance, Vol. 2. New York: Wiley and Sons, 1986:39-1/39-60.
40. Wickens CD. Processing resources and attention. In: Damos D, ed. Multiple task performance. London: Taylor & Francis, 1991.
41. Wickens CD. Engineering psychology and human performance, 2nd ed. New York: Harper Collins, 1992.
42. Wickens CD, Bellenkes AH, Kramer AF. Expertise and attention control in pilot visual scanning (UIUC-BI-HPP-95-5). University of Illinois Beckman Institute of Advanced Science and Technology Technical Report. Urbana, IL: Beckman Institute, 1995.
43. Wickens CD, Gopher D. Control theory measures of tracking as indices of attention allocation strategies. Hum Factors 1977; 19:349–65.
44. Wiener EL. Controlled flight into terrain accidents: system-induced errors. Hum Factors 1977; 19:171–81.