Training Simulation for Helicopter Navigation by Characterizing Visual Scan Patterns

Sullivan, Joseph

Aviation, Space, and Environmental Medicine, Vol. 82, No. 9, September 2011
http://hdl.handle.net/10945/41765

Downloaded from NPS Archive: Calhoun
Training Simulation for Helicopter Navigation by Characterizing Visual Scan Patterns

Joseph Sullivan, Ji Hyun Yang, Michael Day, and Quinn Kennedy

**Introduction:** Helicopter overland navigation is a cognitively complex task that requires continuous monitoring of system and environment parameters and years of training. This study investigated potential improvements to training simulation by analyzing the influences of flight expertise on visual scan patterns. **Methods:** There were 12 military officers who varied in flight expertise as defined by total flight hours who participated in overland navigation tasks. Their gaze parameters were tracked via two eye-tracking systems while subjects were looking at out-the-window (OTW) and topographic Map views in a fixed-based helicopter simulator. **Results:** Flight performance measures were not predicted by the expertise level of pilots. However, gaze parameters and scan management skills were predicted by the expertise level. For every additional 1000 flight hours, on average, the model predicted the median dwell will decrease 28 ms and the number of view changes will increase 33 times. However, more experienced pilots scanned more OTW than novice pilots, which was contrary to our expectation. A visualization tool (FEST: Flight and Eye Scan visualization Tool) to replay navigation tasks and corresponding gaze data was developed. Qualitative analysis from FEST revealed visual scan patterns of expert pilots not only looking ahead on the map, but also revisiting areas on the map they just flew over to retain confidence in their orientation. **Discussion:** Based on the analysis provided above, this work demonstrates that neurophysiological markers, such as eye movements, can be used to indicate the aspects of a trainee’s cognitive state that are useful for cueing an instructional system. **Keywords:** expertise, gaze, terrain association, cognition.

**HELCOPTER OVERLAND** navigation is a complex task that normally is assigned to the nonflying pilot, who is responsible for providing verbal instructions to the flying pilot to reach navigation checkpoints. As described in Wright (23), navigation is never the sole aim of a mission. In addition to navigation, pilots are required to complete a higher level task such as logistics support, intelligence, surveillance and reconnaissance, or combat search and rescue. The nonflying pilot has additional responsibilities, including terrain and obstacle avoidance, monitoring and managing engine and system performance, and communications. Consequently, skilled pilots oftentimes stay off-track intentionally to stay oriented or to achieve a higher level task. In this particular case, a high root mean square (RMS) error of flight trajectory, a commonly used performance measure, does not necessarily indicate a pilot’s poor performance. Rather, it is a result of a pilot’s ability to monitor key environmental parameters. This ability is accrued through extensive helicopter overland navigation training.

During helicopter overland navigation training, instructors often face the dilemma of knowing when and how to provide the best feedback to the student. Busy flying in a challenging environment, avoiding terrain and obstacles, the instructor has few opportunities to understand what mistakes were made by the student, when the mistakes were made, and how to explain those mistakes to the student to provide an opportunity to learn from them, all within a few minutes of the mistake being made. When a student deviates from the planned course, the instructor is missing critical information to maximize the learning potential of the real training opportunity, that of the student’s cognitive state.

In order to provide the best feedback to the student in the right form at the right time, the instructor pilot needs to know the student’s cognitive state. Currently, observing the student’s cognitive state is guesswork and there are few real-time cues available for the instructor to assess how the student monitors key parameters, perceives flight routes, and implements navigation strategies. We investigated whether visual scan differences between expert and novice pilots in helicopter navigation tasks could act as a useful cue to indicate student’s cognitive states and aid navigation instruction. If visual scan measures provide us some estimates on students’ cognitive states and if there exist differences in scan patterns between different expertise levels, our study suggests that providing real-time eye scan data can be a potential method to enhance helicopter navigation training.

A common goal in training research is to train novices to behave and think like experts (11). It is known that pilots with more flight experience consistently perform better on aspects of flight control than those with less experience (1,10,20). When making aviation-related decisions, experts described more elaborate scenario problems than novices, indicating greater depth of understanding of the issues (14). As stated above, helicopter
overland navigation is complex and entails additional demanding tasks above and beyond flight control. Thus, a common flight performance measure, such as RMS error of flight trajectory, may not predict expertise levels as it does in other aviation tasks. Additionally, from a training perspective, it is not clear which cognitive strategies improve with expertise level. The improvement is considered to be accrued with practice, but the learning mechanism is yet to be uncovered. In order to better explain why pilots’ performance differ by expertise level and to find cues for assessing their cognitive states, we suggest observing human behaviors (e.g., where they look) which influence their performance (e.g., how they navigate).

Recent eye-tracking technology provides nonintrusive devices to collect ocular data (4) and this technology can be applied to real operational environments as well as laboratory settings. Common visual scan measures collected from eye tracking are saccades (rapid movements of eyes), dwell duration (also called fixation duration and is the interval between two successive saccades), and blink rate. For example, Marshall (13) modeled different cognitive states, i.e., engaged vs. relaxed, normal vs. distracted, and fatigued vs. alert, from eye movement and pupil size; Van Orden et al. (21) showed eye-tracking measurements can be used to determine level of visual processing load. Using eye scan behavior to detect expertise differences has been successfully used in several domains, such as video gaming and driving (15,18). In the aviation domain, pilots exhibit different visual scanning patterns during various phases of flying under instrument flight rules, with the shortest dwell times during the landing (1,9).

Scanning differences between novice and expert pilots occur, and experts’ scanning patterns are correlated with better performance as measured by reduced flight path error on all axes and faster reaction times (1,7,8). In pilot decision-making, experts had longer dwell times during the landing (1,9). Scanning differences between novice and expert pilots occur, and experts’ scanning patterns are correlated with better performance as measured by reduced flight path error on all axes and faster reaction times (1,7,8). In pilot decision-making, experts had longer dwell times during the landing (1,9).

The previously mentioned studies did not investigate expertise and visual scan differences in helicopter terrain navigation tasks, which are considered to be more cognitively demanding and continuously complex than fixed wing aircraft operating tasks. Learning to fly helicopters is generally considered to be the most challenging and difficult aviation task (16). We focused on improving our understanding of cognitive processing associated with helicopter terrain navigation by analyzing visual scanning differences between experts and novices, with the goal of enhancing helicopter pilot training methods. We have designed overland navigation tasks in a flight simulator integrated with eye-tracking systems and performed human-in-the-loop experiments with pilots with various flight hours. The simulated navigation tasks entailed ‘flying’ to 12 waypoints depicted on a map.

We made the following hypotheses for helicopter overland navigation tasks:

1. Flight expertise is not associated with flight performance, such that RMS error and flight log time (FLT) error should not be predicted by flight expertise.
2. Flight expertise is associated with visual scan parameters, such that more expert pilots will have shorter dwell duration, more OTW-Map view changes, less number of fixations per view, and larger OTW scanning percentage.

As part of this project we developed a visualization tool, Flight and Eye Scan visualization Tool (FEST), designed to provide a representation of spatial and temporal correspondence among features scanned in OTW (3D) and Map (2D) views in relation to the actual aircraft location. FEST can be used as a teaching or debriefing tool as well as visual scan pattern visualization device.

METHODS

Subjects

There were 19 male military personnel, 29 to 40 yr of age, who participated in the study. For four subjects, substantial data were missing due to eye-tracking device calibration issues. Out of the remaining 15 subjects with complete data, 12 flew up to waypoint 5 (out of 12 waypoints) without experimenter intervention. Therefore, these 12 subjects were used in the analyses and this study only analyzed the experimental data between waypoint 2 and waypoint 5. By limiting data analysis to these waypoints, we hoped to remove confounding factors that can be generated from different route difficulties. Minimum skill requirement for the study was completion of at least one overland navigation class. Among the 12 subjects, 3 subjects were helicopter flight instructors and 2 subjects had other navigation-related instructing experiences. Expertise was defined by the total flight hours (TFH), in which higher TFH values indicate increased expertise of the pilot. TFH varied from 0 to 3100 h (avg = 1488 h, SD = 1104 h) and overland flight hours varied from 0 to 2500 h (avg = 612 h, SD = 853 h). Eight subjects are from the U.S. Navy, three from the U.S. Marine Corp, and the rest are from the U.S. Air Force, U.S. Army, and Brazilian Navy. No special neurological, visual acuity, or spatial ability tests were performed. The study was approved by the Naval Postgraduate School Institutional Review Board. Subjects were recruited from e-mail advertisement through Naval Postgraduate School e-mail account holders. All the subjects were given written informed consent to participate, with the right to withdraw at any time.

Equipment

The basic elements of the apparatus included the flight simulator X-Plane 8.6, a 43” by 24” screen to present the OTW view, a 33” by 33” display for the Map and instrument display, two stereo cameras and associated faceLAB 4.6 software for collecting eye data, and cockpit-style seat with sided mounted joystick. We had two separate faceLAB systems (two sets of stereo cameras with 12.5-mm lenses, three infrared strobe lights) for tracking eye gaze for OTW and Map displays. Data from X-Plane 8.6 and faceLAB were sent to an image generator in
60 Hz, which provided an OTW and a Map view combining an OpenSceneGraph terrain model of Twenty-nine Palms, CA.

The helicopter was designed to be on an automated terrain-following mode at fixed 150 ft (45.7 m) above ground level flying at 60 kn. However, the pilot was able to control the heading of the aircraft using the lateral control of the joystick. The joystick pitch control (up/down) was programmed to change the up/down view of the OTW, not the actual pitch angle of the aircraft. The Map display presented a 1:50,000 topographical land map typically used for flight planning and execution. The map was fixed in position about the pair-wise mean of the waypoints, whereas the orientation of the map was synchronized to the aircraft’s heading to maintain a track-up orientation. The bottom portion of the screen contained instruments to support a navigation task: the left-most instrument display was a compass typical of legacy Navy H-60 (SH/HH-60F/H) displays. To the right of the compass display were typical barometric and radar altimeters. The rightmost portion of the instrument cluster contained a digital-style elapsed time clock.

Navigation Task

The navigation task was to fly over 12 waypoints (indicated as black circles on Fig. 1) after studying the area using the Falcon View flight planning software system widely employed by diverse communities within DoD. The first waypoint was located slightly south of the map, so it is not shown in the figure. Each waypoint pair has a “doghouse” that indicates (from top to bottom): the next waypoint number, the recommended heading to reach that waypoint from the previous one, the distance between waypoints, and the amount of time it takes to traverse the distance assuming a speed of about 60 kn. The task was created to be challenging. Waypoints were very close together and the terrain tended to be ambiguous, so subjects needed to make course corrections based on visual cues from both the OTW and Map screens (their goal being to bring their perceived location closer to their actual location).

Procedure

After a brief introduction to our study, subjects were asked to read an informed consent form and sign the form before the experiment started. Then we asked them to fill out a demographic/background questionnaire which included questions regarding their flight time and skill sets. The next step was a calibration of faceLAB stereo cameras to verify that the visual scan data was usable (error less than 3°) before subjects started the navigation tasks. Subjects were asked to sit in the simulator chair where eye-tracking cameras had been mounted in between the chair and the simulator screen. Once the calibration was done, the simulated flight environment was explained to the subjects (e.g., altitude and speed maintained by autopilot, forward/backward movement of the flight stick controlled the view of the helicopter, the digital map stayed oriented automatically, etc.) and then they flew a practice route. The practice run took about 7 to 8 min, giving subjects enough time to become familiar with the simulated environment and the simulator itself.

Following the calibration phase and equipment familiarization navigation route exercise, subjects were briefed on the main navigation route (Cleghorn-West, Fig. 1) for up to 20 min. After the briefing, subjects were directed back to the flight simulator and evaluators re-verified calibration. Subjects then flew the main route (6 min long) while evaluators collected eye-scan data and flight information. If a subject got too far off course, the experimenter would verbally intervene, giving them a course to guide them back to a waypoint. Subjects were debriefed upon task completion and total experiment time varied from 1 to 1.5 h.

Statistical Analyses

We used linear regression in which the predictor variable was TFH to see if expertise can predict flight performance and/or visual scan characteristics. A similar pattern of results was found when we used nonparametric regression, specifically rank regression. We decided to provide linear regression analysis in this paper because the results from linear regression are more intuitive. The main outcome measures for the flight performance were: 1) RMS error of the flight trajectory; and 2) FLT error, i.e., actual flight duration minus ideal flight duration. The RMS error was defined as a sum of Euclidian distance error between the shortest path and actual flight path, i.e.:
where $x_a$ = actual flight position, $x_o$ = corresponding reference trajectory point, $n$ = number of data points between waypoint 2 and waypoint 5, $t_0$ = initial time, and $t_f$ = terminal time.

The main outcome measures for visual scan patterns were: 1) median and standard deviation of dwell duration, 2) OTW and Map scan time ratio, 3) number of OTW-Map view changes, and 4) number of fixation points. Dwell duration (or the duration of fixations) is calculated as a period between consecutive saccades (12). Because the navigation tasks had two different views (OTW and Map), the variables, OTW and Map scan time ratio and number of OTW-Map view changes, were included to explain how many features pilots scanned per view. All variables were transformed into z-scores. Data from faceLAB, X-plane, and the image generator were combined into a text file and all measurements were not. Spearman’s rank correlation coefficient between the dependent variables indicated that RMS and FLT error were positively correlated with each other ($\rho = 0.664$, $P = 0.022$). However, neither of the flight performance variables was correlated with any of the scan parameters.

Regression analysis indicated that TFH was not a significant predictor of either of the flight performance measures, RMS and FLT error. However, TFH was associated with several scan parameters. First, TFH was associated with shorter overall and OTW median dwell durations. Thus, when the coefficients are converted from z-scores to raw units, we found that for every additional 1000 flight hours, on average, the model predicted that median of overall and OTW dwell will decrease 28 ms and 26 ms, respectively ($b = -0.649$, $SE(b) = 0.241$, $P = 0.002$; and $b = -0.693$, $SE(b) = 0.228$, $P = 0.001$ in z-scores, respectively). TFH also was associated with more OTW-Map view changes, in which more experienced pilots changed views between OTW and Map more frequently (33 times more with each additional 1000 flight hours of experience, i.e., $b = 0.602$, $SE(b) = 0.252$, $P = 0.004$ in z-scores). Table III summarized our linear regression model, showing only statistically significant findings. Post hoc statistical power analysis (2) showed that the power of our linear model ranges between 0.67 and 0.87 given sample size $= 12$, $\alpha = 0.05$, and observed $R^2 = 0.35 - 0.48$.

FEST was developed and it provided a virtual playback of gaze and flight data any time in the navigation by showing an over-the-shoulder view of the experimental setup, i.e., synchronized OTW display, Map display, and virtual human gaze (see Fig. 2). The pilot’s gaze direction is indicated by two vectors originating at the operator’s eyes. The replay provides typical video type controls: rewind, pause, stop, and fast forward. In the tool, OTW gaze is displayed in both the OTW and Map display in green dots where fixation duration is linked to the size of the green dots. For example, a big green dot in the FEST tool represents long OTW fixation. Similarly, Map gaze is displayed in red dots, corresponding helicopter position as a blue cross, and waypoints as black numbers. FEST provided valuable insight and inspired additional metrics for in-depth evaluation of the trainee’s cognitive state.

**DISCUSSION**

The statistical results showed that for an overland helicopter navigation task, many gaze parameters were associated with flight expertise, whereas flight performance measurements were not. Flight performance variables, RMS and FLT error, were not correlated with any scan.
parameters, although many of the visual scan parameters were correlated with each other. Together, these results support our hypothesis that measuring only flight performance may not be an informative indicator of underlying cognitive processes used by experts.

This result supports our first hypothesis, i.e., for a helicopter overland navigation task, flight expertise is not associated with flight performance such as RMS and FLT error. We found some support for the second hypothesis. The only result that did not support our hypothesis was the percentage of time spent scanning OTW. We expected that experts would spend more time scanning OTW than novices do. Analysis indicated the opposite pattern. That is, experts on average spent proportionally less time looking out the window. Thus, for every additional 1000 flight hours, on average, the model predicts that OTW scanning time will decrease 5%.

Correlation between gaze parameters could be originated from physical constraints among them. For example, more OTW-Map view changes in expert pilots should provide shorter scan time allocated to each view, which can result in fewer number of fixations per view, although expert’s dwell durations were shorter than novices. To explore why this pattern occurred, we conducted post hoc analyses in which we looked at the overall time participants spent in dwell during their simulated flight task. Pilots with more TFH spent a smaller proportion of their flight time in dwell, i.e., for each additional 1000 flight hours, on average, pilots spent 2.9% less of their flight time in dwell (b = −0.633, SE(b) = 0.245, P = 0.027 in z-scores). This could be associated with more time spent in saccade and/or transition between views. On the other hand, expert pilots had better ideas about where they should look during

TABLE III. RELATIONSHIP OF FLIGHT HOURS (TFH) ON SCAN CHARACTERISTICS.

<table>
<thead>
<tr>
<th>Variables</th>
<th>TFH</th>
<th>F</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic dwell characteristics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median dwell duration</td>
<td>−0.649 (0.241)*</td>
<td>7.25*</td>
<td>0.420</td>
</tr>
<tr>
<td>Median OTW dwell duration</td>
<td>−0.693 (0.228)*</td>
<td>9.23*</td>
<td>0.480</td>
</tr>
<tr>
<td>Scan management characteristics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of OTW-Map view changes</td>
<td>0.602 (0.252)*</td>
<td>5.69*</td>
<td>0.362</td>
</tr>
<tr>
<td>Number of fixations per OTW view</td>
<td>−0.534 (0.267)†</td>
<td>3.99†</td>
<td>0.285</td>
</tr>
<tr>
<td>OTW scanning percentage</td>
<td>−0.592 (0.255)*</td>
<td>5.40*</td>
<td>0.350</td>
</tr>
</tbody>
</table>

OTW = out the window; TFH = total flight hours.

Summary of linear regression model results [parameter estimates shown in parentheses (SE), whereas intercept coefficients are not shown in the table since they are all close to 0]. All variables shown in the table are standardized in z-score and only statistically significant results were presented.

1 P < 0.10, *P < 0.05.
navigation and that ability could result in fewer fixations per view. Previous studies (e.g., 8) reported that expert pilots scan more features than novices in a given amount of time in different aviation environments.

The initial version of FEST (Fig. 3) depicted representative scan patterns between pilots with different expertise level, expert and novice, respectively. The novice scan data depicts less coherent scan patterns and little overlap between OTW (dots) and Map (square) scans. Although this depiction was extremely helpful for exploring differences in correlation between spatial and temporal Map and OTW eye scan data, it does not include the field of view information.

Revision of FEST provided insight into how expert scan patterns are organized. Based on the cognitive task analysis (3) and subject matter expert insights from previous studies (5,19), experts would be expected to consistently scan the map well ahead of the aircraft’s position. This pattern was demonstrated frequently by expert performers. Visual analysis through FEST provided insight into performance and also revealed patterns that were not expected. At least one expert followed a very consistent pattern that involved spending substantial time scanning the map behind the aircraft’s current position. This pattern can be seen clearly in Fig. 4.

The squares immediately behind the aircraft position represent recent map scans. This pattern is repeated consistently from waypoint 1 through 6. Based on visualization of scan and debrief comments, this pattern appears to be a deliberate and helpful strategy as noted from a description of navigation past waypoint 6. Almost immediately after making the left turn at waypoint 6, the subject perceived that he had made the turn early and was south of the course. He was, in fact, only slightly north of the course. He immediately turned right nearly 90° to correct. Soon after making the turn, he entered a valley perpendicular to the flight path and wider than expected. He turned left to follow the valley with the intent of disambiguating between the two east-west oriented valleys. Shortly after making this turn he identified a unique terrain formation in the floor of the valley. This terrain formation was only present in the northernmost valley. On his next map scan, he located the formation on the map. Correlating this key feature allowed the subject to recover from his perceived error and continue successful navigation. With only raw performance data...
Fig. 4. Subject map scan behind as well as ahead of current position. Black line with circles shows intended route, thin dashed line indicates overall flight trajectory, thick black link indicates current aircraft position, dots represents OTW gaze, and squares are map scan points.

such as RMS error or timing data, an observer would have no way of knowing if the course deviations were intentional. Without scan data, there would be no way at any point in that evolution to know if recovery was likely and if effective navigation and learning were taking place.

Even after the subject had reoriented himself and was successfully navigating along course, his tendency was to focus back to the area of uncertainty. In at least one case, an expert appeared to dedicate any additional mental resources to map study in an attempt to solve the puzzle of how the mismatch might have occurred. This observation tends to support the earlier assertion that the division of scan between the OTW and Map view was impacted by a floor effect. It seems highly likely that if the navigation or other combined tasks were beyond moderately challenging, individuals would have dedicated less time to studying the map. While this analysis is highly anecdotal and difficult to substantiate, post-event debrief comments strongly support this premise. The deliberate strategy of checking the map immediately behind the presumed aircraft position may have played a role in the subject’s ability to recover. From the experts’ debrief comments, it appears that they tend to maintain multiple possible navigation solutions simultaneously. They continually challenge these solutions against the widest set of evidence they can gather, including verifying the terrain they recently covered does in fact correspond to terrain represented on the map. Checking recently covered terrain against the map depiction could provide an opportunity for hypothesis confirmation and alternative hypothesis generation. In the example above, the subject’s debrief comments indicated that when faced with unexpected terrain features in view, his alternative possible location was based on a map scan in the vicinity of recently covered terrain. Although the subject was mistaken in the assumption that he was off course, he could not have mismatched terrain in view with a plausible alternative location if he had not been comparing terrain covered against multiple possible locations on the map. Without this robust hypothesis generation and testing, including terrain recently covered, it would not have been possible for the subject to identify the point at which he thought he had made the turn early. Of the expert navigators in this study, two made minor errors and corrected themselves; each described similar structures for recovery in debrief comments.

In summary, we have designed a helicopter overland navigation task and collected eye-tracking data in a flight simulator. Based on the analysis provided above, this work demonstrates that neurophysiological markers can be used to indicate aspects of a trainee’s cognitive state that are useful for cueing an instructional system. For the complex cognitive task of helicopter overland navigation via terrain association, two types of visual scan measures emerged as informative predictors: basic scan parameters and high-level scan management measures. According to the degree of consciousness and voluntary gaze control, gaze parameters are grouped into two different categories, i.e., “basic dwell characteristics” and “scan management characteristics.” For example, OTW-Map view change is considered to be a scan management variable because pilots initiate a view change with an intention. On the other hand, dwell duration is considered to be basic due to the lack of consciousness in the eye movement. Results show that both intentional and unintentional aspects of gaze measurements are affected by pilots’ expertise level. Studying how these different degrees of gaze automaticity shape trainees’ cognitive states will help provide a specific navigation training guide. In addition, visualization of scan pattern is useful for informing instructors of a trainee’s strategy and can reveal unexpected strategies of experts.

The statistical results supported our hypotheses except for the OTW scanning time, in which TFH was associated with less OTW scanning time. One possible explanation on why experts spent less time on OTW is that the navigation task resulted in a floor effect. It is feasible that, particularly for experts, the task was not overly taxing. This is likely considering that navigation is normally one of several tasks being conducted concurrently. In the absence of additional tasks commensurate with real-world conditions, experts may have spent much more time reviewing the map than they normally would. Given that navigation was expressly briefed as the point of the study and several expert pilots completed the route with relative ease, this seems highly probable. This possibility is supported by the qualitative visual analysis of scan patterns from FEST.

Scan patterns can be augmented by adding more neurophysiological markers such as EEG, EMG, or GSR into
our experiment. Having rich observation data, we should be able to better understand cognitive states associated with overland navigation tasks. Modeling cognitive states using Hidden Markov Models (6) or Bayesian Network (24) will follow. The implications of the results are that they may provide important information about the strategies that pilots use to be situationally aware. Additionally, the techniques developed here could be used to examine the effects of practice and of different helicopter technology on pilot behavior.

ACKNOWLEDGMENTS
This work is funded by Naval Modeling Simulation Office (NMSO) and Office of Naval Research (ONR). Dr. Sullivan appreciates his Thesis Committee members for providing us insightful advice to the project and paper. Prof. Ron Fricker reviewed the statistical analysis in this paper and we are very thankful. We are grateful to Noah Llyod-Edelman for helping us in calibrating the experimental device.

Authors and affiliation: Joseph A. Sullivan, Ph.D., Ji Hyun Yang, Ph.D., Michael Day, B.S., and Quinn Kennedy, Ph.D., Naval Postgraduate School, Monterey, CA.

REFERENCES