Presented at the 11th International Symposium on Aviation Psychology.
Columbus, OH: The Ohio State University. 2001.

PILOT VISUAL WORKLOAD AND TASK MANAGEMENT IN FREEFLIGHT: A MODEL OF VISUAL SCANNING

Christopher D. Wickens, Xidong Xu, John Helleberg, and Roger Marsh
University of Illinois, Aviation Human Factors Division
Savoy, Illinois

ABSTRACT

Seventeen instrument rated pilots flew a high fidelity simulator in two experiments designed to assess changes in visual attention allocation strategies brought about by self separation (free flight) responsibilities. Seven pilots maneuvered to avoid traffic, using a cockpit display of traffic information (CDTI). Ten pilots flew the same maneuvers in a "baseline" condition when identical maneuvers were commanded by ATC. Visual scanning data were recorded between three areas of interest: an outside view (135 degree projection upon which traffic could be depicted), the instrument panel, and the CDTI. Occasional traffic was presented in the outside view that was not announced on the CDTI. Visual scanning data were fitted by an optimal sampling model, revealing the influences of task priority (aviate > navigate), and channel bandwidth. The model accounted for 79% of the variance in visual attention allocation between conditions. Additional evidence was obtained for the influence of scanning effort and event salience. The CDTI required 25% of the pilot's visual resources diverted from scanning the outside world, but this diversion dud not penalized detection of the unannounced traffic.

INTRODUCTION

A key concept underlying the study of cockpit task management is that of **task priority**. For example most pilots are taught the intrinsic "aviate navigate communicate systems-management" (ANCS) priority hierarchy (Schutte & Trujillo, 1996); however ample evidence may be supplied for incidents in which this hierarchy has been violated, for example, as pilots become engaged in map study while failing to fly the aircraft (navigation dominates aviation), or fault diagnosis while failing to navigate safely (Strauch, 1997). The issue of task prioritization concerns the extent to which pilots emphasize more important tasks, while de-emphasizing to a greater extent (or delaying) performance of those of lower priority.

There are numerous ways of examining pilot task prioritization. For example, one can study the delays, degradations or total failures to perform critical tasks (Chou, Madhavan, & Funk, 1996; Raby & Wickens, 1994; Laudeman & Palmer, 1995; Dornheim, 2000). One can also study the extent to which certain tasks are **interrupted** by others (Latorella, 1998; Damos, 1997). Our emphasis in the current paper will be on the role of visual scanning to reflect the pilots' allocation of attention, and therefore their task management strategies and task prioritization schemes. In this regard, measures of visual scanning have both strengths and weaknesses, which enable them to complement, rather than replace other measures of task management. On the positive side, scanning measures allow the scientist to infer the degree of importance (or neglect) of tasks like monitoring or maintaining situation awareness that may have few overt behavioral manifestations. Thus for example a long period of time during which an altimeter is never viewed, is fairly strong evidence for neglect of vertical control, in spite of the fact that this failure may not lead to direct altitude deviations.

On the negative side, one must realize that it is possible to attend to regions that are not in foveal vision (the key diagnostic measure of visual scanning analysis), just as it is possible to fixate on things that are not "attended", as attention is allocated instead to cognitive activity or to auditory information. Finally, in many applied environments, there is a many-to-one mapping of fixation regions to "tasks". For example the "task" of traffic monitoring is associated with scanning across the entire outside world (and cockpit traffic display, if such is available). This latter problem can be partially addressed by defining "areas of interest" (AOIs) to be of relatively large extent, and is also addressed through modeling, as we do in this paper.

We present below a model of how task prioritization affects the allocation of visual attention, as reflected by visual scanning to various AOIs, where visual information is acquired. The model contains both prescriptive components of how scanning **should** be influenced, based upon expected value modeling (Carbonnell, Ward, & Senders, 1968), as well as descriptive components of two negative influences that may counteract optimal scanning behavior. This model is then evaluated against data collected in two experiments examining pilot performance and scanning in a free flight scenario.

The model contains four basic components. The first two components are prescriptive regarding how scanning should be done, and are topdown components, indicating the influence of knowledge driven factors in allocating visual attention. Component 1 is the **bandwidth** (BW) of an area of interest, and is reflected in Sender's (1964) original work on instrument scanning. AOIs that contain more frequently or rapidly changing events should be sampled more frequently (Senders, 1964; Moray, 1986). Component 2 define the **value** (V) of accessing information at an AOI (or correspondingly the cost of missing such information). Value can be assessed either in terms of the benefits of detecting events on an AOI, or in terms of the relative **priority** of performing tasks, whose information is represented in the AOI in question. Note that multiplying the influence of bandwidth and value will yield a measure of the **expected value** (EV) of sampling an AOI. Thus, in theory, the combined influence of the two top down factors of value and bandwidth can define a prescriptive value of "how often a supervisor should sample" (Sheridan, 1972).

(3) The **salience** (S) of an information source is a bottom up, rather than top down influence, and describes the extent to which events on the AOI in question are salient, and therefore attract attention when they occur (Yantis, 1993). (4) The **information access effort** (E) of an AOI characterizes the cognitive and physical effort necessary to switch attention from one AOI to another (Gray & Fu, 2001). This could involve scanning, head movement (Wickens, 1993), or even additional keyboard interaction (Wickens, 2000). Note that, unlike value, bandwidth and salience, effort is not only the unique property of a single AOI, but rather, can also be the property of a pair of AOIs, and so influences fixation transitions (Ellis & Stark, 1986).

It should also be noted that neither salience nor effort should influence optimal scanning, and indeed, both may disrupt those optimal patterns. However designers can, and often do capitalize upon both influences to reinforce the influences of the two expected value parameters. For example, in a well designed cockpit, events known by the designer to be of value (important) are often coupled with salient alerts or highlights (thus compensating for the rarity or low bandwidth of those events). Correspondingly, information sources that are frequently used and/or are related, in good design are generally placed close together, to minimize the effort of their mutual access (Wickens, Vincow, Schopper, & Lincoln, 1997; Wickens & Carswell, 1995).

While several studies have examined how scanning is influenced by some of these components in isolation (e.g., Senders, 1964; Gray & Fu, 2001; Ellis & Stark, 1986; Bellenkes, Wickens, & Kramer, 1997), few have examined them in combination, and fewer still have done so in the context of the aircraft cockpit (Carbonnell et al., 1968, Mumaw et al., 2000).

The model of scanning that we evaluated can be represented in terms of two equations, one descriptive (how scanning **is** carried out), and the other prescriptive, (regarding the optimal influence of expected value).

1. Descriptive: $P(AOI_j) =$

$$(bBW_j)\ (vV_j) + sS_j - eE(k[n-1] - jn)$$

That is, the probability that the jth AOI will be fixated is determined by the weighted sum of the positive influences of bandwidth and value (the expected value term) and salience as "attractor" forces, each modified by their weight, and , inhibited by the effort required to access AOI_j, given that the eye was previously fixated on AOI_k. It is easy to postulate additional influences on such a model. In particular, concurrent task workload could have a substantial influence on the weighting of effort. High workload would appear to enhance the negative contributions of the effort parameter, to minimize long distance scans. Correspondingly, the model makes no distinction between objective values and bandwidths, and those that are subjectively represented in the pilots' mental model through experience (and may therefore be biased).

$$\text{(2) Prescriptive: } P(AOIj) = \sum_{t=1..n} (bB_j)(rR_j \backslash T_n)(VT_n)$$

Model 2 states that the probability of fixating on an AOI **should be** determined by the bandwidth (B) of the AOI in question (here subscripted j), multiplied by the relevance of information contained within that AOI to performance of task n, multiplied by the value of performing task n, as this value might, for example, be established in an "ANCS" task hierarchy. This product is summed overall tasks. The applicability of model 2 to scanning behavior in the experiments will be described following a brief presentation of the experimental methodology.

EXPERIMENTAL METHODS

Twenty one pilots participated in two experiments designed to evaluate the changes in visual attention allocation within the cockpit, imposed by the requirement to maintain self separation in a "freeflight" scenario, and the to establish baseline measures of the sorts of maneuvers that pilots tended to select, in order to avoid traffic conflict (see Wickens, Helleberg, & Xu, 1999, 2000). All experiments were conducted in a Frasca single engine simulator, with the flight dynamics of a light general aviation aircraft, and 135 degrees display of the outside world (Figure 1).

Figure 1. The simulation environment, showing the CDTI to the left and the instrument panel (IP) to the right.

Ten licensed pilots participated in the freeflight experiment, conducted first (only 7 provided usable eye movement data). These pilots were provided with a CDTI seen in Figure 1, in order to monitor for traffic and, if the traffic presented a pending loss of separation, to maneuver around it, using lateral, vertical or airspeed control. Pilots were instructed to call out "traffic in sight" as soon as it was visible in the forward view of the outside world (OW). Pilots flew four scenarios over four sessions; each scenario containing ten flight legs, with 6 conflict legs (requiring a traffic avoidance maneuver) and 4 non conflict legs. Traffic was also present in the latter legs, but never became a threat. On one of the non conflict legs in each session, a single traffic aircraft could be seen in the OW view but was not present on the CDTI, simulating the presence of an "unequipped" aircraft, or one with its transponder turned "off".

Fourteen pilots of equivalent flight experience participated in the baseline experiment, conducted second (10 provided usable eye movements). The traffic avoidance maneuver patterns flown by the freeflight pilots were captured, and used to generate ATC commands that would create the same maneuvers (hence the same flight control activity) for the baseline pilots. The only difference is that these pilots did not have a CDTI, and only maneuvered following the ATC instructions. The particular conflict geometry (location, aspect angle) of the "transponder off" aircraft in the freeflight experiment was captured, and detection performance of this same traffic aircraft was the focus of analysis in the baseline experiment, even though here the label was meaningless to the pilot.

RESULTS: SCAN MODELING

Full results of the two experiments concerning the maneuvers chosen, and the detailed analysis of visual scanning can be found in Wickens et al. (2000). Here we focus only on three elements relevant to the modeling efforts: the allocation and dwell of visual scanning, and traffic detection performance. Table 1a presents the data for the percentage of time that visual fixation was within each of the three primary AOIs – the instrument panel (IP), the outside world (OW) and the CDTI, as broken down by leg type (conflict, no conflict) and by experiment (freeflight, baseline). (Naturally the CDTI contains no data for the baseline experiment.) These data can be assumed to reflect a measure of the allocation of visual attention. Table 1b presents the equivalent data for the mean dwell duration on an AOI. That is, the mean duration that the scan remains within an AOI before it leaves. Thus a long dwell on the IP could represent the sequential fixation on a number of instruments within the IP.

Table 1a. Percent dwell time in each area of interest.

<u>Expt</u>	IP	CD	OW	IP	CD	OW
Baseline	63	XX	37	57	XX	43
Freeflt	52	28	20	57	16	27

(a) conflict trials (b) non-conflict trials

Table 1b. Mean dwell time (sec).

<u>Expt</u>	IP	CD	OW	IP	CD	OW
Baseline	6.1	XX	2.8	5.0	XX	3.3
Freeflt	3.3	2.1	1.6	4.1	1.5	2.0

(a) conflict trials (b) non-conflict trials

We tested the prescriptive model (2) in the following way, as shown by the parameter values contained in Tables 2a, 2b, and 2c, representing Bandwidth (B), Relevance (R) and Value (V) respectively. In each case we estimated the simplest set of digital values that would preserve an ordinal relation across trial type, AOI and experiment, as described below.

Table 2. Multipliers for the model in equation (2).

(a) Bandwidth Multipliers

	Leg	IP	OW	CDTI
Free Flight	Confl.	3	2	1
	N.C.	2	1	0.50
Baseline	Confl.	3	2	
	N.C.	2	1	

(b) Relevance Multipliers

	Task	IP	OW	CDTI
Free Flight	Aviate	3	1	0
	Navigate	1	2	2
Baseline	Aviate	3	1	
	Navigate	1	2	

(c) Priority Multipliers

	Aviate	Navigate
Free Flight	3	2
Baseline	3	1

Table 2a presents the bandwidth associated with the AOI in question. BW is highest (=3) for the IP, given that this AOI contains both the most rapidly changing instrument (the attitude indicator) as well as the most different sources of information (the additional five instruments on the panel). B is intermediate (=2) for the OW primarily because this AOI contains a single dynamic element, the visual horizon, which captures the high bandwidth of the attitude indicator, but does not contain easily visible changes corresponding to the other 5 instruments. B is lowest (=1) for the CDTI because this AOI presents only aircraft location and ownship heading information, all of which show only slow visible changes. In addition, we assume, plausibly, that IP and OW bandwidths will both decrease on the straight & level non-conflict trials, compared to the conflict trials in which lateral and vertical maneuvering (involving high bandwidth roll and pitch changes) is nearly always present (Wickens, Helleberg, & Xu, 1999). The coefficients preserve this relation. The bandwidth of these AOIs is not assumed to be influenced by the type of experiment flown (baseline vs. freeflight), since both were designed to yield equivalent maneuvers.

Table 2b presents the relevance of information contained in each AOI, for the two primary tasks of "aviate" (maintaining the aircraft in a stable attitude and avoiding stall), and "navigate". We defined navigate, in this context, primarily in terms of flying to avoid traffic hazards. Across both experiments, the relevance to aviating is clearly highest on the instrument panel (r=3). Aviating is however given some non-zero value (=1) for the OW to the extent that pilots may be flying by some reference to the true horizon (we simulated VMC conditions), as well as to the artificial horizon on the ADI. The relevance of the AOIs to navigation shows a very different picture. The IP is relevant to this task only insofar as it presents heading information (and this is redundantly conveyed in the CDTI) (r=1). Both the OW and the CDTI contain information more relevant to navigating around traffic. Hence both double the coefficient (r=2). A similar pattern is shown in the baseline, as in the free flight experiment, except that the CDTI is excluded.

Finally, Table 2c presents the critical variables from the standpoint of task management: the inferred **importance** of the two tasks which scanning supports, aviating and navigating. Adhering to the standard ANCS hierarchy (Schutte & Trujillo, 1996), we always provide greater weighting to aviating (=3). However in free flight the task of navigation has greater priority than in baseline conditions (=2 vs. 1), because in freeflight air traffic control has a reduced role in traffic monitoring and routing decisions. Even in freeflight however, we still assume aviating (=3) to be more important than navigating (=2).

It should be noted that the parameter values above were specified independently of the data contained in Table 1, by some fairly straightforward consensual rules. Applying the equation (2) to the parameter values in Table 2, we obtained a set of predicted optimal attention values, in each AOI across the 2x2 combination of trial type and experiment. Prior to model fitting we normalized the values across the three AOIs within each row, to reflect a predicted percentage of scan (which added to 100). Figure 2 presents the data of predicted versus obtained percentage allocation for the ten cells in Table 1a, yielding a strong linear fit (r=.92, p<.05) and accounting for 85% of the variance across trials.

The model fit appears to be quite good, suggesting reasonably "optimal" scanning behavior. However, there are other ways that our initial assumptions that generated the coefficient in Table 2 might be charged, to obtain a better fit. Assuming, for example, that the CDTI has **some** relevance for aviating in free flight, and raising all of the relevant coefficients in the baseline so that they are equal in total value to those in free flights, yields a final correlation of 0.96.

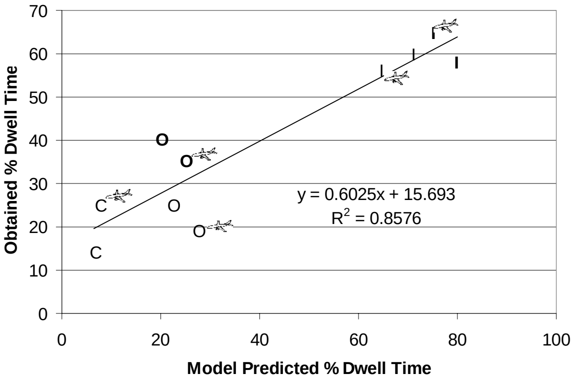

Figure 2. Model predicted versus obtained attention allocation (percent dwell time) measures. I = Instrument Panel, O = Outside World, C = CDTI. Heavy points are from the baseline condition. Airplane symbols designate conflict trials.

The modeling analysis above has not considered the influence of effort nor salience, as neither of these factors should be considered in an optimization model. Rather, they are "forces" to be overcome through altered design, to the extent that they may inhibit optimal scanning. Although we do not address the role of salience here (but see Mumaw et al., 2000), the influence of effort can be seen in two respects. In Table 1b (visual dwell) we note that the mean dwell duration on the instrument panel is substantially (2 seconds) and significantly ($p<.01$) longer in the baseline experiment than it is in the freeflight experiment. A long dwell may be taken to mean that the pilot chooses to do a great deal of sequential scanning in nearby regions (adjacent instruments/ short scans) before returning outside, hence conserving overall access effort. Thus the long dwells on the IP are diagnostic of the constraining influence of effort. The marked shortening (by 2 seconds) of these dwells when free flight responsibility is imposed is diagnostic of the increased influence of "navigate" task priority in free flight, on counteracting the effects of effort conservation.

PERFORMANCE VALIDATION

Does increased visual attention allocation to an area of interest improve the performance of tasks supported by that AOI? We can answer this question by looking at performance differences in traffic detection both between conditions and between pilots. Performance validation contrasts between conditions however are somewhat difficult to make because the information necessary to detect traffic in free flight conditions is distributed across both the CDTI and the OW, whereas in the baseline condition it is only contained in the outside world. In this regard, it is noteworthy that the total percent time dwelling on "traffic information AOIs" is approximately equivalent in both groups, and in both, the mean traffic detection latency is approximately equivalent (22.5 sec in freeflight, 23 sec in baseline). This equivalence is understandable. The baseline condition benefits from availing more time for OW scanning (since there is no CDTI to compete for visual attention); but the freeflight condition compensates for this by allowing the CDTI to direct visual attention to the region of space where the conflict is likely to occur (a contextual information cueing that is formally equivalent to the bandwidth parameter in the model).

The freeflight deficit in OW scanning does however provide a strong prediction for detection of the uncued, "transponder off" aircraft. Since this is not depicted on the CDTI, it will not benefit from any contextual cueing. Thus the 28% advantage in OW scanning for the baseline condition (Table 1a) should support a clear advantage. However such an advantage was not found. Apparently the greater priority for traffic detection (in freeflight: see table 2c) compensated for the reduced availability of OW visual resources, in sustaining effective OW monitoring

The OW detection results did indicate a strong role of salience, when aircraft were classified by their visibility (aspect angle). Those aircraft of low visibility required over twice as long to detect, given that they were unannounced by a CDTI (for both baseline conditions and in the transponder off condition in freeflight).

Finally, it is possible to obtain some performance validation by examining correlations of RT and attention allocation between subjects, rather than between conditions. Here we observed a strong positive correlation between IP attention and RT in both baseline ($r=.80$), and in free flight ($r=.75$) experiments, as well as a corresponding (although weaker) negative correlation between OW attention and detection RT (baseline $r=-0.60$, freeflight, $r=-0.65$). Thus, differences between pilots in attention allocated to the OW do influence detectability of OW traffic.

ACKNOWLEDGMENTS

Experiments supporting this research were funded by a grant from the FAA #DTFA 98-G-022. Dr. Dennis Beringer was the scientific/technical monitor. The modeling effort was supported by a grant from NASA Ames Research Center #NASA NAG 2-1120. Dr. David Foyle was the scientific/technical monitor. Also,

thanks to Paul Kroft for assistance with many aspects of data analysis.

REFERENCES

Bellenkes, A. H., Wickens, C. D., & Kramer, A. F. (1997). Visual scanning and pilot expertise. Aviation, Space, and Environmental Medicine, 68(7), 569-579.

Carbonnell, J. R., Ward, J. L., & Senders, J. W. (1968). A queuing model of visual sampling. IEEE Tran. on Man-Machine System, MMS-9, 82-87.

Chou, C., Madhavan, D., & Funk, K. (1996). Studies of cockpit task management errors. The International Journal of Aviation Psychology, 6(4), 307-320.

Damos, D. L. (1997). Using interruptions to identify task prioritization in Part 121 air carrier operations. Proceedings of the 9th Symposium on Aviation Psychology. Columbus: Ohio State Univ.

Dornheim, M. A. (2000, July 17). Crew distractions emerge as new safety focus. Aviation Week and Space Technology, 58-65.

Ellis, S. R., & Stark, L. (1986). Statistical dependency in visual scanning. Human Factors, 28, 421-438.

Gray, W. D., & Fu, W. T. (2001). Ignoring perfect knowledge in-the-world for imperfect knowledge in-the-head. CHI Letters, 3(1).

Latorella, K. (1998). Effects of modality on interrupted flight deck performance. Proceedings of the 42nd Annual Meeting of the HFES. Santa Monica, CA.

Laudeman, I. V., & Palmer, E. A. (1995). Quantitative measurement of observed workload in the analysis of aircrew performance. IJAP, 5(2), 187-198.

Moray, N. (1986). Monitoring behavior and supervisory control. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), Handbook of perception and performance, Vol II. New York: Wiley & Sons.

Mumaw, R., Sarter, N., Wickens, C. D. et al. (2000). Analysis of pilots' monitoring and performance on highly automated flight decks (Final NASA Project Report). Moffett Field, CA: NASA Ames Research Center.

Raby, M., & Wickens, C.D. (1994). Strategic workload management and decision biases in aviation. IJAP, 4(3), 211-240.

Schutte, P. C., & Trujillo, A. C. (1996). Flight crew task management in non-normal situations. Proceedings of the 40th Annual Meeting of the HFES. Santa Monica, CA: HFES.

Senders, J. (1964). The human operator as a monitor and controller of multidegree freedom systems. IEEE Trans. on Human Factors in Electronics, HFE-5, 2-6.

Sheridan, T. (1972). On how often the supervisors should sample. IEEE Trans. on Systems, Sciences, and Cybernetics, SSC-6, 140-145.

Strauch, B. (1997). Automation and decision making: Lessons from the Cali accident. Proceedings of the 41st Annual Meeting of the Human Factors and Ergonomics Society. Santa Monica, CA: HFES.

Wickens, C. D. (1993). Cognitive factors in display design. Journal of the Washington Academy of Sciences, 83(4), 179-201.

Wickens, C. D. (2000). Imperfect and unreliable automation and its implications for attention allocation (ARL-00-10/NASA-00-2). Savoy, IL: University of Illinois, Aviation Res. Lab

Wickens, C. D., & Carswell, C. M. (1995). The proximity compatibility principle. Human Factors, 37(3), 473-494.

Wickens, C. D., Helleberg, J., & Xu, X. (1999). Maneuver choice in free flight. Proceedings of the World Aviation Congress (1999-01-5591). Warrendale, PA: Society of Automotive Engineers, Inc.

Wickens, C. D., Helleberg, J., & Xu, X. (2000). Decision and workload implications of free flight and the cockpit display of traffic information. Proceedings of the IEA 2000/HFES 2000 Congress. Santa Monica, CA: HFES.

Wickens, C. D., Vincow, M. A., Schopper, A. W., & Lincoln, J. E. (1997). Computational models of human performance in the design and layout of controls and displays (CSERIAC SOAR Report 97-22). Wright Patterson AFB, OH.

Yantis, S. (1993). Stimulus-driven attentional capture. Current Directions in Psychological Science, 2, 156-161.