FLIGHT DECK INTERVAL MANAGEMENT AVIONICS: EYE-TRACKING ANALYSIS

Kara Latorella
NASA Langley Research Center
Hampton, VA 23681

John W. Harden
Old Dominion University
Norfolk, VA 23508

Interval Management (IM) is one NexGen method for achieving airspace efficiencies. In order to initiate IM procedures, Air Traffic Control provides an IM clearance to the IM aircraft's pilots that indicates an intended spacing from another aircraft (the target to follow – or TTF) and the point at which this should be achieved. Pilots enter the clearance in the flight deck IM (FIM) system; and once the TTF's Automatic Dependent Surveillance-Broadcast signal is available, the FIM algorithm generates target speeds to meet that IM goal. This study examined four Avionics Conditions (defined by the instrumentation and location presenting FIM information) and three Notification Methods (defined by the visual and aural alerts that notified pilots to IM-related events). Current commercial pilots flew descents into Dallas/Fort-Worth in a high-fidelity commercial flight deck simulation environment with realistic traffic and communications. All 12 crews experienced each Avionics Condition, where order was counterbalanced over crews. Each crew used only one of the three Notification Methods. This paper presents results from eye tracking data collected from both pilots, including: normalized number of samples falling within FIM displays, normalized heads-up time, noticing time, dwell time on first FIM display look after a new speed, a workload-related metric, and a measure comparing the scan paths of pilot flying and pilot monitoring; and discusses these in the context of other objective (vertical and speed profile deviations, response time to dial in commanded speeds, out-of-speed-conformance and reminder indications) and subjective measures (workload, situation awareness, usability, and operational acceptability).

Background

Interval Management (IM) is one NexGen method for achieving airspace efficiencies. In order to initiate IM procedures, Air Traffic Control provides an IM clearance to the IM aircraft's pilots that indicates an intended spacing from another aircraft (the target to follow – or TTF) and the point at which this should be achieved. Pilots enter the clearance in the flight deck IM (FIM) system; and once the TTF's Automatic Dependent Surveillance-Broadcast (ADS-B) signal is available, the FIM algorithm generates target speeds to meet that IM goal. The algorithm generating these speeds [1] is based on the standard terminal arrival route (STAR) in use, conforms to standard speed constraints in the terminal environment, and is adaptive to forecasted winds. When conducting FIM operations, in accordance with the concept of operations for NASA's technology demonstration efforts [2], the crew operates with autothrottles on, with autopilot engaged, and the auto-flight system in Vertical Navigation (VNAV) and Lateral Navigation (LNAV). In the tested concept of operations, the IM speeds are presented in the flightdeck, and the crew is responsible for selecting the new speed in the Speed Window of the Mode Control Panel (MCP), instructing the aircraft to achieve this new speed. To support FIM, the crew is responsible for safely flying the aircraft while maintaining situation awareness of their ability to follow FIM speed commands and to achieve the FIM spacing goal.

The objective of this investigation was to assess different FIM Avionics configurations based on objective data (flightpath and speed profile deviations, and response times), subjective assessments and ratings, and eye-tracking data. This paper discusses other results, but focuses on the eye-tracking metrics of performance used to characterize crew performance.

Methods

Participants

Twelve crews participated in the study, each with two experienced (between 19 and 40 years of flying, mean of 28.9 years) commercial pilots who were type rated in the same class as the simulated aircraft. Eleven of these subjects reported demographic data.

Apparatus & Scenarios

The study was conducted in NASA Langley's Integration Flightdeck (IFD) simulator, which approximated a Boeing 757 aircraft. The standard flightdeck was augmented with two Electronic Flight Bags (EFBs), and two ADS-B guidance displays (AGDs). Figure 1 shows these displays (in the aft position for the EFB) for the left side; where

1

the position was mirrored for the right side. Scenarios required crews to fly from approximately 25,000 feet to land at Dallas-Fort Worth International Airport (KDFW). Scenarios began in level flight prior to top of descent, in VNAV Path autoflight mode engaged. The aircraft were in an unconstrained vertical path descent from at or near Top of Descent until reaching the first altitude constraint at 11,000 feet. The aircraft operated in VNAV Speed with the MCP speed window open, until the flaps were extended, and the autoflight mode reverted to VNAV Path. The last speed target given was the reference speed for flaps at 30, plus five knots to enable stabilized approach by 1000 feet above ground level. Scenarios concluded typically after roll out on touchdown, but occasionally in advance of that in order to save time (always after aircraft configuration for a stabilized approach was complete). Subjects were instructed to fly as they typically would, as though they had passengers in the back of the airplane, to respond to speed targets in a timely manner, to try to maintain speed conformance within seven knots, and to remain within 400 feet of the VNAV path. Confederate Air Traffic Controllers provided realistic communications to both the IFD and to roughly 20 other simulated aircraft in the environment. Prerecorded Automatic Terminal Information Service (ATIS) messages were available on the appropriate frequency.

Figure 1. The Integration Flightdeck Simulator, showing the EFB in the Aft position, and the AGD.

Experimental Conditions and Design

The Avionics Configurations tested were defined by an Avionics Condition (display devices and locations) and a Notification Method (whether events were indicated only visually, or were augmented with aural indications). Each crew evaluated four Avionics Conditions: (1) Integrated –FIM target speeds were presented in the upper left corner of the primary flight display (PFD) and speed profile deviation information was implicitly indicated as the deviation between current speed and an instantaneous speed profile bug on the PFD speed tape. The FIM page in the MCDU displayed numeric speed profile deviation (in knots). Significant deviations from the speed profile triggered a message on the EICAS system. (2) EFB-Aft –speed targets, speed deviation information and messages, and all elements of the IM clearance were presented on an EFB in the position shown in Figure 1. (3) EFB-Fore – all information was presented on the EFB, but this display was located in a more forward location, just under the outboard window. (4) EFB-Aft-AGD – in which the EFB-Aft condition was augmented with the ADS-B Guidance Display (AGD). The AGD repeats the same FIM target speed and speed deviation information given on the EFB.

Crews received notifications when conditions required their attention, *i.e.*, when a new FIM target speed occurred (*target speed onset*), if the current aircraft speed significantly deviated from the FIM target speed (*conformance deviation*), and if they failed to enter a new FIM target speed within a reasonable time period (*reminder*). A conformance deviation indicator was provided when the aircraft current speed was more than seven knots different from the instantaneous speed on the FIM speed profile, the speed changed more than five seconds ago, and aircraft current speed was not converging to the FIM target speed. A reminder was provided if the crew did not dial in the correct FIM target speed within 10 seconds. If the speed was still not dialed in, the reminder indication was repeated at most two more times at 10 second intervals. This study evaluated three notification methods defined by the modality (V for visual, A for aural) associated with the triplet of implementations: *target speed onset*, *conformance deviation*, and *reminder* events. The VVV method provided only visual (V) cues for all three events. The AAA method augmented these visual indications with an aural (A) tone, again for all three events. The VAV method included visual indications for all three events, and presented the tone only if pilots significantly deviated from the speed profile.

Each crew member had the opportunity to fly an arrival and approach with each of the Avionics Conditions twice, once as pilot flying (PF) and once as pilot monitoring (PM). Avionics condition and Crew Role were within-crew variables. Notification Method was a between-crew variable. Order of Avionics conditions were counterbalanced over crews, and the assignment of scenarios to Avionics conditions was also counterbalanced.

Data Collection & Analysis

The study collected objective (vertical and speed profile deviations, response time to dial in commanded speeds, out-of-speed-conformance, and reminder indications) and subjective ratings (workload, situation awareness, usability, and operational acceptability) for each run. Post-experiment questionnaire items asked subjects to consider pairwise preference comparisons and to also rate (using 9-point scales with anchoring cues) the operational acceptability of the Avionics conditions in the context of the notification method they received, the utility of aural indications, and factors associated with operational acceptability (workload, situation awareness, and crew coordination).

Oculometer data was collected using two 6-camera Smarteye (SE) eye-trackers (SE Pro software, version 5.8) and recorded in Smarteye logfiles at 60Hz, corresponding to a nominal frame rate of about 17 msec. Oculometer data was also sent to simulation files, which were recorded at 5Hz. This experiment resulted in 192 eye-tracker logfiles (12 crews x 2 pilots/crew x 8 runs/crew). Complete logfile data was available for 168 (87.5%) of the data. The majority of missing data pertained to the first speed target, after pruning these from consideration in all datafiles, only seven logfiles (approximately 3.6%) remained affected by significant datafile loss.

In addition to incomplete data files, recorded data may be of questionable quality. SE software reports a head and gaze quality value for each reported point of gaze (POG), defined by the system's confidence in head and eye position assessment, normalized over the data previously acquired in that session. SE's Gaze Direction Quality metric ranges from 0.0 to 1.0; where 0 corresponds to the 1st percentile of all quality values experienced to that point, and 1.0 corresponds to the 99th percentile. As such, this value is individual-dependent and only useful as a general guide to the degree to which the eye-tracker has sufficient information upon which to base a POG determination. SE recommends that the system be given some time to "fill up" the buffer for this measure so that its reported values stabilize. Therefore, removing data associated with the beginning of the runs had the added benefit of stabilizing the quality measures. Unless specified otherwise, the following analyses were conducted on only those POG data that were associated with a gaze direction quality of 0.7 or greater. Regretably, data loss and insufficient data quality can not be considered random errors. Situations in which pilots gaze was extreme (downward or to the side) was more likely to result in lost or poor quality data. As such, data from the EFB_Aft condition was disproportionally affected.

Generalized linear models, with compound symmetry covariance structures (assuming heterogeneous variances and constant correlations among repeated measures) were used to model this mixed factor study with repeated measures. These models employed robust estimation of variances (to handle violations of model assumptions) and Satterthwaite adjusted degrees of freedom (to mitigate issues associated with missing data). Statistics were calculated with respect to Gamma distributions using a log link function, as most data were defined by non-negative values, and all distributions were positively skewed. Models included terms for main effects associated with Avionics Condition (EFB_Aft, EFB_Forward, EFB_Aft+AGD, Integrated) and Notification Method (VVV, VAV, AAA); and the two-way interactions of these main effects. For some measures, each pilot provided data (e.g., Noticing Time); whereas for others, the crew served as the experimental unit (e.g., Minimum Noticing Time). When the experimental unit was a pilot, the Role (Pilot Flying (PF) or Pilot Monitoring (PM)) and interactions of Role with Avionics Condition and with Notification Method were included in analyses. Significant fixed effects were further investigated with Sidak-adjusted sequential pairwise comparisons; which protect for inflated alpha, and are more powerful than Bonferroni-adjusted tests. Results were interpreted at alpha=0.10, but p-values are provided for the reader who choses to consider more stringent criteria.

Oculometer Results

Oculometer data was taken to help characterize the attentional sampling pilots used in response to the different Avionics Configurations for presenting FIM information. These metrics included those that addressed: the frequency with which pilots sampled the FIM display(s); the degree to which pilots' points-of-gaze were "Heads-Up," that is, looking out the window; and the time for pilots to notice IM events on the FIM displays. In addition, eye-tracker data was used to analyze metrics related to workload: the length of time pilots dwelled on the FIM display on first regard following an IM event, and an entropy-based measure that has been proposed to be related to workload. The selection of data appropriate for consideration of each measure is presented per section.

Sampling the FIM Display

FIM Display Sampling analyses were conducted on the portion of each scenario from 99 seconds into the run, until 19 seconds following the eighth speed target encountered. This period was determined by attempting to maximize data used, minimize disproportionate lost data, include an equivalent number of speed target changes, and attempt to have roughly equivalent task durations. While the number of data points taken in these windows differed only slightly, counts were normalized by data frames per scenario. Results show only the Avionics Condition significantly predicted differences in counts of POGs on FIM displays (p<0.001). On average, pilots were most likely to sample the Integrated FIM Display; of the retrofit conditions, more likely to sample the FIM Display(s) associated with the EFB_Fore, then EFB_Aft+AGD, and least likely to sample the FIM Display associated with the EFB_Aft condition (all pairwise comparisons, p<0.032). Notification Method was not significant.

Heads-Up Sampling

The set of data used in this assessment was defined in the same manner as for the FIM Display Sampling analysis. However, whereas the FIM Display analysis used only data in which gaze direction quality was sufficient, this analysis employs a technique developed at NASA Langley [3] to define Heads-Up gazes from head pitch data when gaze quality is questionable. Avionics Conditions significantly affected Heads-Up POGs (p=0.005). Pairwise tests show only one significant comparison; this indicating that pilots experienced significantly more Heads-Up POGs in the Integrated Condition than for the EFB_Aft Condition (p=0.016), where all other comparisons were not significant (all p>0.106). The Avionics Condition and Notification Method interaction term was significant (p=0.011), but pairwise comparisons did not reach significance (all p>0.438).

Noticing Times

This analysis addresses the time (Noticing Time) for the PF and PM, separately, to first attend to the display containing information about FIM speeds. Appropriate display(s) are defined by Avionics Condition, as previously described. For the EFB_Aft+AGD condition, the first Noticing Time was identified as a POG on either the EFB or the AGD. Noticing times were identified in logfile data, and conducted on periods following each of eight speed targets per run. These were defined as the first POG that landed on the appropriate display(s) for which the data quality was 0.7 or greater, and for which there was a second such subsequent gaze, with fewer than five frames (nominally 88msec of data) of intervening missing or poor quality data in the same display. Noticing Times were significantly affected by Avionics Conditions (p < 0.001), the interaction of Avionics Condition and Notification Method (p=0.001), and Role (PF v. PM) (p=0.077). Noticing Times, and the variability in these, tended to decrease across conditions in this order: EFB_Aft, EFB_Fore, EFB_Aft+AGD, Integrated. Noticing time with the Integrated condition was, on average, over five times faster than with the EFB_Aft condition. The Integrated condition was significantly faster than all other conditions, and the EFB_Aft+AGD condition was significantly faster than both the EFB_Fore and the EFB_Aft conditions (all pairwise, p< 0.013). This main effect contains a significant interaction of the Avionics Condition and the Notification Method which shows that, for the EFB_Fore condition, Noticing time for the AAA condition was significantly longer than for the VAV condition (p=0.084). For the other three Avionics Conditions, pairwise comparisons of Notification Methods did not significantly differ (all p≥0.250), but means suggest that pilots with the VVV method were slowest to notice new speed targets. PFs were faster to notice commanded speed changes than PMs, by about 200msec.

The same data set was similarly analyzed to investigate how the crews' Minimum Noticing Time, and the absolute difference of pilots' Noticing Times were affected by experimental conditions. Factors of significance for these variables are similar to findings observed for each pilot's Noticing Times. Avionics Condition (p<0.001), and the interaction of Avionics Condition and Notification Method (p<0.001) significantly affected the crews' first notice of a new speed target. With regard to the main effect, means followed the same order as for Noticing Times, but were more sensitive to differences in conditions. Pairwise comparisons showed only significantly faster Minimum Noticing Times for the Integrated condition than other conditions (all p<0.065). Pairwise tests of interaction terms show that the AAA Method was associated with significantly faster Minimum Noticing Times than the VAV Method (with the EFB_Aft Condition, p=0.086) and the VVV Method (with the Integrated Condition, p=0.006).

For the same data periods used to assess pilot and crew Noticing Times above, when a Notice was detected, a count was kept for how many times the PM was the pilot to notice first. Analysis as a Poisson distribution with a log link function shows only a significant effect of Notification Method, whereby PMs were more likely to be first to notice new commanded speeds than PFs when using the VVV method than the AAA method.

Indicators of Pilot Workload

Two measures postulated to reflect workload are examined: Dwell Time and a measure related to scan path entropy. Initial Dwell Time on a display has been associated with the difficulty of processing visual stimuli and therefore extracting meaning from it [4], and has been associated with pilot workload [5]. Dwell Time was calculated from the data frame of first Notice (a POG on an appropriate FIM display) until either the frame before a POG on another display was reported, or five frames of missing or poor quality data occurred; then this number of frames was multiplied by the nominal frame rate. Data was framed by the eight speed targets encountered starting with the second of these and, as for Noticing Times, used logfile data. Dwell times were significantly affected by the Avionics Condition factor (p<0.001). The Integrated and EFB_Aft+AGD conditions did not significantly differ from each other, but they both supported significantly shorter Dwell Times than either the EFB_Fore or EFB_Aft conditions (and these last two did not significantly differ from each other) (all pairwise, p <0.002).

In theory, more information in dense, confusing, or unintitive presentations should require longer dwell times to detect and extract pertinent information. Based on initial work by [6], this measure was applied to characterize distribution of POGs [7]. These authors and others [8] found that as pilots' workload increased, visual sampling became more systematic and entropy decreased. The Nearest-Neighbor Index (NNI) measure of entropy was found to be consistent with both objective (p300 EEG responses) and subjective (NASA-TLX score) measures of workload [9]. The NNI metric investigated here, is the ratio of the average observed minimum distances among POGs, and the mean distance expected if the distribution were random. The NNI is therefore equal to one when the distribution is completely random, and higher values suggest more systematic search – presumably induced by higher workload conditions [10]. Total entropy measures were calculated from software developed for this purpose at NASA Langley [11], based on Di Nocera's publications [9,10]. The 5 Hz eye-tracker data was used for this analysis due to processing complexity. Data included in this analysis was from the eight speed targets beginning with the second speed target occurring in logfiles for each run. NNIs were calculated for good quality data following the occurrence of a new speed target, and for the following 19 seconds. Notification Method (p=0.099) and Role (p=0.024) significantly affected NNIs. While pairwise comparisons on Notification Methods failed to reach significance (all p≥0.132), observation of means shows clearly higher entropy (higher workload) when pilots had VVV notifications than other Notification Methods. NNIs were higher for pilots when in the PM role.

Discussion

The Integrated condition supported better heads up time than the EFB_Aft condition, fastest noticing times than all other conditions, shorter first dwell times than both the EFB_Aft and EFB_Fore conditions, and was sampled most frequently. The finding that this Avionics Condition was sampled more frequently than others is not surprising. The Integrated condition presented FIM information on the PFD, and obviously other information on this display is crucial to flight operations. Regrettably, the eye-tracker data did not provide sufficient resolution to distinguish between POGs to FIM information vs. other PFD content. However, in concert with other findings, this result indicates this condition most effectively supports FIM operations with minimal disruption to scan. Subjective ratings of Situation Awareness, distraction, and pairwise preference comparisons as reported elsewhere [12] are consistent with this finding.

Subjective commentary and ratings were least complimentary of the EFB_Aft condition, and eye-tracker findings are again consistent. When in this position, FIM information was sampled least frequently (based on means, though not significantly different from the other retrofit solutions), was slowest to notice (not significantly different than the EFB_Fore condition – but this was hampered inordinately when paired with the AAA Notification Method), and was one of the conditions that caused longer initial looks on the FIM display to extract information (where the EFB_Fore condition did not statistically differ). While the deleterious impacts of the EFB_Aft condition may not be surprising to this community, this study assessed it because the EFB has been implemented in this position in some cases. The aforementioned results, and those that show that both the EFB_Aft+AGD had faster noticing times and shorter dwell times than the EFB_Fore, seem to indicate superiority of the EFB_Aft+AGD over the EFB_Fore condition. However, other results show the reverse order – FIM information was sampled more frequently in the EFB_Fore condition, and other results based on pilots' awarness of new speeds and overall acceptability ratings were higher.

Most of the significant results associated with these analyses pertained to differences among the Avionics Conditions – that is, the placement and type of display used to present the FIM information, rather than the type of aural/visual Notification Method used. It is, however, important to consider this finding in light of the experimental design: whereas Avionics Condition was considered as a within-subject/crew variable, Notification Method was a between-subject/crew variable – and therefore was subject to greater noise in the data from individual differences across levels. Notification Method did not statistically affect differences in Sampling Frequency, Heads-Up

5

Sampling, or Dwell Times; and had only interaction effects with Avionics Condition for pilot's Noticing Times (where the AAA method seemed to significantly delay noticing in the EFB_Fore condition), minimum crew Noticing Times (showing superiority of the AAA method over the VAV method for the EFB_Aft condition and over the VVV method for the Integrated condition), and weak effects on the NNI (where means indicate higher workload for the VVV condition).

When in the PF role, pilots were generally faster in regarding the FIM display after a speed change, and had more systematic scan patterns (higher NNIs). However, when aural indications were available for all FIM events (the AAA method), PMs were more likely to be the first to notice speed changes than PFs. While decreasing responsiveness to FIM events by some small degree may advantage FIM operations, integration of new technologies and procedures must consider the full context of performance and cohesive job design – and the disruption of PF scan may be more costly to overall operations than the benefit to FIM operations.

References

[1] Abbott, T.S. (2012). *An Overview of a Trajectory-Based Solution for En Route and Terminal Area Self-Spacing: Third Revision*. NASA/CR-2012-217786. NASA Langley Research Center: Hampton, VA.

[2] Baxley, B. T., Swenson, H. N., Prevot, T., & Callantine, T. J., (2012). NASA's ATM Technology Demo-1: Integrated Concept of Arrival Operations, *31st Digital Avionics Systems Conference,* Williamsburg, VA, Oct.

[3] Ellis, K.K.E., Arthur III, J.J., Latorella, K. A., Kramer, L., Shelton, K J., Norman, R.M. & Prinzel, L.J. (2012). Quantifying Pilot Visual Attention in Low Visibility Terminal Operations. In *Proceedings of the AHFE International Conference on Applied Human Factors and Ergonomics*: San Francisco, CA.

[4] Callan, D. J. (1998). Eye movement relationships to excessive performance error in aviation. In *Proceedings of Human Factors and Ergonomics Society Annual Meeting*: Santa Monica, CA.

[5] Becker, S.I. (2011). Determinants of Dwell Time in Visual Search: Similarity or Perceptual Difficulty? *PLoS ONE 6*(3): e17740. doi:10.1371/ journal.pone.0017740.

[6] Shannon, C.E. (1948). A Mathematical Theory of Communication. Bell System Tech Journal 27(3): 379–423.

[7] Tole, J., Stephens, A., Vivaudou, M., Harris, R. & Ephrath, A. (1982). entropy, Instrument Scan, and Pilot Workload. In *Proceedings of the IEEE Conference on Systems, Man & Cybernetics*: Seattle, WA.

[8] Harris, Sr., R.L., Glover, B.J. & Spady, Jr. A. (1986). *Analytical Techniques of Pilot Scanning Behavior and Their Application.* NASA-TP-2525 19860018448. NASA Langley Research Center: Hampton, VA.

[9] Camilli, M., Terenzi, M., & DiNocera, F. (2007). Concurrent validity of an ocular measure of mental workload. In D. deWaard, G.R.J. Hockey, P.Nickel, and K.A. Brookhuis (Eds.), *Human Factors Issues in Complex System Perform*ance. Shaker Publishing: Maastricht, the Netherlands.

[10] Di Nocera, F. (2007). Cognitive Aspects and Behavioral Effects of Transitions Between Levels of Automation. *European Office of Aerospace Research & Development Report* (Contract Number FA8655-05-1-3021).

[11] Harden, J.W. & Latorella, K.A. (2014). *Software for the analysis of Smarteye Oculometer Data for Workload and Scan Path Comparions.* Internal Report.

[12] Latorella, K. (2015) *Avionics Configuration Assessment for Flightdeck Interval Management: A Comparison of Avionics and Notification Methods.* NASA-TM-in press.