In Proceedings of the 15th International Symposium on Aviation Psychology
April 27-30, 2009
Wright State University, Dayton, Ohio

EFFECTS OF VISUAL, SEAT, AND PLATFORM MOTION DURING FLIGHT SIMULATOR AIR TRANSPORT PILOT TRAINING AND EVALUATION

Judith Bürki-Cohen, Andrea L. Sparko, and Young Jin Jo
USDOT/RITA/Volpe National Transportation Systems Center
Cambridge, Massachusetts
Tiauw H. Go
Nanyang Technological University
Singapore, Singapore

Access to affordable and effective flight-simulation training devices (FSTDs) is critical to safely train airline crews in aviating, navigating, communicating, making decisions, and managing flight-deck and crew resources. This paper provides an overview of the Federal Aviation Administration-Volpe Center Flight Simulator Human Factors Program examining the requirements for the qualification and use of FSTDs. We will summarize past research investigating the need for a full hexapod-platform motion system, describe regulatory and industry developments, and report on current activities.

Aviation Training Challenges

As we are writing this paper in early 2009, the global economy has entered a down-turn that may temporarily ease the world-wide pilot and aviation personnel shortages that have plagued the aviation industry in recent years. However, the challenge of training crewmembers with increasingly different backgrounds may remain, as the recent merger of several major airlines would suggest. At the same time, the proportion of certain types of accidents appears to be increasing (Boeing, 2007). One airliner disabled by a bird-strike was able to successfully land in the Hudson River, saving everybody on-board and sparing many that could have been killed on the ground. The crew had been trained under the rules of the Advanced Qualification Program (AQP), an alternative scenario- and proficiency-based, team-building curriculum targeting not only technical and procedural skills, but also cognitive and human-factors awareness [Federal Aviation Administration (FAA), 2005]. The recently published Notice of Proposed Rulemaking regulating traditional air-crew training takes a similar approach (Federal Register, 2009). Scenario-based full-mission training will play a critical role as the United States (US) and the European Community (EC) are overhauling their air-traffic-management (ATM) systems to improve current operations and accommodate future increases in traffic volume and complexity. This requires an update of the existing ATM infrastructure and the introduction of advanced tools to increase the efficiency of the air-transportations system (ATS) both in the air and on the ground without compromising safety.

For training, this all boils down to an increased need for effective and affordable Flight-Simulation Training Devices (FSTDs). FSTDs are the only means to efficiently train crews in aviating, navigating, communicating, making decisions, and managing their resources all at once. In fact, AQP is not possible without access to FSTDs, and the proposed new crewmember-qualification rules mandate it. Airplane mergers increase the need for a team-training approach. The US's Next Generation (NextGen) ATS and the EC's SESAR (Single European Sky ATM Research) may radically change the roles and collaborations of air- and ground crews and automation. FSTDs will have to be able to simulate all these interactions, in addition to increased emphasis on accurate simulation of environmental hazards and Loss of Control (LOC) situations. Given that airlines have limited resources for supporting such training under any circumstances, but especially now, this update on the FAA-Volpe Center Flight Simulator Human Factors Program is provided for the benefit of training developers. We will summarize past research investigating the need for a full hexapod-platform motion system, describe regulatory and industry developments, and report on current activities.

FAA-Volpe Center Flight-Simulator Requirements Research

As the FAA launched the "One Level of Safety" initiative for major and regional airlines and more and more airlines adopted AQP in the mid-nineties, the FAA launched the FAA-Volpe Flight Simulator Human Factors Program to identify potentially unnecessary obstacles to universal access to the benefits of flight-simulator training. The purpose of the program was to systematically examine each requirement for simulators used in zero-flight-time training and evaluation of airline pilots for its contribution to the training value of those simulators. The focus of this examination was the training value added by each requirement. A review of the literature and consultation with many subject matter experts (SMEs), including two FAA industry workshops (proceedings available), revealed that the scientific basis for requiring a hexapod-motion platform for airline pilot training and evaluation was shaky. While simulator motion does have great face value (after all, the airplane moves), decades of research have failed to show an effect of motion on transfer of training to the airplane or on reverse transfer from the airplane to the simulator for pilot evaluation (Bürki-Cohen, Longridge, and Soja, 1998). The program thus initiated a series of carefully controlled studies to answer the following questions: 1) Are there maneuvers in airline-pilot training where platform motion cues, in addition to the visual cues from a wide field-of-view (FOV) out-the-window view and instruments, result in an operationally relevant improvement of transfer between the simulator and the airplane? 2) Do airline pilots need to be trained to avail themselves of motion cues? 3) Are motion cues from a hexapod-platform representative of those experienced in the airplane? 4) Can alternative systems provide onset cues and perception of realism?

To ensure valid and replicable studies, we did everything possible to achieve the statistical power to find an effect of motion. For the first three studies,[1] we tested the extremes, i.e., comparing FAA Level C or D full flight simulators (FFS) with the motion turned on or off. We calibrated all cueing and measurement systems. We counterbalanced across groups any other factors that we could not keep constant on all aspects that could affect their flying skills. We also tested more pilots than in most other studies to randomly balance any differences that we were unable to systematically control or counterbalance. We carefully selected maneuvers where motion cues were most likely to serve an alerting function according to the literature and SMEs, i.e., high-workload maneuvers with unpredictable mechanical or weather disturbances. We measured nearly 80 dependent variables directly from the simulator representing pilot control inputs, flight precision, and simulator motion performance at a high sampling rate (at least 30 Hz). We also asked all participants for their opinions on all aspects of the simulator and the behavior of the pilots flying. We started out with experienced pilots, who according to the literature would be most likely to rely on motion cues (Young, 1967). We kept the motion status and in fact the purpose of the study secret to prevent any bias for or against motion from affecting the results.

We performed three studies and did not find any operationally relevant effects of motion. The first two studies examined the effect of motion on recurrent training and evaluation. Study 1 tested regional airline captains in a Level C FFS of a turboprop “powerhouse” airplane with wing-mounted engines. The simulator had a 60 inch stroke with a 1.7 Hz heave bandwidth. The maneuvers consisted of engine failures with continued and rejected takeoffs (V1 cuts and RTOs) at a quarter mile Runway Visual Range and with 10 knots crosswind. For the V1 cut, we found no differences between the pilots that flew the simulator with motion and those that flew it without motion at “First Look,” i.e., when pilots flew the simulator the first time after having flown the airplane for twelve months.[2] For the RTO, the motion pilots showed lower yaw activity, but motion did not affect their heading compared to the no-motion pilots. After additional training (when necessary) and transfer to the simulator with motion of all pilots, pilots having been trained with motion showed lower speed exceedance that may have been achieved with higher pitch standard deviation (STD). They also showed higher yaw activity with more pedal reversals, but without directional control benefit. There were no differences between the two groups for the RTO. In general, all differences were small and operationally irrelevant. However, the data also showed a relatively large discrepancy between the lateral acceleration of the simulator compared to the lateral accelerations from the flight model during the first few seconds following the engine failure. Was the motion of our test simulator just unusually ineffective? With the help of the FAA’s National Simulator Program, we obtained data from eight other Level C and D simulators and found that in terms of lateral acceleration, the test simulator used appears to be representative of other FAA qualified FFS.

Given these findings, in collaboration with the National Aeronautics and Space Administration (NASA), we conducted a second study using the FAA-NASA Level D B747-400 simulator with its motion re-engineered so as to provide greatly improved lateral acceleration and somewhat improved heave motion, trading off rotational motion which, according to some studies, is less important than translational motion (see AIAA-2003-5678 for references). For this study, we tested both captains and first officers, and we replaced the rejected takeoff with a V2 cut, arguing that it was a more diagnostic maneuver requiring multi-axis control. We also had pilots return the one-

[1] Due to the six-page space limit, the link to all studies performed within the FAA-Volpe Center Flight Simulator Human Factors Program is given in the references.

[2] Reported effects have a probability of $p<0.05$ to have occurred by chance. Effects of $0.05<p<0.10$ are reported as trends.

engine-out airplane to the airport for two difficult landings, one a side-step landing with a microburst (SSL) and the other a hand-flown Precision Instrument Approach and landing with quartering head- and tailwinds (PIA). During First Look, we finally found the V1-cut pedal reaction time advantage with motion that may arise from faster perception of vestibular motion compared to visual motion, but it was less than half a second and had no effect on flight precision. Most importantly, it did not transfer to the simulator with motion; once the pilots that were trained without motion were given motion cues, they responded as fast as the pilots trained with motion. Also at First Look, the motion pilots had lower heading STD with lower root mean square (RMS) pedal and yaw activity, but higher pedal bandwidth. They also had lower pitch STD. For the V2 cut, the motion group had more pedal reversals than the no-motion pilots at First Look, and when all pilots transferred to the simulator with motion, the motion-trained pilots had a 0.8 s slower mean pedal reaction time. There were no differences between the two groups for the V1 cut at transfer. Again, all differences between the two groups were small and operationally irrelevant. The only difference that approached operational significance, according to our SME, was found with the one-engine-out PIA during the approach-fix-to-decision height phase, where the motion-trained group had worse horizontal directional control than the no-motion trained group both before and after transfer to motion (by 0.19 dots higher localizer STD and 0.16 dots higher localizer exceedance). The motion group had also higher STD for heading and bank angle. For the SSL, the motion group had lower pedal bandwidth and landed more softly (by 42 ft/min) but less precisely (by 225ft, but both groups landed well within the landing box). All these latter differences were small.

Having found no operationally relevant advantage of providing motion cues for recurrent training and evaluation with two very different groups of pilots flying a very different airplane, the question remained whether it was perhaps their familiarity with the motion of the real airplane that masked an effect regardless of a heightened sensitivity of experienced pilots to motion cues that was reported in the literature. We thus tested a group of new-hires fresh from ground school in a Level D B717-200 simulator with a 54 inch stroke of platform motion and a 90 degree phase lag at 8.3 Hz. Due to time constraints, we focused on the PIA and the V1 cut. For the First Look V1 cut, we found a trend for a faster pedal response with motion (again, less than half a second), and again, the effect disappeared once all pilots transfer to motion. Also, this trend did not translate in an operationally relevant effect on flight precision. Both during First Look and Transfer to motion, the motion group had steadier column RMS, smaller airspeed exceedance, but higher pitch STD during the V1 cut and higher pedal RMS for the PIA. Again, all effects were small and operationally irrelevant.

In summary, it appears that the answers to the first three research questions are “No.” Consulting the literature and SMEs, we had carefully chosen maneuvers within and even somewhat beyond the air-transport-pilot curriculum that should have depended on an alerting function of physical motion, if there were one. In the one maneuver where pilots did respond minimally faster with motion, even pilots that had not experienced motion during training were able to use the motion cue once they received it, indicating motion was not necessary for training. Finally, we obtained the same results even in a simulator that was tuned to provide the best possible motion cues for the maneuvers trained. Our follow-up to Study 1 found large differences between the equations of motion and the real motion for several simulators available for examination, indicating that a good overall safety record had been obtained despite negligible motion cues in at least some devices (see Bürki-Cohen et al., 2001 manuscript for summary of follow-up study).

Regulatory and Industry Impact

The lack of finding an operationally relevant effect of costly motion platforms, even with state-of-the-art motion and visual systems and with demonstrated power to find an effect ($\beta<0.20$, Cohen, 1988), has received a mixed reception. The FAA tried to introduce tighter motion standards in its proposal for Part 60, regulating FSTD qualification and use, but removed the minimum excursions, velocities, and accelerations in the final version based on industry questions regarding their effect on training (see Bürki-Cohen et al., 2005, and FAA, 2008). In general, the experimental rigor and the reliability of the studies were not questioned. Some critics wanted transfer to be tested in the airplane instead of the simulator with motion as a stand-in for the airplane. Aside from safety, cost, and experimental control issues, however the fact that during testing with motion pilots trained without motion performed no differently than pilots trained with motion appears to preclude any conclusion that motion plays a major role in training. Others have questioned whether the results would generalize to other maneuvers, pilot populations, or airplanes. However, while our research began by testing the potentially most motion-reliant maneuvers, pilots, and simulated airplane (i.e., disturbance maneuvers, experienced pilots, wing-mounted powerhouse), our subsequent studies did take into account other aspects that may increase the need for motion. We still found no impact on training effectiveness. We therefore feel that these results would hold for other maneuvers, pilots, and airplanes trained in air-transport operations.

Internationally, a European turboprop airplane manufacturer responding to its world-wide customers' urgent need for access to the benefits of flight-simulator training used this research to specify a high-level "Full Flight Trainer" (FFT X™). The FFT uses the same data package as a Level D simulator to simulate the manufacturer's 74 passenger high-wing twin-engine turboprop airplane providing "motion cueing without a motion base." Motion is simulated via a wide FOV collimated visual system and a dynamic seat providing heave motion-onset cues via electric jacks. Vibration cues are provided via loud-speakers. The simulator thus takes advantage of the fact that humans are very proficient in perceiving motion via multiple perceptual systems, including not only the vestibular, but also the visual, tactile, and proprioceptive sensory systems. The existence of this simulator, and the fact that it was granted training and checking credits similar to a Level B FFS (including recurrent training and checking) by the French National Aviation Authorities (NAA, a member of the Joint Aviation Authorities) critically influenced the drafting of the International Civil Aviation Organization's Doc. 9625 Edition 3 draft: Manual of Criteria for the Qualification of Flight Simulators (Royal Aeronautical Society, 2008). Despite taking a generally strong stance for platform motion in its summary table with device examples, the draft takes a training-task-based approach and provides for an alternative, albeit cumbersome, à la carte approach to developing cost-effective simulators tailored to airlines' needs.

Current Research Activities

The FAA-Volpe Center heard about the FFT while participating in the international working group drafting Doc. 9625 and offered to help with its evaluation in an attempt to answer the final research question, on whether there is an affordable alternative to full platform motion. A first "proof-of-concept" phase concluded with the successful type rating of six pilots, the first time in the world that pilots were type-rated using a simulator without a hexapod-motion base. All pilots were employees of the NAA. Two were designated as "Experienced Pilots" due to the fact that they held a multipilot-crew license and had airline experience. The four "Non-Experienced" pilots held single-pilot licenses, and in one case had flown as little as 563 hours. Those latter pilots underwent a somewhat expanded curriculum in the classroom and in an FFT without seat motion that was used for both groups before using the FFT-X with seat motion. We collected opinion data throughout the course of training, which are reported in Bürki-Cohen et al., 2007. The most interesting comparisons were those collected after participants had flown the actual airplane. In those, instructors reported the pilots to be the "same as typical trainees" with regard to performance, control strategy and technique, workload, and ease of learning. The trainees themselves declared the FFT to be the "same as the airplane" with regard to handling qualities, feel and response of controls, ease of learning, comfort, workload, and overall simulator cues. They said that they used a "somewhat similar" control strategy, and rated the acceptability of the simulator as "satisfactory as is." During a final debriefing session, the trainees and instructors agreed that the transition between the FFT and the airplane had been successful. The NAA decision maker concluded that the strategy of using an FFT to focus on effective stimulation of the pilot, instead of rote simulation of the airplane, had been validated, and authorized the next phase of the FFT evaluation, which resulted in the successful type rating of 16 more pilots and also served as the set-up phase for Phase 3, a systematic comparison of the training values of the FFT with those of the FFS of the same airplane. For this FFT/FFS comparison phase, all pilots were first prepared for type-rating in the FFT. Next, they were divided into an FFT and FFS group, keeping the experience level constant between the two groups. The pilots were then brought into their assigned training device and each flew two take-offs and landings to familiarize themselves with the device. After familiarization, they were trained on V2 cuts followed by PIAs with quartering head and tail winds, and on V1 cuts followed by SSLs with microbursts. Pilots were trained in each scenario three times. After training, they filled out questionnaires. The next day, both groups were brought into the FFS and tested on a V1 cut followed by a PIA with quartering head and tail winds and on a V2 cut followed by an SSL with microburst. After testing, they filled out another questionnaire. Instructors also filled out questionnaires after training and testing. During both training and testing, the two pilots took turns flying the scenarios. This experiment is taking place in France and directed via e-mail and telephone communications with no direct supervision by the experimenters. To date, the data of 7 FFT- and 5 FFS-trained crews have been analyzed and a few interesting trends seem to be emerging. We will first describe the pilots' opinions collected in questionnaires.

This is the first experiment in our program where the motion condition (and thus perhaps the purpose of the experiment) could not be concealed from pilots. In Studies 1 through 3, pilots were tested and trained in the exact same simulator. During training without motion, we still lifted the bridge leading to the simulator and initialized the motion, before washing it out. Therefore, many pilots never realized or realized only after several maneuvers that the physical motion cues were missing. We also found no marked preference for either condition in extensive questionnaires administered to all study participants. However, in the present study, pilots knew exactly whether they entered the FFT via a few steps or the FFS over a bridge. Looking at some "pilot" questionnaires completed by pilots participating in Phase 2 and at questionnaires from two crews participating in the current Phase 3, we noted a preference for the FFS that may have been the expression of a true preference or, alternatively, a bias for the motion device arising from the knowledge that the airplane moves. After rejecting countermeasures such as blindfolding or administering pre-experiment questionnaires to diagnose bias as impractical or potentially inducing additional bias, we settled on removing the detailed questions on simulator cues, leaving only questions on simulator properties assumed to be affected by all cues, namely handling qualities, feel and response of controls, comfort, workload, acceptability, control strategy and technique, and ease of learning.

To date, we have completed preliminary analyses on the last three assessments, using the SAS General Linear Model procedure with the factors Group (FFT vs. FFS trained) and Session (Training vs. Testing). For acceptability, we found no effects of any factors, with Least Squares Means (LSMean) ratings of 3.9 by the FFS-trained group for both sessions and 3.8 vs. 4.1 by the FFT-trained group for training and testing, respectively [all F, including Group by Session interaction, $F(1,65)<1$]. The rating scale ranged from 1 (uncontrollable) to 5 (excellent). For control strategy and technique, we just missed a Group effect [$F(1,187)=2.61$, $p>.10$], but found a trend of a Group by Session interaction [$F(1,187)=3.69$, $.05<p<.10$]. According to a Tukey-Kramer multiple pairwise comparisons test on the LSMeans, this was due to a trend ($p<.10$) of the FFT-trained group to rate the simulator higher during testing than the FFS-trained group [LSMeans 3.34 vs. 3.08 on a scale of 1 (very different from the airplane) to 4 (same as airplane)]. None of the other comparisons were significant, which meant that the transition to the FFS did not increase the rating of the FFT-trained group. Finally, for ease of gaining proficiency, we found an overall effect of Group, with the FFT-trained group rating the simulators slightly lower than the FFS-trained group regardless of session [2.9 vs. 3 on a scale of 1 (very hard) to 4 (very easy), $F(1,65)=5.42$, $p<.05$].[3] In summary, the pilot opinions described to date do not indicate a marked preference for either the FFT or the FFS, despite the fact that in this study, the absence of a motion platform and thus substantial physical motion cues was obvious.

For all of our studies, a first objective analysis looked at the number of success rates, defined as successful take-offs or landings without LOC or abnormal ground contact. To be considered a success, maneuvers also had to be flown within four STDs of the most important performance variables (go-arounds, where a pilot forgoes landing in favor of another try, were eliminated from all analysis). As in all our other studies, unsuccessful maneuvers were rare and did not differ between the two groups in this study. The V1 and V2 cuts to date were flown with a 100% success rate regardless of group. The PIA and SSL success rates for the FFT group were 96 percent. The respective success rates for the motion group were 97 percent for the PIA and 92 percent for the SSL. With respect to the detailed data on pilot control inputs or workload and flight path precision, we have performed some preliminary MANOVAs on the V1 cut and on the first stage of the PIA (approach fix to decision height), the two maneuvers with the most interesting results across the studies. It appears that we have again replicated the effect of platform motion on the pedal reaction time to the V1 cut during training. It took the pilots in the FFT an average of 1.73 seconds to respond to the engine failures, 0.42 seconds longer than it took the pilots in the FFS [$F(1,68)=4.18$, $p<.05$]. This reaction time advantage, however, did not affect the heading STD [$F(1,68)<1$]. Most importantly, it again disappeared after transfer of all pilots to the FFS [$F(1,21)<1$], indicating that even very-low time pilots do not have to be trained to use motion cues. For the first stage of the PIA, the only significant difference discovered so far is a slightly lower RMS wheel response for the FFT-group during training of 6.04 degrees vs. the 7.79 degrees of the FFS-group [$F(1,67)=16.02$, $p<.01$], with no effect on localizer STD [$F(1,67)<1$]. Again, this effect disappeared when all pilots were tested in the FFS.[4]

Conclusions and Future Plans

Pending further data collection and analyses, it appears that the answer to the final research question—whether alternatives to platform-motion systems can provide onset cues and perception of realism—is yes. However, we do not really know what the contribution of the seat-motion system is to training effectiveness aside

[3] We are always reporting the results based on Type III Sum of Squares (SS).
[4] In this preliminary analysis, we performed the analyses separately for the two sessions and thus do not have any interaction data.

from added face validity compared to a system with only visual motion cues. We are currently planning to conduct similar operational testing for recurrent training and testing, again using a high-level simulator with the motion turned off and on, with the intent to pursue skill maintenance of pilots trained with and without motion over a longer period of time than just the experiment itself.

Acknowledgments

The authors thank their FAA Program Manager, Dr. Eleana Edens of the Air Traffic Organization Human Factors Research and Engineering Division, and Dr. Thomas Longridge, Manager of the FAA Flight Standards Service Voluntary Safety Programs, for their critical insight and support of this work. We also thank Dr. Shuang Wu (Computer Sciences Corporation) for her emergency statistics advice. Dr. Stephen Popkin, Director of the Volpe Center Human Factors Research and System Applications Center of Innovation, and Dr. Michael Zuschlag, Engineering Psychologist at the Center, stood by with excellent advice as usual.

References

All references to work conducted within the FAA-Volpe Center Flight Simulator Human Factors Program can be found at http://www.volpe.dot.gov/hf/pubs.html or by contacting the first author at Judith.Burki-Cohen@dot.gov

Boeing Company (2007). Statistical Summary of Commercial Jet Airplane Accidents: Worldwide Operations 1959-2006.

Cohen, J. (1988). Statistical Power Analysis for the Behavioral Sciences. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

FAA (2005). Advanced Qualification Program. (http://www.faa.gov/education_research/training/aqp/)

FAA (2008). Code of Federal Regulations Title 14. Part 60. Flight simulation training device initial and continuing qualification and use. (http://ecfr.gpoaccess.gov/cgi/t/text/text-idx?c=ecfr&sid=32a30a7ce06986548dd94d1b635acfe9&rgn=div5&view=text&node=14:2.0.1.1.1&idno=14)

Federal Register (2009). Vol. 74, No. 7 / Monday, January 12, 2009 / Proposed Rules. Pages 1288-1453. ([Docket No. FAA–2008–0677; Notice No. 08–07]). (http://edocket.access.gpo.gov/2009/pdf/E8-29584.pdf)

Pasztor, A. & Lunsford, J.L. (2009). Crash probe turns to pilot training. *Flight Safety Information 19FEB09-0*44. (from *Wall Street Journal*).

Royal Aeronautical Society (2008). Manual of Criteria for the Qualification of Flight Simulators (Doc. 9625) (draft, submitted to ICAO Aug. 29, 2008). (http://www.raes-fsg.org.uk/cgi-bin/yabb2i/YaBB.pl)

Young, L. R. (1967). Some effects of motion cues on manual tracking. *Journal of Spacecraft and Rockets, 4 (10),* 1300-1303.