

The Effect of Simulator Motion Cues on Initial Training of Airline Pilots

Judith Bürki-Cohen*

USDOT/RITA/Volpe National Transportation Systems Center, Cambridge, MA 02142

Tiau H. Go†

Massachusetts Institute of Technology, Cambridge, MA 02139

Two earlier studies conducted in the framework of the Federal Aviation Administration/Volpe Flight Simulator Human Factors Program examining the effect of simulator motion on *recurrent* training and evaluation of airline pilots have found that in the presence of a state-of-the-art visual systems, motion provided by a six-degree-of-freedom platform-motion system only minimally affected evaluation, and did not benefit training, of pilots that were familiar with the airplane. This paper gives preliminary results of a study on the effect of simulator platform motion on initial training of airline pilots that have never flown the simulated airplane.

Nomenclature

<i>AC</i>	=	Advisory Circular
<i>AGL</i>	=	Above Ground Level
<i>ARC</i>	=	Aviation Rulemaking Committee
<i>ASP</i>	=	Advanced Simulation Program
<i>FAA</i>	=	Federal Aviation Administration
<i>FAR</i>	=	Federal Aviation Regulations
<i>FD</i>	=	Flight Director
<i>FFS</i>	=	Full Flight Simulator
<i>FSTD</i>	=	Flight Simulation Training Device
<i>FSD</i>	=	Flight Simulation Device
<i>FTD</i>	=	Flight Training Device
<i>I/E</i>	=	Instructor/Evaluator
<i>ILS</i>	=	Instrument Landing System
<i>KIAS</i>	=	Knots Indicated Airspeed
<i>MAC</i>	=	Mean Aerodynamic Chord
<i>MSL</i>	=	Mean Sea Level
<i>NASA</i>	=	National Aeronautics and Space Administration
<i>NPRM</i>	=	Notice of Proposed Rule Making
<i>p</i>	=	probability of null hypothesis (i.e., no effect of motion)
<i>PF</i>	=	Pilot Flying
<i>PNF</i>	=	Pilots Not Flying
<i>RVR</i>	=	Runway Visual Range
<i>STD</i>	=	Standard Deviation
<i>t(v)</i>	=	value of <i>t</i> in the “Student” distribution of <i>t</i> with <i>v</i> degrees of freedom
<i>V₁</i>	=	take-off decision speed; the minimum speed in the take-off, following a failure of the critical engine, at which the pilot can continue the take-off and achieve the required height above the take-off surface within the take-off distance
<i>V_{1 cut}</i>	=	engine failure at or above <i>V₁</i> with continued take-off
<i>V₂</i>	=	take-off safety speed; a speed that will provide at least the gradient of climb required by the airplane

* Engineering Psychologist, Human Factors Division, DTS-79, 55 Broadway. Member AIAA.

† Aerospace Research Engineer, Center for Transportation and Logistics, 37-219, 77 Massachusetts Avenue. Member AIAA.

V_R = certification rules with the critical engine inoperative
= rotation speed; a speed that will ensure that V_2 is reached before the airplane reaches 35 ft above the runway

I. Purpose and Background

THE overall goal of this research is to ensure that flight simulators used by airlines for zero-flight-time training and evaluation of airline pilots are effective. Zero flight time refers to total training and evaluation in the simulator. When pilots graduate to the airplane for their supervised Initial Operating Experience, they are already entrusted with the safety of passengers on the airplane. Successful zero-flight-time training implies that the competencies acquired in the simulator transfer to the airplane. Vice versa, for accurate zero-flight-time evaluation, the competencies a pilot demonstrates in the airplane must transfer to the simulator. We define competency not only as the result a pilot achieves, such as compliance with flight-precision standards, but also what actions lead to this result. In other words, to assess transfer, both pilot-airplane performance (flight precision) and pilots' control inputs (behavior or workload) have to be considered.

Simulators have been used for airline-pilot training in the United States since the 1950s. Before 1980, simulators were used mainly for procedural training, whereas stick-and-rudder skills were acquired in the airplane. In 1975, the Federal Aviation Administration (FAA) instituted the Advanced Simulation Program (ASP),¹ which allowed zero-flight-time training and evaluation in simulators qualified by the FAA. Zero-flight-time brought important safety benefits such as safe training of emergency maneuvers—“[i]t's a long time since I lost a buddy in a training accident.”² Also, training and evaluation in the simulator offers the opportunity to present scripted scenarios addressing not only the motor, but also the cognitive skills increasingly needed for flying (e.g., crew resource and task management). To ensure that simulators are effective for their intended purpose as a substitute for the aircraft in training, it is critical that the required simulator cues not only are sufficient but are necessary for full transfer of performance and behavior between simulator and airplane.

That the required cues are sufficient *and* necessary is especially pertinent given the FAA's efforts to codify the existing guidelines for the evaluation, qualification, and maintenance of Flight Simulation Training Devices (FSTDs) as Federal Aviation Regulations (FAR) Part 60. These guidelines are currently contained in advisory circulars providing standards for Full Flight Simulators (FFS) Levels A through D and Flight Training Devices (FTDs) 1 through 6.^{2,3} The codification effort also endeavors to harmonize FAA and international standards. The Notice of Proposed Rule Making (NPRM) for Part 60, “Flight Simulation Device Initial and Continuing Qualification and Use,” was published September 25, 2002.⁴ This NPRM attempted to specify, in its “Table of Objective Tests,” the required motion performance of a full flight simulator by indicating minimum excursions, accelerations, and velocities in all six degrees of freedom. It also specified a required frequency response. This triggered questions from the industry on “[w]hat deficiencies in training have been recorded[...]for the currently qualified simulators” and on “the added training value by ‘higher’ proposed minimum requirements.”⁵ To address these and other comments from the industry, the FAA established a Flight Simulation Device Aviation Rule Making Committee (FSD ARC). The ARC convened representatives not only of the FAA, but also the regulated community, namely, airlines, training organizations, airplane and FSTD manufacturers, and pilot associations. In November 2003, the FSD ARC submitted recommendations to the FAA that no longer contained the motion specifications found in the original NPRM.⁶ The ARC recommendations have been published in February 2004 and commented upon by the industry. Change 1 of the NPRM is scheduled for publication in November 2005. The anticipated effective date is November 2006.⁸

The fate of the proposed motion requirements for Part 60 underscores the need for objective data derived from scientific experiments on the effects of simulator motion in airline-pilot training and evaluation. The FAA/Volpe Flight Simulator Human Factors Program provides such data. At a Joint FAA/Industry Symposium on Level B Airplane Simulator Motion Requirements, subject matter experts from FAA, industry, and academia agreed that in the absence of visual-reference cues, simulator motion should provide an early alert for sudden disturbances caused by system failures or weather.⁷ Research from multiple sources indicates that the vision-induced illusion of self-motion (vection) may indeed be too slow to develop to function as an alert.⁸ The experts at the symposium conceded, however, that there is no scientific evidence for an effect of simulator motion on transfer of flying skills between the airplane and the simulator. An extensive literature review⁹ confirmed that in some studies simulator motion helped pilots fly the simulator, but none of these effects transferred to the airplane. The same review,

²Rolfe, J., 26 March. 2001, URL: <http://www.raes.org.uk/fl-sim/FSG%2025%20Years.htm> [cited August 9, 2005].

⁸http://www.faa.gov/safety/programs_initiatives/aircraft_aviation/nsp/part60 [cited 1 August, 2005].

however, raised a number of problems with many of the studies examined, such as using visual and motion systems that are outdated today. “Bad” motion, most often motion that was insufficiently synchronized with the visual system, has been shown to have a worse effect than receiving no motion cues.¹⁰ Also, a pilot is more likely to experience vection when “enclosed” by a wide field-of-view providing peripheral visual cues.¹¹ Other aspects that might have affected the outcome of previous work are failure to control for pilot or evaluator bias, maneuvers with motion providing feedback on pilots’ control inputs rather than on outside disturbances, pilots too inexperienced to rely on motion cues, incomplete measurements (regarding the variables recorded and sampling rate), or, most often, too few participants to prevent individual differences from obscuring an effect (insufficient statistical power).⁹

II. Summary of Previous Experiments

Two earlier studies conducted in the framework of the FAA/Volpe Flight Simulator Human Factors Program examined the role of motion for training and evaluating airline pilots that were familiar with the motion of the airplane. The two studies on recurrent training and evaluation did not support the notion that motion improves transfer of training. Both used simulators qualified by the FAA for use in zero-flight-time recurrent training and evaluation. The first study used an “as is” Level C simulator of a turboprop airplane with wing-mounted engines. The procedure was to evaluate and train captains on engine failures before and after take-off in the simulator either with or without motion, and then compare their control-input behavior and flight precision for the same maneuvers in the simulator with motion as a stand-in for the airplane (quasi-transfer design). No systematic differences between the two groups were found during evaluation, training, and transfer testing in the simulator with motion, despite the fact that power analyses showed sufficient resolution to reveal operationally relevant effects. However, the analyses also showed that especially for the V_1 cut, the motion stimulation generated by the simulator was rather benign.¹² A comparison with nine other FAA qualified simulators showed that such attenuation of the motion cues for the V_1 cut is not uncommon.¹³

This first study may have legitimately raised the question of whether simulator-motion standards require further specification to ensure adequate simulation of the motion cues experienced in the airplane. Given the concomitant cost to both the users and the regulators of simulators, however, a logical follow-up study was to establish whether it was at all feasible to improve transfer between simulator and airplane by increasing the fidelity of the motion cues. This meant pushing some of the inherent limits of the six degree-of-freedom Stewart platform-motion systems used for FFSs, especially in simulating sustained lateral accelerations. In collaboration with the National Aeronautics and Space Administration (NASA), we tested the effect of enhanced motion in an FAA/NASA Boeing 747-400 Level D simulator on recurrent training and evaluation. Based on findings in the literature that translational motion cues are more important than rotational cues,^{14,15} we traded off some of pitch and roll and all of yaw to improve heave and especially sway accelerations.

As predicted in the literature, with the enhanced motion system pilots responded to the V_1 cut faster with motion than without, but the effect was less than half a second. Moreover, it minimally affected flight precision. Most importantly, although pilots without motion were always slower than pilots with motion even when they were told to expect a V_1 cut, once all pilots transferred to the simulator with motion, any difference between the two groups disappeared—regardless of whether they had been trained with or without motion. To increase both the duration of the test maneuvers and the number of control axes, this experiment also included landing maneuvers with severe weather. For these maneuvers, any effects found during evaluation and training did transfer to the simulator with motion. Interestingly, the effects found did not attest to a training benefit from motion. In fact, pilots trained without motion on an engine-out approach and landing with shifting crosswinds tracked the localizer better and with fewer wheel inputs throughout than pilots trained in the simulator with motion. The difference of about one quarter dot may not yet be operationally relevant, however. On the other hand, the fact that there were differences between groups already at the first presentation of the maneuver could be interpreted as a reason to require motion for evaluation. The questionable operational relevance of the effects and the uncertainty about which evaluation, the one with or the one without motion, more accurately reflected a pilot’s precision/behavior in the real airplane, cast doubt on this interpretation, however.^{16,17}

Having found that pilots who are very familiar with the large amplitudes and the sustained accelerations of the motion of the real airplane may not benefit from the inherently attenuated simulator motion during their recurrent training and evaluation, the study reported herein examined whether simulator motion affected airline pilots in their initial training, i.e., before they were familiar with the motions of the real airplane or even its simulator.

III. Method

A. Design Philosophy

As in the experiments on recurrent airline pilot training and evaluation, our design goals for examining the role on initial training/evaluation were twofold: First, we made sure to capture any operationally relevant effect of motion; second, we sought to avoid any confounding variables that would result in an invalid effect of motion.

With regard to the first design goal, how could anyone be sure, if no effects were found, that there really was none hidden somewhere? This is an age-old problem in scientific inquiry dating at least back to David Hume.^{18,19} The critical word here, however, is “operationally relevant.” It has been our practice, in this and previous work, to provide two pieces of information with regard to our results. First, we determine whether there is a statistically significant difference between the two groups. To accept a difference as statistically significant, we require the probability of a chance occurrence of this difference to be lower than five percent (if it is lower than 10 percent, we consider it a “statistical trend”). Second, we use power analyses to determine the smallest difference between the two groups that we could have detected given the masking effect of the idiosyncratic differences between participants.¹⁹ So, we never conclude that there is no effect of motion, but only that if there were an effect of larger than a certain size, we should have likely found it. Of course, we try to minimize the size of the detectable effect and thus increase the power of our experiments. There are two ways of achieving this: minimize the idiosyncratic differences between participants or, if that is not possible, increase the number of participants. The individual differences between participants can be reduced by selecting a homogeneous group. Any variables that could not be avoided by selection should be counterbalanced across groups to achieve overall homogeneity.

Other precautions to ensure a valid result included using a state-of-the-art wide field-of-view visual system that would indeed induce vection. We calibrated the simulator each morning before collecting data to ensure that there was no drift in simulator performance. We concealed the purpose of the experiment as much as possible to prevent participant bias from affecting the result. We chose maneuvers that were diagnostic based on the literature and our past experiments. We measured both control-input and performance variables at a sampling rate of at least 30 Hz. Although participants were assured that the experiment was non-jeopardy, we were confident that as new-hires, they were highly motivated to perform well.

When considering these design goals, please keep in mind that in order to gain access to the relevant pilot population, we needed to fit our experiment into a busy operational training environment of both an airline and its training center. This brought along not only significant time constraints, but also some other limitations with regard to simulator operation, data collection/storage, and success in maintaining the purpose of the study confidential especially for the Pilots Not Flying (PNFs) and Instructor/Evaluators (I/Es).

B. Participants

Forty-nine newly hired pilots from a large low-fare airline participated as Pilots Flying (PFs) after completion of ground school, but before they had ever entered the simulator. Four different classes presented themselves in approximately one-month intervals. PFs varied widely in experience, ranging from less than 3000 hours in a small twin-engine turboprop to almost 17000 hours mainly in jet airplanes. To achieve maximum homogeneity, the pilots were matched into “counterbalancing pairs,”** with one member of the pair being assigned to training with motion, the other without motion. The matching was accomplished by the fleet manager and the chief pilot for the simulated airplane. Based on years of experience in training new-hires for the simulated airplane, they predicted pilots’ performance based on their résumés. First, they categorized pilots as having low, medium-low, medium, medium high, and high jet experience (using as many of these categories as necessary). From these categories, they pulled pairs based on airplanes flown, positions held (captain, first-officer, check airman), and employers (airlines and/or branches of the armed forces). Very rarely, outliers with no match in their class had to be counterbalanced with pilots from a subsequent class. If a pilot lost his or her match due to data loss, he/she was counterbalanced with a pilot from a later class.

Fourteen qualified B717-200 simulator instructors participated as PNFs or as I/Es. They were randomly assigned to group (motion/no-motion) and function (PNF or I/E) so that no one PNF or I/E could effect a difference between groups. The difference between the times a PNF or I/E served in the two groups never exceeded two.

** Note that matching was used as a heuristic to achieve an overall balance, the balance within each pair was insufficient to treat the experiment as a matched-pairs design.

C. Equipment

A CAE FFS of a Boeing 717-200 airplane with two rear-fuselage-mounted jet engines and the capacity to carry approximately 106 passengers was used in the experiment. It was certified at the FAA Level D according to AC120-40B.² It was equipped with a nine-channel digital sound system. The hydraulic control-loading system ran on a digital microprocessor at an iteration rate of 3 kHz. The hydraulic digital six-degrees-of-freedom motion system had a mechanical stroke length of 54 inches with 48 inches available during operations. The motion frequency response had a 90 degrees phase lag at 8.3 Hz (45 degrees at 4 Hz and 180 degrees at 13 Hz). The three-channel Vital VIII+ collimated visual system with cross-cockpit viewing had a 180 x 40 degrees (17/23 split) field of view. The transport delays, measured from the activation of the flight control inputs to the system response, were 60 and 80 ms for motion and instruments, respectively. For the visual system, it was typically 120 ms at the 40 Hz image-update rate for dusk, dawn, and night scenes and about 100 ms at the 60 Hz image-update rate for day scenes. This is well within the FAA recommendation of 150 ms or lower for Level D flight simulators.² In comparison, the transport delay recommended for driving simulators, which have higher variations in acceleration than civil transport airplanes due to road contact, is 50 ms or lower.²¹

The simulator was used as is. To reduce I/E workload, the test maneuver scenarios (including weather, airport, and starting point of the scenario) were programmed so that the I/E only needed to select a specific test maneuver from the simulator-operator touch screen to run it. This would also activate recording of the desired variables for each maneuver. Similarly, the daily calibration tests could also be selected from the touch-screen menu.

D. Maneuvers and Airplane/Airport Variables

In our investigations on the effect of motion on recurrent training, we found that the alerting advantage of motion was most likely to manifest itself with an engine failure on take-off but still very close to the ground, where a seriously delayed response would result in a wing or tail scrape. Pilots also had difficulties handflying a single-engine raw-data Instrument Landing System approach and landing (ILS approach) with shifting crosswinds.^{††} Interestingly, it had been the no-motion pilots who appeared to fly this maneuver more precisely and with fewer control inputs. To examine these effects in initial training, we trained and tested a V_1 cut with the right or left engine failed at V_1 and a hand-flown single-engine ILS approach with shifting winds.

The simulation scenario was defined so that pilots departed from and landed at LaGuardia airport, NY, on Runway 4. Runway 4 is 7000 ft long at an elevation of 22 ft MSL. Both maneuvers were flown during the day and the temperature was 15 degrees Celsius. The zero-fuel weight of the simulated airplane was 77700 lbs. The altimeter was set at 29.92 mmHg. The autopilot was always inoperable.

For the **V_1 cut**, the ceiling was set to 100 ft, with a Runway Visual Range (RVR) of 600 ft. A Southwesterly wind blew from 220 degrees at 10 knots. The take-off weight of the simulated airplane was 87200 lbs, with the center of gravity at 11.5% MAC. The recommended flap setting was 5. The V speeds were 134 (V_1), 139 (V_R), and 135 (V_2) KIAS. The experimental run was stopped at an altitude of 1000 ft AGL.

For the simulated single-engine **ILS approach**, in addition to one engine and the autopilot, the flight director (FD) was also inoperative. The airplane was positioned about seven nautical miles from the runway at an altitude of 2250 ft AGL. The ceiling was 300 ft and the RVR 4,000 ft. During the approach, 10-knot winds shifted continuously from a 45-degree quartering head- to a 45-degree quartering tailwind.

E. Procedures and Design

1. Calibration

Every day of data collection, an automated normal approach and landing maneuver was flown to check the sensitivity of the simulator and the day-to-day consistency of the motion system. To check the sensitivity of the simulator, we measured the control responses of the simulator to automated control inputs. Then the simulator response to these control inputs was compared to the airplane data. To check the motion system, we measured the excursions of the simulator actuators to derive the roll and pitch angle motions of the simulator.

2. Briefings

All participants were provided with a written briefing giving the general purpose of the experiment. To prevent participant bias from influencing the results, the briefings did not refer to simulator motion. For the same reason, all participants were asked to treat their experiences as confidential. All briefings contained the flight plan and airport, weather, and airplane information and the specific information described below. PFs were always briefed orally by the experimenters, and PNFs and I/Es were briefed orally whenever possible.

^{††} Note that in previous work, we referred to this maneuver as a Precision Instrument Approach.

PFs were told that they would fly very challenging maneuvers to test the simulator. Their performance would reflect on the simulator, not them, and remain confidential. They were asked to fly as precisely as possible, i.e., follow the assigned heading after take-off, follow the glide slope and localizer during approach, touch down within the touch-down zone. They were told that they would evaluate the quality of the simulator in questionnaires and were given a rough idea of the sequence of flying and questionnaires.

PNFs were asked to perform their “regular Pilot-Not-Flying” duties. They were informed that they would be asked to evaluate both the quality of the simulator and PF performance/behavior in questionnaires and were given some information on the performance criteria and the flying/questionnaire sequence.

I/Es were asked to familiarize the PFs with the simulator but not to let them fly before data collection to prevent adaptation to the simulator. They were asked to talk them through the procedures so that the data collected would reflect PFs “stick and rudder” skills as opposed to procedural knowledge. To minimize instructor effect, they were asked to reserve feedback on the training runs until completion of the maneuver instead of providing in-flight coaching. The briefing contained a checklist listing the required actions throughout the experiment with a column to enter the time and comments. Finally, it also contained the performance standards²² for both maneuvers.

3. Experiment

The experiment followed a so-called quasi-transfer design. In a quasi-transfer experiment, transfer of training is tested in the simulator with motion as a stand-in for the airplane.^{23,24} This procedure was necessary to be able to fly diagnostic maneuvers in safety and to expose each pilot to exactly the same conditions. To avoid the training and adaptation effects that would have been encountered with a within-subjects design, the comparison of the effects of motion on training itself and on transfer of training was made between groups, i.e., pilots were assigned to only one of two carefully counterbalanced groups, the motion or the no-motion group. The comparison of the phases (training and transfer of training), however, was made within groups. In summary, the experiment was testing two factors, one between and one within subjects, and each factor had two fixed levels.

For the first nine pilots, the experimenter stayed off the platform to minimize interference with the proceedings. Because of the high instructor workload, however, for the rest of the pilots, the experimenter monitored the experiment procedures from the jump seat and administered the questionnaires.

The experiment sequence, as seen by the I/E, was as follows:

Phase 1: Training

[Set up appropriate simulator configuration for training (motion vs. no-motion dependent on group).]

1. Train V_1 cut (engine 1 failed):

- a. Announce V_1 cut of engine 1, fly, and give feedback.
- b. Announce V_1 cut of engine 1, fly, and give feedback.
- c. Announce V_1 cut of engine 1, fly, and give feedback.

2. Train hand-flown single-engine ILS approach (engine 1 failed):

- a. Turn FD off, announce ILS approach (engine 1 failed), fly, and give feedback.
- b. Turn FD off, announce ILS approach (engine 1 failed), fly, and give feedback.
- c. Turn FD off, announce ILS approach (engine 1 failed), fly, and give feedback.

PF, PNF, I/E complete Questionnaire 1.

Phase 2: Transfer of Training

[Set up simulator configuration for testing (motion on for all pilots).]

1. Test 1:

- a. Test V_1 cut (engine 2 failed) on take-off: Do not announce engine failure.
- b. Test ILS approach (engine 2 failed): Do not announce engine failure and turn off FD.

PF, PNF, I/E complete Questionnaire 2.

2. Test 2:

- a. Test V_1 cut (engine 2 failed) on take-off: Do not announce engine failure.
- b. Test ILS approach (engine 2 failed): Do not announce engine failure and turn off FD.

PF, PNF, I/E complete Final Questionnaire.

IV. Results

A. Data from Simulator

Close to 80 variables on pilot-airplane performance and pilots' control inputs were recorded directly from the simulator. Simulator-data collection was fully successful for 29 pilots (although for two of them, the data for the second V_1 cut transfer-of-training test was lost). For four pilots, complete data was recorded for one of the maneuvers only. For seven pilots, the file labels were missing, so we could only compare motion vs. no-motion without consideration of phase (of course, all maneuvers flown without motion were from training). For nine pilots, no simulator data was recorded. All labeled files were analyzed in two-by-two mixed Analyses Of Variance (ANOVAs). Significant interactions were examined with Bonferroni post-hoc tests on least-squares means adjusted for multiple comparisons.

1. V_1 Cut After Engine Failure to 800 ft

For the V_1 cut, data collection started at the point of engine failure and lasted up to 800 ft. The following pilot-airplane performance variables were analyzed: standard deviation (STD) of heading deviation, bank- and pitch-angle STD, indicated airspeed

deviation (average of absolute exceedance of 5 knots band around V_2), and mean absolute roll and yaw rates. The control-input variables examined included pedal-reaction time and pedal, wheel, and column responses [root mean square (RMS) of square root of the total area under the control position power-spectral-density curve].

The most important result for the V_1 cut is illustrated in Fig. 1, which shows a significant interaction between the effects of experiment phase (training vs. transfer of training) and group for pilots' pedal-reaction time [F(1,58)=4.27, $p < 0.05$]. On average, pilots responded 0.47 s faster to the engine failure during training with motion than without motion. However, this effect was not quite statistically significant ($p < .10$). Most importantly, it did not transfer to the simulator with motion: when all pilots received motion cues during the transfer-of-training tests, all pilots responded equally fast regardless of the simulator configuration during training ($p = 1.0$). This was due to a significant improvement of the response time of the no-motion pilots during transfer of training by 0.52 s, presumably because the motion cues alerted them of the engine failure ($p < 0.05$).

Figures 2 through 4 show motion effects on variables reflecting longitudinal control. Figure 2 shows that the motion group kept the RMS of their column response an average of 0.16 inches steadier than the no-motion group [F(1,58)=7.94, $p < .01$]. This effect did not interact with phase [F<1], so this difference held even once all pilots transferred to motion.

The steadier column control may have helped the motion pilots to comply better with the airspeed. Figure 3 shows a 1.34 knots smaller average airspeed exceedance [F(1,59)=5.38, $p < 0.01$] for the motion group. This effect again did not interact with phase [F<0.1].

As can be seen in Fig. 4, however, the pitch angle was kept marginally steadier by the no-motion group [F(1,57)=3.60, $p < 0.10$]. On average, the no-motion group had a 0.38 degrees smaller pitch-angle STD than the motion group (0.24 degrees during training and 0.52 degrees during transfer). This effect did not interact with phase [F(1,57)=1.56, $p > 0.10$].

None of the other variables examined showed any significance.

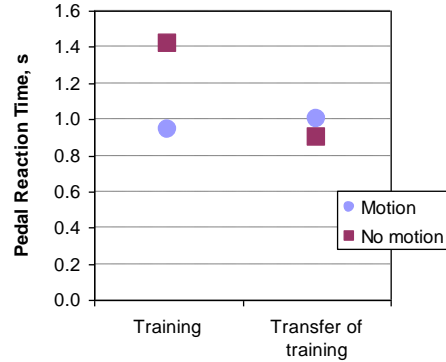


Figure 1. V_1 cut pedal-reaction time by phase.

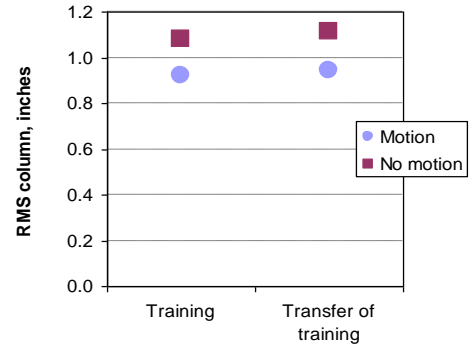


Figure 2. V_1 cut RMS column by phase.

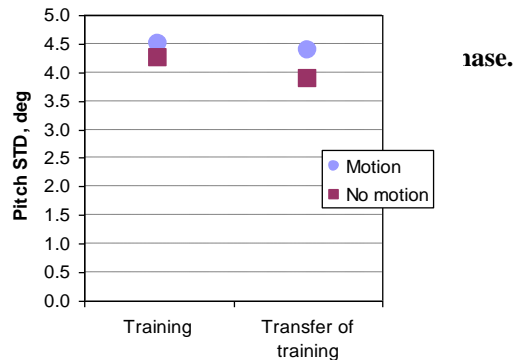
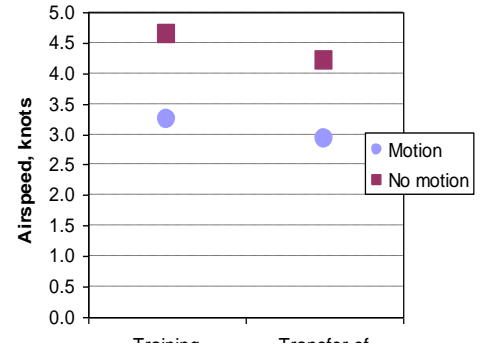


Figure 4. V_1 cut pitch STD by phase.

2. ILS Approach From 1450 ft AGL to Decision Height

For the ILS approach, the data from 1450 ft AGL to the decision height at 250 ft AGL were examined. For the pilot-airplane performance, the following variables were analyzed in addition to the ones analyzed for the V_1 cut: STDs of the localizer and glide-slope deviations and the averages of the localizer and glide-slope exceedances (absolute deviation exceeding ± 0.5 dots around the reference). For the control-input behavior, the variables examined were the same as for the V_1 cut excluding pedal-response time.

Most importantly, there were no interactions between the variables of group and phase, so all phase effects reported were identical for the motion and the no-motion group and all group effects transferred to the transfer-of-training phase when all pilots were tested with motion.

Figure 5 shows the only ILS-approach group effect, namely, the no-motion group held the pedal 0.08 inches steadier than the motion group, with no apparent effect on any of the performance variables [$F(1,60)=5.94$, $p<0.05$].

Table 1 shows that there were many phase effects, all indicating improvement between training and transfer of training. There were no interaction effects (all $F<2.27$, $p>0.10$), so these improvements occurred regardless of whether pilots were trained with or without motion.

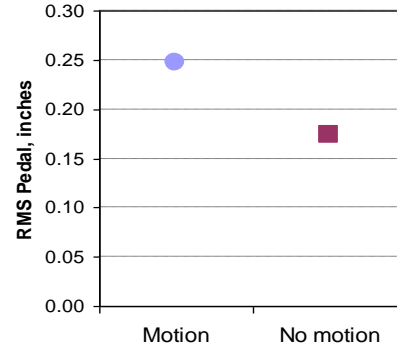


Figure 5. ILS pedal RMS by group.

Table 1. ILS approach—Effects of experiment phase.

Variable	Mean		Statistics	
	Training	Transfer	F	p<
STD heading (degrees)	6.06	4.69	(1,60)=4.53	0.05
STD bank (degrees)	5.01	4.29	(1,60)=3.58	0.10
STD pitch (degrees)	1.59	1.17	(1,60)=9.28	0.01
Yaw activity (deg/s)	0.73	0.62	(1,60)=3.80	0.10
Roll activity (deg/s)	1.70	1.41	(1,60)=4.69	0.05
STD glide slope (dot)	0.41	0.30	(1,59)=4.80	0.05
Localizer exceedance (dot)	0.36	0.20	(1,59)=2.81	0.10
RMS wheel (degrees)	11.29	9.25	(1,60)=10.20	0.01
RMS column (inches)	0.67	0.51	(1,60)=13.53	.001

3. Power Analyses

At transfer, the only effects of motion found in this experiment were an improvement of longitudinal control for the V_1 cut and an increase in pedal RMS for the ILS approach. This raises the question of whether the resolution of the experiment was good enough to be reasonably certain that any other effects that would be operationally relevant could have been found. As in our previous research, we follow the general convention in defining “reasonable certainty.” Table 2 shows the smallest effect sizes that should have been found accepting a risk of 5 percent of falsely rejecting the null hypothesis of no difference between the two groups ($p<0.05$) and a risk of 20 percent of falsely rejecting the hypothesis that there is a difference between groups (power of 0.80).²⁰

Table 2. Group and phase minimum detectable effect sizes.

Maneuver	Selected Measures	Effect Size for power=0.80
V ₁ cut	Pedal reaction time (seconds)	0.38
	STD bank angle (degrees)	0.63
	STD heading (degrees)	0.52
	RMS wheel (degrees)	1.90
	RMS pedal (inches)	0.07
ILS Approach	STD glide slope (dot)	0.14
	STD localizer (dot)	0.20
	STD heading (degrees)	1.85
	RMS wheel (degrees)	1.84
	RMS pedal (inches)	0.09

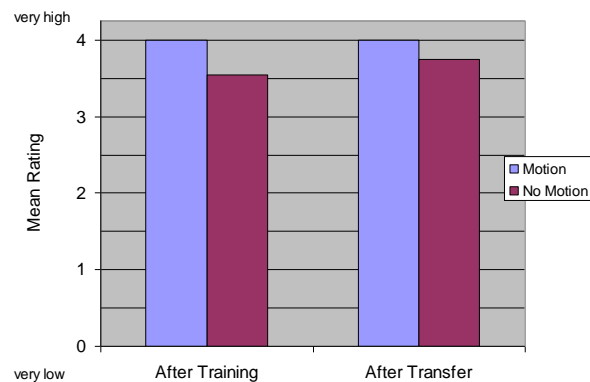
B. Participants' Perceptions

Participants perceptions were collected via structured questionnaires before and after transfer to the simulator with motion. A final open-ended questionnaire is not presented here. PFs were asked to communicate their perceptions of the simulator's acceptability, handling qualities, cue fidelity and their own comfort, mental and physical workload, and ability to gain proficiency in the simulator via questionnaires. PNFs were asked about the simulator cues and acceptability, PFs' performance, workload, and ability to gain proficiency, and their own comfort. I/Es compared PFs' control-input behavior, performance, workload, and ability to gain proficiency to that of an average student. All ratings were on a five-point scale. A rating of one was anchored with some appropriate expression for "unacceptable." A rating of five was anchored with an appropriate version of "excellent," as indicated along the y-axes in Fig. 6 through 9.^{**} To gain valid insights from the questionnaire data, it was critical that the sequence indicated in the procedures was followed exactly. Therefore, only the questionnaires administered by the experimenters were analyzed (N=40).

Four two-tailed t-tests were performed for each (sub)question. The first two tests examined differences *between* the responses of the motion and the no-motion group, one after training, the other after transfer of training. The third and the fourth test examined whether there were differences *within* each group in how they responded before and after transfer of training, when both groups received motion cues. Although "multiple t-tests" are often frowned upon because they increase the chances of falsely rejecting the null hypothesis that there is no difference between the motion and the no-motion group, they serve the purpose of this preliminary paper to pinpoint any possible effects of motion.

1. Pilots Flying

Only one single statistically significant difference was found between the responses of the pilots trained with motion and those trained without: During training, the pilots experiencing motion found the simulator more acceptable than the pilots flying the simulator without motion [$t(38)=2.33, p<0.05$].

**Figure 6. Acceptability ratings by phase.**

^{**} Note that although the anchor for the highest possible rating of 5 is given, the y-axes are labeled only up to 4. Some participants chose to give half-point ratings (e.g., 3.5).

As can be seen in Fig. 6, after training the motion pilots gave the simulator an average acceptability rating of 4, whereas the no-motion rated it at 3.55. Interestingly, however, the transferring to motion did not quite significantly improve the acceptability rating of the no-motion [$t(19)=1.75, p<.10$], despite the fact that the ratings two groups were now statistically equivalent [$t(38)=1.35, p>0.10$].

Figure 7 illustrates the only other between-difference in the ratings by the PFs that approached statistical significance, and this only for the V_1 cut [$t(37)=1.83, p<0.10$]: Before transfer to motion, the motion group rated the “overall” quality of the simulator cues at 4.08, whereas the no-motion group gave it a lower rating of 3.74. Again, however, the no-motion group did not significantly increase their ratings after they experienced motion [$t(19)=0.30, p<0.10$]. Also, for the ILS approach, there were no differences between the average “overall” cue ratings as a function of motion configuration or phase. This is especially remarkable given that some pilots were aware of the absence of motion, either because they noticed it, or because the purpose of the experiment had leaked.

As in our previous work, there was no evidence of sensory conflict inducing discomfort in the no-motion condition. As shown in Fig. 8, PFs and PNFs rated their comfort similar in the two simulator configurations, if anything, the rating of the no-motion group decreased after transfer to motion [all $t(38)<1.67; p>.10$]. The average rating across phases and motion conditions was 3.78.

Figures 9 and 10 show two within-group differences found for the motion group only: After transfer to motion, the motion group’s perception of physical workload improved from 3.15 to 3.5 for the V_1 cut and from 2.35 to 2.7 for the ILS approach [$t(19)>2.39, p<0.5$]. (Note that the higher any rating, the better. Thus, as the workload rating increased, the workload itself was perceived as being lower). There were no differences in workload ratings within the no-motion group, perhaps because any adjustment to flying the simulator was counteracted by now having to deal with motion. Note that there were no between-group differences in workload ratings for either of the questionnaires [$t(37$ and $38)<1.25, p>0.10$].

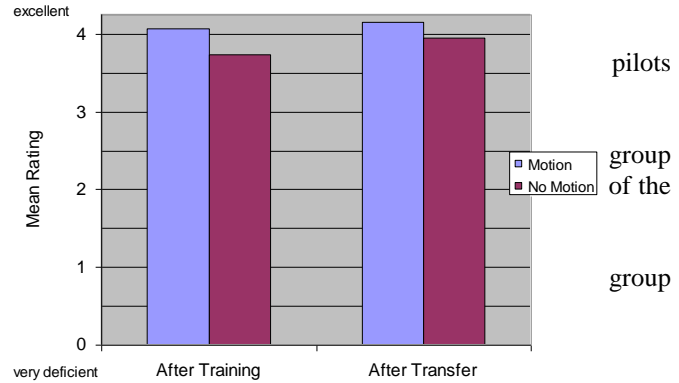


Figure 7. V_1 cut “overall” simulator-cue ratings.

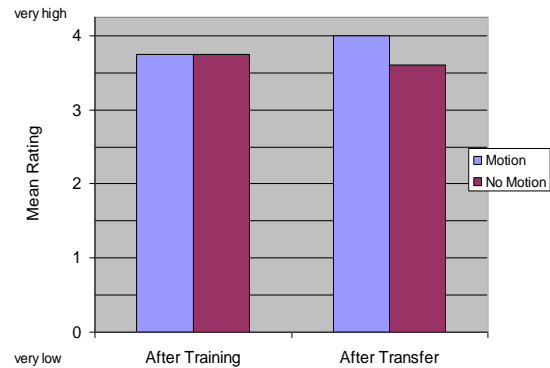


Figure 8. Ratings of comfort in simulator

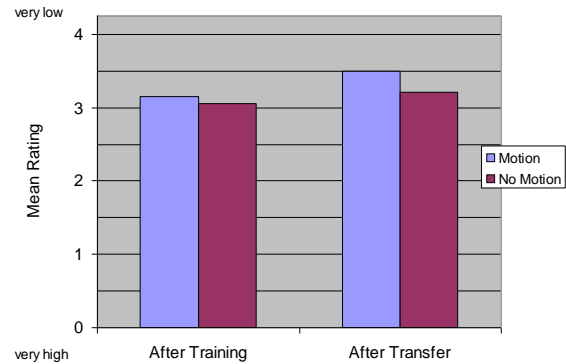


Figure 9. V_1 cut physical workload ratings.

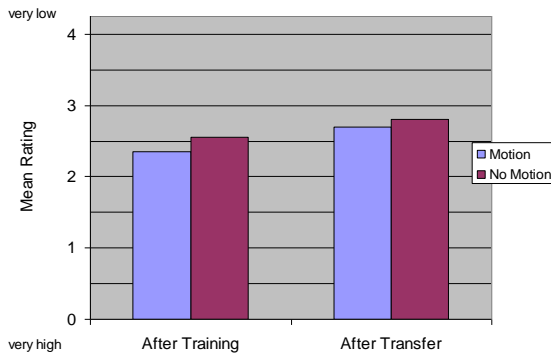


Figure 10. ILS physical workload ratings.

higher with than without motion [V_1 cut: 4.15 vs. 3.74, $t(37)=2.10, p<.05$; ILS approach: 4.15 vs. 3.8, $t(38)=2.04,$

$p < 0.05$]. These effects are confirmed by corresponding within-group effects: PNFs rated the cues for both maneuvers higher after transfer to motion [$t(18 \text{ and } 19) > 2.72$, $p < 0.05$].

For the I/Es, there were no significant between-group effects of motion, with the exception of a statistical trend to rate the ILS-approach performance of the motion group higher than the performance of the no-motion group before transfer [3.7 vs. 3.05, $t(38) = 1.72$, $p < 0.10$]. Once both groups transferred to motion, however, I/Es rated their performance almost identically [3.65 vs. 3.7, $t(38) < 1$]. This trend is confirmed by a within-group difference: The performance ratings for the no-motion group significantly increased after transfer to motion [$t(19) = 3.20$, $p < 0.05$].

There was only one more within-group effect of phase: The PNFs rated the V_1 cut workload of the no-motion group lower after transfer to motion [3 vs. 3.34, $t(18) = 2.65$, $p < 0.05$]. Finally, there was a statistical trend for the PNFs to rate the acceptability of the simulator higher after transfer to motion [4.4 vs. 4.55, $t(19) = 1.88$, $p < .10$].

V. Summary and Discussion

Below is a summary of the preliminary results from this study of the effect of motion on initial training of a V_1 cut and an engine-out ILS approach in a simulator of a single-aisle twinjet with aft-fuselage mounted engines. The results are discussed with reference to our previous studies.

First, the study confirmed the small but statistically significant alerting effect of motion found in the recurrent study with enhanced motion,^{16,17} although for initial training, the effect was only marginally significant. Even forewarned of an engine failure, pilots without motion cues remained unable to respond to an engine failure on take-off as fast as pilots with motion cues. It also showed, however, that like experienced pilots, pilots unfamiliar with the motion cues encountered in the airplane were able to catch up immediately once they receive motion cues, in other words, they did not have to be trained with motion to recognize the cues signaling an engine failure on takeoff. During the transfer portion of the study, all pilots responded equally fast for the V_1 cut, regardless of the simulator configuration employed during training. With platform motion, the no-motion trained pilots improved significantly in response time, presumably because the motion cues alerted them to the engine failure.

Second, for the V_1 cut only, motion appeared to help pilots to keep the column steady, which in turn helped them with airspeed—but not pitch angle—control. Recurrent pilots in the simulator with enhanced motion had controlled pitch angle better with motion, but only during the very first exposure to the V_1 cut. Already with the second V_1 cut, which was still flown without motion by the no-motion group, the difference between groups was gone. For both studies, the effects were small, and their operational relevance will need to be assessed by the operators themselves.

Third, although both groups improved on many variables for the ILS approach between training and transfer of training, the only group effect found was steadier pedal control for the no-motion group throughout. The recurrent study with enhanced motion had also found an overall steadier control strategy for the no-motion group, but for the wheel, not for the pedal. Also, the improved flight precision without motion found for recurrent pilots was not replicated with initial pilots.

Fourth, participants' perceptions did not indicate a marked preference for either of the two conditions. Most importantly, again there was no evidence that the sensory conflict between eyes and vestibular apparatus induced discomfort in the no-motion condition.

Whether the overall statistical power of the experiment was sufficient to find all operationally relevant effects will need to be decided by the operators based on Table 2, which lists the smallest detectable effects. They show that the resolution of this experiment lies somewhere in-between those of the two previous studies with recurrent pilots.

Moreover, a few additional analyses need to be performed before coming to final conclusions on this study. Two more segments of the ILS approach remain to be analyzed. Multivariate analyses will give us a better overall feel of the effect of motion and will give us a better estimate of power. The individual training progress of pilots also needs to be considered by examining the individual training runs (for this study, we computed the average value for each pilot). More detail can be extracted from participants' perceptions.

Nevertheless, despite all the difficulties of conducting an experiment in an operational training environment and with a very heterogeneous group of pilots, this study of the effect of the motion provided by a Level D FFS on initial airline training was powerful enough to find a number of interesting effects that fit in well with our previous research on the effects of simulator platform motion in recurrent training.

Acknowledgments

The authors thank Dr. Eleana Edens from the FAA Air Traffic Organization Human Factors Research and Engineering Division and Dr. Thomas Longridge from the FAA Flight Standards Service Voluntary Safety programs, for helpful comments and support of this work. We also thank “who” has volunteered “his” time to review the paper and provide valuable suggestions. Finally, we thank all the participants and their employers: the

PFs, who accepted to fly extremely stressful maneuvers without any previous experience in the simulator and did exceptionally well, the PNFs and I/Es, who were working on a tight schedule, and the simulator technicians, who really didn't need all the additional work we loaded on them.

References

- ¹Federal Aviation Regulation, "Advanced Simulation Plan," 14 Code of Federal Regulations, Part 121, Appendix H, U.S. Government Printing Office, Washington, DC, 1980.
- ²Federal Aviation Administration, "Airplane Simulator Qualification," Advisory Circular No. 120-40B, U.S. Department of Transportation, Washington, DC, 1991.
- ³Federal Aviation Administration, "Airplane Flight Training Device Qualification," Advisory Circular No. 120-45A, U.S. Department of Transportation, Washington, DC, 1992.
- ⁴U.S. Department of Transportation/FAA, Notice of Proposed Rulemaking, "Flight Simulation Device Initial and Continuing Qualification and Use," Docket Management System FAA-2002-12461, filed September 25, 2002.
- ⁵Air Transport Association of America, Inc., Simulator Technical Issues Group, "Comments," Docket Management System FAA-2002-12461-29, filed February 19, 2003.
- ⁶U.S. Department of Transportation/FAA, Report, "FSD ARC Recommendations of Part 60 Notice of Proposed Rulemaking," Docket Management System FAA-2002-12461-84, filed February 3, 2004.
- ⁷Bürki-Cohen, J. (Ed.), "Joint FAA/Industry Symposium on Level B Airplane Simulator Motion Requirements," Washington Dulles Airport Hilton, June 19-20, 1996, electronic transcript, URL: <http://www.volpe.dot.gov/opsad/valreqju.html> [cited 15 July 2005].
- ⁸Howard, I.P., "The Perception of Posture, Self Motion, and the Visual Vertical," *Handbook of Perception and Human Performance: Vol. 1. Sensory Processes and Perception*, edited by K.R. Boff, L. Kaufman, and J. Thomas, New York, 1986, pp. 18-1 to 18-62.
- ⁹Bürki-Cohen, J., Soja, N., and Longridge, T., "Simulator Platform Motion—The Need Revisited," *International Journal of Aviation Psychology*, Vol. 8, No. 3, 1998, pp. 293-317.
- ¹⁰Chung, W.Y., Schroeder, J.A., and Johnson, W.J., "Effects of Vehicle Bandwidth and Visual Spatial-Frequency on Simulation Cueing Synchronization Requirements," *AIAA Atmospheric Flight Mechanics Conference*, Collection of Technical Papers (A97-37244 10-08), AIAA, Washington, DC, 1997.
- ¹¹Mack, A., "The Perception of Posture, Self Motion, and the Visual Vertical," *Handbook of Perception and Human Performance: Vol. 1. Sensory Processes and Perception*, edited by K.R. Boff, L. Kaufman, and J. Thomas, New York, 1986, pp. 17-1 to 17-38.
- ¹²Go, T.H., Bürki-Cohen, J., and Soja, N.N., "The Effect of Simulator Motion on Pilot Training and Evaluation." *AIAA Modeling and Simulation Technologies Conference* (AIAA 2000-4296), AIAA, Washington, DC, 2000.
- ¹³Bürki-Cohen, J., Soja, N.N., Go, T.H., Boothe, E.M., DiSario, R., and Jo, Y.J., "Simulator Fidelity: The Effect of Platform Motion," DOT/FAA/RD-01/XX, 2001, unpublished.
- ¹⁴Schroeder, J.A., "Helicopter Flight Simulation Motion Platform Requirements," NASA TP-1999-208766, July 1999.
- ¹⁵Bray, R.S., "Initial Operating Experience with an Aircraft Simulator Having Extensive Lateral Motion," NASA TM X-62155, 1972.
- ¹⁶Bürki-Cohen, J., Go, T.H., Chung, W.Y., Schroeder, J.A., Jacobs, S., and Longridge, T., "Simulator Fidelity Requirements for Airline-Pilot Training and Evaluation Continued: An Update on Motion Requirements Research," Proceedings of the 12th International Symposium on Aviation Psychology, April 2003.
- ¹⁷Go, T.H., Bürki-Cohen, J., Chung, W.Y., Schroeder, J.A., Saillant, G., Jacobs, J., "The Effects of Enhanced Hexapod Motion on Airline Pilot Recurrent Training and Evaluation," *AIAA Modeling and Simulation Technologies Conference* (AIAA-2003-5678), AIAA, Washington, DC, 2003.
- ¹⁸Morgan, P.L., "Null Hypothesis Significance Testing: Philosophical and Practical Considerations of a Statistical Controversy," *Exceptionality*, Vol. 11, No. 4, 2003, pp. 209-221.
- ¹⁹Streiner, D.L., "Unicorns Do Exist: A Tutorial on 'Proving' the Null Hypothesis," *Canadian Journal of Psychiatry*, Vol. 48, No. 11, 2003, 756-761.
- ²⁰Cohen, J., *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed., Lawrence Erlbaum Associates, Hillsdale, New Jersey, 1988.
- ²¹Kemeny A., "Perception and Simulation," *Proceedings of the Driving Simulation Conference*, Paris, July 1999, pp.13-28.
- ²²Federal Aviation Administration, "Airline Transport Pilot and Aircraft Type Rating: Practical Test Standards for Airplane," FAA-S-8081-5D with change 1, Flight Standards Service, U.S. Department of Transportation, Washington, DC., 2001.
- ²³Boldovici, J.A., "Simulator Motion," ARI Technical Report 961, U.S. Army Research Institute, Alexandria, VA, 1992.
- ²⁴Taylor, H.L., Lintern, G., and Koonce, J.M., "Quasi-Transfer as a Predictor of Transfer From Simulator to Airplane," *The Journal of General Psychology*, Vol. 120, No. 3, 1995, pp. 257-276.