

ACQUIRING MANUAL FLYING SKILLS IN A VIRTUAL REALITY FLIGHT SIMULATOR

Wietse D. Ledegang

Erik van der Burg

Ivo V. Stuldreher

Mark M.J. Houben

Eric. L. Groen

TNO, Human Performance, The Netherlands

Danny van der Horst

Erik A.M. Starman

Guido Almekinders

Royal Netherlands Air Force, The Netherlands

In this study, we explored the possibility of objectively assessing the progress in manual flying skills by student pilots using Virtual Reality (VR). Using a VR flight simulator of the Pilatus PC-7 training aircraft, fifteen participants without flying experience practiced basic flight maneuvers based on self-study and without receiving feedback. Relevant flight performance measures were normalized and a learning curve was fitted, representing learning speed and end-level. During some runs an N-back task was included as a secondary task to quantify the participants' cognitive capacity. Interestingly, performance on the N-back was not a good predictor of someone's learning curve. The correlation between performance measures and flight instructor gradings confirmed that, for a limited set of maneuvers, we were able to objectify the students' learning behavior of acquiring a set of manual flying skills in a VR flight simulator. The results of this study show the potential of measuring learning performance in VR.

Aspiring military pilots within the Royal Netherlands Air Force (RNLAf) undergo Elementary Military Pilot Training (in Dutch: Elementaire Militaire Vlieger Opleiding, EMVO) in the Pilatus PC-7 turboprop training aircraft. Because of its lifetime, the PC-7 aircraft will be replaced by a new training capacity in 2026. In addition to a new aircraft and high-fidelity simulation training, Virtual Reality (VR) is identified as a potential training means to accomplish part of the training objectives in the future training syllabus. VR has already been introduced in initial pilot training within the Royal Air Force (RAF), United States Air Force (USAF) and Royal Australian Air Force (RAAF) (Pope, 2019; Air Education and Training Command, 2020; Lewis & Livingston, 2018; Pennington et al., 2019). Ross (2022) concludes that, based on the results of eighteen studies performed between 2018 and 2021, student pilots trained in VR performed at least as well as students trained with traditional means. Furthermore, a combination of VR and traditional flight training can decrease the training time required (Lewis & Livingston, 2018; McCoy-Fisher et al., 2019; Pope, 2019; Sheets & Elmore, 2018; Pennington et al., 2019; Mishler et al., 2022).

In this study, we explored the possibility of using a VR flight simulator to objectively measure the learning performance of student pilots while acquiring manual flying skills and associated visual behavior. In VR, objective performance measures can be recorded, which may

be helpful for monitoring the student's progression. These measures may be derived from control inputs, flight performance, and the gaze behavior as recorded by a built-in eyetracker.

Learning theory shows that during learning the ability to execute tasks evolves from slow and effortful controlled processing to fast and less effortful, or automatic, processing (Tinga et al., 2019; Schneider en Chein, 2003). In this way, learning improves task proficiency while cognitive demands decrease. We therefore hypothesized that an increase in so-called 'cognitive spare capacity' can indicate learning. These considerations led us to define two research questions: 1) Can we determine an overall learning curve for the acquisition of technical flying skills, based on various performance measures obtained across a limited set of basic flight maneuvers, and 2) does the cognitive spare capacity of student pilots correlate with their ability to learn these basic flying skills? Note that the learning of associated visual behavior is described in a separate paper, see Stuldreher et al. (*in press*).

Method

Participants

Fifteen military cadets (12 males and 3 females) of the Royal Military Academy participated in this study (mean age: 23.7 years, $\pm$ standard deviation of 2.4 years). They had an average of 3.6 ± 7.8 hours of flight experience on powered- and glider aircraft and 2.4 ± 7.7 hours on flight simulators. Prior to the experiment, all pilots signed an informed consent, stating that the details of the experiment had been sufficiently explained and that they participated voluntarily. The experiment was conducted with the approval of the institutional ethics committee and was in accordance with the (revised) Helsinki Declaration.

Materials

The simulator environment (see Figure 1), developed by multiSIM BV, consisted of a fixed-base cockpit (front-seat) of a Pilatus PC-7 turboprop trainer aircraft and control devices with control loading. A VARJO-Aero VR headset with built-in eye-tracker was used to present the cockpit and virtual environment near Woensdrecht Air Force Base, The Netherlands, rendering at 90Hz. The flight model characteristics were comparable to the PC-7 aircraft and were validated by EMVO flight instructors.

Procedure

The participants repeatedly practiced three flight maneuvers in a fixed order: Straight-and-Level flight (SAL); Speed Change (SC); and Level Turn (LT). Each maneuver was performed three times during runs of 210 seconds each, followed by a fourth run in which the same maneuver was performed while simultaneously executing an additional memory task as a measure of cognitive spare capacity. Each block of four consecutive runs was repeated three times, spread over two days, thus cumulating to twelve runs per maneuver (i.e., 36 runs overall).

The primary task consisted of manual aircraft control, including the instrument scan and lookout. The secondary task during each fourth run consisted of an auditory 2-back memory task (Kirchner, 1958), which required the participant to continuously update their working memory (i.e., remembering the last two letters of an auditory sequence of continuously changing letters at a fixed 3-seconds interval with a 25% repetition probability). The participants were instructed to

respond self-paced by pressing a button on the throttle if the letter heard was identical to the letter two trials back and to withhold a response if the letter was different.

Prior to each block, the participants studied standardized instruction material that included a video of each flight maneuver in which a flight instructor explained the task in the same simulation environment. During the experiment participants received no feedback on their performance.



Figure 1. Setup of the VR simulator during the experiment, with the participant inside the cockpit mock-up and the experimental test leader behind the instructor station.

Measurements

For the analysis of the SAL maneuver the parameters during an entire run were used, while for SC and LT maneuvers the parameters were extracted during phases of deceleration and turning, respectively. To compare errors in performance (i.e., the deviation of a parameter from a target value) across different flight parameters, these measures were normalized in relation to the largest error observed across all participants and combined into an overall performance measure (ranging from 0: worst performance to 1: perfect performance).

Although very simplified, learning curves were estimated by fitting two linear functions representing a ‘learning part’ and an ‘end-level’ on the runs without the N-back task. Learning speed is quantified by the number of runs needed to reach end-level performance. It is assumed that the combination of a high end-level and fast learning resembles a high learning performance.

For the N-back task the percentage of errors (i.e., miss or false hit) was calculated. A baseline-corrected number of errors was calculated by subtracting the number of errors that were made in the N-back task prior to the experiment (i.e., without flying).

After the experiment, a flight instructor graded a semi-random selection of 27 runs based on a video replay of each recording. For each of these runs the instructor rated Overall performance, Basic aircraft control, and Multi-tasking according to EMVO grading categories (i.e., Unsatisfactory 1-3, Fair 4-6, and Good 7-9).

Statistical analysis

For each maneuver, an explorative analysis was conducted to examine which performance measures varied significantly over the twelve runs. This was done in separate repeated-measures Analyses of Variance (ANOVAs) with run (1-12) as within-subjects variable. The normalized performance in runs with the N-back task (runs 4, 8, and 12) was compared to the preceding runs without the N-back task (3, 7, and 11, respectively) by means of three separate two-tailed t-tests. A repeated-measures ANOVA was conducted on the overall normalized performance, averaged over the SAL, SC and LT maneuvers together, as function of the different runs for all participants. In all analyses, alpha was set to .05.

Results

The results show that learning to fly SAL is related to the errors in airspeed, roll, altitude and heading as these measures show significant effects as function of run. Learning to execute the SC is related to the errors in altitude, airspeed and heading, while learning of the LT is related to the errors in altitude, roll and side slip. See Table 1 for an overview.

After normalizing the flight performance measures and combining these into an overall mean normalized performance, repeated measures ANOVA shows significant main effects of run for SAL ($F(11, 154) = 17.556, p < .001$), LT ($F(11, 154) = 8.042, p < .001$) and SC ($F(11, 154) = 12.827, p < .001$). This indicates that for all maneuvers the participants were able to improve their performance with more repetitions.

Table 1.

Relevant performance measures and learning curve fit details per flight maneuver.

	Straight-and-Level (SAL)		Speed Change (SC)		Level Turn (LT)	
Performance errors	<i>p</i>	<i>F</i> (1,11)	<i>p</i>	<i>F</i> (1,11)	<i>p</i>	<i>F</i> (1,11)
Airspeed	<.001	7.06	.004	2.68		
Roll	<.001	3.81			.022	2.12
Altitude	<.001	6.25	<.001	8.45	<.001	7.61
Heading	<.001	9.68	<.001	7.42		
Side slip					.042	1.91
Learning curve fit	Mean (Std)	Range	Mean (Std)	Range	Mean (Std)	Range
R ²		.93		.83		.69
Start level	.56 (.16)	.21-.79	.59 (.21)	.19-.82	.69 (.17)	.30-.87
End level	.80 (.08)	.58-.89	.86 (.09)	.63-.94	.86 (.08)	.67-.94
Time to end level	8.36 (1.78)	5.2-11.0	5.86 (2.89)	1.96-11.0	6.78 (3.25)	2.04-11.0
Learning speed	.69 (.09)	.58-.89	.73 (.11)	.63-.94	.74 (.09)	.67-.94
Learning performance	1.06 (.06)	.94-1.19	1.14 (.06)	.99-1.25	1.14 (.09)	.95-1.30

Even though it is very simplified to fit a linear learning curve on the normalized performance measures, the fits show good results (i.e., R² varies from .69 to .93). Participants were able to improve their manual flying skills up to a normalized end-level of .80, .86 and .86 within, on average, 8.36, 5.86 and 6.78 runs for the SAL, SC and LT maneuvers respectively. Due to the normalization procedure, the impression can be given that SAL was the most difficult to learn (i.e., most runs needed to achieve end-level). However, because SAL was quite easy,

only small performance improvements could be achieved with each repetition, which took longer to reach end-level. See Table 1 and Figure 2 for more details.

Comparing flight performance between runs with the N-back task (i.e., the fourth run of a session) and their preceding run (i.e., the third run) yielded no significant effect for the SAL and LT maneuver, indicating that flight performance did not improve when the participants performed the additional N-back task. Since flight performance did improve across the first three runs without the N-back task, it appears that the additional N-back task interfered with learning. For the SC maneuver the analysis even yielded a significant drop in performance between run 3 and 4, $t(14) = 2.205$, $p = .045$, and between run 7 and 8, $t(14) = 3.385$, $p = .004$. There was no difference between runs 11 and 12, $t(14) = 0.790$, $p = .443$. During the SC the N-back task thus not only caused ‘stagnation’ of the learning, but even a performance decline.

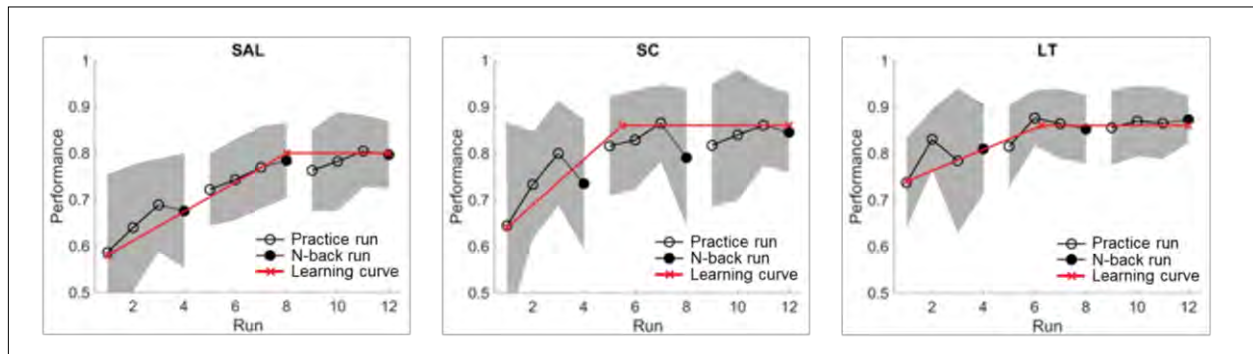


Figure 2. Normalized flight performance as function of run for the Straight-and-Level (SAL), Speed Change (SC) and Level Turn (LT) maneuvers. The filled symbols correspond to the runs with N-back task, while the shading reflects the standard deviation. The optimal fit of a learning curve, in terms of a ‘learning part’ and ‘end-level’, is shown with a red line.

The ANOVA on the baseline corrected N-back task with run and flight maneuver as within-subject variables showed a two-way interaction, $F(4, 56) = 2.842$, $p = .032$, as well as main effects of flight maneuver, $F(2, 28) = 22.757$, $p < .001$, and run, $F(2, 28) = 5.599$, $p = .009$. Participants made more errors in SC (7.7%) than in SAL (3.7%) and LT (4.1%) maneuvers, $t(14) = 5.231$, $p < .001$, and $t(14) = 5.309$, $p < .001$, respectively. The two-way interaction was further examined by separate ANOVAs for each flight maneuver. This yielded a significant effect of run for SAL, $F(2, 28) = 5.989$, $p = .007$, but not for LT, $F(2, 28) = 2.370$, $p = .112$, and SC, $F(2, 28) = 2.687$, $p = .086$. For the SAL maneuver, the baseline-corrected N-back performance improved over runs. Separate two-tailed t-tests showed that the baseline-corrected value significantly differed from zero in run 4, $t(14) = 3.129$, $p = .007$, but not for run 8 and 12 (p values $\geq .074$). This indicates that the participants were able to perform the N-back task while flying SAL after a few runs, confirming that SAL allowed some degree of cognitive spare capacity.

Computing correlations between N-back task performance and learning curve metrics yielded only one significant negative correlation ($r = -.52$, $p = .045$) in the LT maneuver, which is driven by the ‘end-level’ component ($p = .051$). There are no significant correlations between the N-back task performance and learning curve metrics in the other flight maneuvers or when averaging over the three flight maneuvers.

Finally, normalized performance measures showed significant positive correlations with the instructor ratings for Overall performance ($r = .76$, $p < .001$), Basic aircraft control ($r = .70$, $p < .001$) and Multi-tasking ($r = .59$, $p < .005$).

Discussion

Our primary interest was how the progress in performance (i.e., the learning ability) of student pilots could be measured using objective measures extracted from a VR flight simulator. Because comparing flight performance across various flight parameters in different maneuvers is not trivial, our data analysis had a strong exploratory character, in particular when estimating the participants' learning performance. The extent to which the learned skills, on the limited set of maneuvers, transfer to actual flying still needs to be investigated.

The data shows that the performance of the participants during the three maneuvers could be described with a limited set of objective performance measures, which were normalized and combined into an overall performance measure on which a learning curve was fitted. Although fitting a learning curve by a linear function is an over-simplification, we obtained good fit coefficients by fitting two separate linear functions to the 'learning part' and 'end-level'.

The additional N-back task hindered the progress on flight performance, indicating that the additional memory task drew cognitive capacity away from the primary task. Vice versa, the N-back performance dropped below baseline scores when it was performed in combination with the flight task. We did not find a statistical correlation between the N-back performance and the learning curve parameters. Hence, we did not find evidence for our hypothesis that the learning performance is related to the student's cognitive spare capacity as measured by the N-back task.

While only one instructor performed the post-experiment grading for a limited set of recordings of the performed maneuvers, the results showed strong correlations between the normalized performance measures and the instructor gradings.

Conclusions

Using a VR flight simulator, fifteen participants without flying experience practiced basic flight maneuvers based on self-study and without receiving feedback. Learning performance was extracted from relevant flight parameters, which were normalized and combined into an overall measure. This measure was fitted with a learning curve representing learning speed and end-level. The high correlation with instructor gradings suggests that, for the limited set of maneuvers, the student's progress in manual flying skills could objectively be assessed in the VR flight simulator. Addition of the N-back task hampered the students' flight performance and their learning progression, indicating that the additional task absorbed cognitive capacity. However, the performance on the N-back was not a good predictor of someone's learning curve. The results of this study show the potential of measuring learning performance in a VR simulator, whereas the transfer of training from VR to the real aircraft has yet to be explored.

Acknowledgements

This research was supported by the Defence Research and Development Programmes V1917 and V1903, and RNLAf AIR. The authors acknowledge all RNLAf subject matter experts from EMVO, FCL and CMA for their guidance and advice in the preparation of the experiment and interpretation of the results. Furthermore, we thank multiSIM BV for their VR simulator developments and all participants for their enthusiastic contribution.

References

Air Education and Training Command. (2020). About PTN

Lewis, J., & Livingston, J. (2018). Pilot Training Next: Breaking institutional paradigms using student centered multimodal learning. Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC), Orlando, FL.

McCoy-Fisher, C., Mishler, A., Bush, D., Severe-Valsaint, G., Natali, M., & Riner, B. (2019). Student naval aviation extended reality device capability evaluation.

Mishler, A., Severe-Valsaint, G., Natali, M., Seech, T., McCoy-Fisher, C., Cooper, T., & Astwood, R. (2022). Project Avenger Training Effectiveness Evaluation. Naval Air Warfare Center Training Systems Division Chief of Naval Air Training.

Pennington, E., Hafer, R., Nistler, E., Seech, T., & Tossell, C. (2019, 26/04). Integration of advanced technology in initial flight training. Systems and Information Engineering Design Symposium (SIEDS), University of Virginia, USA.

Pope, T. (2019). A cost-benefit analysis of Pilot Training Next Air University. Wright-Patterson Air Force Base, OH, USA.

Ross, G.A. (2022) Extended Reality Flight Simulators as an Adjunct to Traditional Flight Training Methods, Thesis, Massey University, 2022.

Schneider, W., and Chein, J. (2003) Controlled & automatic processing: behavior, theory, and biological mechanisms, Cognitive Science.

Sheets, T. H., & Elmore, M. P. (2018). Abstract to action: Targeted learning system theory applied to adaptive flight training.

Stuldreher, I.V., Van der Burg, E., Ledegang, W. D., Houben, M.M.J., Groen, E. L., Van der Horst, D., Starmans, E.A.M., Almekinders, G. (in press). Measuring the Lookout Behaviour of Student Pilots in a Virtual Reality Flight Simulator. International Symposium on Aviation Psychology.

Tinga, A.M., de Back, T.T., and Louwerse, M.M. (2019) Non-invasive neurophysiological measures of learning: A meta-analysis, Neuroscience & Biobehavioral reviews.