

Comparison Between Virtual Reality and Physical Flight Simulators for Cockpit Familiarization

Stefan Auer

stefan.auer@fh-hagenberg.at
University of Applied Sciences Upper Austria
Hagenberg, Austria

Harald Reiterer

harald.reiterer@uni-konstanz.de
University of Konstanz
Konstanz, Germany

Jens Gerken

jens.gerken@w-hs.de
Westphalian University of Applied Sciences
Gelsenkirchen, Germany

Hans-Christian Jetter

jetter@imis.uni-luebeck.de
University of Lübeck
Lübeck, Germany



Figure 1: Our study compared consumer-grade VR equipment (left) with a full-scale physical flight simulator (PFS) of a Boeing 737-800NG (right) to determine whether VR can supplement or even replace a PFS during cockpit familiarization.

ABSTRACT

Airlines and flying schools use high-end physical flight simulators (PFS) to reduce costs and risks of pilot training. However, such PFS with full-scale cockpits have very high acquisition and operation costs. In contrast, recent consumer-grade and off-the-shelf soft- and hardware can be used to create increasingly realistic virtual reality flight simulators (VRFS) that could potentially serve as cost-efficient and flexible alternatives. We present a user study with 11 participants to determine whether consumer-grade VRFS can supplement or even replace a PFS during cockpit familiarization training (CFT). We compared a full-scale Boeing 737-800NG PFS with a VRFS based on off-the-shelf flight simulator software

combined with a consumer-grade head-mounted display and either finger tracking or a handheld controller as input device. Participants performed instrument reading tasks and check procedures from the aircraft's operating manual. We did not observe statistically significant differences in successful instrument reading tasks, error rates and task completion between PFS and VRFS during CFT. However, we found that VRFS' Mental Demand, Physical Demand, Effort, task completion times, and levels of simulator sickness were significantly higher and exceeded acceptable levels. We conclude that future consumer-grade VRFS will need to improve soft- and hardware for interacting with simulated switches and reduce simulator sickness before they can serve as PFS alternatives for CFT.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

MuC '21, September 5–8, 2021, Ingolstadt, Germany

© 2021 Association for Computing Machinery.

ACM ISBN 978-1-4503-8645-6/21/09...\$15.00

<https://doi.org/10.1145/3473856.3473860>

CCS CONCEPTS

• **Computing methodologies** → **Simulation evaluation**; *Interactive simulation*; • **Human-centered computing** → **Virtual reality**; *Human computer interaction (HCI)*; User studies; • **Software and its engineering** → *Virtual worlds training simulations*.

KEYWORDS

Aviation, Pilot, Flight Simulation, Cockpit, Training

ACM Reference Format:

Stefan Auer, Jens Gerken, Harald Reiterer, and Hans-Christian Jetter. 2021. Comparison Between Virtual Reality and Physical Flight Simulators for Cockpit Familiarization. In *Mensch und Computer 2021 (MuC '21), September 5–8, 2021, Ingolstadt, Germany*. ACM, New York, NY, USA, 15 pages. <https://doi.org/10.1145/3473856.3473860>

1 INTRODUCTION

Physical flight simulators (PFS) play a central role in the training of a professional pilot. Essential airmanship skills such as aircraft handling, cockpit manipulation, and the correct behavior during emergencies can be trained with great realism and without risk. However, the costs for a fully equipped PFS that is certified by an aviation authority and accurately represents a specific aircraft type has historically approached the aircraft's cost [33] and today still requires between \$6 to \$8 million with another \$400-500 an hour to operate¹, resulting in high training costs per pilot. Furthermore, PFS are inflexible by nature as they can only simulate a single aircraft and cockpit layout, are physically constrained to a particular location, and require additional staff to train a single cockpit crew. Therefore, flying schools and airlines try to reduce the resulting costs of PFS [12, 48] and have been seeking affordable and realistic substitutions that can support or even replace specific parts of basic pilot training [26], e.g., cockpit familiarization.

In recent years, virtual reality (VR) technology has rapidly grown in popularity, both as a consumer product and research topic in human-computer interaction (HCI). The increasing number of available devices and their technical capabilities, such as field of view (FOV), latency, and pixel per degree (PPD) lead to an enhanced perceived quality of the virtual experience [9]. The current generation of consumer-grade VR thereby enables very cost-efficient simulations of settings in which both physical actions and spatial awareness are of key importance, e.g., rock climbing [45], interior design [22], designing interactive rooms with futuristic displays [21], or driving a car [16].

In our work, we explore the potential that a virtual reality flight simulator (VRFS) based on consumer-grade and off-the-shelf VR soft- and hardware could have for professional pilot training in commercial aviation. This domain is traditionally dominated by using a high-end PFS with a full-scale replica of an aircraft cockpit [50] and, in some cases, robotic motion platforms [6, 10]. In contrast, a fully functional virtual aircraft and cockpit can be realized at a high level of detail with comparably low costs using today's much improved consumer-grade VR technology. By integrating hand or finger tracking to mimic the physical and spatial interactions of real-world cockpit manipulation, VR could also have an advantage over learning with desktop-based 2D non-VR flight simulators [38] or printed cockpit pictures or diagrams [27], as muscle memory – one of five major attributes required for a safe flight [17] – is supported.

To understand if a cost-efficient consumer-grade VRFS could supplement or even replace a PFS for cockpit familiarization training (CFT), we conducted a within-subjects user study with 11 participants to compare their performance, task load, and simulator sickness during CFT tasks in a PFS of a Boeing 737-800NG with those in a VR counterpart. Tasks consisted of repeated instrument reading tasks (at three different VR display resolutions) and the execution of basic interactions with cockpit controls in the course of three real-world check procedures from the aircraft's operation manual. For user input in VR, we additionally compared a handheld controller and camera-based finger tracking.

The study revealed several findings about the potentials and problems of consumer-grade VRFS for CFT. Based on our quantitative analysis of objective measurements such as task completion time, instrument reading performance, and error rates, our analysis of self-assessment questionnaires about task load and simulator sickness, and our concluding semi-structured interviews with participants, we found that:

- (1) the percentage of successful instrument reading tasks both in high-resolution VRFS and PFS was equally high with 99.64%. There were no statistically significant differences between PFS and the high-, medium-, or low-resolution VRFS. Participants successfully compensated for lower resolutions with increased head and body movements to reduce viewing distances;
- (2) the participants were able to successfully conduct check procedures, both in PFS and VRFS, without any sequence errors and without significant differences in switch position error rates, despite the lack of haptic feedback in VRFS;
- (3) the mean task completion times for VRFS were always higher than for PFS. This difference became statistically significant with increasing task difficulties (checks 2 & 3), resulting in an overall lower efficiency for VRFS during CFT;
- (4) the reported Mental Demand, Physical Demand, and Effort for VRFS was significantly higher, caused by the weight of the VR devices and time-consuming interactions;
- (5) the participants reported problematic levels of simulator sickness after exposure to VRFS and unanimously preferred the PFS in user interviews.

We conclude that the low error rates in instrument reading tasks and check procedures show that VR can be successfully used for CFT. However, VRFS based on currently available consumer-grade VR cannot yet fully replace PFS for CFT, since the Mental Demand, Physical Demand, Effort, task completion times, and simulator sickness remain above acceptable levels. Therefore, future consumer-grade VRFS will need to improve soft- and hardware for interacting with simulated switches and reduce simulator sickness before they can serve as alternatives to PFS for CFT.

2 RELATED WORK AND BACKGROUND

Our work is positioned in the context of three areas of related work: Virtual reality flight simulation, input methods in VR cockpits, and readability in VR at different display resolutions. The fourth subsection of this chapter provides background information about real-world requirements for cockpit familiarization training.

¹Los Angeles Times, Jan 3rd 2020, <https://www.latimes.com/business/story/2020-01-09/boeing-737-max-simulators>

2.1 Virtual Reality Flight Simulation

Within the virtual simulation arena, flight simulation is perhaps the most pervasive and successful part [39]. The application areas of the different VRFS range from professional training devices [43, 52] and simulations with a focus on testing flexible cockpit layouts [4, 53] to entertainment and gaming on mobile phones [47].

Flight simulators help to reduce the complexity of flying, as they allow the learning of specific tasks under safe conditions. While historic non-digital examples made use of adapted parts from sewing machines [37], digital simulators employ 3D graphics or VR technology for more realistic training conditions and thus new ways of aviation pilot training [34, 40]. This resonates with Podgorski's [36] emphasis of VR's promising results as a support tool in acquiring and transferring tacit knowledge. Consequentially, there is sustained interest in using VR in aviation, as it enables new and flexible ways of training [12]. For example, Peysakhovich et al. [35] focused on cockpit familiarization by learning check procedures from viewing 360° videos on a VR head-mounted display (HMD). However, this happened without displaying an interactive VR cockpit and without physical interaction with the virtual instruments. Unlike in our study, this inhibited the potential benefits of muscle memory and kinesthetic cues during learning.

VRFS can be used as training devices but also as design tools to test cockpit designs. For example, Oberhauser et al. [32] compared the fidelity of a VRFS to a hardware cockpit mockup during flying tasks to evaluate the possible role of a VRFS in early phases of the cockpit design process. They found that VRFS can serve as a reliable low-cost addition in the early development process of cockpit human machine interaction technologies [31]. Also, most pilots were able to successfully perform the requested flying tasks in the VRFS, but they showed a degradation of performance metrics and an increase of workload [30].

2.2 Input Methods in VR Cockpits

In VR, the user's physical interactions for manipulating virtual elements need to be technologically mediated, e.g., by holding input devices or tracking hand and finger positions. One of the resulting drawbacks is that interacting with virtual instruments or switches happens without the haptic or tactile feedback pilots experience when interacting with physical cockpits. For example, Aslandere et al. [4] used an optical system for finger tracking to generate a virtual hand without haptic feedback and with which users achieved an average hit rate of only 77%. Furthermore, the visual representation of the virtual hand also has an influence on interaction, e.g., a transparent hand negatively influences depth perception [3]. However, a transparent hand also enables pilots to see underlying virtual objects such as switches without occlusion. To avoid such trade-offs and also to provide haptic feedback at least for selected controls, other researchers [28, 52] and commercial products² integrated physical joysticks and thrust levers into their VRFS. However, this cannot be applied to all the numerous and complex physical interfaces (e.g. switches, levers, knobs) necessary for CFT for a commercial airliner such as a Boeing 737.

In the future, novel input devices with haptic feedback could be used for improved user input in VR cockpits, such as the *Haptic*

Revolver [49] or robotic arm based systems like *Snake Charmer* [2]. Glove based approaches typically use vibrators [46, 51] or actuators [14] to provide haptic feedback in VR. Other technologies use ultrasound as feedback device [8]. However, for the purpose of our study of consumer-grade VR technologies, we decided for off-the-shelf finger tracking with the Leap Motion³ and HTC Vive handheld controllers, both without haptic feedback, because they have a high availability at relatively low costs.

2.3 Readability in VR at different display resolutions

The display resolution of a VR HMD greatly influences the virtual performance and experience [9, 42]. For example, Dowling et al. [13] used different display resolutions for a walking test with obstacles, resulting in significantly lower walking speed at lower resolutions. In the medical context, low resolutions in VR are also used to simulate low visual acuity, e.g., due to the loss of photoreceptor cells [20], to provide a better understanding of individuals with visual impairments. Especially aviation pilots are required to have sufficient visual fitness⁴ to carry out their many supervisory tasks [5]. Therefore, the display resolution within a VRFS and the resulting visual acuity plays a critical role as it defines the readability of virtual instruments and labels within the simulated cockpit.

Previous work focused on determining optimal text parameters for reading in VR without a specific focus on VRFS. A study by Grout et al. [18] focused on the readability of curved virtual panels in VR. Their results show that it is possible for users to perform traditional reading tasks inside a immersive virtual environment with near-baseline performance under ideal circumstances. Kojic et al. [25] studied the user experience of reading in VR and let users choose the best and the worst text parameters for reading in virtual reality. However, both studies do not give a clear recommendation for minimum text size and resolution. In contrast, Schiefele et al. [44] performed a readability test for VRFS with different display resolutions based on Landolt rings and text on different backgrounds. They recommend a minimum FOV of 80° and a minimum letter-size of 8.25mm at a distance of 55cm (equals 0,859° letter-size) for virtual cockpits. However, their HMD from 1999 had a maximum resolution of 1,280 x 1,024 pixels with a FOV varying from 30° to 100°.

In our research, we used the consumer-grade but high-resolution *Pimax5k+ HMD*⁵ with a maximum display resolution of 2,672 x 2,692 pixels per eye and a FOV of 150° (further details see below). To better understand the influence of different display resolutions on instrument reading performance, we decided to include own instrument reading tasks for three different resolutions into our study.

³Leap Motion <https://www.ultraleap.com/>

⁴EASA Medical Requirements For Air Crew: <https://www.easa.europa.eu/sites/default/files/dfu/AnnexItoEDDecision2019-002-R.pdf>

⁵Pimax <https://www.pimaxvr.com/>

²Adams Group Multi-Task-Trainer <https://adamsgroup.de/sit/>

	PFS	High-End VR with physical cockpit	High-End VR with virtual cockpit and haptic feedback	consumer-grade off-the-shelf VR
Acquisition Costs	>\$1 Mio	>\$50k	>\$15k	>\$5k
Operation Costs	high ^{M,L,P}	medium ^{M,L}	medium ^{M,L}	low
Portability	none	low	high	high
Flexibility of Cockpit Layout	none	medium	high	high
Potential Simulator Sickness	low	medium	medium	high
Haptic Feedback	high	high	medium	low/none

Table 1: Potential advantages and disadvantages of different technologies used for CFT. The row Operation Costs comprise costs for (M)aintenance, recurring (L)icense fees, and additional (P)ersonnel.

2.4 Cockpit Familiarization Training (CFT)

An essential part of pilot training is CFT. During CFT, pilots learn the position, use and purpose of switches, controls, and instruments [15]. One objective of CFT is the correct handling of the pilots' checklist, as improper use of checklists is a major contributing factor in aircraft accidents [11].

In our informal conversations with commercial providers of flight training technology, we learned that a high-end PFS is considered too costly and inefficient for such basic training. Also, low-cost alternatives such as desktop-based 2D non-VR flight simulators [38] or printed cockpit pictures or diagrams [27] lack the desired spatial and physical qualities. Instead, future VRFS with tracking of pilots' head, arm, and hand positions in order to learn spatial relations based on visual and kinesthetic perception were considered a promising and cost-effective alternative to PFS-based CFT. However, providers were concerned about the lack of haptic feedback in a virtual cockpit and especially about simulator sickness caused by VR technologies.

Table 1 summarizes potential advantages and disadvantages of three VRFS variants that could be used for CFT: (1.) Using a high-end VR headset with a physical cockpit mockup helps to provide haptic feedback and might reduce the level of simulator sickness for a student pilot but it also has higher acquisition costs⁶, decreased portability, and flexibility. (2.) A high-end VR with a virtual cockpit and haptic feedback (e.g. XTAL 8k⁷ with Manus Prime II Haptic Gloves⁸) increases haptic realism, portability, and flexibility but still has increased acquisition and operation costs. (3.) A consumer-grade VRFS based on commercial off-the-shelf technology has low acquisition and operation costs, high portability and flexibility, and greatest potential for cost savings, especially in classroom settings with many installations. Therefore, despite the obvious lack of haptic feedback and the possibility for simulator sickness, we decided to further explore such a consumer-grade VRFS in our study.

3 METHOD AND STUDY DESIGN

We compared a consumer-grade VRFS to a Boeing 737-800NG PFS that featured a full-scale cockpit replica with original instruments

⁶According to *Adams Group* hardware costs for their Multi-Task-Trainer (see <https://adamsgroup.de/sit/>) are in the range of €40k to €60k for a single cockpit (video call, Mar 12th 2021).

⁷XTAL <https://vrgineers.com/xtal/>

⁸Manus <https://www.manus-vr.com/>

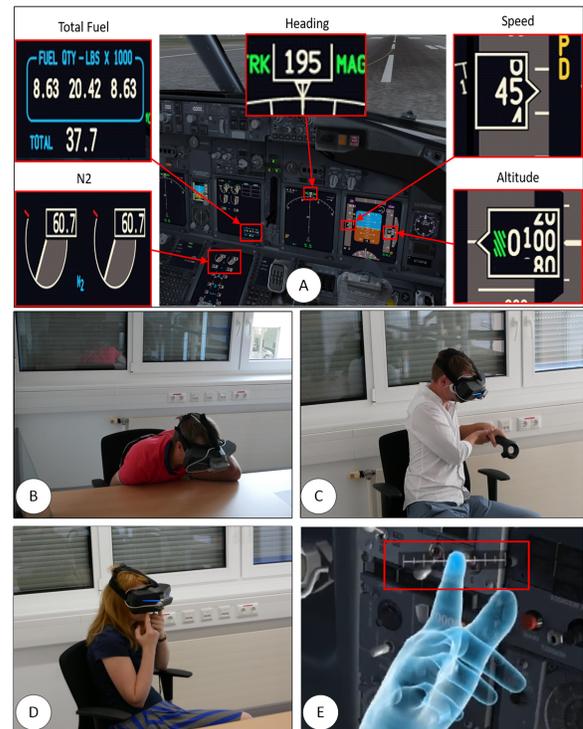


Figure 2: Details about the tasks and observed behaviors from the user study: (A) The five aircraft parameters for the instrument reading task. (B) P10 leans forward in order to read the values on the virtual cockpit displays in VR_{low} . (C) P6 uses both hands to compensate the weight of the VR controller during longer use and for increased precision. (D) P7 uses her fingers to compensate the HMD's weight on her nose. (E) The finger tip of the virtual index finger and the virtual pop-up control occluding switch and switch position.

and was built mostly from original parts by flight simulator enthusiasts as part of a commercial flight simulation attraction⁹. This PFS was not certified by the European Union Aviation Safety

⁹Synthetic737, <https://www.synthetic737.at/>

Agency (EASA)¹⁰. However, except for the absence of fully functional circuit breakers, it would fulfill the requirements for a flight and navigation procedures trainer (FNPT) Level II simulator¹¹. Therefore it supports the development of fundamental skills of flying and can be considered fully capable for CFT.

The goal of our study was a quantitative comparison of both simulator technologies in terms of instrument reading capabilities and user performance (i.e. task completion times and error rates), self-reported task load, and self-reported simulator sickness during CFT. We deliberately decided against flying tasks to keep the complexity of the user tasks at an introductory level and also to exclude potential influences through differences between PFS and VRFS in the visual or kinesthetic perception of aircraft motion. The simulated aircraft remained motionless during the whole study. At the end of the task, we collected qualitative feedback from participants during semi-structured interviews to gain additional information that helped us contextualize and explain our quantitative results.

We conducted the study using commercial flight simulator software and consumer-grade hardware for the VRFS to assess to what extent this readily available technology is already capable of satisfying real-world demands. This had implications on the VRFS's frame rate: While a minimum of 90 fps (frames per second) is recommended for VR applications [1], the maximum supported frame rate of the commercial simulator software Prepar3D by Lockheed Martin¹² was 60 fps. This frame rate was further reduced by the necessary rendering times when using a highly detailed cockpit model of a Boeing 737. These details of the cockpit model resulted in a further drop of the frame rate to around 35 fps and occasionally even below. A further factor was the rendering resolution. The HMD's display hardware enabled high resolutions per eye with a good readability of instruments but higher resolutions also resulted in increased rendering times and thus lower frame rates. For keeping adverse effects of lower frame rates within reason (e.g. in terms of simulator sickness), we ensured 25 fps as an absolute minimum frame rate. This is comparable to 3D motion pictures in cinemas (e.g. RealD 3D cinema standard with 24 fps per eye, very wide FOV in front rows, seated position).

We entered our study with a set of research questions and hypotheses about differences between PFS and VRFS for CFT. These were based on the obvious limitations of a VRFS compared to a PFS in terms of visual fidelity, resolution, frame rates, use of input devices, and lack of haptic feedback:

- *Instrument Reading Performance*: How strongly is the reading performance in the VRFS affected by decreasing display resolutions and how does it compare to the PFS?
- *Error Rate and Task Completion Time*: Are the error rates and task completion times for the VRFS higher than for the PFS? Are participants able to complete their tasks in the VRFS at all? Are there differences between controller-based and hand-tracking input for VRFS?
- *Perceived Workload*: How strongly does the perceived workload increase for the VRFS compared to the PFS?

¹⁰European Union Aviation Safety Agency <https://www.easa.europa.eu/>

¹¹Specifications for Aeroplane Flight Simulation Training Devices: [https://www.easa.europa.eu/sites/default/files/dfu/CS-FSTD\(A\)%E2%80%94Issue2.pdf](https://www.easa.europa.eu/sites/default/files/dfu/CS-FSTD(A)%E2%80%94Issue2.pdf)

¹²Prepar3D by Lockheed Martin, <https://www.prepar3d.com/>

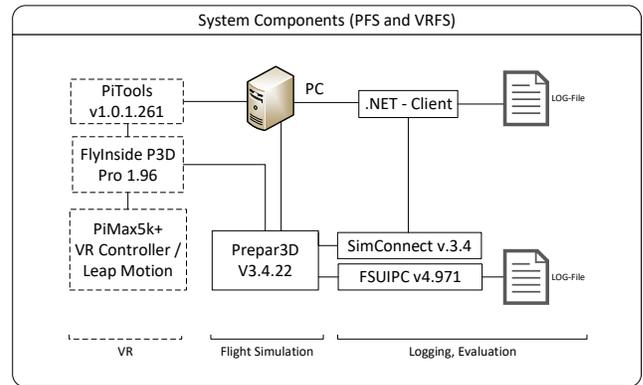


Figure 3: Both the VRFS and PFS share Prepar3D as common simulation software that uses FSUIPC and SimConnect to enable monitoring, measurements, and manipulation of the virtual Boeing737. They also provide an interface for the self-developed logging client. VR output used a Pimax5k+ HMD, PiTools, FlyInside, and a Vive VR controller or Leap Motion sensor.

- *Simulator Sickness*: How strongly are users affected by simulator sickness in the VRFS compared to the PFS?

3.1 Participants

We invited 17 study participants who were recruited from students of our university and previous customers of the PFS flight attraction. None of the participant had experience as flight student or professional pilot, neither with real nor simulated aircraft. We still excluded two participants because they were very regular users of the PFS and had collected over 2,000 flying hours in the simulator and therefore can be regarded as comparable to professional pilots in the context of our study. Three further participants dropped out of the study between the test sessions for the first and second condition (both were separated by either 1 or 2 weeks, see below). Furthermore, one participant was excluded from the VRFS condition, as his glasses could not be worn together with the VR HMD. The remaining 11 participants (24-45 years, $M = 31.64$, $SD = 7.13$, 3 female, 8 male)¹³ were split into two groups. Group A (5 persons) started with the PFS session, group B (6 persons) with the VRFS session. Due to the limited availability of the PFS, group A had two weeks, and group B had one week in between the sessions.

3.2 Apparatus

3.2.1 Hardware configuration of PFS. The PFS software was executed on a PC with Windows 10 and an Intel i7 3.6 GHz, an Nvidia GeForce GTX 1080Ti GPU, and 32 GB RAM. The non-stereoscopic 3D rendering of the external environment outside the cockpit (e.g., the runway) could be seen through the cockpit's windows and was projected with three HD projectors on a 180° cylindrical screen in front of the cockpit. However, our study contained no tasks that

¹³Percentage of female participants in our study: 27.27%. Global percentage of female commercial pilots: 5.18% (Source: <https://www.bbc.com/news/business-46071689>)

required viewing the external environment. Our PFS did not include a full-motion platform for simulating kinesthetic stimuli.

3.2.2 Hardware configuration of VRFS. The VRFS used a desktop PC with Windows 10, an Intel i7 3.7 GHz CPU, an Nvidia GeForce RTX 2080 GPU, and 16 GB RAM. As HMD for the VRFS, we chose a *Pimax5k+* with positional tracking (six degrees of freedom). We decided for the Pimax and against the more popular Oculus or HTC Vive since it supports higher display resolutions and therefore enabled us to examine the readability of instruments for a wider range of resolutions. Also, to ensure a fair comparison between VRFS and PFS, we opted for the Pimax's wide FOV. In principle, the Pimax5k+ HMD combined with PiTools v1.0.1.261 provides a display resolution of 2,672 x 2,692 px per eye, a maximum horizontal FOV of 170°, and a screen refresh rate of 200 Hz. However, to ensure a minimum rendering performance of at least 25 frames per second, we needed to reduce the FOV to 150° and the refresh rate to 120 Hz.

To mimic natural interactions with physical PFS switches, we also included two different input technologies into the VRFS. In a first condition, participants used a front-facing Leap Motion optical finger tracking sensor attached to the HMD to interact with virtual switches by moving a real-time representation of their hand and finger positions in VR. In a second condition, participants used a Vive handheld controller to control the position and orientation of a virtual hand in VR with an extended virtual index finger to interact with the switches (Fig. 2E).

Depending on the type of switch, the VRFS software FlyInside (see below) also displayed virtual pop-up controls in front of the virtual switch to provide an enlarged view of the switch's current position for easier manipulation (highlighted with a red border in Fig. 2E). However, in practice, this still resulted in sometimes cumbersome interaction and problems with occlusion.

While both VR input systems were expected to equally support users during the tasks by utilizing proprioception and muscle memory, the Leap Motion has a smaller tracking area and tracking range that is limited by the FOV and resolution of the front-facing Leap Motion camera. Hand and finger tracking can therefore get lost when hands or arms are moved outside of the Leap Motion's FOV and optimal viewing distance. The handheld Vive VR controller, however, uses wall- or ceiling-mounted IR-emitters to track positions and therefore achieves a room-sized tracking area around the participant with higher precision but requires holding the controller at all times. Therefore, we were interested if these differences also would result in notable differences in performance, task load, or user preference during our study.

3.2.3 Simulation Software Stack for PFS and VRFS. Our implementation of the VRFS aimed to create the greatest possible similarity between the VRFS and PFS to ensure high internal validity of the study. Both the VRFS and PFS therefore used the same aircraft type, cockpit layout, and were based on the same simulator software. The entire software stack of both simulators was identical except the additional VR layer in the case of the VRFS (Fig. 3). Both simulators were based on the Prepar3D simulation software. For supporting

hand and finger tracking as user input in VR, we used the VR add-on software FlyInside¹⁴. PMDG¹⁵ supplied the model of the Boeing 737-800NG for the PFS and ProSim737¹⁶ for the VRFS. SimConnect is a part of Prepar3D and the FSUIPC¹⁷ provides an application programming interface (API) and software development kit (SDK) to monitor and control the simulation. A simple, self-developed .NET-client additionally measured and logged data during the study.

3.3 Independent Variables

To compare the differences between PFS and VRFS, we used two within-subject variables: the mode of visual presentation of the cockpit and the input methods for interacting with the cockpit's switches.

3.3.1 Visual Presentation Mode. The mode of visual presentation was either PFS or a consumer-grade stereoscopic VRFS whereas the former represents the state-of-the-art and the latter a potential cost-efficient alternative. Furthermore, within VRFS, we compared three different HMD display resolutions per eye to compare the readability of instruments in VR to the PFS: 2,672 x 2,696 pixels (100%, noted as VR_{high}), 2,296 x 2,320 pixels (74%, VR_{med}), and 1,600 x 1,572 pixels (34%, VR_{low}). This resulted in four possible modes of visual presentation: PFS , VR_{high} , VR_{med} , and VR_{low} . The order of presentation was counterbalanced.

We chose to vary the VR display resolution during the study in order to ideally identify the lowest possible boundary for regular use. Obviously, it would be best to always use the highest possible resolution but, in practice, the resulting frame rate drops rapidly with growing resolutions. This can be a cause of simulator sickness. Therefore, our goal of finding a sweet spot or trade-off with both acceptable readability and frame rate has high practical relevance.

3.3.2 Input Methods. The input method had three modalities: PFS with real finger input, VRFS with Leap Motion sensor (noted as VR_L), and VRFS with a handheld controller (noted as VR_C). All input methods were counterbalanced and the tasks for their comparison were always performed at the highest display resolution VR_{high} .

3.4 Tasks

Participants first conducted a simple instrument reading test. The participants were requested to read aloud altitude, speed, heading, total fuel, and N2. Within the cockpit, these five values were displayed in different fonts and font sizes (Fig. 2A) but consistently in PFS and VRFS. The smallest font had a physical letter-size of 8.2 mm for a viewing distance of 50 cm (this corresponds to 0.93°, which is above the recommendation of Schiefele et al. [44] mentioned in section 2.3).

For repeated readings with different values but traceable results, the experimenter put the simulated aircraft into five predefined scenarios (including ground and airborne). Participants were asked to read aloud the five different aircraft parameters displayed in

¹⁴FlyInside by FlyInside Inc., <https://flyinside-fsx.com/>

¹⁵PMDG Simulations LLC, <https://pmdg.com/>

¹⁶ProSim737 by ProSim Aviation Research B.V., <https://prosim-ar.com/prosim737/>

¹⁷FSUIPC by John Dowson, <http://www.fsuiipc.com/>

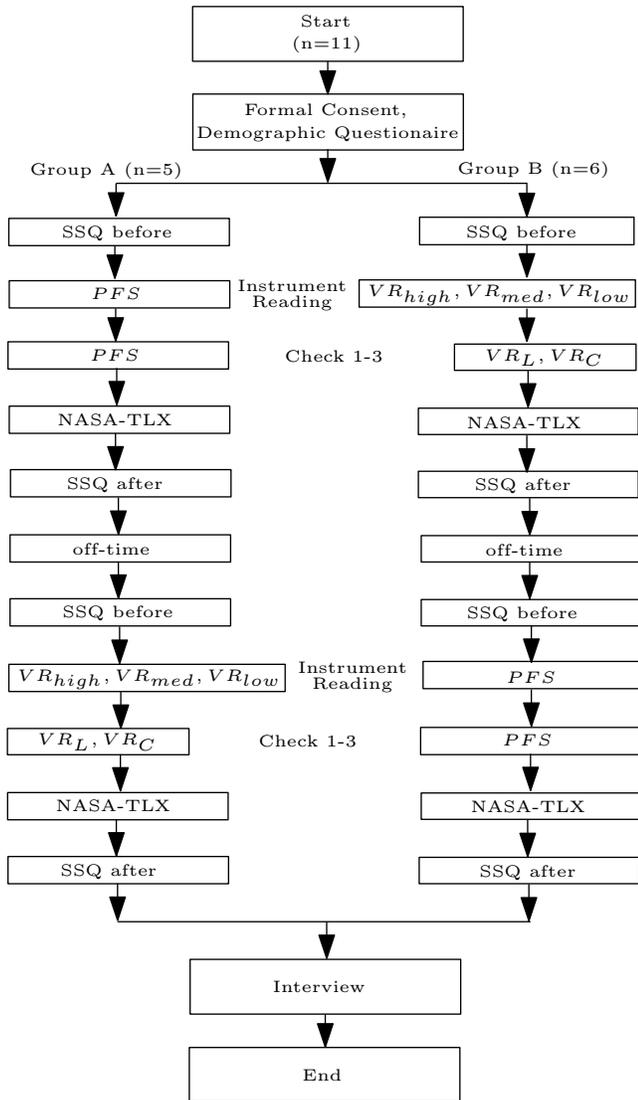


Figure 4: Visualization of the study procedure. In VR, the participants had to perform the Instrument Reading tasks with all three resolutions (VR_{high} , VR_{med} , VR_{low}), before they performed the Check Procedure tasks with either the VR Controller (VR_C) and the Leap Motion Sensor (VR_L). The order of the VR resolutions and input methods was counter-balanced.

the cockpit for each of the five predefined scenarios for each presentation mode. This resulted in a total of 100 reading tasks per participant with 25 readings tasks in PFS and 75 reading tasks in VRFS (25 reading tasks for each of the three VR display resolutions VR_{high} , VR_{med} , and VR_{low}).

Thereafter, the participants performed CFT based on three check procedures from the Operations Manual of the Boeing 737-800NG¹⁸. All three checks were executed with the highest HMD resolution VR_{high} . Participants were seated at the right seat of the cockpit and executed checks with their left hand. After learning them by heart during an initial training phase, participants had to perform the checks without interventions by the experimenter or call outs. The tasks had an increasing difficulty:

- Check1 (ENGINE START, page NP.20.24f): manipulation of 5 switches in correct sequence with a total of 14 possible switch positions
- Check2 (AFTER ENGINE START, NP.20.27f): manipulation of 8 switches in correct sequence with a total of 21 possible switch positions
- Check3 (ENGINE SHUTDOWN, NP.20.41f): manipulation of 14 switches in correct sequence with a total of 32 possible switch positions

3.5 Dependent Variables

As dependent variables we chose different indicators for user performance, self-reported task load, and self-reported simulator sickness.

3.5.1 Reading Performance [%]. Reading Performance was measured as the percentage of correct instrument reading tasks, i.e., the percentage of correct values that were read by the participants. There were 25 values that were read per mode of visual presentation. A reading task was rated as correct only if the entire value was correctly read to the experimenter. Even in cases of only a single incorrect digit (e.g., 14.3 instead of 12.3), the task was still rated as incorrect. With regard to the usability of a flight simulator in the sense of ISO-9241-11 (i.e. usability as *effectiveness*, *efficiency*, and *satisfaction*), the reading performance should be considered a measure of *effectiveness*.

3.5.2 Error Rate [%]. The participants had to memorize and perform three check procedures from the checklist of the Boeing 737-800NG Operations Manual. These procedures unambiguously define a sequence of switches that have to be set to specific switch positions. Two types of errors were recorded: first, an incorrect order of the switches during the interaction was considered a *sequence error*; second, a *switch position error* occurred whenever a switch was left in an incorrect position at the end of each check. Similar to the reading performance, the error rate can also be considered a measure of *effectiveness* in the sense of ISO-9241-11's usability definition.

3.5.3 Task Completion Time [sec]. The task completion time was measured while performing each of the three checks. It was determined by the time span between the switch manipulations of the first and last action of a check.¹⁹ Since check 1 always included

¹⁸Boeing provides the Operation Manual only directly to airlines but declassified versions can be found for reference online, e.g., at <http://toulouse747.com/wp-content/uploads/2018/12/Boeing-B737-700-800-900-Operations-Manual.pdf>

¹⁹ During the study, we encountered a bug in Flyinside that resulted in two switch groups that could not be switched by participants in VRFS. As a workaround, the experimenter manually changed the state of these switches using an external control console whenever a participant moved their virtual hand there and tried to interact. The reaction time of the experimenter was therefore subtracted from the task completion time.

	Reading performance				p-values			
	n	Mean	Median	SD	PFS	VR _{high}	VR _{med}	VR _{low}
PFS	11	99.64%	100%	1.21	PFS	-	1.000	1.000
VR _{high}	11	99.64%	100%	1.21	VR _{high}	1.000	-	1.000
VR _{med}	11	99.27%	100%	1.62	VR _{med}	1.000	1.000	-
VR _{low}	11	97.09%	100%	7.18	VR _{low}	1.000	1.000	1.000

Table 2: Reading Performance [%] with post-hoc Wilcoxon-Signed Rank Test and Bonferroni corrected p-values.

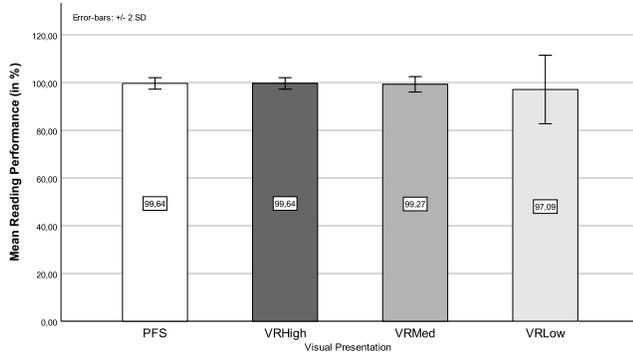


Figure 5: The percentage of successful instrument reading tasks in PFS and VR_{high} was equally high with 99.64%. There were no statistically significant differences between the different visual presentations.

waiting for the engines to start and there was no user input required during this spin-up procedure, the constant spin-up time of 101.6 seconds was always subtracted from the task completion time during data analysis. Following the usability definition in ISO-9241-11, the task completion time should be considered a measure of *efficiency*.

3.5.4 Perceived Workload. The participants' task load was measured using the NASA Task Load Index (NASA-TLX) [19] without paired comparisons of the subscales [7, 29], also known as *Raw TLX*. After performing all three checks, the participants rated their Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration on a scale ranging from *very low* (0) to *very high* (+10). ISO-9241-11 defines *satisfaction* as "the extent to which attitudes related to the use of a system, product or service and the emotional and physiological effects arising from use are positive or negative". Accordingly, we consider the TLX subscale Frustration as a measure for a negative emotional and physiological effect and thus opposed to *satisfaction* in ISO-9241-11.

3.5.5 Simulator Sickness. The Simulator Sickness Questionnaire [24] was applied before and after each PFS and VRFS session. This standardized, subjective questionnaire measures 16 symptoms on a Likert-scale ranging from *not at all* (0) to *severe* (3). These symptoms are *General Discomfort, Fatigue, Headache, Eyestrain, Difficulty Focusing, Increased Salivation, Sweating, Nausea, Difficulty Concentrating, Fullness of Head, Blurred Vision, Dizziness (eyes open), Dizziness (eyes closed), Vertigo, Stomach Awareness, and Burping* which are assigned to the categories *nausea, oculomotor,* and

disorientation. As some symptoms are associated with more than one category, the categories are not disjunctive. We consider these symptoms of simulator sickness as negative emotional and physiological effects in the sense of ISO-9241-11 that negatively affect users' *satisfaction*.

3.6 Procedure

Participants gave informed consent, filled out a demographic questionnaire, and a pre-exposure Simulator Sickness Questionnaire (SSQ). The study strictly adhered to all ethical guidelines of our university as well as national guidelines and legal regulations concerning COVID-19.

During an initial training phase, the experimenter (first author of this paper and a former military jet-fighter pilot with 18 years of experience in aviation, who also designed, executed, and evaluated this study) supported the test persons during learning to perform each check correctly by providing helpful background knowledge about the aircraft and engines as well as pointing out errors. All participants could decide for themselves when they wanted to complete the learning phase and felt able to perform each check under test conditions, which meant as quickly and error-free as possible and without any intervention from the experimenter.

After performing all three checks, the participants filled out the NASA-TLX. During the VR condition, the participants were asked to repeat the checks with the second input method after a short break. The second input method was followed by filling out a second NASA-TLX.

Each test session was concluded by filling out a post-exposure SSQ and, at the end of both sessions, answering the questions of the semi-structured interview. The interview gave participants the opportunity to informally share their experiences and comments. The following initial set of questions was used as a conversation starter:

- *Concerning cockpit familiarization: do you think that a VR cockpit can replace a hardware simulator?*
- *Do you prefer the hardware or software cockpit?*
- *What do you think about VR in flight simulation?*
- *In VR: How did the different resolutions influence your VR-experience?*
- *Is there anything else you would like to share?*

On average, a complete test session took 45 minutes in the PFS and 75 minutes in the VRFS. A visual presentation of the study procedure is shown in Fig. 4.

Input	n	Sequence Errors			p-values			Switch Position Errors			p-values			
		Mean	Median	SD	PFS	VR _L	VR _C	Mean	Median	SD	PFS	VR _L	VR _C	
Check1	PFS	11	0.00	0.00	0.00	-	1.000	1.000	0.00	0.00	0.00	-	1.000	1.000
	VR _L	11	0.00	0.00	0.00	1.000	-	1.000	0.00	0.00	0.00	1.000	-	1.000
	VR _C	11	0.00	0.00	0.00	1.000	1.000	-	0.00	0.00	0.00	1.000	1.000	-
Check2	PFS	11	0.00	0.00	0.00	-	1.000	1.000	0.00	0.00	0.00	-	0.471	0.951
	VR _L	11	0.00	0.00	0.00	1.000	-	1.000	0.18	0.00	0.40	0.471	-	1.000
	VR _C	11	0.00	0.00	0.00	1.000	1.000	-	0.09	0.00	0.30	0.951	1.000	-
Check3	PFS	11	0.00	0.00	0.00	-	1.000	1.000	0.27	0.00	0.65	-	1.000	0.471
	VR _L	11	0.00	0.00	0.00	1.000	-	1.000	0.27	0.00	0.47	1.000	-	0.471
	VR _C	11	0.00	0.00	0.00	1.000	1.000	-	0.09	0.00	0.30	0.471	0.471	-

Table 3: While the average number of switch position errors slightly increases with difficulty of the check, the pairwise comparisons are not statistically significant (Wilcoxon-Signed-Rank and Bonferroni corrected p-values).

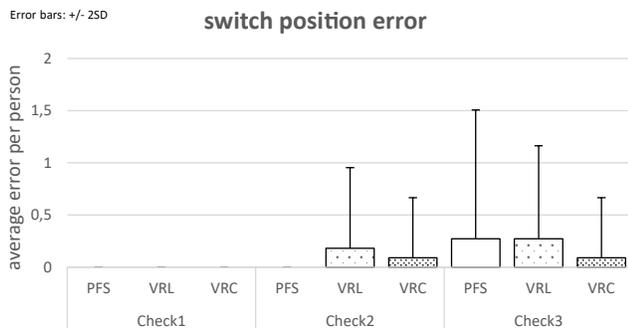


Figure 6: We did not observe any sequence error. The switch position error rate increased with difficulty of the check. None of the pairwise comparisons between the different visual presentations was statistically significant.

4 RESULTS AND DISCUSSION

4.1 Reading Performance

To our surprise, the reading performance was very high among all visual presentations, even including VR_{low} (Fig. 5, Table 2). Accordingly, data was not normally distributed and we analyzed the data with the non-parametric Friedman test. This showed no significant difference among the four different visual presentations ($\chi^2(3) = 1.571; p = 0.666; W_{Kendall} = 0.048$). We observed only one reading error (out of 275 reading tasks²⁰) among all participants for both PFS (99.64%, 1 error) as well as VR_{high} (99.64%, 1 error). For lower display resolutions, participants managed to successfully complete the task with only three errors for VR_{med} (99.27%, 3 errors) and seven errors for VR_{low} (97.09%, 7 errors) – of the latter, to be noted, five errors can be attributed to a single participant who failed to have a single correct reading at VR_{low} .

Judging from our observations, the display resolution in VR_{low} required intense compensation activities such as moving the head very close to the virtual instruments in order to reduce viewing distance and increasing local resolution (Fig. 2B). Consequentially,

some users reported negative experiences, mentioning that the values in VR_{low} were hardly readable (P3). P11 outright rejected VR_{low} as “No-Go” for a successful task completion. Further participants reported severe difficulties in differentiating between the digits 6 and 8 (P6), 0 and 8 (P7), or 2 and 7 (P9). These problems were not observed nor mentioned for VR_{med} or VR_{high} .

Finding 1 - Instrument Reading: The percentage of successful reading tasks both in PFS and high-res VRFS was equally high with 99.64%. There were no significant differences between PFS and all display resolutions of VRFS. Overall, the effect of resolution on correctness of instrument readings was much lower than expected. However, for VR_{low} this was only possible due to compensation strategies by the participants, by reducing the viewing distance through intense head and body movements. This additional demand led some participants to reject VR_{low} as unsuitable for the task.

Implications: Based on our results, we can recommend resolutions of 2,296 x 2,320 pixels per eye (VR_{med}) or above to provide a sufficient reading performance while preventing intense head and body movements.

4.2 Error rate

The error rates were calculated based on the success of the CFT tasks, which consisted of three different check procedures (check 1-3) as sub-tasks. For the CFT tasks, PFS was compared to the two VRFS input modes Leap Motion VR_L and Controller VR_C . The VRFS resolution was always VR_{high} .

Interestingly, we did not record any *sequence errors* neither during the PFS nor the VRFS conditions, and we observed an overall small error rate with just few and isolated *switch position errors* (see Fig. 6, Table 3). Accordingly, for each check procedure, a Friedman test, applied on the non-normally distributed data, did not show significant differences between the conditions at check 1 ($\chi^2(2) = 0.000; N = 11; p = 1.000; W_{Kendall} = 0.000$), and check 2 and 3 (each with $\chi^2(2) = 2.000; N = 11; p = 0.368; W_{Kendall} = 0.091$). These low error rates, especially the absence of sequence errors, indicate that both PFS and VRFS enabled participants to spatially

²⁰25 instrument reading tasks per visual presentation * 11 participants = 275

Check	Input	Task Completion Time				p-values		
		n	Mean	Median	SD	PFS	VR_L	VR_C
Check1	PFS	11	7.59	7.65	1.01	-	0.150	0.078
	VR_L	11	15.72	11.53	11.60	0.150	-	1.000
	VR_C	11	13.13	13.54	5.96	0.078	1.000	-
Check2	PFS	11	7.73	7.50	1.82	-	0.009**	0.009**
	VR_L	11	34.45	35.95	7.12	0.009**	-	1.000
	VR_C	11	34.20	32.09	9.54	0.009**	1.000	-
Check3	PFS	11	19.81	17.90	8.19	-	0.009**	0.009**
	VR_L	11	56.91	57.33	8.53	0.009**	-	0.150
	VR_C	11	50.49	50.11	8.21	0.009**	0.150	-

Table 4: Task Completion Time [sec], p-values are based on Wilcoxon-Signed-Rank and are Bonferroni corrected.

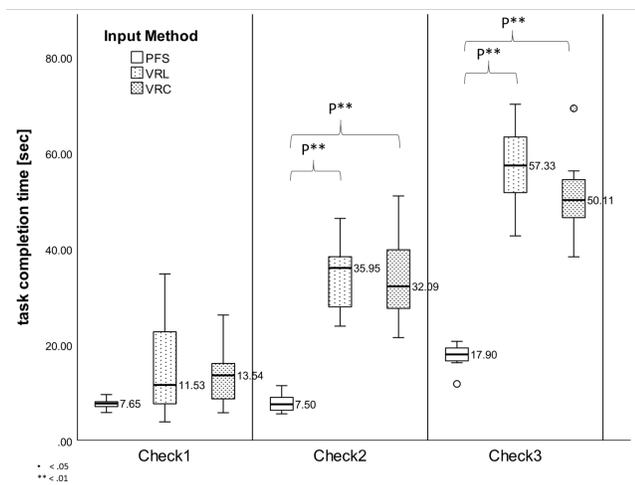


Figure 7: Box plots for Task Completion Time of all three checks with different input methods. The medians of the physical flight simulator (PFS) is always lower than the medians of Leap Motion (VRL) and Vive Controller (VRC).

memorize the correct order of switches and their correct end position, which is the essence of learning check procedures during CFT.

Finding 2 - Error Rate: Despite the lack of haptic feedback in VR_L and VR_C , participants were still able to successfully learn and reproduce check procedures in VR without any sequence error and without significant differences in switch position error rates. Overall, there were no statistically significant differences in error rates between PFS , VR_L , VR_C .

Implications: Based on our results, we conclude that the absence of haptic feedback or physical switches did not impair participants' task completion and correctness and thus should not be considered a necessity for successful CFT in a VRFS.

4.3 Task Completion Time

The task completion times for the cockpit familiarization tasks (Fig. 7, Table 4) were not normally distributed, and we applied a

non-parametric test, accordingly. Overall, descriptive statistics show that for PFS the average and median task completion time were faster compared to VR for all three check procedures. The Friedman test also identified a significant difference among the input methods (PFS , VR_L , VR_C) for check 2 ($\chi^2(2) = 16.909$; $N = 11$; $p < 0.001$; $W_{Kendall} = 0.769$) and check 3 ($\chi^2(2) = 17.636$; $N = 11$; $p < 0.001$; $W_{Kendall} = 0.802$), but not for check 1 ($\chi^2(2) = 2.364$; $N = 11$; $p = 0.307$; $W_{Kendall} = 0.107$). The post-hoc pairwise comparisons (Wilcoxon-Signed-Rank, Bonferroni corrected) confirm these results. For check 1, the differences between PFS and both VR_L ($p = 0.150$; $r = 0.417$)²¹ as well as VR_C ($p = 0.078$; $r = 0.474$) are not significant. Accordingly, the differences for check 2 and check 3 are statistically significant (both $p = 0.009$; $r = 0.626$). Differences between the two VRFS conditions are not statistically significant.

As the low error rates already indicated, the VRFS environment enabled participants to effectively complete the check procedures. However, the input techniques in VRFS caused a substantial increase in task completion times and thus an overall lower efficiency. Based on our observations and participants' comments during the task and interview, there are several potential explanations for this increase. Participants mentioned the "unnatural way of user input in VRFS" (P6, P10). Also, while the Vive VR controller has higher tracking precision, it felt increasingly heavy after longer use requiring compensation strategies (P3, P6, Fig. 2C). On the other side, the Leap Motion camera interrupted user interactions whenever the real-world hand accidentally left the Leap's FOV (P6). These unsolved challenges are also reflected in the very mixed preferences regarding the VR input method, with 6 participants preferring Leap Motion and 5 participants preferring the Vive controller.

Another contributing factor was the implementation-specific design of the UI of the FlyInside simulation software in VRFS. Interacting with a certain switch is done through an overlay slider control. This visually occludes the switch itself, making it sometimes difficult to select the desired position, which can be blocked from the participants' view (P5, Fig. 2E). These observations resonate with previous studies [4] that concluded that current forms of virtual hand-button interaction might be sufficient for virtual prototyping but are not ready for pilot training.

²¹The effect size r is calculated by $r = \frac{Z}{\sqrt{n}}$ as described by R. Rosenthal [41, p.239].

TLX category	NASA TLX					p-values		
	Input	n	Mean	Median	SD	PFS	VR _L	VR _C
Mental Demand	PFS	11	32.73	30.00	18.76	-	0.099	0.036*
	VR _L	11	50.00	60.00	20.49	0.099	-	0.306
	VR _C	11	53.64	60.00	18.04	0.036*	0.306	-
Physical Demand	PFS	11	8.64	5.00	6.74	-	0.024*	0.009**
	VR _L	11	38.18	40.00	24.72	0.024*	-	1.000
	VR _C	11	43.18	30.00	25.42	0.009**	1.000	-
Temporal Demand	PFS	11	15.46	10.00	12.93	-	0.459	0.276
	VR _L	11	21.36	20.00	13.25	0.459	-	0.774
	VR _C	11	23.64	20.00	13.43	0.276	0.774	-
Performance (inverted)	PFS	11	90.00	90.00	10.49	-	0.903	1.000
	VR _L	11	87.27	90.00	13.30	0.903	-	1.000
	VR _C	11	89.09	90.00	10.20	1.000	1.000	-
Effort	PFS	11	21.82	15.00	14.88	-	0.030*	0.045*
	VR _L	11	40.91	40.00	21.19	0.030*	-	1.000
	VR _C	11	42.73	30.00	22.40	0.045*	1.000	-
Frustration Level	PFS	11	3.18	0.00	5.13	-	0.108	0.138
	VR _L	11	13.64	10.00	11.20	0.108	-	0.192
	VR _C	11	21.36	15.00	21.80	0.138	0.192	-

Table 5: Raw TLX values with Wilcoxon-Signed Rank pairwise comparisons, Bonferroni corrected p-values. For Physical Demand, Mental Demand, and Effort the differences between VRFS conditions and PFS are statistically significant. Overall, the average Raw TLX values are equal or higher in the VRFS environment compared to PFS.

Finding 3 - Task Completion Time: The mean task completion times for VRFS were always higher than PFS and this difference was statistically significant for increased task difficulties (checks 2 & 3). This led to a generally lower efficiency of VRFS for CFT.

Implications: Based on our observations and user comments, the primary source of longer task times were implementation- or design-specific problems when interacting with virtual switches resulting in time-consuming and sometimes cumbersome interactions. A secondary source were the weight of the VR controller after longer use and the limited FOV of the Leap Motion camera. We believe the design of future VRFS could already greatly benefit from comparably small changes to the audiovisual design and feedback of virtual switches before focusing on greater technological challenges such as better hand or finger tracking, haptic feedback, or introducing physical switches.

4.4 Perceived Workload

Raw TLX was used for measuring task load during CFT. The data was again not normally distributed and we applied a non-parametric test, accordingly. The Friedman test identified a significant difference among the input methods (PFS, VR_L, VR_C) for the subscales Mental Demand ($\chi^2(2) = 6.343; N = 11; p = 0.042; W_{Kendall} = 0.288$), Physical Demand ($\chi^2(2) = 15.048; N = 11; p < 0.001; W_{Kendall} = 0.684$) as well as Effort ($\chi^2(2) = 9.897; N = 11; p = 0.007; W_{Kendall} = 0.450$).

Looking at post-hoc pairwise comparisons (Wilcoxon-Signed-Rank, Bonferroni corrected), we saw that for Mental Demand the

differences between PFS and VR_C ($p = 0.036; r = 0.533$) were statistically significant (Table 5). This is also the case for the differences related to Physical Demand (VR_L: $p = 0.024; r = 0.569$ and VR_C: $p = 0.009; r = 0.626$) and Effort (VR_L: $p = 0.030; r = 0.550$ and VR_C: $p = 0.045; r = 0.520$).

We observed two possible reasons for the increased Physical Demand. First, the participants reported that over time, the weight of the HMD became increasingly uncomfortable (P6, P8: after 20 minutes) (Fig. 2D). Second, the already reported issue of the weight of the VR controller (see section 4.3) resulted in users requiring both hands to hold it in mid-air to still achieve enough input precision. Furthermore, the already mentioned time-consuming interactions in VRFS very likely increased the Mental Demand and Effort.

Finding 4 - Workload: The NASA-TLX showed statistically significant differences between PFS and VRFS for Mental Demand, Physical Demand, and Effort. The increased Physical Demand and Effort was very likely caused by the weight of the VR HMD and VR controller as well as longer interactions with extended arms. The problematic design of interactions with virtual switches in VRFS possibly also added to Mental Demand.

Implications: Like in finding 3, we recommend better audiovisual representation and more natural mappings for virtual switches to accelerate task completion. This could reduce the perceived workload, especially if reducing the weight of VR HMDs and VR controllers is not possible.

Simulator Sickness Questionnaire (Difference: after - before)						p-values		
SSQ category	vis.	n	Mean	Median	SD	vis.	PFS	VRFS
Nausea	PFS	11	6.94	0.00	11.36	PFS	-	0.046*
	VRFS	11	3.47	0.00	11.51	VRFS	0.046*	-
Oculomotor	PFS	11	-5.51	0.00	9.03	PFS	-	0.011*
	VRFS	11	16.48	9.54	21.37	VRFS	0.011*	-
Disorientation	PFS	11	-3.80	0.00	9.00	PFS	-	0.017*
	VRFS	11	18.98	13.92	18.96	VRFS	0.017*	-
Total Score	PFS	11	-1.02	0.00	8.54	PFS	-	0.013*
	VRFS	11	14.59	15.93	18.87	VRFS	0.013*	-

Table 6: This table lists the differences of SSQ-values. The *after* value was subtracted from the *before* value.

4.5 Simulator Sickness

The SSQ questionnaire was completed before and after each of the two conditions (*PFS* and *VRFS*), so four times for each participant. SSQ literature associates a total SSQ score between 5-10 with "minimal symptoms", 10-15 with "significant symptoms", 15-20 with "symptoms are a concern", and values above 20 with a "problem simulator" [23]. Looking at the descriptive statistics (see Table 7), the total SSQ score after exposure to the *PFS* ($Med = 3.74, Mean = 7.48, SD = 10.84$) was therefore low in absolute terms. In contrast, the total score after exposure to the *VRFS* was notably high ($Med = 21.60, Mean = 23.81, SD = 22.68$) and entered the "problem simulator" range.

For further analysis, we applied a non-parametric test (Wilcoxon-Signed Rank) since the SSQ data was again not normally distributed. To address potential in-person variability, which could exist due to measurements for *PFS* and *VRFS* being taken on different days, we analyzed and compared the relative changes of before and after measurements within the different SSQ dimensions. This revealed that for *VRFS* changes in SSQ scores are mostly to the worse and significantly larger compared to the *PFS* condition. The differences are statistically significant for Nausea ($z = 2.000; p = 0.046; N = 11; r = 0.426$), Oculomotor ($z = 2.552; p = 0.011; N = 11; r = 0.544$), Disorientation ($z = -2.379; p = 0.017; N = 11; r = 0.507$), and the SSQ Total Score ($z = -2.497; p = 0.013; N = 11; r = 0.532$), see Fig. 8, Table 6.

Table 7 shows the pairwise before/after comparisons within each of the two visual presentations and indicates significant differences for *VRFS* in the same dimensions for Oculomotor ($z = -2.099; p = 0.036; r = 0.447$), Disorientation ($z = -2.388; p = 0.017; r = 0.509$), Total Score ($z = -2.201; p = 0.028; r = 0.469$), but not for Nausea ($z = -0.962; p = 0.336; r = 0.205$).

In conclusion, the total SSQ scores and the relative changes after exposure to *VRFS* clearly indicate that the employed VR technologies were prone to simulator sickness. Therefore, they yet cannot replace a *PFS* for CFT – in particular compared to the low total SSQ scores and the absence of relative changes to the worse after exposure to *PFS*. This was also confirmed during the user interviews in which all participants unanimously preferred the *PFS* and considered the *VRFS* rather a supplement and not a replacement of the *PFS*.

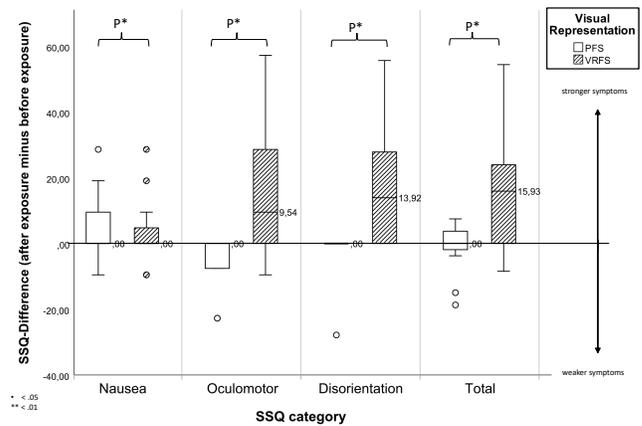


Figure 8: Box plot of the perceived change in Simulator Sickness due to the exposure to either *PFS* or *VRFS*. The change was statistically significant for all SSQ dimensions.

Finding 5 - Simulator Sickness: The SSQ revealed problematic levels of simulator sickness after exposure to *VRFS*. Furthermore, changes from before exposure to after were significantly larger than for *PFS*, and the participants unanimously preferred the *PFS* in user interviews. In conclusion, our consumer-grade *VRFS* cannot replace *PFS* for CFT yet.

Implications: Current consumer-grade *VRFS* need to be further optimized to reduce simulator sickness. During optimization, higher frame rates should be given priority over higher rendering resolution or greater level of detail for cockpit models. In general, for performing CFT in a consumer-grade *VRFS*, it is necessary to find a sweet spot between frame rates, rendering quality, and the cost of high-performance VR and graphics hardware.

5 LIMITATIONS AND IMPLICATIONS FOR FUTURE WORK

Due to the complexities of our study, which involved a high-end full-scale *PFS* that was part of a commercial flight simulation attraction and took place during the COVID-19 pandemic in July and August

Simulator Sickness Questionnaire						p-values				
SSQ category	vis / time	n	Mean	Median	SD					
						vis / time	before	after	before	after
Nausea	PFS / before	11	3.46	0.00	4.81	PFS / before	-	0.071	-	-
	PFS / after	11	10.41	0.00	13.12	PFS / after	0.071	-	-	-
	VR / before	11	6.94	9.54	7.50	VR / before	-	-	-	0.336
	VR / after	11	10.41	0.00	13.12	VR / after	-	-	0.336	-
Oculomotor	PFS / before	11	10.34	7.58	12.35	PFS / before	-	0.063	-	-
	PFS / after	11	4.82	0.00	9.14	PFS / after	0.063	-	-	-
	VR / before	11	10.34	7.58	11.38	VR / before	-	-	-	0.036*
	VR / after	11	21.36	15.16	20.84	VR / after	-	-	0.036*	-
Disorientation	PFS / before	11	7.59	0.00	13.00	PFS / before	-	0.180	-	-
	PFS / after	11	3.80	0.00	9.00	PFS / after	0.180	-	-	-
	VR / before	11	5.06	0.00	9.38	VR / before	-	-	-	0.017*
	VR / after	11	24.04	27.84	23.37	VR / after	-	-	0.017*	-
Total Score	PFS / before	11	8.50	3.74	9.47	PFS / before	-	0.943	-	-
	PFS / after	11	7.48	3.74	10.84	PFS / after	0.943	-	-	-
	VR / before	11	9.18	7.48	9.52	VR / before	-	-	-	0.028*
	VR / after	11	23.80	18.70	18.42	VR / after	-	-	0.028*	-

Table 7: Descriptive statistics for SSQ measurements. Before/After comparison for each dimension with Wilcoxon-Signed Rank test.

2020, we had to make concessions with regard to study design and execution. These resulted in limitations of our study.

First, an obvious limitation is the comparably small sample size of 11 participants. While it is not possible to infer whether our findings would have been confirmed also for larger samples, the very low error rates, substantial task completion time differences between PFS and VRFS, high total SSQ scores for VRFS ("problem simulator"), and unanimous preference of PFS seem very unlikely to substantially change or even reverse for more participants.

Second, differences in "off times" between both conditions (Group A: two weeks, Group B: one week) were unavoidable to meet the operating schedule of the PFS and the personal schedules of the participants. We ran a further analysis with this "off-time" as grouping variable and did not find a significant effect between these two groups of participants for neither error rate and task completion time, nor for instrument reading performance.

Third, there were two implementation-specific limitations of VRFS that reduce the generalizability of our findings:

(A.) As already discussed, frame rates were comparably low and changed for the different VR display resolutions (i.e., a higher resolution resulted in a lower frame rate) and they could not be precisely controlled. We could only ensure that frame rates never dropped below 25 frames per second which is comparably low for VR systems. Especially, for VR_{high} this was a contributing factor to the high total SSQ scores. However, in order to keep the simulation software stack in both conditions as similar as possible for internal validity, we could not use alternative and possibly faster simulator software. We tried to account for this by using a powerful PC and graphics card from the top of the consumer range.

To better address the problem of simulator sickness in consumer-grade VRFS, future research should attempt to identify the sweet spot between high frame rates, display resolutions, and hardware costs, i.e., the minimum frame rate and hardware costs with still acceptable levels of simulator sickness for sustained use of VRFS. Furthermore, future research should attempt to more precisely identify the sources of simulator sickness for VRFS in CFT. For example, in our study the SSQ values for VRFS increased for Oculomotor, Disorientation, and Total Score, but not for Nausea.

(B.) As discussed, one reason for the low efficiency of the VRFS were the input methods for interacting with switches and their specific implementation within the FlyInside software. This cannot be attributed to VR in general, since, at least in principle, it would be possible to replace them by improved and customized implementations in cooperation with the software manufacturers. Currently, they remain a key factor why interaction with consumer-grade VRFS is not yet efficient enough for cockpit familiarization training in commercial settings. Interestingly, the absence of haptic feedback in VR did not have a significant influence on error rates in our study. Therefore, it is promising to explore how close a VRFS can come to a PFS without costly equipment for haptic feedback. Future research should focus on software-based solutions for switch interaction in VRFS, e.g., improving the interaction design by using smarter and occlusion-free visual overlays and providing additional visual and/or acoustic feedback.

Nonetheless, as we discussed in related work, a further step could be improved hardware that also provides better haptic or tactile feedback and possibly a higher efficiency than VR_L or VR_C . Such hardware solutions are mostly research prototypes or niche products that are not available as consumer-grade off-the-shelf

products yet. Still, it seems promising to explore their effects on a VRFS's usability and on the pilot's learning success.

Fourth, based on our results, the implications for future Cockpit Familiarization Training in a Virtual Reality Flight Simulator are to

(1) use a minimum of $2,296 \times 2,320$ pixels per eye to provide an appropriate reading performance and prevent compensation strategies,

(2) use virtual cockpit models with a sufficient level of detail in order to increase frame rate and decrease levels of cybersickness,

(3) avoid occluding, non natural GUI elements during interaction with the virtual cockpit, and enhance the interaction by using additional feedback (e.g. acoustic),

(4) find a sweet spot between frame rates, rendering quality, and the cost of high-performance VR and graphics hardware, in order to get acceptable levels of simulator sickness.

6 CONCLUSION

In this paper, we presented the results of a user study that compared a full-scale physical flight simulator of a Boeing 737-800NG with a cost-efficient consumer-grade virtual reality flight simulator for the purpose of cockpit familiarization training. The overall effectiveness in terms of task completion and error rates of the virtual simulator was equivalent to the physical simulator. The absence of haptic feedback or physical switches did not impair participants' task completion and correctness and thus should not be considered a necessity for successful CFT. However, there were substantial and statistically significant differences in efficiency (i.e. longer task completion times for VR) and satisfaction (i.e. higher task load for VR). Also, the observed simulator sickness reached problematic levels, despite the fact that the simulated aircraft remained motionless at all times.

Based on our five main findings and their implications, we conclude that recent consumer-grade VR cannot fully replace PFS for cockpit familiarization training yet. However, we believe that software improvements with regard to the interaction and audiovisual representation of the virtual switches as well as increasing frame rates will be able to solve the observed problems. Still, since VR input and haptic output techniques are an active field of research, it seems promising to also further explore such technologies in future work on VRFS.

ACKNOWLEDGMENTS

This paper is a part of the X-PRO project. The project X-PRO is financed by research subsidies granted by the government of Upper Austria. The source code of the .NET-client is based on PMDG-Connect²² created by Noah Pell. Many thanks to the crew of Synthetic737²³ for the support during the user study.

REFERENCES

- [1] Gae Ae Ryu and Kwan-Hee Yoo. 2020. Key Factors for Reducing Motion Sickness in 360° Virtual Reality Scene: Extended Abstract. In *The 25th International Conference on 3D Web Technology* (Virtual Event, Republic of Korea) (*Web3D '20*). Association for Computing Machinery, New York, NY, USA, Article 28, 2 pages. <https://doi.org/10.1145/3424616.3424723>
- [2] Bruno Araujo, Ricardo Jota, Varun Perumal, Jia Xian Yao, Karan Singh, and Daniel Wigdor. 2016. Snake Charmer. In *TEI '16*, Saskia Bakker, Caroline Hummels, and Brygg Ullmer (Eds.). ACM, New York, 218–226. <https://doi.org/10.1145/2839462.2839484>
- [3] Turgay Aslandere, Daniel Dreyer, and Frieder Pankratz. 2015. Virtual hand-button interaction in a generic virtual reality flight simulator. In *2015 IEEE Aerospace Conference*. IEEE, Big Sky, MT, 1–8. <https://doi.org/10.1109/AERO.2015.7118876>
- [4] Turgay Aslandere, Daniel Dreyer, Frieder Pankratz, and Rene Schubotz. 2014. A generic virtual reality flight simulator. In *Virtuelle und Erweiterte Realität, 11. Workshop der GI-Fachgruppe VR/AR*. Shaker Verlag, Bremen, Germany, 1–13.
- [5] Gordon Baxter, Denis Besnard, and Dominic Riley. 2007. Cognitive mismatches in the cockpit: will they ever be a thing of the past? *Applied ergonomics* 38, 4 (2007), 417–423. <https://doi.org/10.1016/j.apergo.2007.01.005>
- [6] Tobias Bellmann, Martin Otter, and Gerd Hirzinger. 2011. The DLR Robot Motion Simulator Part II: Optimization based path-planning. In *2011 IEEE International Conference on Robotics and Automation*, Antonio Bicchi (Ed.). IEEE, Piscataway, NJ, 4702–4709. <https://doi.org/10.1109/ICRA.2011.5979914>
- [7] James C Byers. 1989. Traditional and raw task load index (TLX) correlations: Are paired comparisons necessary? *Advances in Industrial Ergonomics and Safety I: Taylor and Francis*. 1, 1 (1989), 481–485.
- [8] Tom Carter, Sue Ann Seah, Benjamin Long, Bruce Drinkwater, and Sriram Subramanian. 2013. UltraHaptics: multi-point mid-air haptic feedback for touch surfaces. In *Proceedings of the 26th annual ACM symposium on User interface software and technology - UIST '13*. ACM Press, St. Andrews, Scotland, United Kingdom, 505–514. <https://doi.org/10.1145/2501988.2502018>
- [9] Yen-Chih Chiang, Shih-Song Cheng, Huei-Siou Chen, Le-Jean Wei, Li-Min Huang, and David KT Chu. 2018. Retinal Resolution Display Technology Brings Impact to VR Industry. In *ACM SIGGRAPH 2018 Posters (SIGGRAPH '18)*. ACM, New York, NY, USA, 43:1–43:2. <https://doi.org/10.1145/3230744.3230781>
- [10] R. Crosbie and D. Kiefer. 1985. Controlling the human centrifuge as a force and motion platform for the dynamic flight simulator. In *Flight Simulation Technologies Conference*. American Institute of Aeronautics and Astronautics, Reston, Virginia, 1742. <https://doi.org/10.2514/6.1985-1742>
- [11] Asaf Degani and Earl L. Wiener. 1993. Cockpit Checklists: Concepts, Design, and Use. *Human Factors* 35, 2 (1993), 345–359. <https://doi.org/10.1177/001872089303500209>
- [12] Kai-Uwe Doer, Jens Schiefel, and W. Kubbat. 2001. *Virtual Cockpit Simulation for Pilot Training*. Technical Report. University of Darmstadt (Germany).
- [13] Jason Dowling, Wageeh Boles, and Anthony Maeder. 2006. Simulated Artificial Human Vision: The Effects of Spatial Resolution and Frame Rate on Mobility. In *Advances in Intelligent IT: Active Media Technology 2006*, Y Li, M Looi, and N Zhong (Eds.). IOS Press, Netherlands, 138–143.
- [14] Joseph Dumas, Andrew Novobilski, Dawn Ellis, and Mark Paschal. 2002. VR on a Budget: Developing a Flight Simulator in a Small Institution with off-the-Shelf Hardware and Open Source Software. *J. Comput. Sci. Coll.* 18, 2 (Dec. 2002), 138–142.
- [15] James Exter. 1989. The modular cockpit approach to aircrew training device development. In *Flight Simulation Technologies Conference and Exhibit*. American Institute of Aeronautics and Astronautics, Reston, Virginia, 225–232. <https://doi.org/10.2514/6.1989-3293>
- [16] David Goedicke, Jamy Li, Vanessa Evers, and Wendy Ju. 2018. VR-OOM: Virtual Reality On-rOad driving siMulation. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*. ACM Press, Montreal QC, Canada, 1–11. <https://doi.org/10.1145/3173574.3173739>
- [17] Jerome N. Gregoire, Celeste M. Alfes, Andrew P. Reimer, and Mary F. Terhaar. 2017. Flying Lessons for Clinicians: Developing System 2 Practice. *Air Medical Journal* 36, 3 (May 2017), 135–137. <https://doi.org/10.1016/j.amj.2017.02.003>
- [18] Cameron Grout, William Rogers, Mark Apperley, and Steve Jones. 2015. Reading Text in an Immersive Head-Mounted Display: An Investigation into Displaying Desktop Interfaces in a 3D Virtual Environment. In *Proceedings of the 15th New Zealand Conference on Human-Computer Interaction* (Hamilton, New Zealand) (*CHINZ 2015*). Association for Computing Machinery, New York, NY, USA, 9–16. <https://doi.org/10.1145/2808047.2808055>
- [19] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Advances in Psychology*. Vol. 52. Elsevier, Amsterdam, 139–183. [https://doi.org/10.1016/S0166-4115\(08\)62386-9](https://doi.org/10.1016/S0166-4115(08)62386-9)
- [20] Jasmine S. Hayes, Vivian T. Yin, Duke Piyathaisere, James D. Weiland, Mark S. Humayun, and Gislin Dagnelie. 2003. Visually guided performance of simple tasks using simulated prosthetic vision. *Artificial organs* 27, 11 (2003), 1016–1028. <https://doi.org/10.1046/j.1525-1594.2003.07309.x>
- [21] Hans-Christian Jetter, Roman Rädle, Tiare Feuchtmeyer, Christoph Anthes, Judith Friedl, and Clemens Nylandstedt Klokmoose. 2020. "In VR, Everything is Possible!": Sketching and Simulating Spatially-Aware Interactive Spaces in Virtual Reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (*CHI '20*). Association for Computing Machinery, New York, NY, USA, 1–16. <https://doi.org/10.1145/3313831.3376652>
- [22] Genovefa Kefalidou, Mirabelle D'Cruz, André Castro, and Rui Marcelino. 2019. Designing Airport Interiors with 3D Visualizations. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland

²²PMDGConnect <https://www.github.com/noah4477/PMDGConnect>

²³Synthetic737 <https://www.synthetic737.at>

- Uk) (*CHI EA '19*). Association for Computing Machinery, New York, NY, USA, 1–8. <https://doi.org/10.1145/3290607.3299045>
- [23] Robert S Kennedy, Julie M Drexler, Daniel E Compton, Kay M Stanney, D Susan Lanham, and Deborah L Harm. 2003. Configural Scoring of Simulator Sickness, Cybersickness and Space Adaptation Syndrome. In *Virtual and adaptive environments*, Lawrence J. Hettinger and Michael W. Haas (Eds.). Lawrence Erlbaum Associates Publishers, Mahwah, NJ, 247–278. <https://doi.org/10.1201/9781410608888.ch12>
- [24] Robert S. Kennedy, Norman E. Lane, Kevin S. Berbaum, and Michael G. Lilienthal. 1993. Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness. *The International Journal of Aviation Psychology* 3, 3 (1993), 203–220. https://doi.org/10.1207/s15327108ijap0303_3
- [25] T. Kojić, D. Ali, R. Greinacher, S. Möller, and J. Voigt-Antons. 2020. User Experience of Reading in Virtual Reality – Finding Values for Text Distance, Size and Contrast. In *2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*. 2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX), Athlone, Ireland, 1–6. <https://doi.org/10.1109/QoMEX48832.2020.9123091>
- [26] Thomas Longridge, Paul Ray, Edward M. Boothe, and Judith S. Burki-Cohen. 1996. *Initiative towards more affordable flight simulators for U.S. commuter airline training*. Royal Aeronautical Society, USA.
- [27] Maxine Lubner, Sharon Devivo, Emerson Allen, Andrew Dattel, and Deb Henneberry. 2013. Preliminary Examination of Simulator-Based Training Effectiveness. *17th International Symposium on Aviation Psychology* 17, 1 (2013), 44–49. https://corescholar.libraries.wright.edu/isap_2013/103/
- [28] R.G. Menendez and J.E. Bernard. 2001. Flight simulation in synthetic environments. *IEEE Aerospace and Electronic Systems Magazine* 16, 9 (Sept. 2001), 19–23. <https://doi.org/10.1109/62.949532>
- [29] W. F. Moroney, D. W. Biers, F. T. Eggemeier, and J. A. Mitchell. 1992. A comparison of two scoring procedures with the NASA task load index in a simulated flight task. In *Proceedings of the IEEE 1992 National Aerospace and Electronics Conference*. IEEE, Dayton, USA, 734–740 vol.2. <https://doi.org/10.1109/NAECON.1992.220513>
- [30] Matthias Oberhauser, Reinhard Brauningl, Daniel Dreyer, and Ioana Victoria Koglbauer. 2016. Pilots' Interaction with Hardware Controls in a Virtual Reality Flight Simulator. In *Proceedings of the 32nd Conference of the European Association for Aviation Psychology*. European Association for Aviation Psychology, Cascais, Portugal.
- [31] Matthias Oberhauser and Daniel Dreyer. 2017. A virtual reality flight simulator for human factors engineering. *Cognition, Technology & Work* 19, 2 (2017), 263–277. <https://doi.org/10.1007/s10111-017-0421-7>
- [32] Matthias Oberhauser, Daniel Dreyer, Reinhard Brauningl, and Ioana Koglbauer. 2018. What's Real About Virtual Reality Flight Simulation?: Comparing the Fidelity of a Virtual Reality With a Conventional Flight Simulation Environment. *Aviation Psychology and Applied Human Factors* 8, 1 (March 2018), 22–34. <https://doi.org/10.1027/2192-0923/a000134>
- [33] Jesse Orlansky, Joseph String, and Paul R Chatelier. 1982. *The cost-effectiveness of military training*. Technical Report. Institute for defence analyses Alexandria.
- [34] Randy Pausch, Thomas Crea, and Matthew Conway. 1992. A Literature Survey for Virtual Environments: Military Flight Simulator Visual Systems and Simulator Sickness. *Presence: Teleoperators and Virtual Environments* 1, 3 (1992), 344–363. <https://doi.org/10.1162/pres.1992.1.3.344> arXiv:<https://doi.org/10.1162/pres.1992.1.3.344>
- [35] Vsevolod Peysakhovich, Louis Monnier, Mélanie Gornet, and Stéphane Juaneda. 2020-03. Virtual reality vs. real-life training to learn checklists for light aircraft. In *Eye-Tracking in Aviation. Proceedings of the 1st International Workshop (ETAVI 2020)*. ISAE-SUPAERO, Université de Toulouse; Institute of Cartography and Geoinformation (IKG), ETH Zurich, Toulouse; Zurich, 47 – 53. <https://doi.org/10.3929/ethz-b-000407648>
- [36] Daniel Podgórski. 2010. The Use of Tacit Knowledge in Occupational Safety and Health Management Systems. *International Journal of Occupational Safety and Ergonomics* 16, 3 (2010), 283–310. <https://doi.org/10.1080/10803548.2010.11076845> PMID: 20828486.
- [37] Savern Reweti. 2014. *PC-based aviation training devices for pilot training in visual flight rules procedures : development, validation and effectiveness*. Doctoral. Massey University.
- [38] R Richards. 2002. Common cockpit helicopter training simulator. In *AVSIM 2002 Conference And Exhibition*. AVSIM 2002 Conference And Exhibition, Bellevue, Washington, 13–14.
- [39] Andrew Robinson, Katerina Mania, and Philippe Perey. 2004. Flight simulation: research challenges and user assessments of fidelity. In *Proceedings of the 2004 ACM SIGGRAPH international conference on Virtual Reality continuum and its applications in industry - VRCAI '04*. ACM Press, Singapore, 261. <https://doi.org/10.1145/1044588.1044644>
- [40] J. M. Rolfe and K. J. Staples. 1997. *Flight simulation* (repr ed.). Cambridge Univ. Press, Cambridge.
- [41] Robert Rosenthal, Harris Cooper, L Hedges, et al. 1994. Parametric measures of effect size. *The handbook of research synthesis* 621, 2 (1994), 231–244.
- [42] J. Ryu, N. Hashimoto, and M. Sato. 2005. Influence of resolution degradation on distance estimation in virtual space displaying static and dynamic image. In *2005 International Conference on Cyberworlds (CW'05)*. IEEE, Singapore, 8 pp.–50. <https://doi.org/10.1109/CW.2005.54>
- [43] Jens Schiefele, Oliver Albert, and Kai Uwe Doerr. 1998. IFR flight simulation in a distributed virtual environment. In *Modeling and Simulating Sensory Response for Real and Virtual Environments (SPIE Proceedings)*. SPIE, Bellingham, United States, 111. <https://doi.org/10.1117/12.317567>
- [44] Jens Schiefele, Oliver Albert, Kai Uwe Doerr, Martin Kelz, and Norman Schmidt-Winkel. 1999. Evaluation of required HMD resolution and field of view for a virtual cockpit simulation. In *Helmet- and Head-Mounted Displays IV*, Ronald J. Lewandowski, Loran A. Haworth, and Henry J. Girolamo (Eds.), Vol. 3689. International Society for Optics and Photonics, SPIE, Bellingham, United States, 143 – 154. <https://doi.org/10.1117/12.352824>
- [45] Peter Schulz, Dmitry Alexandrovsky, Felix Putze, Rainer Malaka, and Johannes Schöning. 2019. The Role of Physical Props in VR Climbing Environments. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland UK) (*CHI '19*). Association for Computing Machinery, New York, NY, USA, 1–13. <https://doi.org/10.1145/3290605.3300413>
- [46] Georg Stevenson, Andreas Riener, and Alois Ferscha. 2013. "TactiGlove" – A Guidance System to Effectively Find Hidden Spots in 3D Space. In *Mobile Computing, Applications, and Services*. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, Vol. 110. Springer Berlin Heidelberg, Berlin/Heidelberg, 80–99. https://doi.org/10.1007/978-3-642-36632-1_5
- [47] Kelvin Valentino, Kevin Christian, and Endra Joelianto. 2017. Virtual reality flight simulator. *Internetworking Indonesia Journal* 9, 1 (2017), 21–25.
- [48] John Vince. 1993. 10 - Virtual Reality Techniques in Flight Simulation. In *Virtual Reality Systems*, R.A. Earnshaw, M.A. Gigante, and H. Jones (Eds.). Academic Press, Boston, 135 – 141. <https://doi.org/10.1016/B978-0-12-227748-1.50018-4>
- [49] Eric Whitmire, Hrvoje Benko, Christian Holz, Eyal Ofek, and Mike Sinclair. 2018. Haptic Revolver: Touch, Shear, Texture, and Shape Rendering on a Reconfigurable Virtual Reality Controller. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*. ACM Press, Montreal QC, Canada, 1–12. <https://doi.org/10.1145/3173574.3173660>
- [50] Guy Williams, Ken Lawrence, and Richard Weeks. 2004. Reconfigurable Flight Simulators in Modeling and Simulation. In *AIAA Modeling and Simulation Technologies Conference and Exhibit (Guidance, Navigation, and Control and Co-located Conferences)*. American Institute of Aeronautics and Astronautics, Providence, Rhode Island. <https://doi.org/10.2514/6.2004-5459>
- [51] Chou Wusheng, Wang Tianmiao, and Hu Lei. 2003. Design of data glove and arm type haptic interface. In *11th Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*. IEEE Computer Society, Los Alamitos, Calif, 422–427. <https://doi.org/10.1109/HAPTIC.2003.1191332>
- [52] I Yavrucuk, E. Kubali, and O. Tarimci. 2011. A low cost flight simulator using virtual reality tools. *IEEE Aerospace and Electronic Systems Magazine* 26, 4 (April 2011), 10–14. <https://doi.org/10.1109/MAES.2011.5763338>
- [53] L. Yu, Z. Xu, and L. Xin. 2019. Research on Reconfigurable Method of Cockpit Simulation for Virtual Training System. In *2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*. IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), Chengdu China, 2007–2011. <https://doi.org/10.1109/ITNEC.2019.8729557>