Original Article

What's Real About Virtual Reality Flight Simulation?

Comparing the Fidelity of a Virtual Reality With a Conventional Flight Simulation Environment

Matthias Oberhauser,[1] Daniel Dreyer,[2] Reinhard Braunstingl,[1] and Ioana Koglbauer[1]

[1]Graz University of Technology, Graz, Austria
[2]Airbus Defence and Space, Manching, Germany

Abstract: With the help of immersive virtual reality technology, novel cockpit systems can be evaluated with pilots in an early design phase. This comparative study investigates the functional fidelity of a virtual reality flight simulator (VRFS) in comparison with a conventional flight simulator. Pilots' movement time to reach cockpit controls, deviation from the ideal flight path, workload, and simulator sickness are evaluated using an operational scenario. The results show statistically significant differences in heading, altitude, and flight path, as well as delays in operating the controls in virtual reality. Yet, most participants could safely and reliably complete the flight task. For use cases in which adaptations to pace, exposure time, and flight task are acceptable, which is often the case in early phases of the design process, VRFSs can be viable tools for human factors engineering.

Keywords: virtual reality, flight simulation, flight performance, movement time

Virtual reality (VR) technology enables the early experiencing and testing of products in a fully immersive virtual environment. Thanks to recent technological advances in this field, new applications of VR technology in academia and industry emerge on a regular basis. All these systems have in common that they can only provide an artificial representation of reality, that is, they offer a limited level of fidelity. The visualization and the possibilities of interaction are limited and biased by various confounding factors. To interpret results that are being gathered in VR systems, this bias must be examined. The level of this bias is influenced by the hard- and software setup and the given task. Hence it is unique for every use case. In publications related to VR, the focus often lies on the development of new systems and the demonstration of novel use cases, whereas the fidelity assessment of these systems has often been ignored. In this article, a flight simulator based on immersive VR technology is presented: the Virtual Reality Flight Simulator (VRFS), which was developed by and is used at Airbus Group Innovations. In order to assess the fidelity of the system and the bias that is introduced by the virtual environment, the system is compared with a conventional hardware flight simulator.

Virtual Reality

Rheingold (1991) defines virtual reality (VR) as an experience in which a person is "surrounded by a three-dimensional computer-generated representation, and is able to move around in the virtual world and see it from different angles, to reach into it, grab it, and reshape it" (p. 17).

This quote describes the conventional view of VR and is the underlying definition in this paper. VR technology presents a fully synthetic environment to a user while completely omitting the real world. Systems that offer any kind of visual merging of the real and the virtual world fall under the term "mixed reality" (MR), with augmented reality (AR) as the best-known technology (Milgram & Kishino, 1994).

The origins of VR technology reach back to the 1960s. In 1962, Morton Heilig patented the Sensorama Simulator, a device similar to a slot machine with a high field of view (FOV) display. In 1965, Ivan Sutherland developed the first head-mounted display (HMD), the anecdotally so-called Sword of Damocles. In the early 1990s, based on this first pioneering research, VR technology was about to enter the consumer market with support from major companies in the consumer electronics industry. Yet, with the limited

https://doi.org/10.1027/2192-0923/a000134

computing power at this time, it was not possible to deliver a satisfying experience at a reasonable price (Robertson & Zelenko, 2014). By the mid-1990s, only selected use cases in industry and academia remained. In his paper "What's Real About Virtual Reality," Brooks (1999) reviewed some of these applications in the field of training, ergonomic assessments, and psychological treatment. He concluded that VR "that used to almost work now barely works" (p. 16). Today, VR is on the verge of entering the consumer market once again with major hard- and software companies currently developing affordable HMDs, controllers, and tracking systems that overcome the limitations of earlier VR devices (Avila & Bailey, 2014).

Flight Simulators

Today, a wide range of flight simulators for training are available: from simple desktop systems to procedure-training devices and motion-based full-flight simulators (Robinson, Mania, & Perey, 2004). The European Aviation Safety Agency (2012) classifies these simulators as basic instrument training devices, flight and navigation procedures trainers, flight training devices (Level 1 and 2), and full-flight simulators (Level A–D) with a given set of requirements that these systems have to fulfil. The use of flight simulators is not limited to training purposes only. With the rise of human factors engineering, the need for a means of evaluating flight deck designs has gained importance (Rehmann, 1995). Conventional simulators consist of a mathematical flight model, a system simulation, a hardware cockpit replica, an outside visual, and an optional motion simulation (Rolfe & Staples, 1986). These sophisticated systems in some cases may exceed the procurement costs of the real system. Hence it is necessary to adapt the level of realism to the use case, that is, aim for an optimal cost/benefit ratio (Farmer, Rooij, Riemersma, Jorna, & Moraal, 1999). A simulator that, as Kaiser and Schroeder (2002) state, makes "a greater use of the virtual to replace the physical" (p. 466) can help to optimize this ratio for some use cases.

Related Work

With the advent of immersive virtual technology in the early 1990s, a new type of flight simulator based on the emerging VR technology was envisioned: a flight simulator based on VR technology that uses an HMD, which is flexible, mobile, and takes up less space than a conventional hardware simulator (Moroney & Moroney, 2009). Early demonstrators of immersive virtual flight simulations were created by McCarty, Sheasby, Amburn, Stytz, and Switzer (1994) and by Persiani, Piancastelli, and Liverani (1997). These simulators suffered from contemporary hard- and software limitations. Later, extensive research at the Technical University of Darmstadt was conducted to connect a virtual cockpit to a flight simulator for the purpose of flight training (Dörr, Schiefele, & Kubbat, 2001). More recent demonstrators are the low-cost VR flight simulator presented by Yavrucuk, Kubali, Tarimci, and Yilmaz (2009) or the Rapidly Reconfigurable Research Cockpit by Joyce and Robinson (2015).

The Virtual Reality Flight Simulator

In this study, the VRFS that was developed with a focus on human factors engineering and system integration will be used (Oberhauser et al., 2015; Oberhauser & Dreyer, 2017). The system is based on a demonstrator developed by Aslandere, Dreyer, Pantkratz, and Schubotz (2014) that mainly uses consumer technology.

The core system of the VRFS consists of an optical head and hand tracking system, an HMD, and a flight simulation, as shown in Figure 1 (Dreyer, Oberhauser, & Bandow, 2014). The HMD that is used for this study has a diagonal FOV of 60° and a resolution of $1{,}280 \times 1{,}024$ pixels for each eye. Based on the tracking input, a three-dimensional scene is rendered by the flight simulation. This scene includes a virtual cockpit, the outside visual, and a representation of the human hand. Also, some essential hardware elements such as a flight stick or flight control panels are integrated into the system. Placed on the exact same position as in the virtual cockpit, these hardware elements create a so-called mixed mock-up and enable an easy interaction in the virtual environment (Aslandere et al., 2014). All these components communicate through an open source network framework, the robot operating system (ROS; Quigley et al., 2009). With this framework, external hard- and software components such as flight displays, hardware elements, or system simulations can easily be integrated (Oberhauser, Dreyer, Convard, & Mamessier, 2016).

This simulator has been actively used for evaluating cockpit concepts and human–machine interface (HMI) components in an early stage of the engineering design process (Oberhauser et al., 2015, 2016). These evaluations have been designed as comparative studies, in which a legacy cockpit is compared with a novel cockpit concept. With this approach, confounding factors that stem from the virtual environment can be ruled out.

Simulator Fidelity

All flight simulators, whether conventional or virtual, provide different levels of fidelity, that is, the extent to which the simulator mimics the behavior of the real system

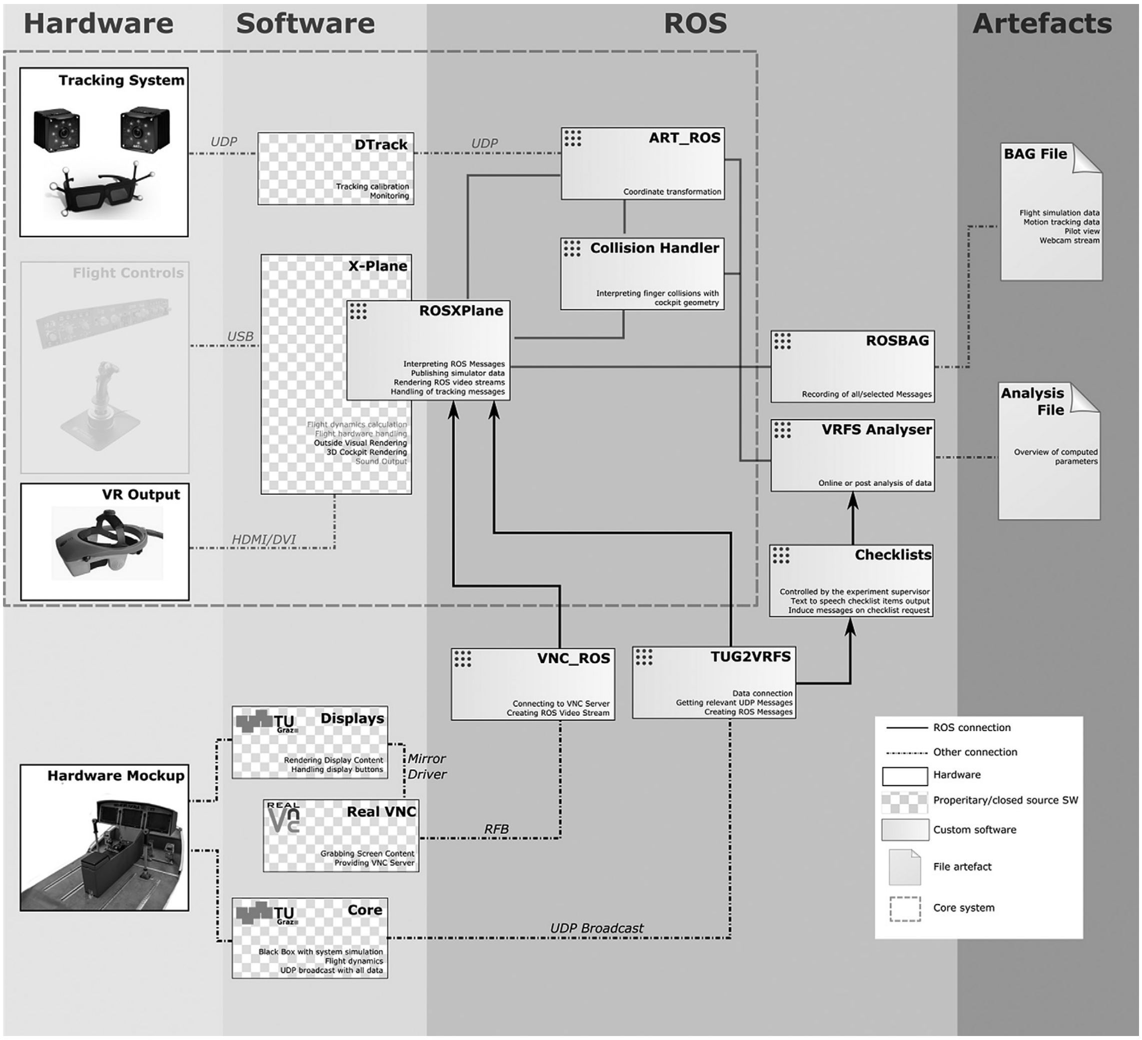

Figure 1. The architecture of the Virtual Reality Flight Simulator.

(Farmer et al., 1999). The term "fidelity" has multiple subcategories such as physical fidelity, functional fidelity, or task fidelity (Liu, Macchiarella, & Vincenzi, 2009). For VR systems, this variety of categories is often condensed to the term "immersion" (McMahan, Bowman, Zielinski, & Brady, 2012). Quantifying the fidelity or immersion with a set of metrics is an almost impossible task as considering all aspects in such an analysis is not feasible (Liu et al., 2009). Another approach is measuring fidelity indirectly by comparing a real-life flying task with a similar task in a simulator. The subjective human perception and objective performance data can be used to assess the fidelity of a simulation environment (Perfect, White, Padfield, & Gubbels, 2013). A similar approach will be used in this study by comparing the VRFS with a conventional hardware simulator. In particular, the task fidelity, that is, the replication of tasks and maneuvers, and the functional fidelity, that is, the replication of the effects of user input, are of interest.

Method

Apparatus

For this study, the VRFS is compared with a general aviation hardware flight simulator located at the Technical University of Graz. This simulator features a 190-degree outside visual, which is rendered in X-Plane, a commercial flight simulation software. It simulates engine and cockpit sound effects and offers a radio simulation. The simulator has a glass cockpit with two primary flight displays (PFD) and one multi-function display (MFD). An existing outside

Figure 2. A user immersed in the Virtual Reality Flight Simulator.

hull including an overhead panel was removed for the trials in favor of an unobstructed view for the tracking system.

This simulator is used for research purposes and is not certified according to the European Aviation Safety Agency (2012). Still, in the absence of a motion system, the simulator would fulfil the requirements for a flight training device Level 2.

The cockpit of the hardware simulator was remodeled and integrated into the VR environment as shown in Figure 2. It illustrates that the pilot is seated in the conventional flight simulator but wearing an HMD and tracking targets attached to the hands. The custom flight model, the hardware components, and the flight displays are connected to the virtual environment through the ROS network framework. In this way, it can be ensured that the functionality and even the haptics of both the hardware as well as the virtual simulator are identical. For the participant, this means that if he or she touches a control element in the VR environment, he or she simultaneously touches this control element in the real hardware mock-up, which leads to the respective haptic sensation.

Independent Variable

The experiment has one within-subject variable: the simulation environment. Therefore, two variants are tested, the conventional hardware simulator and the VR environment. The order of presentation was counterbalanced in the experiment.

Procedure

The flight task takes place at Graz Airport (LOWG). It starts with a short taxiing phase from the parking position, a takeoff from Runway 17C, and a left-hand traffic pattern at an altitude of 2,000 ft. During the scenario, the subject receives instructions via pre-recorded audio commands, as shown in Figure 3A, for operating all cockpit elements except for the throttle lever and the flight stick, which have to be controlled permanently. Figure 4 shows the cockpit elements that are used during the traffic pattern. The heading bug and the altitude bug are controlled via rotary knobs in the flight control unit (FCU). The flaps lever, the parking brake, and the throttle are located in the pedestal. The flight stick is controlled with the left hand.

The subjects must report abeam threshold; the base turn is commanded 30 s after this report. After a touch-and-go and a second left-hand pattern, the subjects have to perform a landing with a full stop. This scenario was conducted once in the virtual environment using the HMD and once in the hardware simulator with the projected outside visual. The order of the simulation environment was counterbalanced to avoid artifacts. Thus, half of the group began with the VR simulation and half of the group began with the hardware simulator. Before the tests, the pilots flew a left-hand pattern for familiarization with each simulation environment.

Dependent Measures

Movement Time

Movement time is used to measure the interaction performance. The movement time is the timespan of the hand traveling toward the control element, the so-called ballistic phase after the pilot has received the audio command. The reaction time is excluded as it highly depends on the subject and the current workload level. The movement time is an indicator of the difficulty to perceive and reach the position of a cockpit control element in the two different simulation environments. In the standard setup of the VRFS, the subject wears optical tracking targets on both hands to allow the optical tracking system to capture the position and orientation, which is necessary to visualize

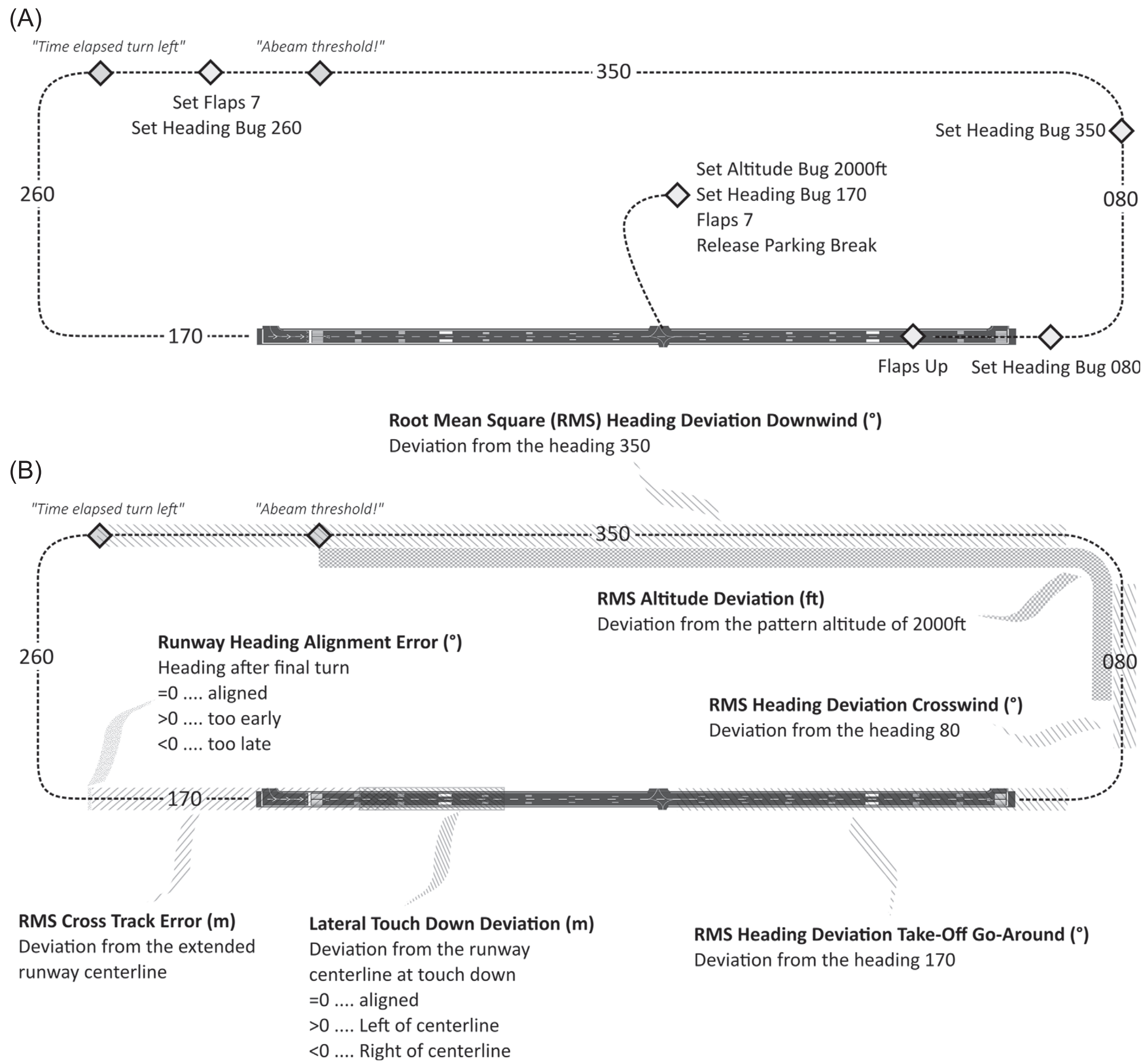

Figure 3. The objective dependent measures of the pilot's performance related to the traffic pattern.

the hands in the virtual environment. This existing positional information is used to derive the interaction times in an automatic process. The system can distinguish between the reaction time, that is, the time the pilot is not moving his or her hand after a command, and the movement time, that is, the time it takes to reach the control element.

Flight Performance

Flight performance was measured as the deviance from the ideal traffic pattern. The pattern was divided into flight segments, similar to a study by Le Ngoc and Kalawsky (2013). Several dependent measures were used to judge the pilots' performance. Figure 3B illustrates the traffic pattern and the corresponding dependent measurements:

- Heading deviation,
- Altitude deviation,
- Runway heading alignment error,
- Final approach cross track error, and
- Lateral touchdown deviation.

Root Mean Square (RMS) Heading Deviation (Degrees)

This parameter was measured for takeoff and touch-and-go, for the crosswind leg and the downwind leg. The participants were commanded to select the appropriate heading

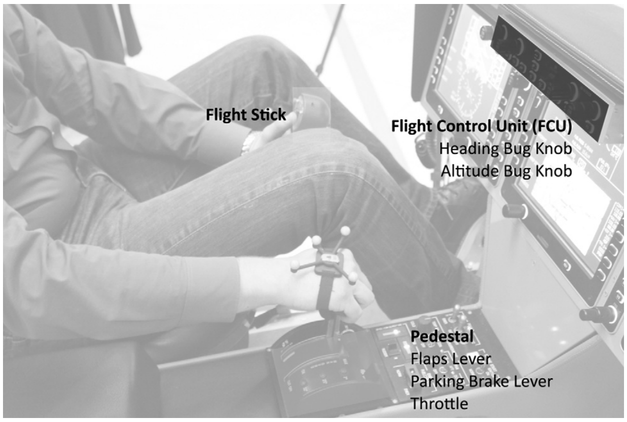

Figure 4. The cockpit elements that are relevant for the scenario.

during the trial and were briefed to keep this heading as exact as possible.

RMS Altitude Deviation (Feet)
This parameter was measured from the point at which the pilot reached the target altitude of 2,000 ft until the "abeam threshold" report. The participants were briefed about the altitude of the traffic pattern. The altitude bug in the PFD was set to this altitude as a reminder before takeoff.

Runway Alignment Error (Degrees)
The runway alignment error is the heading deviation after the roll-out of the final turn. If the value is close to zero, the aircraft is aligned. If it is smaller than zero, the turn was stopped too early; if it is bigger than zero, it was finished too late.

RMS Cross Track Error (Meters)
This is the lateral deviation of the aircraft from the extended runway centerline: The more stable the final approach, the smaller the RMS cross-track error.

Lateral Touchdown Deviation (Meters)
This is the lateral deviation from the runway centerline at touchdown. The longitudinal deviation was not analyzed as the participants did not receive instructions on a target touchdown point.

Workload
Pilots' workload was measured using the NASA-Task Load Index (NASA-TLX; Hart & Staveland, 1988). After each trial, the pilots rated their mental, physical, and temporal demand, effort, frustration, and performance on a scale ranging from *very low* (−5) to *very high* (+5).

Simulator Sickness
The Simulator Sickness Questionnaire (SSQ) was applied after each trial. This standardized, subjective questionnaire measures simulator sickness by asking participants to rate the intensity of 27 symptoms associated with simulator sickness. These symptoms are then attributed to three categories: nausea, oculomotor, and disorientation (Kennedy, Lane, Berbaum, & Lilienthal, 1993).

Participants

In all, 28 private and commercial pilots with a mean age of 42.5 years (minimum = 25, maximum = 72, *Mdn* = 37 years) participated in this trial. The average total experience of the participants was 2,485 flight hours (*Mdn* = 575 flight hours). The average flight experience of the participants in the past 90 days before the experiment was 33 flight hours (*Mdn* = 11 flight hours). The average flight simulator experience of the participants was 95 hr (*Mdn* = 21 flight hours). None of the pilots had experience with VR. Every participant signed an informed consent form. Two pilots did not complete the trials owing to simulator sickness and problems with the vision in the virtual environment.

Data Analysis

A repeated-measures analysis of variance with two within-subjects factors was conducted: the simulation environment (VR vs. conventional) and the trial (first vs. second traffic pattern). The analysis of kinematic deviation data used absolute values. To analyze the relationship between age, total flight hours, and hours in the past 90 days and individual differences in movement time, flight performance, and workload, we calculated Spearman's of correlation. Alpha was set at .05.

Results

Movement Time

The movement time for the rotary knobs was significantly longer in VR than in the conventional flight simulation as shown in Table 1. The mean difference between VR and hardware simulation varied between 1.113 and 1.681 s on the ground and varied between 0.945 and 1.165 s during flight.

The pilots needed a significantly longer movement time for setting the flaps for landing in VR than in the conventional flight simulation. However, there were no significant differences in the movement time for retracting the flaps

Table 1. Movement time (s) for the first traffic pattern

Control element	Type	Mean difference (*SE*)	Repeated measures ANOVA *F*	*p*	η^2	95% CI *LL*	*UL*
In flight	Heading bug 080	0.945 (0.204)	*F*(1, 22) = 21.43	.0001	.49	.522	1.369
	Heading bug 350	0.957 (0.199)	*F*(1, 24) = 23.11	.0001	.49	.546	1.367
	Heading bug 260	1.165 (0.274)	*F*(1, 25) = 18.13	.0001	.42	.602	1.729
	Flaps up	0.863 (0.603)	*F*(1, 22) = 2.04	.17	.09	–	–
	Flaps final	0.536 (0.219)	*F*(1, 23) = 6.00	.02	.20	–	–
On ground	Heading bug 170	1.113 (0.314)	*F*(1, 24) = 12.55	.002	.34	.465	1.762
	Altitude bug 2000	1.681 (0.460)	*F*(1, 24) = 13.35	.001	.36	.732	2.631
	Flaps takeoff	1.010 (0.203)	*F*(1, 24) = 24.68	.0001	.51	.590	1.430

Note. Mean differences, standard errors, and 95% confidence intervals refer to differences between the movement time in virtual reality and in the conventional simulation. CI = confidence interval. *LL* = lower limit. *UL* = upper limit.

Table 2. Results of the statistical analysis of kinematic data

Dependent measures	Env.	*M*	*SE*	*F*(1, 25)	*p*	η^2	95% CI *LL*	*UL*
RMS heading TO/GA	*VR*	3.446	0.425	15.21	.001	.37	2.572	4.319
	Conv.	1.657	0.148				1.353	1.961
RMS heading crosswind	*VR*	5.392	0.554	6.16	.02	.20	4.251	6.533
	Conv.	3.575	0.610				2.318	4.831
RMS heading downwind	*VR*	4.981	0.443	4.92	.036	.17	4.070	5.893
	Conv.	3.912	0.614				2.648	5.176
RMS altitude deviation	*VR*	120.183	16.989	14.64	.001	.37	85.194	155.172
	Conv.	59.526	8.916				41.164	77.888
Runway alignment error	*VR*	9.871	1.226	23.89	.0001	.49	7.345	12.396
	Conv.	3.330	0.468				2.367	4.294
RMS cross track error	*VR*	37.808	8.284	14.56	.001	.37	20.746	54.870
	Conv.	5.885	0.597				4.655	7.115
Lateral touchdown dev.	*VR*	4.872	0.575	16.56	.0001	.40	3.687	6.056
	Conv.	2.335	0.373				1.566	3.103

Note. Env. = Environment; VR = virtual reality; Conv. = conventional. CI = confidence interval. *LL* = lower limit. *UL* = upper limit.

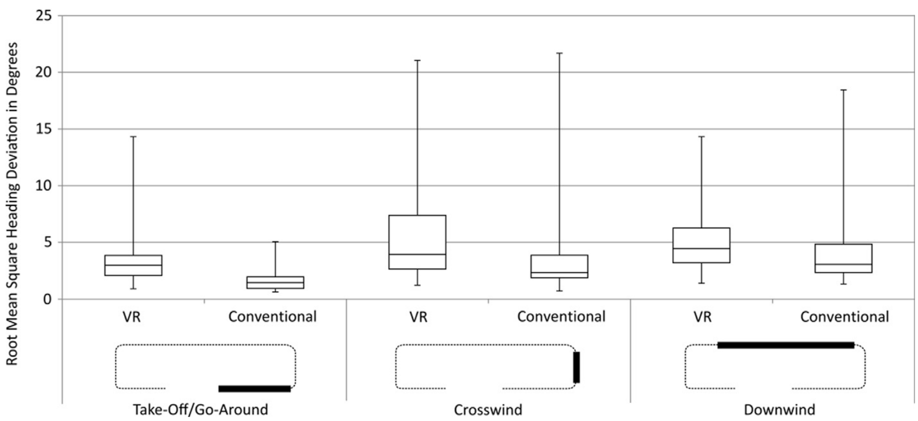

Figure 5. The RMS heading deviation.

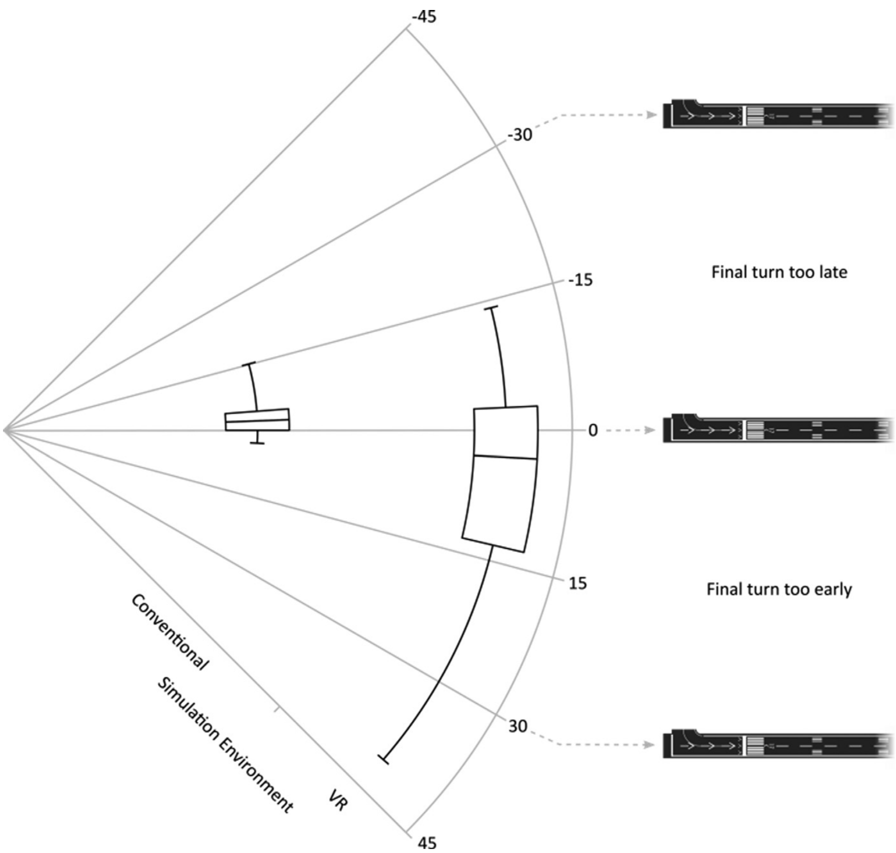

Figure 6. Angular box plot of the runway alignment error.

after takeoff. In the population, the mean difference between the movement time in VR and in the conventional simulation is expected to be within the 95% confidence interval.

Flight Performance

The results of kinematic data are presented descriptively in Table 2. The RMS heading deviation shows a significant difference between the two environments in all three legs analyzed: takeoff/touch-and-go, $F(1, 25) = 15.21$, $p < .001$, crosswind, $F(1, 25) = 6.16$, $p < .02$, and downwind, $F(1, 25) = 4.92$, $p < .036$. The heading deviation in the virtual environment was larger compared with the conventional simulator, as shown in Figure 5.

The same effect applies to the RMS altitude deviation. Here, a significant difference between the two environments, $F(1, 25) = 14.64$, $p < .036$, was also observed.

The results of the runway alignment error are depicted in Figure 6. There is a significant difference between the two environments, $F(1, 25) = 23.89$, $p < .0001$. It can be observed that in the VR environment the runway alignment

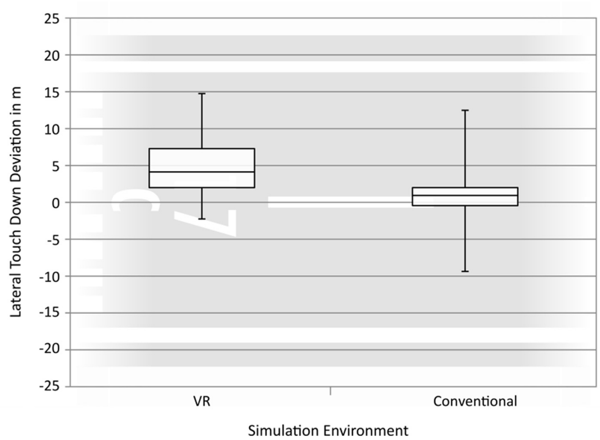

Figure 7. Deviation from the centerline on touchdown.

error is more scattered with a standard deviation of 1.226. In addition, there is a tendency toward performing the final turn too early, that is, undershoot.

The RMS cross-track error also shows a significant difference between the two environments, $F(1, 25) = 14.56$, $p < .001$. With a mean of 37,808 me in the VR environment

Table 3. The results of the workload analysis

NASA-TLX dimension	Mean difference (*SE*)	Repeated measures ANOVA *F*(1, 27)	*p*	η^2	95% CI *LL*	*UL*
Mental demand	1.750 (0.311)	31.61	.0001	.54	1.111	2.389
Physical demand	2.357 (0.458)	26.51	.0001	.50	1.418	3.296
Temporal demand	0.857 (0.373)	5.26	.03	.16	0.091	1.623
Performance	−1.750 (0.426)	16.87	.0001	.39	−2.624	−0.876
Effort	2.500 (0.387)	41.81	.0001	.61	1.707	3.293
Frustration	2.214 (0.533)	17.26	.0001	.39	1.121	3.308

Note. Mean differences, standard error, and 95% confidence intervals refer to differences between the simulation environments (VR hardware). CI = confidence interval. *LL* = lower limit. *UL* = upper limit.

Table 4. The analysis of simulator sickness

Dependent measures	Env.	*M*	*SE*	*F*(1, 25)	*p*	η^2	95% CI *LL*	*UL*
Nausea	*VR*	28.961	6.482	11.351	.002	.296	15.662	42.260
	Conv.	8.177	2.179				3.707	12.647
Oculomotor	*VR*	40.931	5.558	12.154	.002	.310	29.526	52.336
	Conv.	19.927	3.978				11.765	28.089
Disorientation	*VR*	53.691	8.402	14.699	.001	.353	36.452	70.931
	Conv.	23.366	5.594				11.887	34.844

Note. Env. = Environment; VR = virtual reality; Conv. = conventional. CI = confidence interval. *LL* = lower limit. *UL* = upper limit.

versus a mean of 5,885 m in the conventional environment, this metric is significantly degraded in the virtual environment.

The lateral touchdown deviation from the centerline (see Figure 7) shows a significant difference for the landings in the VR and the conventional simulator, $F(1, 25) = 16.56$, $p < .0001$. The subjects tended to land left of the centerline in the virtual environment.

Pilots' Workload

The results of the statistical analysis of the NASA-TLX scores are summarized in Table 3. The pilots' mental, physical, temporal demand, effort, and frustration were significantly higher in the VR than in the conventional flight simulation. Pilots' self-rating of performance was also significantly lower in VR than in the conventional flight simulation; an observation that coincides with the objective trajectory data. The population parameters are expected to be included within the reported 95% confidence intervals.

In addition, significant positive correlations between various workload scales in VR and in the conventional simulation show that higher workload in the conventional flight simulation environment was associated with higher workload in VR.

Individual differences in movement time, flight performance, and workload between the simulation environments did not significantly correlate with pilots' age, total flight experience, flight experience in the past 90 days, or flight simulator experience.

Simulator Sickness

Pilots' symptoms of simulator sickness were significantly stronger in VR. As shown in Table 4, the scores for nausea were higher in VR ($M = 28.961$) than in the conventional simulator ($M = 8.177$). The scores for oculomotor symptoms were higher ($M = 40.931$) in VR than in the conventional flight simulation ($M = 19.927$). The scores for disorientation were higher in VR ($M = 53.691$) than in the conventional simulation ($M = 23.366$). The statistical analysis shows that these differences are significant for all three categories of symptoms.

Discussion

Berg and Vance (2016) reviewed recent advances and applications in the field of VR following the publication by Brooks (1999). They conclude that the technology that "barely worked" almost 20 years ago now finally "works." This technological advance enables a higher number of VR applications that are affordable and accessible. In this research, one of these applications in the field of aviation is presented – the VRFS. In order to use this tool for human factors engineering, an assessment of the simulation fidelity in comparison with a conventional hardware simulator was made. In particular, two aspects of pilots' performance in the simulator were investigated: the movement time and the precision of the flight path.

During flight in VR, the movement time to reach the rotary knob for setting the heading bug (080, 350, 260) and the "flaps final" position was significantly longer than in the conventional flight simulator. On ground in VR, the movement time for reaching the rotary knobs for setting the heading bug and the altitude bug as well as the time to reach the flap lever for setting the "flaps takeoff" position was also significantly longer. No significant difference was found for setting "flaps up" in flight, perhaps because this was an extreme position and the action relies more on muscle memory than on visual feedback. In summary, these results show that the time needed to reach a cockpit element in the VR environment is significantly longer than in the hardware simulator. This behavior is in accordance with previous research that investigated the interaction in VR systems (Liu, Liere, Nieuwenhuizen, & Martens, 2009; Viau, Feldman, McFadyen, & Levin, 2004). The movement time is longer owing to confounding factors of the VR environment, such as an artificial spatial vision, a limited FOV, and inaccuracies in the virtual hand model. This slower movement leads to longer task completion times, which should be considered when designing experiments in the virtual environment. It is important to keep in mind that a time-critical sequence of interactions can cause problems when completing a task in the virtual environment.

The deviation in flight performance was significantly larger in VR than in the conventional flight simulation for all measures used: heading deviation (takeoff/touch-and-go, crosswind and downwind), altitude deviation, runway alignment error, cross-track error, and lateral touchdown deviation. In summary, the analysis of the kinematic data of the flight pattern shows less precision in the objective metrics of the traffic pattern in the virtual environment compared with the conventional hardware simulator. This is also confirmed by the pilots' performance self-rating, which is significantly lower in the virtual environment. Yet, this degradation is not critical to safely conduct the given flight task in the VR environment.

This difference in flight performance might be influenced by various confounding factors that stem from the VR environment as well as the hardware used. The HMD only offers a limited FOV. Yet, this hardware choice is necessary as consumer devices with a high FOV, such as Oculus Rift or HTC Vive, do not offer the necessary resolution to read flight displays in the virtual environment. With a limited FOV and in the absence of peripheral stimuli, the spatial situational awareness is limited. This is a factor that might have caused pilots to undershoot the final turn and have led to an unstable final approach in some cases. In addition, depending on the pilot's viewpoint, a slight head movement is necessary in order to read flight parameters from the head-down displays. This changes the familiar scanning pattern and probably contributed to the altitude and heading deviations that were observed.

Besides the described perceptual limitations, the pilots reported higher mental, temporal, and physical demand as well as higher effort and frustration introduced by the virtual environment. This additional workload might affect the performance negatively as it binds more mental resources, which are then not available for the flight task. In addition, the pilots reported a higher level of simulator sickness (nausea, disorientation, and oculomotor symptoms) in the virtual environment. Simulator sickness in virtual environments is mainly caused by a mismatch of sensory and visual information. This mismatch stems primarily from latency of the head tracking. In this experiment, the external tracking system has a system latency of approximately 17 ms, the scene takes 16.67 ms to render and it takes the same time to present the complete image to the user through the HMD. This leads to an overall delay from a head movement to the corresponding visual representation of at least 50 ms. This is well above the 20 ms threshold under which a head tracking delay is believed to be imperceptible and is a magnitude that, according to Carmack (2013), "will feel responsive, but still subtly lagging." Ergonomic aspects of the HMD also contribute to the symptoms. This simulator sickness might also negatively influence the pilot's performance.

Future Work

The current interest in VR technology has led to a highly dynamic market with new VR technologies emerging on a regular basis. Two technological trends are currently visible: (a) high-fidelity, low-latency consumer systems mostly used in gaming that require significant computing power provided by a PC or a gaming console, for example, Oculus Rift or HTC Vive; and (b) mobile devices that use smartphones as a display and for rendering, such as Google Daydream or Samsung GearVR.

The technological progress in the field of high-fidelity consumer devices can directly benefit the VRFS presented here: HMDs with a high FOV and high resolutions can be expected soon. As suggested by previous research, this will increase the level of immersion, in particular the peripheral stimuli that are an important input for the vestibular system (Lin, Duh, Abi-Rached, Parker, & Furness, 2002). Most likely, this will also increase the flight performance and decrease the difference between the VRFS and the conventional simulator.

Another recent trend is the rising interest in augmented reality technologies, in particular Microsoft's Holographic initiative. With these technologies, two- or three-dimensional elements can be augmented into the real-world view of the user. A simulator based on this technology could consist of a semi-functional hardware cockpit mock-up. With the AR goggles, the outside visual and the display content could be augmented with synthetic content. In comparison with a fully immersive system as presented in this research, an AR solution could offer a more natural interaction as the user would be able to see his or her own hands. However, contemporary see-through devices offer only a very limited FOV for the augmented reality content, which is a considerable disadvantage.

Today, the VRFS is mainly used for human factors engineering but, even with the described shortcomings and depending on use cases, it could already be used as a flight and navigation procedures trainer or as a basic instrument training device like the system by Dörr et al. (2001). With the current hard- and software advances in the field of VR, the VRFS, as well as other systems, could evolve into more sophisticated training simulators. Whether and how such systems can be certified as flight training devices or even full-flight simulators is subject to further research. The advancement of mobile VR devices enables new areas of application.

Conclusion

In this research, the fidelity of a VRFS was assessed by comparing the system with a conventional hardware simulator using an operational scenario. Dependent measures regarding the pilot's interaction with cockpit elements and the precision in flying the traffic circuit were taken into account.

Most participants were able to safely and reliably complete the flight task after a short acclimatization phase in the virtual environment. From the 28 pilots participating in the trials, two did not complete the task owing to simulator sickness and problems with the vision in the virtual environment. The limitations in the functional fidelity and task fidelity had a significant impact on the flight performance and movement time of the remaining 26 pilots. Thus, the degradations in flight performance could be quantified and will be considered in the experimental design of future studies with the VRFS.

Another factor that was measured in the trials was the movement time when interacting with cockpit elements. The results show that the interaction in the virtual environment is challenging and takes additional time. As a consequence, future test scenarios have to be created according to this knowledge. The pace of scenarios must be adapted, the exposure time has to be limited, and the flight task should be feasible even with additional workload from the VR environment.

In summary, the results show that the fidelity of the VRFS is significantly lower compared with the fidelity of a conventional flight simulator and thus cannot substitute the latter, particularly if the pilot's behavior and performance are to be as close to reality as possible. For use cases in which adaptations to pace, exposure time, and flight task are acceptable, which is often the case in early phases of the design process, the VRFS can be a viable tool for human factors engineering. Here, the ability for rapid prototyping and usability testing is important because the implementation of design changes is easier and more cost-effective in early than in more mature stages. Yet, the performance and workload differences reported in this paper should be considered carefully when conducting such studies using the VRFS. Nevertheless, the further development of VR technology offers the potential to minimize some of these differences and thus to present pilots with an even more accurate flying experience in VR.

Acknowledgments

This is an extended version of two papers by Oberhauser, Dreyer, Braunstingl, and Koglbauer (2017a, b) published in the Proceedings of the 32nd Conference of the European Association of Aviation Psychology, Cascais, Portugal.

This paper was edited by Associate Editor André Droog.

References

Aslandere, T. I., Dreyer, D., Pantkratz, F., & Schubotz, R. (2014) A generic virtual reality flight simulator. In G. Zachmann, R. Weller, & A. Hinkenjann (Eds.), *Virtuelle und Erweiterte Realität, 11. Workshop der GI-Fachgruppe VR/AR* (pp. 1–13). Aachen, Germany: Shaker.

Avila, L., & Bailey, M. (2014). Virtual reality for the masses. *IEEE Computer Graphics and Applications, 34*(5), 103–104. https://doi.org/10.1109/MCG.2014.103

Berg, L. P., & Vance, J. M. (2016). Industry use of virtual reality in product design and manufacturing: A survey. *Virtual Reality*, 1–17. https://doi.org/10.1007/s10055-016-0293-9

Brooks, F. P. (1999). What's real about virtual reality? *IEEE Computer Graphics and Applications, 19*(6), 16–27. https://doi.org/10.1109/38.799723

Carmack, J. (2013). *Latency mitigation strategies*. Retrieved from https://web.archive.org/web/20140719053303/http://www.altdev.co/2013/02/22/latency-mitigation-strategies/

Dörr, K.-U., Schiefele, J., & Kubbat, W. (2001). Virtual cockpit simulation for pilot training. In North Atlantic Treaty Organization. (Eds.), *RTO meeting proceedings: Vol. 58. What is essential for virtual reality systems to meet military human performance goals? RTO Human Factors and Medicine Panel (HFM) workshop* (p. 11). Neuilly-sur-Seine, France: RTO, Research and Technology Organization.

Dreyer, D., Oberhauser, M., & Bandow, D. (2014). HUD symbology evaluation in a virtual reality flight simulation. In *Proceedings of the International Conference on Human-Computer Interaction in Aerospace '14* (pp. 9–14). New York, NY: ACM. https://doi.org/10.1145/2669592.2669652

European Aviation Safety Agency. (2012). *Certification specifications for aeroplane flight simulation training devices. (CS-FSTD(A))*. Cologne, Germany: Author.

Farmer, E., Rooij, J. van., Riemersma, J., Jorna, P., & Moraal, J. (1999). *Handbook of simulator-based training*. Aldershot, UK: Ashgate.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in Psychology, 52*, 139–183.

Heilig, M. L. (1962). *US Patent No. 3,050,870*. Washington, DC: U.S. Patent and Trademark Office.

Joyce, R. D., & Robinson, S. K. (2015). The rapidly reconfigurable research cockpit. In *AIAA Aviation, AIAA Modeling and Simulation Technologies Conference*. Reston, VA: American Institute of Aeronautics and Astronautics. https://doi.org/10.2514/6.2015-2336

Kaiser, M. K., & Schroeder, J. A. (2002). Flights of fancy: The art and science of flight simulation. In P. S. Tsang & M. A. Vidulich (Eds.), *Principles and practices of aviation psychology* (pp. 435–471). Mahwah, NJ: Lawrence Erlbaum.

Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993). Simulator Sickness Questionnaire: An enhanced method for quantifying simulator sickness. *The International Journal of Aviation Psychology, 3*(3), 203–220. https://doi.org/10.1207/s15327108ijap0303_3

Le Ngoc, L., & Kalawsky, R. S. (2013). Visual circuit flying with augmented head-tracking on limited field of view flight training devices. In *Proceedings of the AIAA Modeling and Simulation Technologies Conference 2013. Guidance, Navigation, and Control and Co-located Conference* (pp. 1–16). Reston, VA: American Institute of Aeronautics and Astronautics. https://doi.org/10.2514/6.2013-5226

Lin, J. J.-W., Duh, H. B. L., Abi-Rached, H., Parker, D. E., & Furness, T. A. (2002). Effects of field of view on presence, enjoyment, memory, and simulator sickness in a virtual environment. In B. Loftin, J. X. Chen, S. Rizzo, M. Goebel, & H. Hirose (Eds.), *Proceedings of the IEEE Virtual Reality Conference 2002* (pp. 164–171). Washington, DC: IEEE Computer Society.

Liu, D., Macchiarella, N. D., & Vincenzi, D. A. (2009). Simulation fidelity. In D. A. Vincenzi, J. A. Wise, M. Mouloua, & P. A. Hancock (Eds.), *Human factors in simulation and training* (pp. 61–73). Boca Raton, FL: CRC Press.

Liu, L., Liere, R. van., Nieuwenhuizen, C., & Martens, J.-B. (2009). Comparing aimed movements in the real world and in virtual reality. In A. Steed, D. Reiners, & R. W. Lindemann (Eds.), *IEEE Virtual Reality Conference, 2009*. Piscataway, NJ: IEEE. https://doi.org/10.1109/VR.2009.4811026

McCarty, D. W., Sheasby, S., Amburn, P., Stytz, M. R., & Switzer, C. (1994). A virtual cockpit for a distributed interactive simulation. *IEEE Computer Graphics and Applications, 14*(1), 49–54. https://doi.org/10.1109/38.250919

McMahan, R. P., Bowman, D. A., Zielinski, D. J., & Brady, R. B. (2012). Evaluating display fidelity and interaction fidelity in a virtual reality game. *IEEE Transactions on Visualization and Computer Graphics, 18*(4), 626–633. https://doi.org/10.1109/TVCG.2012.43

Milgram, P., & Kishino, F. (1994). A taxonomy of mixed reality visual displays. *IEICE Transactions on Information Systems, E77-D*(12), 1321–1329.

Moroney, W. F., & Moroney, B. W. (2009). Flight simulation. In J. A. Wise, V. D. Hopkin, & D. J. Garland (Eds.), *Handbook of aviation human factors* (2nd ed., pp. 355–388). London, UK: CRC Press.

Oberhauser, M., & Dreyer, D. (2017). A virtual reality flight simulator for human factors engineering. *Cognition, Technology & Work, 19*(2), 263–277. https://doi.org/10.1007/s10111-017-0421-7

Oberhauser, M., Dreyer, D., Braunstingl, R., & Koglbauer, I. (2017a). Pilots' flight performance in a virtual reality flight simulator. In M. Schwarz & J. Harfmann (Eds.), *Proceedings of the 32nd Conference of the European Association of Aviation Psychology* (pp. 322–335). Groningen, The Netherlands: European Association for Aviation Psychology.

Oberhauser, M., Dreyer, D., Braunstingl, R., & Koglbauer, I. (2017b). Pilots' interaction with hardware controls in a virtual reality flight simulator. In M. Schwarz & J. Harfmann (Eds.), *Proceedings of the 32nd Conference of the European Association of Aviation Psychology* (pp. 565–575). Groningen, The Netherlands: European Association for Aviation Psychology.

Oberhauser, M., Dreyer, D., Convard, T., & Mamessier, S. (2016). Rapid integration and evaluation of functional HMI components in a virtual reality aircraft cockpit. In F. Rebelo & M. Soares (Eds.), *Advances in ergonomics in design. Proceedings of the AHFE 2016 International Conference on Ergonomics in Design* (pp. 17–24). Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-319-41983-1_2

Oberhauser, M., Dreyer, D., Mamessier, S., Convard, T., Bandow, D., & Hillebrand, A. (2015). Bridging the gap between desktop research and full flight simulators for human factors research. In D. Harris (Ed.), *Lecture notes in computer science, engineering psychology and cognitive ergonomics* (pp. 460–471). Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-319-20373-7_44

Perfect, P., White, M. D., Padfield, G. D., & Gubbels, A. W. (2013). Rotorcraft simulation fidelity: New methods for quantification and assessment. *The Aeronautical Journal, 117*(1188), 1–48.

Persiani, F., Piancastelli, L., & Liverani, A. (1997). Virtual flight simulator. In *Proceedings of the 10th ADM International Conference on Design Tools and Methods in Industrial Engineering* (pp. 693–701). Bologna: Associazione Nazionale Disegno e Metodi.

Quigley, M., Conley, K., Gerkey, B., Faust, J., Foote, T., Leibs, J., . . . Ng, A. Y. (2009). ROS: An open-source Robot Operating System. *ICRA Workshop on Open Source Software, 3*(3), 5.

Rehmann, A. J. (1995). *A Handbook of flight simulation fidelity requirements for human factors research: DOT/FAA/CT-TN95/46*. Springfield, VA: National Technical Information Service.

Rheingold, H. (1991). *Virtual reality*. New York, NY: Summit Books.

Robertson, A., & Zelenko, M. (2014). *Voices from the past: An oral history of a technology whose time has come again*. Retrieved from http://www.theverge.com/a/virtual-reality/oral_history

Robinson, A., Mania, K., & Perey, P. (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* (pp. 261–268). Singapore: ACM. https://doi.org/10.1145/1044588.1044644

Rolfe, J. M., & Staples, K. J. (1986). *Flight simulation (1st ed.). Cambridge aerospace series*. Cambridge, NY: Cambridge University Press.

Sutherland, I. E. (1965). The ultimate display. In *Proceedings of the International Federation of Information Processing (IFIP)*. London, UK: Macmillan.

Viau, A., Feldman, A. G., McFadyen, B. J., & Levin, M. F. (2004). Reaching in reality and virtual reality: A comparison of movement kinematics in healthy subjects and in adults with hemiparesis. *Journal of NeuroEngineering and Rehabilitation, 1*(1), 11. https://doi.org/10.1186/1743-0003-1-11

Yavrucuk, I., Kubali, E., Tarimci, O., & Yilmaz, D. (2009). A low cost flight simulator using virtual reality tools. In American Institute of Aeronautics and Astronautics. (Eds.), *Proceedings of the AIAA Modeling and Simulation Technologies Conference. Guidance, Navigation, and Control and Co-located Conferences* (pp. 1–9). Reston, VA: American Institute of Aeronautics and Astronautics. https://doi.org/10.2514/6.2009-5832

Received April 24, 2017
Revision received November 7, 2017
Accepted November 13, 2017
Published online April 23, 2018

Matthias Oberhauser
Graz University of Technology
Kopernikusgasse 24/IV
8010 Graz
Austria
matthias.oberhauser@gmail.com

Matthias Oberhauser worked at Airbus Group Innovations on the application of virtual reality technologies for human–machine interface design and human factors evaluations. In his current role at the Mixed Reality Lab at BMW, he is developing new methods for the use of virtual reality gaming technologies in the automotive developing process. He is currently pursuing a PhD at the Technical University of Graz, Austria.

Daniel Dreyer studied aeronautical engineering at the Technical University of Munich and holds a PhD in Ergonomics. He is currently working at Airbus Defence and Space as Human Factors Systems Engineer. He is involved in the development of cockpits and ground control stations of manned and unmanned aircraft.

Reinhard Braunstingl received a PhD degree in mechanical engineering from Graz University of Technology, Graz, Austria, in 1991. He is a professor with the Graz University of Technology, as well as the head of the research platform Flight Simulation. In addition, he is an experienced pilot, flight instructor, and flight examiner.

Ioana Koglbauer is an aviation psychologist and pilot affiliated with Graz University of Technology, Austria. She received her PhD and habilitation in psychology from the University of Graz, Austria. Her teaching and research activities are focused on human-machine systems in the aviation, automotive, and emergency monitoring domains.