Effects of Motion Fidelity on Pilot Performance using Mixed-Reality Visuals

Peter M. T. Zaal*, Matt L. Blanken†, Samuel P. Orth‡, and Supreethi V. Penmetcha§
NASA Ames Research Center, Moffett Field, CA 94035

Paul Slade¶
Quantum3D, Inc., Milpitas, CA 95035

This paper investigates the effects of simulator motion fidelity on pilot performance using mixed-reality visuals. In NASA's Vertical Motion Simulator, eleven pilots flew a lift-plus-cruise electric vertical takeoff and landing vehicle in a vertiport approach and landing task with four different motion conditions: no motion, small hexapod motion, large hexapod motion, and full motion. The simulator motion condition statistically significantly affected: 1) the severity of cybersickness, 2) motion ratings, 3) final approach and take-off boundary crossing groundspeed, 4) longitudinal touchdown point, and 5) battery time remaining after landing. The findings suggest that motion systems are essential in simulators with virtual- or mixed-reality visuals when it is required to have pilot flying performance similar to that in real flight.

I. Nomenclature

df	=	degrees of freedom	t_{bat}	=	battery time remaining, s
h_{AT}	=	height above threshold at FATO crossing, ft	V_{AT}	=	groundspeed at FATO crossing, kts
H_m	=	motion washout filter transfer function	x_{td}	=	longitudinal touchdown deviation, ft
IQR	=	interquartile range	y_{td}	=	lateral touchdown deviation, ft
K_m	=	motion gain	α	=	probability of Type I error
M	=	median value	β	=	probability of Type II error
MR	=	motion fidelity rating	Δ_{GS}	=	glideslope deviation, deg
p	=	probability value	Δ_{LOC}	=	localizer deviation, deg
$Q1$	=	first quartile value	ζ_m	=	motion filter damping ratio
$Q3$	=	third quartile value	η_p^2	=	effect size
s	=	Laplace variable	λ	=	Box-Cox transformation parameter
SSQ_{ts}	=	simulator sickness questionnaire total score	χ^2	=	statistical test statistic
t	=	t-test statistic	ω_m	=	motion filter break frequency, rad/s

II. Introduction

This paper investigates the effects of simulator motion fidelity on pilot performance in a demanding piloting task while wearing a mixed-reality (MR) headset for displaying out-the-window (OTW) visuals. An experiment was conducted in the NASA Ames Vertical Motion Simulator (VMS) with four motion conditions: no motion, small hexapod motion, large hexapod motion, and full motion.

Virtual reality (VR) technologies have been a topic of research for decades. One of the first VR head-mounted displays (HMD) was developed at NASA Ames Research Center in 1985 [1]. In addition, there is a rich history of training astronauts with VR at NASA Johnson Space Center [2]. In the last decade, VR and MR technologies have seen major advancements and are now readily available to consumers due to products from major technology companies.

*Principal System Architect, Metis Technology Solutions, Inc., SimLabs, AIAA Associate Fellow, email: peter.m.t.zaal@nasa.gov.
†Simulation Engineer, NASA Ames Research Center, SimLabs, AIAA member, email: matthew.l.blanken@nasa.gov.
‡Simulation Engineer, Symvionics, Inc., SimLabs, AIAA member, email: samuel.p.orth@nasa.gov.
§Simulation Engineer, NASA Ames Research Center, SimLabs, email: supreethi.v.penmetcha@nasa.gov.
¶Principal Software Engineer, Quantum3D, Inc., email: pslade@quantum3d.com.

Furthermore, VR and MR are increasingly used in flight simulators and training devices for pilot training, and training facilities are keen on increasing the use, due to the lower cost and smaller footprint of HMD systems compared to conventional visual systems.

Even though VR/MR technologies have improved significantly in recent years, and issues such as reduced visual acuity have been greatly reduced by implementing higher-resolution displays, limitations remain, including increased visual fatigue and discomfort due to vergence-accommodation conflict [3]. In addition, the implications of using VR/MR for pilot training are still largely unknown as research is limited [4]. Specific gaps in knowledge include:

1) The effects of VR/MR on pilot well-being and, in turn, the quality of training, are largely unknown. For example, nausea and other discomfort are experienced significantly more using HMDs compared to conventional visual systems.
2) The effects of the quality of the visual cues are still largely unknown. This includes characteristics such as brightness, resolution, and field of view (FOV). Studies on these characteristics in the context of pilot performance have been conducted for conventional displays, but results might not directly transfer to HMDs.
3) Guidelines for the fidelity of simulator motion needed with VR/MR visuals for effective pilot training, and the synchronization of HMD visuals with physical motion, have not been researched. Again, these topics have been investigated for conventional simulator visuals, but the findings might not transfer to HMDs.

The last two gaps in knowledge are specifically important in the training of skill-based tasks, such as manually flying an aircraft, where the quality and synchronization of visual and motion cues affect performance [5]. Due to these gaps in knowledge, very limited regulatory standards for using VR/MR in training programs currently exist, which limits the use of the technology for training. Regulations that do exist focus on helicopter simulation and use conventional visual systems as the norm [6]. Furthermore, in cases where HMD technologies can be used for training, the quality of the training could potentially be reduced compared to current standards due to the lack of knowledge.

This paper adds data to help address the third knowledge gap and provides the first insights into the effects of simulator motion fidelity on pilot performance in VR/MR simulations. The paper first describes the experiment setup in Section III. Results are provided in Section IV, followed by a discussion in Section V. Section VI provides conclusions and recommendations for future work.

III. Experiment Setup

A. Method

1. Vertiport Approach and Landing Task

This study used a simulated lift-plus-cruise (LPC) electrical vertical takeoff and landing (eVTOL) vehicle in a vertiport approach and landing task. The LPC conceptual model (Fig. 1) was developed by NASA's Revolutionary Vertical Lift Technology project [7]. This vehicle is capable of taking off and landing vertically using lifting rotors, and cruising in forward flight with the lifting rotors stowed. Simplified vehicle control concepts for this vehicle were developed and tested in previous simulations. This study used the standard simplified control command response types detailed in [8] without assisted hover automation.

The vertiport approach and landing task in this study was adapted from the Federal Aviation Administration's Handling Qualities Test Guide for Powered Lift Vertical Take-Off and Landing (VTOL) Capable Aircraft [9]. Test runs with a test pilot were used to optimize the task for LPC vehicle performance constraints and to create a more challenging task by modifying the Handling Qualities Test Guide's original performance requirements. Pilots performed the landing at a landing pad (gray and white areas in Fig. 2) of a vertiport on the Fifth & Mission Parking Garage in San Francisco. The vertiport was set in a San Francisco visual database surrounded by buildings with high-resolution textures.

A diagram of the approach and landing task is provided in Fig. 3. The task consists of four segments: 1) glidepath capture, 2) glidepath tracking, 3) deceleration, and 4) transition to hover for landing. Fig. 3 also indicates a hypothetical vehicle trajectory in blue. The task began in straight and level flight at 60 kts, at an altitude of 880 ft mean sea level (MSL), and 1 nmi downrange of the target landing area. Pilots captured and maintained a 9-degree target approach glidepath angle starting at the Final Approach Fix (FAF). A pulse light approach slope indicator (PLASI), located next to the touch-down and lift-off (TLOF) area of the target landing pad of the vertiport, was available to guide pilots along the glideslope. In addition, standard glideslope and localizer indicators were provided on the primary flight display (PFD).

Pilots maintained 60 kts while on the glidepath up to the deceleration (DECEL) point at an altitude of 380 ft MSL. At the DECEL point, pilots were required to begin a smooth deceleration while continuing on the glidepath to cross the

Fig. 1 Visual presentation of the simulated lift-plus-cruise electrical vertical takeoff and landing vehicle.

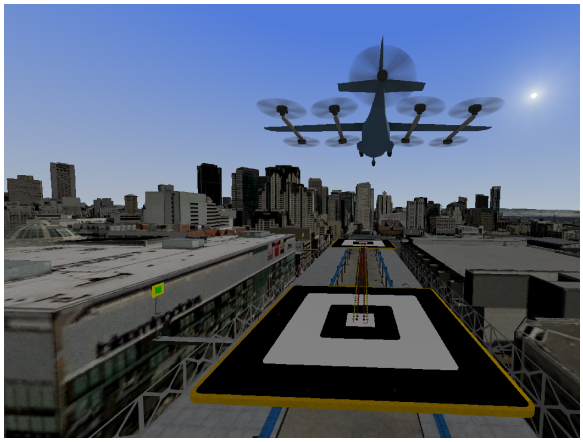


Fig. 2 Virtual out-the-window scene showing the Fifth & Mission Parking Garage with vertiport in San Francisco.

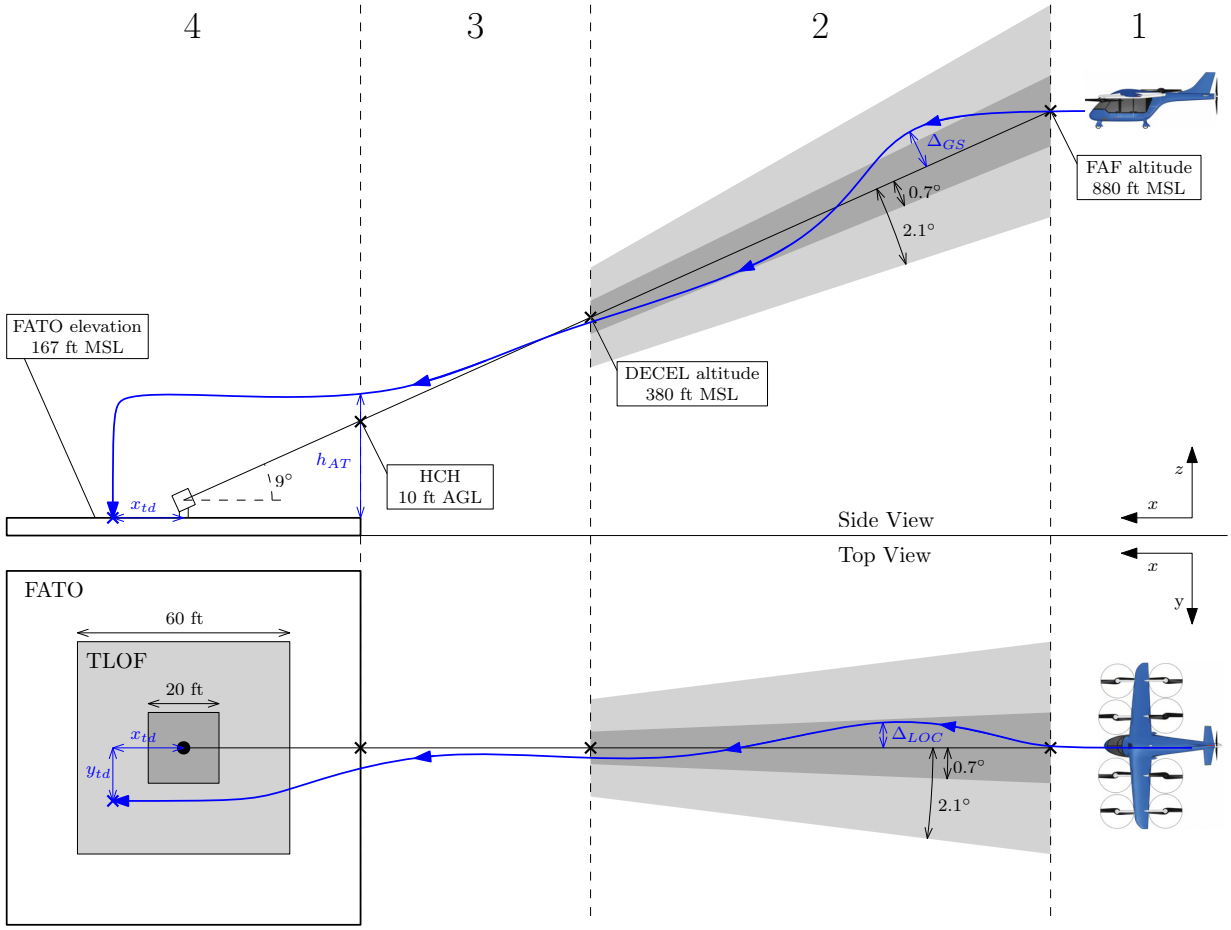


Fig. 3 Approach task diagram.

final approach and take-off (FATO) threshold at the helipad crossing height (HCH) of 10 ft above ground level (AGL) and with 5 kts groundspeed. After crossing into the FATO area, pilots transitioned to a 10 ft AGL hover above the center of the TLOF area before touching down smoothly on the landing pad. Pilots were instructed to land with a certain level

Table 1 Vertiport-approach task performance criteria.

Criterion	Performance		Dependent
	Desired	Adequate	Measure
Maintain glideslope from FAF to DECEL within:	±0.7°/±1 dot	±2.1°/±3 dots	RMS Δ_{GS}
Maintain localizer from FAF to DECEL within:	±0.7°/±1 dot	±2.1°/±3 dots	RMS Δ_{LOC}
Altitude from 10 ft HCH at FATO crossing within:	±10 ft	±20 ft	h_{AT}
Groundspeed from 5 kts at FATO crossing within:	±5 kts	±10 kts	V_{AT}
Longitudinal touchdown deviation from TLOF center within:	±20 ft	±60 ft	x_{td}
Lateral touchdown deviation from TLOF center within:	±20 ft	±60 ft	y_{td}
Battery time remaining at touchdown:	≥ 1,200 s	≥ 1,140 s	t_{bat}

of battery time remaining in reserve. Since the LPC vehicle drains the battery significantly more during hover, the transition from forward flight to hover and subsequently landing had to be performed expeditiously to meet the criteria. Light turbulence according to a Dryden-based turbulence model was introduced throughout the task to increase pilot workload.

Criteria for desired and adequate performance were provided to the pilots, and their performance against these criteria was evaluated and communicated following each landing (Table 1).

2. Apparatus

Data collection took place in the VMS (Fig. 4). The VMS provides motion in all three rotational and all three translational degrees-of-freedom, including 60 ft of vertical travel and 40 ft of lateral travel [10]. The LPC vehicle accelerations drove the motion system of the simulator. This study used the Rotorcraft Cab (R-Cab) with a single seat (Fig. 5). Two active electric sidesticks, positioned on the left and right sides of the seat, controlled the LPC vehicle. Right stick fore-aft movements controlled vertical acceleration below 34 kts and flight path angle rate above 34 kts, left-right movements controlled bank angle below 40 kts and roll rate above 40 kts, and twist movements controlled heading rate below 40 kts and sideslip angle above 40 kts. Left stick fore-aft movements controlled forward acceleration based on groundspeed below 40 kts and forward acceleration based on indicated airspeed above 40 kts. Transitions between the control modes of this simplified vehicle controls scheme were smooth and without indication to the pilot. Pedals were installed in the cockpit but not used for aircraft control. Three head down displays (HDDs) were positioned in front of the pilot providing primary-flight (center display), navigation (left display), and powertrain (right display) information.

Fig. 4 Vertical Motion Simulator motion system.

The out-the-window visuals were provided by a Varjo XR-4 Secure Focal Edition headset, worn on the pilot's head, using MR visuals [11]. This meant that the virtual OTW visual scene was overlaid on an image of the real simulator cockpit provided by passthrough cameras on the headset (Fig. 6). The real cockpit view included the HDDs and the two sidesticks. The boundary between the virtual and real cockpit scenes was defined by a cutout in the visual model of the vehicle.

The Varjo XR-4 (Fig. 7) has a 4K mini-LED display per eye with 200 nits of brightness and 1:10000 contrast ratio, offering a 120° horizontal by 105° vertical field of view with full binocular range vision at 51 pixels per degree. 20-megapixel cameras with autofocus on the outside of the headset facilitate MR simulation. The headset also features a built-in eye tracker with 200 Hz eye tracking to allow for foveated rendering; that is, the scene will be rendered at a higher resolution around where a user is looking. The weight of the headset is 1021 grams. A Quantum3D image generator provided the OTW visuals for the MR headset using Quantum3D's Mantis real-time scene management software and an NVIDIA RTX 4090 video card.

To determine the location and orientation of the MR headset in the 3D environment, this study used an LP-Research LPVR-DUO tracking system, which consists of four principal components: 1) a headset tracking marker attachment; 2) an ART SmartTrack3 optical sensor; 3) an LP-IG1 inertial measurement unit (IMU); and 4) the LPVR-DUO sensor

Fig. 5 Simulator Cockpit.

Fig. 6 Mixed-reality visuals showing the virtual out-the-window scene and a passthrough of the real cockpit environment.

fusion algorithm and driver that integrates with Varjo Base, the HMD's configuration software.

The basic tracking algorithm works as follows: The SmartTrack3 captures the position and orientation of the HMD within the cockpit, using the relative positioning and known geometry of the tracking marker attachment orbs. The LP-IG1 IMU captures the acceleration and orientation of the VMS cab, which can be backed out from the Varjo XR-4's internal IMU to isolate the head motion independent of any VMS accelerations. This is then combined with the SmartTrack3's fixed positioning to determine the final OTW view to render in the HMD (Fig. 8).

To enable this capability, the tracking marker attachment was clipped to the front of the Varjo XR-4. The SmartTrack3 sensor was installed on the R-Cab glare shield (Fig. 5), immediately in front of the pilot. It was angled 10 degrees upwards to capture the whole cockpit space in its FOV, and it sampled data at 150 Hz. The LP-IG1 IMU was bolted to the top of the SmartTrack3, such that it captured the acceleration and orientation of the SmartTrack3 with no relative position offset. Pilots did not have to interact with the tracking system.

Pilots wore a David Clark noise-canceling headset at all times allowing for direct communication with simulator operators and researchers. This headset masked most external sounds from the VMS motion drives. In addition, audio cueing representing eVTOL rotor noise was presented via cab-mounted speakers to both enhance the sense of immersion

Fig. 7 Varjo XR-4 mixed-reality headset.

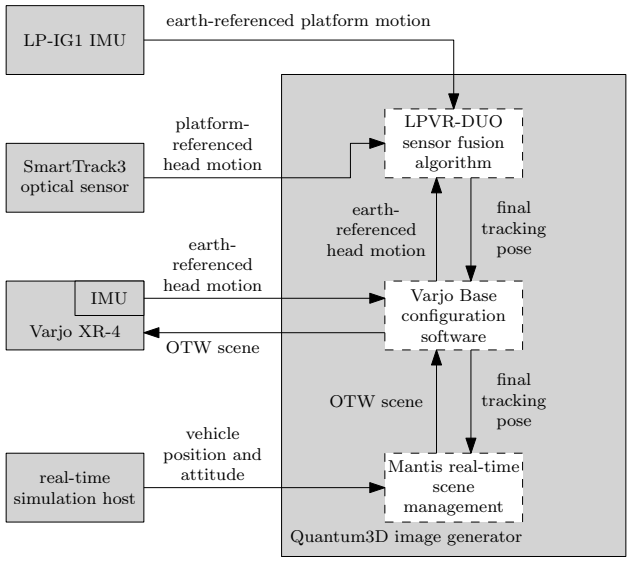


Fig. 8 Mixed-reality system network diagram.

and further mask external sounds.

Transport delay measurements were conducted for the Varjo XR-4 MR visuals, HDD visuals, and simulator motion response using a custom time delay measurement system. This measurement system utilizes a photo sensor to measure the visual responses and cab-mounted accelerometers to measure the VMS motion response. The transport delay of the Varjo XR-4 visuals was 51.1 ms, and the delay of the motion response was 48.2 ms. These measurements include the transport delay of the control inceptors and host computer. The delays being approximately equal mitigated the need to temporally equalize visual and motion cues further to minimize motion sickness. The transport delay of the HDDs was 85.1 ms.

3. Independent Variable

The only independent variable in this experiment was the motion of the VMS cab with four levels: 1) no motion (NOM), 2) small-hexapod motion (SHX), 3) large-hexapod motion (LHX), and 4) full motion (FLM). In the NOM condition, pilots did not experience any motion. The SHX motion condition represented the motion envelope of a small six-inch legged hexapod motion system often used for single-occupancy simulators with smaller visual systems. The LHX condition represented the motion envelope of a 60-inch legged hexapod motion system typically used for Level D simulators used in airline pilot training. The simulated large hexapod was exactly ten times the size of the small hexapod. The full motion condition used the entire VMS motion envelope and approximated actual vehicle motion.

More details on the motion conditions, and associated motion fidelities as determined by the Sinacori criterion, are provided in the appendix. All motion conditions were flown in the VMS with its large-amplitude motion system, but the motion settings were set differently for each condition to achieve the highest motion fidelity within that particular (simulated) motion envelope. For the no-motion condition, the simulator motion system was turned on, and pilots and simulation operators followed the same procedures as during a run with motion; however, the simulator did not move during the task. Pilots were not told which motion condition would be present during any of the simulation runs to make sure pilots were not preconditioned.

The study had a within-subjects repeated-measures design, with each participant performing the approach task under four levels of the independent variable motion. The order of the simulator motion fidelity conditions was randomized using a partial Latin square design (independent variable level vs. repetition). Due to time constraints, instead of four repetitions for a full Latin square design, only three repetitions for each motion condition, after three training runs with full motion, could be performed.

4. Participants and Procedures

The study protocol was reviewed and approved by the NASA Institutional Review Board (ClinicalTrials.gov ID NCT07129876). A power analysis for multiple repeated measures using an effect size of $\eta_p^2 = 0.14$ and strong correlation between measures of 0.7 determined a required sample size of 10 pilots for $\alpha < 0.05$ significance and $\beta = 0.2$ (i.e., 0.8 statistical power). The effect size and correlation between measures were conservatively estimated based on previous research [12]. By distributing a flyer by email, 13 local pilots were recruited to allow for dropouts. Candidate pilots needed to be 18 years or older, be in general good health, have normal or corrected to normal vision, not have claustrophobia, and not be susceptible to motion sickness. Pilots were not offered any remuneration for their involvement in the study.

The experiment session started in the pilot ready room at the VMS facility, where pilots provided their informed consent after they were briefed and given an opportunity to ask questions. Here, they also completed a demographics questionnaire and a short-form motion sickness susceptibility questionnaire (MSSQ-Short) [13]. Pilots then received a briefing on VMS safety and evacuation procedures before entering the simulator cab for the first simulator session. Once seated inside the simulator cab, a simulator operator helped pilots put on the Varjo MR headset, ensuring a secure and comfortable fit. The headset's built-in eye tracker was then calibrated to ensure that the foveated rendering would work accurately. This calibration was initiated by the pilot pressing a button on the headset. Simulator operators verified the final calibration by asking pilots to look at different icons on the primary flight display (PFD) and confirming the tracked position using Varjo's gaze marker. In the event of poor calibration, this procedure was repeated until a satisfactory calibration was reached. Following successful calibration, pilots were given a flight card attached to a knee board with information on the flight task. Finally, the David Clark communication headset was put on over the Varjo MR headset, and the cab door was closed. Following confirmation that pilots were ready to continue, the simulator cab was moved to its starting location at the center of the travel range, where it was ready to begin motion.

Pilots started with a familiarization session with the aircraft in a hover at the southeast end of runway 28L at San

Francisco International Airport (SFO). This session allowed them to get familiar with the simplified vehicle controls scheme. The familiarization was unstructured and pilots were free to fly around trying different maneuvers; however, it was suggested to fly at least one complete circuit. After familiarization, the task started with a training session consisting of three training runs to acquaint pilots with the task criteria and procedures. The familiarization and training runs were all flown with the full motion condition (FLM). After training, each motion condition was repeated three times for a total of twelve measurement runs.

After each training and measurement run, the left HDD switched from showing the navigation display to showing a performance display providing pilots with their performance during the last run in relation to the criteria (Table 1). Pilots were instructed to always pursue desired performance for all performance metrics. Similarly, the center HDD changed from showing the PFD to showing a Simulator Sickness Questionnaire (SSQ) and motion fidelity rating scale at the end of each run. Pilots verbally completed the questionnaire and rating scale through the communication headset. HDDs were switched back to in-flight displays prior to the start of the next run. After all measurements were completed, the VMS cab was brought back to the dock. The entire session lasted for approximately one hour and 15 minutes.

The study protocol also included a second task completed in a second simulator session, which commenced after a break. This task is not reported on in this paper (see Section V).

5. Dependent Measures

A total of nine dependent measures were determined for each experiment run: the SSQ total score, SSQ_{ts}; motion fidelity rating, MR; root mean square (RMS) glideslope deviation, RMS Δ_{GS}; RMS localizer deviation, RMS Δ_{LOC}; FATO crossing height above threshold, h_{AT}, and associated crossing speed, V_{AT}; longitudinal and lateral touchdown deviations from the TLOF center, x_{td} and y_{td}; and the battery time remaining at touchdown, t_{bat}.

Pilots completed the SSQ [14] and a motion fidelity rating [15] at the end of each run. The SSQ required pilots to rate the presence of 16 different cybersickness symptoms with values of none, slight, moderate, or severe for the completed run. The ratings for each symptom were used to calculate a single cybersickness total score (SSQ_{ts}), which represents the overall severity of cybersickness experienced during the run [14]. The motion fidelity rating required pilots to rate the simulator motion experienced during the run (MR) as: low - motion sensations are noticeably different from flight and are objectionable; medium - motion sensations are noticeably different from flight, but not objectionable; or high - motion sensations are like those of flight.

Pilot performance in the vertiport approach task was determined by seven dependent measures directly linked to the performance criteria in Table 1. The RMS glideslope deviation (RMS Δ_{GS}) and RMS localizer deviation (RMS Δ_{LOC}) in degrees were calculated from the FAF to DECEL (Fig. 3). The height above threshold (h_{AT}) in feet and the associated crossing groundspeed (V_{AT}) in knots were calculated at the FATO boundary crossing. The longitudinal and lateral touchdown deviations (x_{td} and y_{td}) from the TLOF center in feet, and the battery time remaining (t_{bat}) in seconds, were calculated at touchdown. The definitions of the geometry-based dependent measures are provided in Fig. 3 in blue based on the hypothetical trajectory also provided in blue. Dark gray areas depict desired performance and light gray adequate performance for the glidepath-deviation and touchdown-location measures.

Mixed-effects models were fit on the dependent measures using motion and repetition, and their interaction (motion $\times$ repetition) as fixed effects. Repetition was included as a fixed effect, as cybersickness often increases with prolonged exposure to VR visuals, and measures might be affected as a result. Repeated measures from each run (i.e., from each repetition of each condition) were used as model outputs (i.e., data were not averaged). The mixed-effects models were progressively built up by adding the different fixed effects one-by-one, starting with the intercept-only model. Likelihood ratio tests between the models with and without a fixed effect or interaction determined the significance of that effect. Outliers were assessed using boxplots. Assumptions of linearity, homoscedasticity, and normality of residuals were checked visually using scatter plots and Q-Q plots of the model residuals.

B. Hypothesis

Based on previous research using conventional simulator visual systems [5, 12, 15], this study tested the following hypothesis: Pilot performance and subjective motion ratings with MR visuals will be significantly different between the different simulator motion conditions.

Table 2 Linear mixed-effects model comparison statistics.

Dependent Measure	Factors								
	Motion			Repetition			Motion × Repetition		
	$\chi^2(6)$	p	Sig.	$\chi^2(8)$	p	Sig.	$\chi^2(14)$	p	Sig.
SSQ_{ts}	15.39	< 0.01	**	0.09	0.96	–	2.86	0.83	–
MR	17.56	< 0.01	**	1.78	0.41	–	6.28	0.39	–
RMS Δ_{GS}	5.89	0.12	–	1.99	0.37	–	3.43	0.75	–
RMS Δ_{LOC}	2.18	0.54	–	2.44	0.30	–	7.94	0.24	–
h_{AT}	1.49	0.68	–	4.08	0.13	–	1.71	0.94	–
V_{AT}	17.73	< 0.01	**	2.80	0.25	–	3.64	0.72	–
x_{td}	8.70	0.03	**	0.01	1.00	–	2.86	0.83	–
y_{td}	5.59	0.13	–	0.37	0.83	–	4.90	0.56	–
t_{bat}	26.63	< 0.01	**	5.27	0.07	–	3.50	0.74	–

** = significant ($p < 0.05$) – = not significant ($p \geq 0.05$)

IV. Results

This section presents the combined results of the 11 pilots who completed the experiment. For all dependent measures, likelihood ratio tests comparing mixed-effects models including and not including a certain fixed effect determined the significance of that effect. Table 2 provides an overview of the statistical test results. These results are detailed in the following subsections.

A. Demographics and Motion Sickness Susceptibility

Of the 13 pilots who started the experiment, 11 pilots (10 males and 1 female) completed the first simulator session. Two pilots withdrew from the experiment due to cybersickness; one pilot dropped out during the familiarization run and one pilot dropped out after the second training run for the approach task. Pilots' ages ranged between 30 and 65 yr (median $M = 54$, inter quartile range $IQR = 17.75$, $n = 13$). Total flying experience ranged from 252 to 8,020 hr ($M = 1,700$, $IQR = 2,033$). Four pilots had rotary-wing flying experience only, while one pilot had fixed-wing flying experience only. The remainder of the pilots had flown both rotary-wing and fixed-wing aircraft with varying ratios. One pilot was a test pilot with 100 flying hours in an eVTOL vehicle with a control scheme similar to the simplified vehicle controls scheme used in this study. The other pilots had no experience with this type of control scheme. Pilots reported their experience in a simulator ranged from 4 to 1,200 hr ($M = 200$, $IQR = 363$), and experience with VR ranged from 0 to 250 hr ($M = 2$, $IQR = 10$).

Pilots' pre-experiment MSSQ responses revealed them to be less predisposed to motion sickness compared to the general population as measured by Golding, whose data were used to develop the MSSQ-Short [13]. Fig. 9 provides histograms of the percentage subjects per MSSQ score range for this study and Golding's study. Approximately 70% of the pilot population in this study scored within the lowest 0-5 susceptibility bin compared to only 24% of Golding's general population. This is mainly because pilots were recruited after self-identifying as not being susceptible to motion sickness. One of the pilots who dropped out was the single pilot in the 15-20 susceptibility bin. The other pilot fell in the 0-5 bin.

B. Simulator Sickness

Fig. 10 provides bar plots of the SSQ total score (SSQ_{ts}) for each run of the different simulator motion conditions (11 pilots times three repetitions per condition). The medians are depicted by the dashed lines. For the mixed-effects model analysis, the model assumptions of homoscedasticity and normality of residuals were not met as the SSQ data were positively skewed. Data were transformed using a Box-Cox transformation to make the data more normally distributed with $\lambda = -0.41$ [16]. The transformed data were used to fit the models. SSQ_{ts} was statistically significantly affected by the independent variable, motion (Table 2). T-test statistics of the final model fixed effect motion reveal that SSQ_{ts} in both SHX and LHX were similar to SSQ_{ts} in NOM ($t(118) = -0.27$, $p = 0.79$ and $t(118) = 0.13$, $p = 0.90$, respectively); however, SSQ_{ts} in FLM was significantly higher than SSQ_{ts} in NOM ($t(118) = 3.20$, $p < 0.01$), meaning that pilots perceived significantly more cybersickness symptoms in the full motion condition compared to all other

motion conditions. Four pilots reported no symptoms at any time (i.e., rated all symptoms with none in every run), accounting for a large portion of the ratings in the lowest bin for each motion condition.

C. Motion Ratings

Motion ratings (MR) are provided in the bar plots of Fig. 11 for every run of the different motion conditions. Dashed lines again depict the medians for each condition. Comparing models, motion was the only fixed effect introducing a statistically significant effect (Table 2). T-test statistics of the final model fixed effect motion indicate that all conditions where the simulator moved (SHX, LHX, and FLM) had statistically higher ratings compared to the no motion condition NOM ($t(118) = 3.22, p < 0.01$, $t(118) = 3.98, p < 0.01$, and $t(118) = 3.03, p < 0.01$, respectively). This can also be observed by the similar distribution of bars for SHX, LHX, and FLM compared to NOM in Fig. 11.

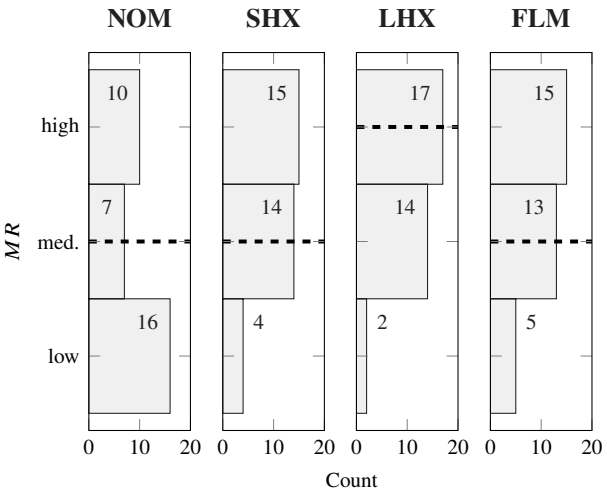


Fig. 11 Motion ratings.

D. Pilot Performance

Task performance metrics are provided in the box plots in Figs. 12 and 13. In these plots, boxes indicate the IQR with a line inside indicating the median. Outliers, any data point below $Q1 - 1.5 \times IQR$ or above $Q3 + 1.5 \times IQR$, are depicted by circles. The means for each condition are represented by an x. Predictions of the mean of the tested pilot population from the models with only significant fixed effects are provided by dashed lines. Finally, desired performance is indicated by dark gray areas and adequate performance by light gray areas (see Table 1). For all task performance metrics, outliers were kept in the data since most were inside the adequate performance boundary, with some just outside and only a few far outside the boundary.

RMS glideslope deviation (RMS Δ_{GS}) and RMS localizer deviation (RMS Δ_{LOC}) are depicted in Figs. 12a and 12b, respectively. Both measures were positively skewed, invalidating the assumptions of homoscedasticity and normality of model residuals. Both measures were transformed using Box-Cox transformations to make them more normally distributed ($\lambda = -0.19$ and $\lambda = -0.04$, respectively). Both glidepath deviation measures were not statistically significantly affected by the independent variable motion. For the majority of the runs, pilots achieved desired performance.

The height above threshold at the FATO boundary (h_{AT}) is depicted in Fig. 12c. This performance metric was

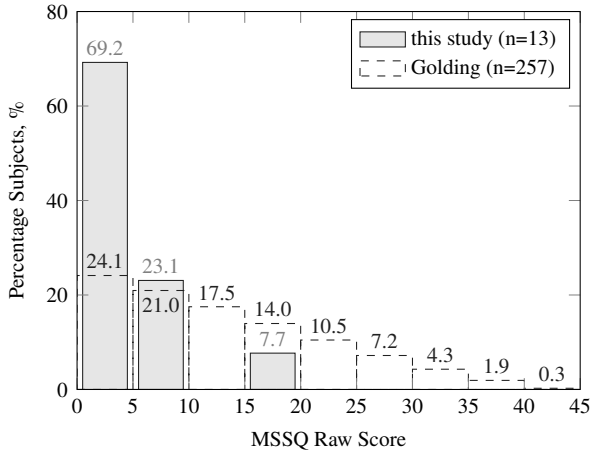


Fig. 9 MSSQ distribution.

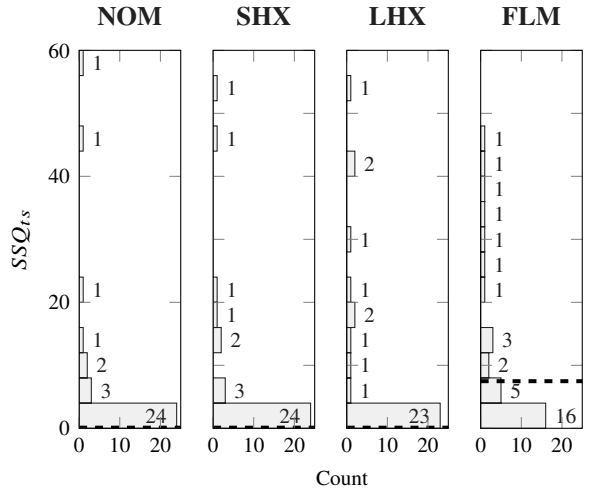
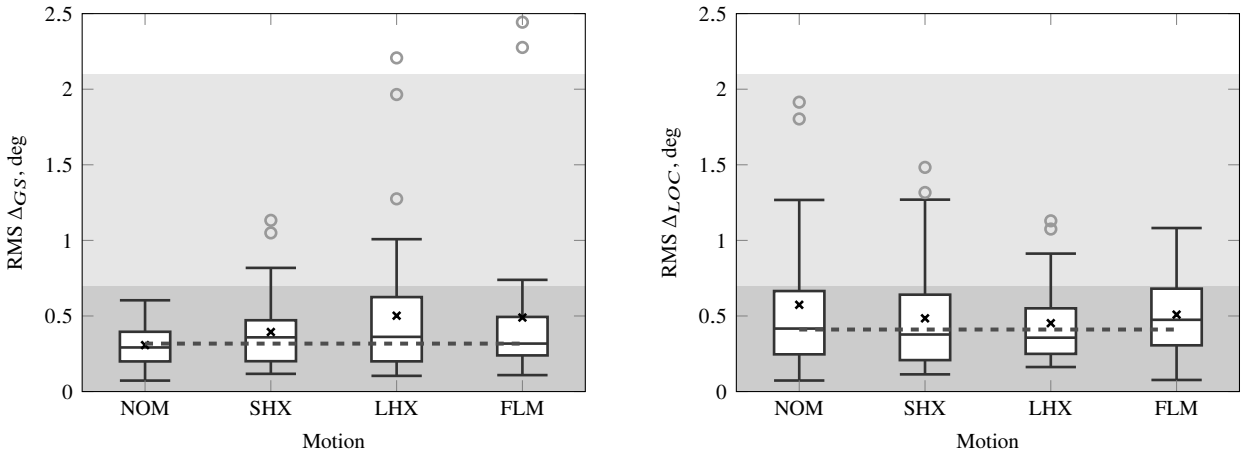
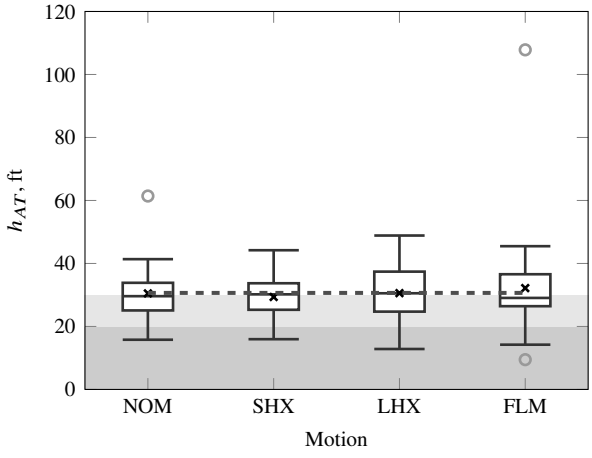


Fig. 10 SSQ ratings.

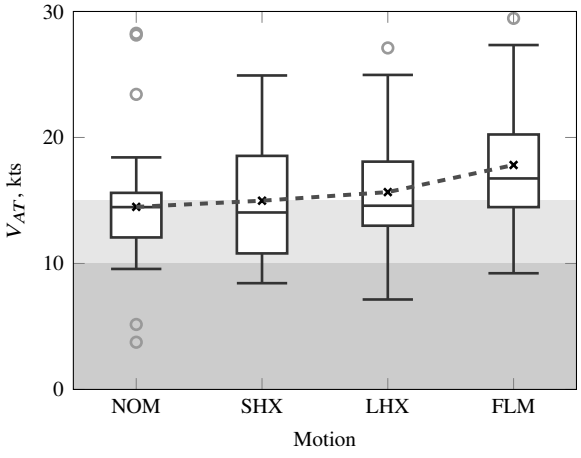


(a) Glideslope deviation.

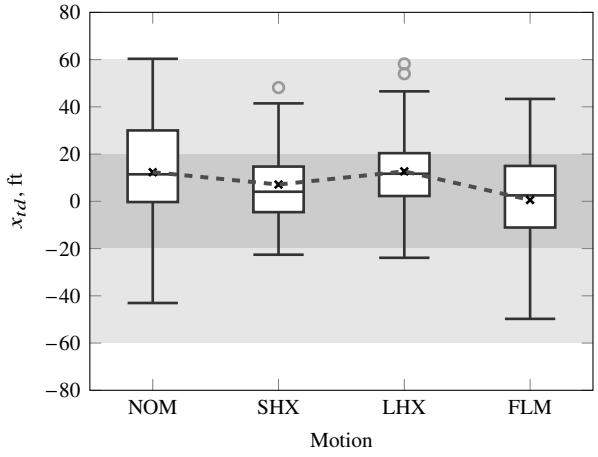
(b) Localizer deviation.



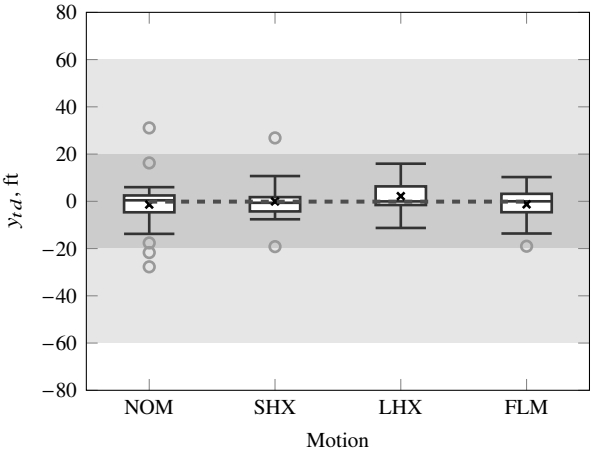
(c) FATO crossing height.



(d) FATO crossing speed.



(e) Longitudinal touchdown deviation.



(f) Lateral touchdown deviation.

Fig. 12 Pilot performance.

not statistically significantly affected by motion (Table 1). Fig. 12d depicts the groundspeed at the FATO boundary crossing. The statistical analysis revealed that V_{AT} was significantly affected by motion as indicated in Table 1. T-test statistics of the final model fixed effect motion reveal that V_{AT} in both SHX and LHX were similar to V_{AT} in NOM ($t(118) = 0.59, p = 0.56$ and $t(118) = 1.41, p = 0.16$, respectively). SSQ_{ts} in FLM was significantly higher than SSQ_{ts} in NOM ($t(118) = 3.99, p < 0.01$), indicating that pilots crossed the FATO boundary significantly faster in the full motion condition. Pilots often did not meet the adequate performance criteria as they instead crossed the FATO boundary too high and fast, with two runs where pilots crossed the boundary above 60 ft.

Longitudinal deviation (x_{td}) and lateral deviation (y_{td}) from the TLOF center at touchdown are depicted in Figs. 12e and 12f, respectively. The longitudinal touchdown deviation was statistically significantly affected by simulator motion (Table 1). Also in this case, x_{td} in both SHX and LHX were similar to x_{td} in NOM ($t(118) = -1.11, p = 0.27$ and $t(118) = 0.08, p = 0.94$, respectively), while x_{td} in FLM was significantly lower than x_{td} in NOM ($t(118) = -2.51, p = 0.01$). Fig. 12e shows that pilots on average overshot the TLOF center under no motion or hexapod motion, but landed significantly closer to the TLOF center longitudinally under full motion. The lateral touchdown deviation was not statistically significantly affected by simulator motion (Table 1). Overall, pilots landed in the desired landing zone in most runs; however, a larger spread in longitudinal touchdown points can be observed.

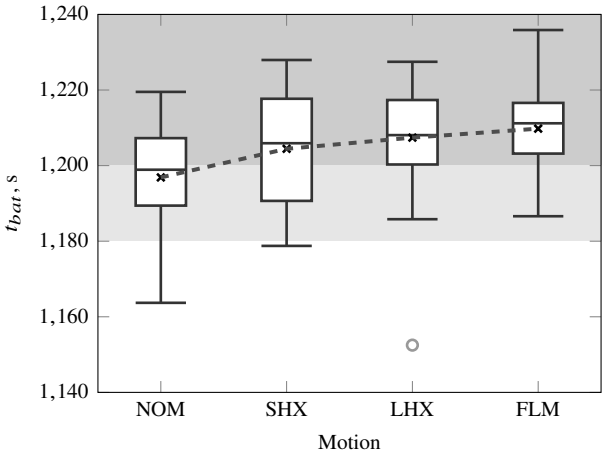


Fig. 13 Battery time remaining.

Finally, the battery time remaining at touchdown (t_{bat}) is provided in Fig. 13. Simulator motion statistically significantly affected battery time remaining at touchdown. T-test statistics of the final model fixed effect motion indicate that the battery time remaining was significantly higher in all conditions where the simulator moved (SHX, LHX, and FLM) compared to the battery time remaining in the no motion condition NOM ($t(118) = 2.97, p < 0.01$, $t(118) = 4.13, p < 0.01$, and $t(118) = 5.06, p < 0.01$, respectively). In the hexapod or full motion conditions, pilots more often achieved a desired battery time remaining at touchdown compared to the no motion condition.

V. Discussion

Eleven pilots completed the experiment flying vertiport approaches and landings with an LPC eVTOL vehicle using a MR headset under different simulator motion conditions. A no motion condition was used as a reference. A 6-inch legged hexapod condition represented smaller hexapod motion systems used for single-seat simulators with small visual systems including MR headsets. A 60-inch legged hexapod motion condition represented Level D hexapod simulators used for airline pilot training. The full motion condition was a stand-in for real aircraft vehicle motion.

A. Effects of Simulator Motion with Mixed-Reality Visuals

The pilot participants tested in this experiment were less susceptible to motion sickness compared to the general population. This was in part by design (participants were asked to participate only if self-identifying as not susceptible to motion sickness), but likely also due to the fact that pilots are conditioned to aircraft and simulator environments with their associated motion cues. Minimal cybersickness was reported by seven pilots through the SSQ administered after each run indicating symptoms related to nausea, oculomotor discomfort and disorientation. Even though the effect size was small, the severity of cybersickness was statistically significantly higher in the full motion condition compared to the hexapod and no motion conditions. Four pilots did not report any occurrence of simulator sickness symptoms.

Subjective motion ratings were collected after each landing asking pilots to rate the quality of the motion in the previous run with low, medium, or high. Motion ratings in all motion conditions were similar and statistically significantly higher compared to the no motion condition. However, the variation in ratings was high, with pilots often rating repetitions of the same motion condition differently. This is the nature of subjective motion ratings and underlines the importance of using behavioral- or outcome-based metrics to evaluate the effects of motion fidelity on pilots.

Simulator motion fidelity significantly affected pilot performance in the final approach and landing while using

a MR headset for OTW visuals. Glideslope and localizer tracking accuracy in the approach up to the DECEL point, height above threshold at the FATO crossing, and lateral touchdown deviation were not statistically significantly affected. Groundspeed while crossing the FATO boundary was statistically significantly higher in the full motion condition compared to the hexapod and no motion conditions. This coincided with longitudinally touching down significantly closer to the center of the landing pad with full motion, and with significantly more battery time remaining. Battery time remaining was similar between all motion conditions but significantly higher with motion compared to no motion. Of the nine dependent measures collected in this study, five were statistically significantly affected by motion. Of those five, three measures - cybersickness severity, FATO boundary crossing speed, and longitudinal touchdown deviation - were similar in the no motion and small- and large-hexapod motion conditions but were significantly different in the full motion condition. Two measures - motion rating and battery time remaining - were similar in the motion conditions but significantly different compared to no motion.

Since cybersickness often increases with prolonged use of VR/MR headsets, repetition and the interaction of motion with repetition were added as fixed effects in the mixed-effects model analysis. However, both the main fixed effect and the interaction did not statistically significantly affect any of the dependent measures. This means cybersickness symptoms, motion ratings, and pilot performance remained consistent for each motion condition during the experiment and were not affected by wearing the MR headset longer.

This study shows that pilot performance using MR visuals is significantly affected by simulator motion, justifying the use of motion systems in simulators with VR/MR visuals. The simplified vehicle controls scheme of the LPC eVTOL vehicle used in this study, an indirect flight control scheme with high-level outer loop control, most likely reduced the effects of motion as pilots were not directly controlling the principal axes of the vehicle. This means that simulator motion might be more important while using VR/MR visuals for aircraft with more conventional control schemes. It should be emphasized that this study does not provide any data on the effects of simulator motion fidelity on pilot training with MR visuals, and the transfer of (negative) training to the real vehicle. More research is required on this topic.

B. Experiment Setup

Accurate and stable head tracking while using a VR/MR headset in any motion-base simulator is a difficult problem. Typically, motion-base simulators utilizing VR/MR headsets do not have an enclosed cab and can use fixed base stations on the ground for tracking. This study, for the first time, used a MR headset in an enclosed cab of the largest vertical motion simulator in the world. This required a novel, commercial off-the-shelf tracking solution using sensor fusion for the cab accelerations measured by a mounted IMU, head accelerations from an IMU on the headset, and head position from an optical sensor. The implemented head tracking solution provided accurate and stable head tracking under all simulator motion conditions without noticeable discrepancies between the virtual and camera passthrough portions of the MR visuals. Only extreme head positions, e.g., when turning the head all the way left or right, resulted in a loss of head tracking and caused the virtual scene to shift relative to the passthrough image of the simulator cab. Loss of head tracking occurred infrequently during the study, but was usually accompanied by an increase in cybersickness symptoms.

The Varjo XR-4 headset used in this study utilized foveated rendering, meaning that the scene was rendered at a higher resolution around a user's gaze point as determined by a built-in eye tracker. This worked well in general, but the foveated rendering noticeably lagged behind for very quick eye movements, resulting in momentarily blurred vision around pilots' gaze point. This was especially an issue when pilots quickly shifted their gaze from looking out the window to scanning the PFD. The momentarily blurred vision as the foveated rendering caught up possibly resulted in difficulty reading instruments. Another source of momentarily blurred vision could possibly be linked to the geometry of the lenses used in the headset. Multiple strategies might be available to reduce blurred vision from these sources, but time constraints did not allow for exploring these. Based on pilot feedback, blurred vision had a minimal effect on the results of this experiment when it occurred.

This study used MR visuals, which are a combination of a virtual OTW scene and a passthrough image of the real simulator cockpit, allowing pilots to see the cockpit HDDs, control inceptors, and their own limbs. However, most simulators that utilize HMD technologies only use VR visuals. MR visuals, such as those used in this study, can increase realism as users can more easily physically interact with different elements in the cockpit, such as buttons and control inceptors. In addition, the combined visual scene in MR can feel more grounded compared to purely VR visuals, potentially reducing cybersickness. These observations are anecdotal and further research is needed to investigate the differences between MR and VR visuals in motion-base simulators.

The approved study protocol included a second task, the lateral reposition and hold task from [9], in addition to

the vertiport approach and landing task. This task was performed by all 11 pilots in a second session following the vertiport approach and landing session with a break in between. This task required simultaneous coordination between the controlled axes of the vehicle, which was difficult for most pilots as they were not familiar enough with the simplified vehicle controls used in this study. This resulted in large variations and outliers in the data for this task. Furthermore, pilots occasionally had to abort and redo runs after they mixed up the controls. A full statistical analysis was done on the results of this task, but no effects of motion were found for any of the dependent measures associated with the task. Due to the issues with the data for this task, it was decided to not include the results in this paper to avoid misinterpretation. This highlights the need for more extensive training to master simplified vehicle controls of novel eVTOL vehicles for pilots who have only flown aircraft with more conventional control schemes.

VI. Conclusion

This study used the NASA Ames Vertical Motion Simulator with an eVTOL vertiport approach and landing task and four different motion conditions to determine the effects of simulator motion while using mixed-reality visuals. Eleven pilots performed the task under all four motion conditions.

Both subjective and objective measures were significantly affected by the motion condition. For the subjective measures, pilots reported more severe symptoms of cybersickness with full motion as compared to no motion, and small and large hexapod motion. In addition, pilots rated the quality of the motion cues higher under all conditions with simulator motion compared to no motion. Objective task performance results indicate that pilots crossed the boundary of the final approach and take-off threshold at higher speeds and touched down closer to the desired touchdown point with full motion as compared to the hexapod and no motion conditions. Finally, more battery time remained in the conditions with motion as compared to no motion.

The statistically significant differences found between motion conditions suggest that motion systems are essential in simulators with virtual- or mixed-reality visuals when pilot flying performance similar to that in real flight is required, but more research is needed to determine the effects on training and transfer of training.

Appendix

The experiment used three conditions with cab motion (SHX, LHX, and FLM), and a condition without cab motion (NOM). The experiment used the standard VMS motion algorithm and hardware for all motion conditions.

A hexapod algorithm took the accelerations from the VMS motion algorithm and calculated the leg extensions for the hexapod motion conditions (SHX and LHX). Six inch was the maximum leg extension for SHX and 60 inch was the maximum leg extension for LHX. The hexapod algorithm accounted for the inherent constraints of trying to simultaneously move a hexapod in several degrees of freedom. Mass and inertia effects of the hexapod system were not taken into account. The hexapod algorithm was used for motion tuning and monitoring purposes only; that is, it was not used to calculate motion inputs for the VMS motion system [17].

The VMS motion logic is completely linear in its motion envelope; that is, the only nonlinear element is the motion limiting near the edges of the envelope. Motion filtering on the VMS is performed in the inertial reference frame. Aircraft model accelerations are transformed from the aircraft body reference frame to the inertial reference frame using small-angle approximations. The logic uses gains and second-order high-pass washout filters to attenuate accelerations in the translational and rotational degrees of freedom:

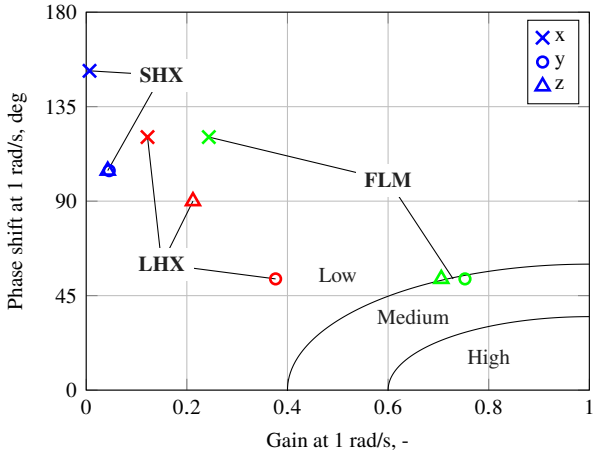
$$H_m = K_m \frac{s^2}{s^2 + 2\zeta_m \omega_m s + \omega_m^2} \tag{1}$$

where s is the Laplace variable, K_m is the motion gain, ζ_m is the filter damping ratio, and ω_m is the filter break frequency. The VMS motion algorithm incorporates residual tilt in the longitudinal and lateral axes and, in the standard configuration used in this experiment, turn coordination in the roll axis. Since the principal motion cues were provided by the high-pass washout filters, residual tilt and turn coordination are not detailed here. More information about the VMS motion algorithm can be found in [10].

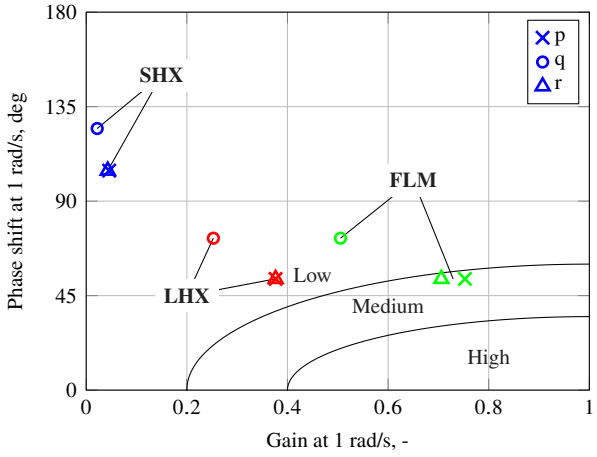
Based on evaluations with a test pilot, the SHX, LHX and FLM motion conditions of the experiment were tuned such that the limits of the simulated hexapods or the VMS motion system were not reached. However, pilots who participated in the experiment for data collection occasionally hit a limit. Table 3 provides the parameters of the high-pass washout filters for all four motion configurations.

Table 3 Motion configuration parameters.

Motion	NOM	SHX	LHX	FLM
Parameter	Motion Gains			
K_{mx}	0.000	0.060	0.300	0.600
K_{my}	0.000	0.080	0.400	0.800
K_{mz}	0.000	0.075	0.300	0.750
K_{mp}	0.000	0.080	0.400	0.800
K_{mq}	0.000	0.060	0.300	0.600
K_{mr}	0.000	0.075	0.400	0.750
	Washout Break Frequencies			
ω_{mx}	-	3.000	1.500	1.500
ω_{my}	-	1.200	0.600	0.600
ω_{mz}	-	1.200	1.000	0.600
ω_{mp}	-	1.200	0.600	0.600
ω_{mq}	-	1.600	0.800	0.800
ω_{mr}	-	1.200	0.600	0.600
	Washout Damping Ratios			
ζ_{mx}	-	0.707	0.707	0.707
ζ_{my}	-	0.707	0.707	0.707
ζ_{mz}	-	0.707	0.707	0.707
ζ_{mp}	-	0.707	0.707	0.707
ζ_{mq}	-	0.707	0.707	0.707
ζ_{mr}	-	0.707	0.707	0.707



(a) Translational.



(b) Rotational.

Fig. 14 Sinacori diagrams.

A Sinacori diagram showing the motion fidelity for all motion conditions with simulator motion (SHX, LHX, FLM) is provided in Fig. 14a for the translational degrees of freedom and in Fig. 14b for the rotational degrees of freedom [15]. Generally, FLM has the highest motion fidelity, followed by LHX, and next SHX. Note that the fidelity of longitudinal motion is low for all motion conditions. No degrees of freedom for any of the motion conditions fall in the high fidelity region.

Acknowledgments

This study was performed under a Space Act Agreement between NASA Ames Research Center and Quantum3D, Inc. (SAA2-403874). The NASA authors thank Quantum3D, Inc. for providing the image generator system and support for this study. The authors are grateful to David Zahn from NASA Ames SimLabs for providing valuable insights to improve the approach and landing task and creating the flight cards while serving as a test pilot for this study. The authors appreciate the assistance of Shawn Naigle from the United States Army Technology Development Directorate for serving as a trial pilot. The authors thank the simulation technicians and operators who helped run the study, and the 13 study participants. Finally, the authors thank the management of the Aerospace Simulation Research & Development Branch and the Aviation Systems Division at NASA Ames Research Center for making this study possible.

References

[1] NASA, *A New Continent of Ideas*, National Aeronautics and Space Administration, 1990, Chap. Computer Technology, pp. 88–91. URL https://ntrs.nasa.gov/citations/20020086961.

[2] Garcia, A. D., Schlueter, J., and Paddock, E., "Training Astronauts using Hardware-in-the-Loop Simulations and Virtual Reality," *Proceedings of the AIAA Scitech 2020 Forum*, American Institute of Aeronautics and Astronautics, 2020. https://doi.org/10.2514/6.2020-0167.

[3] Kramida, G., "Resolving the Vergence-Accommodation Conflict in Head-Mounted Displays," *IEEE Transactions on Visualization and Computer Graphics*, Vol. 22, No. 7, 2016, pp. 1912–1931. https://doi.org/10.1109/TVCG.2015.2473855.

[4] Ross, G., and Gilbey, A., "Extended reality (xR) flight simulators as an adjunct to traditional flight training methods: a scoping review," *CEAS Aeronautical Journal*, Vol. 14, No. 4, 2023, pp. 799–815. https://doi.org/10.1007/s13272-023-00688-5.

[5] Schroeder, J. A., and Grant, P. R., "Pilot Behavioral Observations in Motion Flight Simulation," *AIAA Guidance, Navigation, and Control Conference and Exhibit*, American Institute of Aeronautics and Astronautics, 2010. https://doi.org/10.2514/6.2010-8353.

[6] EASA, "FSTD Special Conditions for the use of Head Mounted Displays (HMD) combined with a motion platform with reduced envelope," Tech. rep., European Union Aviation Safety Agency, 2023. URL https://www.easa.europa.eu/en/downloads/137723/en.

[7] Silva, C., Johnson, W., Antcliff, K. R., and Patterson, M. D., "VTOL Urban Air Mobility Concept Vehicles for Technology-Development," *2018 Aviation Technology, Integration, and Operations Conference*, American Institute of Aeronautics and Astronautics, 2018. https://doi.org/10.2514/6.2018-3847.

[8] Kaneshige, J., Lombaerts, T., Shish, K., and Feary, M., "Simplified Vehicle Control Concept for a Lift Plus Cruise eVTOL Vehicle," *AIAA AVIATION FORUM AND ASCEND 2024*, American Institute of Aeronautics and Astronautics, 2024. https://doi.org/10.2514/6.2024-4207.

[9] Klyde, D., Mitchell, D., Shubert, M., and Jones, M., "Handling Qualities Test Guide for Powered Lift VTOL Capable Aircraft with Indirect Flight Controls and Operating in Day Time Visual Flight Rules," Tech. Rep. DOT/FAA/TC-23/59, Federal Aviation Administration, 2024. URL https://rosap.ntl.bts.gov/view/dot/74025.

[10] Beard, S. D., Reardon, S. E., Tobias, E. L., and Aponso, B. L., "Simulation System Optimization for Rotorcraft Research on the Vertical Motion Simulator," *Proceedings of the AIAA Modeling and Simulation Technologies Conference, Minneapolis (MN)*, 2012. https://doi.org/10.2514/6.2012-4634.

[11] Zaal, P. M. T., Orth, S. P., Blanken, M. L., Penmetcha, S. V., Zahn, D. I., Khan, H. A., Sharma, V., and Comelli, L., "Evaluation of a Mixed-Reality Headset Using the SIMulated Day Useable Cue Environment Paradigm," *70th Annual Symposium & Banquet*, Society of Experimental Test Pilots, 2025.

[12] Zaal, P. M. T., Schroeder, J. A., and Chung, W. W., "Objective Motion Cueing Criteria for Commercial Transport Simulators," *2018 Modeling and Simulation Technologies Conference*, American Institute of Aeronautics and Astronautics, 2018. https://doi.org/10.2514/6.2018-2935.

[13] Golding, J. F., "Predicting Individual Differences in Motion Sickness Susceptibility by Questionnaire," *Personality and Individual Differences*, Vol. 41, No. 2, 2006, pp. 237–248. https://doi.org/10.1016/j.paid.2006.01.012.

[14] Kennedy, R. S., Lane, N. E., Berbaum, K. S., and Lilienthal, M. G., "Simulator Sickness Questionnaire:An Enhanced Method for QuantifyingSimulator Sickness," *The International Journal of Aviation Psychology*, Vol. 3, No. 3, 1993, pp. 203–220. https://doi.org/10.1207/s15327108ijap0303_3.

[15] Schroeder, J. A., "Helicopter Flight Simulation Motion Platform Requirements," Tech. Rep. NASA/TP-1999-208766, National Aeronautics and Space Administration, 1999. URL https://ntrs.nasa.gov/citations/19990080926.

[16] Box, G. E. P., and Cox, D. R., "An Analysis of Transformations," *Journal of the Royal Statistical Society*, Vol. 26, No. 2, 1964, pp. 211–252. https://doi.org/10.1111/j.2517-6161.1964.tb00553.x.

[17] Zaal, P. M. T., Schroeder, J. A., and Chung, W. W., "Objective motion cueing criteria investigation based on three flight tasks," *The Aeronautical Journal*, Vol. 121, No. 1236, 2017, pp. 163–190. https://doi.org/10.1017/aer.2016.119.