Evaluation of a Mixed-Reality Headset Using the SIMulated Day Useable Cue Environment Paradigm

Peter M. T. Zaal*, Samuel P. Orth†, Matt L. Blanken‡, Supreethi V. Penmetcha§, David I. Zahn¶, and Haroon A. Khan‖
NASA Ames Research Center, Moffet Field, CA, 94035, USA

Vikrant Sharma** and Lorenzo Comelli††
National Test Pilot School, Mojave, CA, 93501, USA

Ground-based simulators with simulated out-the-window visuals have long been used for handling qualities evaluations of new aircraft. In recent years, there has been an increased interest to replace conventional projector-based out-the-window visual systems with virtual- or mixed-reality headsets as these devices have the potential to increase realism, reduce cost, and make simulators more accessible. Although the quality of visual cues has improved significantly as headset technologies have advanced, it is still largely unknown if the visual cues provided by these headsets are suitable for handling qualities evaluations. To quantify the quality of simulated day visual cues for handling-qualities evaluations, the SIMulated Day Useable Cue Environment concept from the MIL-DTL-32742 standard has previously been applied for conventional visual systems. This paper is the first to use this concept for a mixed-reality headset. Nine pilots from the National Test Pilot School participated in the evaluation in the Vertical Motion Simulator at NASA Ames Research Center. The evaluation used a simulated electric vertical takeoff and landing vehicle with three Handling Qualities Task Elements from the FAA eVTOL Handling Qualities Test Guide: pirouette, lateral reposition and hold, and vertical reposition and hold. Some limitations make the direct interpretation of the Simulated Day Useable Cue Environment results from this study difficult; however, this study shows that the methodology is a valuable concept to help evaluate the quality of visual cues in mixed-reality headsets for handling qualities evaluations.

I. Introduction

This paper uses the SIMulated Day Useable Cue Environment (SIMDUCE) paradigm as specified in the Handling Qualities for Military Rotorcraft (MIL-DTL-32742) standard [1] to evaluate the useable cue environment (UCE) of a mixed-reality (MR) headset in a motion-base simulator, providing baseline data and lessons learned for future handling qualities evaluations with MR headsets.

The use of simulators in handling qualities evaluations for both civilian and military rotorcraft, as well as for emerging vehicles such as electric vertical takeoff and landing (eVTOL) aircraft, has allowed for a more safe and cost-effective approach to testing. Many requirements for military rotorcraft handling qualities, for example, are allowed to be verified with piloted simulations throughout the development phase, up until the First Flight Readiness Review (FFRR) [1]. In recent years, there has been an increased interest in using virtual-reality (VR) or MR headsets instead of conventional projector-based out-the-window (OTW) visual systems in flight simulators as they have the potential to reduce cost and make simulators more accessible [2, 3]. In addition, these headsets can potentially provide improved visual cues and realism, for example, by allowing pilots to directly look through side cockpit windows. However, despite the potential of the technology, there are currently very few standardized specifications and evaluation tests for the use of VR/MR visuals in simulators, limiting the use for handling qualities evaluations and training. Regulations that do exist focus on helicopter simulation and use conventional visual systems as the norm [4]. SIMDUCE is a standardized

*Principal System Architect, Metis Technology Solutions, Inc., peter.m.t.zaal@nasa.gov.
†Simulation Engineer, Symvionics, Inc., samuel.p.orth@nasa.gov.
‡Simulation Engineer, NASA Ames Research Center, matthew.l.blanken@nasa.gov.
§Simulation Engineer, NASA Ames Research Center, supreethi.v.penmetcha@nasa.gov.
¶Test Pilot, NASA Ames Research Center, david.zahn@nasa.gov, SETP Member.
‖Simulation Engineer, NASA Ames Research Center, haroon.a.khan@nasa.gov.
**Chief Theoretical Knowledge Instructor, National Test Pilot School, vsharma@ntps.edu.
††Flight Test Engineer, National Test Pilot School, lcomelli@ntps.edu.

methodology to evaluate and calibrate simulator visuals for handling qualities testing, but there are no published data or lessons learned for its use with VR/MR headsets. To facilitate the use of these headsets for handling qualities evaluations, this paper adds data to fill this gap.

Nine pilots participated in SIMDUCE evaluations of an MR headset in the Vertical Motion Simulator (VMS) at NASA Ames Research Center. They controlled a Lift-Plus-Cruise (LPC) eVTOL aircraft with Simplified Vehicle Controls (SVC) in three Handling Qualities Task Elements (HQTEs), or flight test maneuvers. The large motion capability of the simulator allowed for near one-to-one replication of the simulated vehicle motions. A second data-collection session was performed without motion and with conventional projector-based visuals to add baseline data.

The paper first describes the method in Section II. Results are provided in Section III, followed by a thorough discussion of the procedure, data limitations, and lessons learned in Section IV. Section V provides conclusions and recommendations for future work.

II. Method

A SIMDUCE evaluation using MR visuals was conducted in the VMS.

A. SIMDUCE Procedure

This study followed the standard SIMDUCE evaluation procedure from the MIL-DTL-32742 standard [1]. Pilots were asked to provide a numeric rating of their ability to be aggressive and precise when flying a specific HQTE. This rating, known as a Visual Cue Rating (VCR), was given on a continuous scale from 1 to 5, where 1 is good and 5 is poor, for three cueing aspects of the simulated environment: attitude, horizontal translational rate, and vertical translational rate. The lower-value (i.e., better) translational rate VCR was then discarded, while both the attitude VCR and higher-value translational rate VCR were averaged across multiple pilots. The average results were plotted to determine the UCE for that specific HQTE. All pilots were given a minimum of two training runs for each HQTE prior to flying one data collection run.

After collecting the three VCRs, a Cooper-Harper Handling Quality Rating (HQR) was also collected and averaged across multiple pilots for each task to determine the vehicle's Handling Qualities (HQ) level [5]. This is recommended per the MIL-DTL-32742 standard as there should be no large discrepancies between the UCE and HQ levels (i.e., the highest UCE level achievable in a specific vehicle is limited by its HQ) [1].

B. Apparatus

1. Vertical Motion Simulator

The VMS is the largest vertical motion simulator in the world (Fig. 1). Its motion system has six independent degrees of freedom, including 60 feet of vertical travel and 40 ft of lateral travel [6]. Five interchangeable cabs are available to simulate a range of different vehicles from helicopters to transport aircraft and lunar landers. The simulator was used in the development of the Cooper-Harper HQR scale and in many handling qualities evaluations over the last decades [6–8].

This SIMDUCE evaluation used the rotorcraft cab (R-Cab) with a single seat in the middle of the cabin. Two sidesticks, positioned on the left and right sides of the seat, were used for aircraft control. Pedals were installed in the cockpit but not used for aircraft control. Three head down displays (HDDs) were positioned in front of the pilot: a navigation display to the left, a primary flight display in the center, and a health and status display with powertrain information on the right (Fig. 2). A projector-based OTW visual system was used in a second data-collection session (see Section II.D)–not the main SIMDUCE evaluations–to gather baseline data for conventional visuals (CV). This system used three collimated displays, driven by an RSi Raster XT-6 image generator, with a total field-of-view (FOV) from +78° to -77° in azimuth and +12° to -17° in elevation. Each collimated display had a resolution of 1600 by 1200 pixels and a 60 Hz refresh rate.

2. Head Mounted Display

A Varjo XR-4 Secure Focal Edition head-mounted display (HMD) provided the MR visuals (Fig. 3), combining a virtual OTW visual scene with a 20 megapixel camera passthrough of the real world. This passthrough allowed pilots to view parts of the physical cab, including the HDDs, the sidesticks, and their own hands (Fig. 4). The HMD featured a 120° horizontal by 105° vertical FOV with a resolution of 3840 by 3744 pixels at a 90 Hz refresh rate [9].

Fig. 1 Vertical Motion Simulator.

Fig. 2 Cockpit setup.

To render the correct OTW scene in the HMD, pilots' head movements were decoupled from the motion of the VMS using a Life Performance Research (LP-Research) LPVR-DUO head tracking system. This system used a sensor fusion algorithm, with an optical sensor and inertial measurement unit (IMU) positioned on the R-cab's glare shield directly in front of the pilot (Fig. 2), to isolate the relative position and orientation of the HMD with respect to the VMS cab [10]. This final pose was passed to a Quantum3D Mantis image generator to render the appropriate OTW visual scene, with the HQTE courses placed within the city of San Francisco and San Francisco International Airport.

C. Flight Tasks

The three HQTEs flown in this SIMDUCE evaluation included the pirouette, lateral reposition and hold, and vertical reposition and hold, with performance requirements based on the FAA eVTOL Handling Qualities Test Guide [11]. The pirouette course was located at the San Francisco International Airport while the other two courses were located in San Francisco. Pilots pressed a button on the right sidestick to mark different parts of the maneuver, such as when starting, translating, and stabilizing. Audio cues informed pilots upon maintaining a stable hover for the appropriate period of time.

1. Lift-Plus-Cruise eVTOL Vehicle

Pilots flew a conceptual NASA design of an LPC vehicle with eight lifting rotors, a pusher propeller, and control surfaces including two ailerons, an elevator, and a rudder (Fig. 5) [12].

The aircraft used an SVC scheme to standardize control axes across the entire speed envelope. At low speeds, with the hover mode disengaged, the right sidestick commanded vertical acceleration with the longitudinal axis, bank angle with the lateral axis, and heading rate with the yaw axis, while the single-axis left sidestick commanded longitudinal acceleration with the longitudinal axis. When hover mode was engaged, the control scheme switched to a translational rate command, where the right sidestick instead commanded vertical speed with the longitudinal axis and lateral velocity with the lateral axis, while the left sidestick commanded longitudinal velocity with the longitudinal axis. Yaw control did not change in hover mode [12]. Both conditions, with and without the hover mode engaged, were used during this experiment, and groundspeeds remained low throughout due to the nature of the maneuvers performed (Table 1).

2. Pirouette

The pirouette HQTE required the pilot to translate around a 100-ft-radius circle, marked with orange and yellow cones delineating the desired and adequate performance boundaries, respectively. The pilot fixed their heading to a

Fig. 3 Mixed-reality headset.

Fig. 4 Mixed-reality visuals.

Fig. 5 Lift-Plus-Cruise eVTOL aircraft.

visual reference in the center of the circle while completing one revolution, before stabilizing to a hover back at the starting position. Compared to the FAA eVTOL Handling Qualities Test Guide, the performance requirement for the altitude deviation was changed from 3 ft to 5 ft (desired) and 5 ft to 8 ft (adequate) [11]. Other performance criteria for the pirouette task were left unchanged (Table 2).

3. Lateral Reposition and Hold

For the lateral reposition and hold HQTE, the pilot began in a stable hover aligned with a visual reference board. The pilot laterally repositioned 400 ft to the right, following a black line with orange cones marking the desired boundary and yellow cones marking the adequate boundary of the performance requirements. The pilot stabilized at a second hover point, aligned with another visual reference board, and maintained a stabilized hover for 5 seconds. This process was repeated by translating left to the first hover point and stabilizing to a hover once again.

Performance criteria for this task were adapted from the FAA eVTOL Handling Qualities Test Guide and adjusted based on performance of the vehicle during the experiment set up (Table 3). The performance requirement for the target groundspeed margin was increased from 2 kts to 5kts (desired) and 4 kts to 8 kts (adequate) for this evaluation [11]. Based on the target groundspeed of 15 kts, an additional requirement was added to translate left or right within a specific amount of time (Table 3).

Table 1 Lift-Plus-Cruise Simplified Control Scheme.

		Hover Mode	
Sidestick	Axis	Disengaged	Engaged
Right	Longitudinal	Vertical Acceleration	Vertical Speed
Right	Lateral	Bank Angle	Lateral Velocity
Right	Twist	Heading Rate	Heading Rate
Left	Longitudinal	Longitudinal Acceleration	Longitudinal Velocity

Table 2 Pirouette Performance Criteria.

	Desired	Adequate
Maintain within the circumference of the circular ground track:	±10 ft	±15 ft
Maintain altitude within:	±5 ft	±8 ft
Maintain heading pointing towards the center of the circle within:	±10 deg	±15 deg
Complete the path within:	≤ 60 s	≤ 80 s
Achieve stabilized hover after returning to the starting point within:	5 s	10 s

Table 3 Lateral Reposition and Hold Performance Criteria.

	Desired	Adequate
Maintain ground track within:	±5 ft	±10 ft
Attain target groundspeed within:	±5 kts	±8 kts
Complete path within:	≤ 25 s	≤ 35 s
Maintain altitude within:	±5 ft	±10 ft
Maintain heading within:	±10 deg	±15 deg
At capture, maintain lateral/longitudinal position:	±5 ft	±10 ft
PIO tendencies in the capture and hold:	No undesirable motions that impact task performance	No divergent PIO (out-of-phase oscillations)

Table 4 Vertical Reposition and Hold Performance Criteria.

	Desired	Adequate
Maintain longitudinal and lateral position within:	±3 ft	±6 ft
Maintain start/finish altitude within:	±3 ft	±6 ft
Maintain heading within:	±5 deg	±10 deg
Complete maneuver within:	≤ 22 s	≤ 27 s
PIO tendencies in the capture and hold:	No undesirable motions that impact task performance	No divergent PIO (out-of-phase oscillations)

Table 5 Experimental conditions.

Condition	Data Collection	Number of Pilots	Simulator Motion	Vehicle Hover Mode	Visual Systems
1	Session 1	8	full motion	disengaged	MR
2	Session 2	3	no motion	engaged	MR
3	Session 2	3	no motion	engaged	CV

4. Vertical Reposition and Hold

The structure of the vertical reposition and hold HQTE was very similar to the lateral reposition and hold HQTE, except the pilot vertically repositioned up and down between a lower hover position and an upper hover position. Additionally, this HQTE was flown with a 15 kt headwind and 15 kt gusts to increase difficulty. The performance requirement for the time to complete the entire maneuver was reduced in this evaluation from 25 s to 22 s (desired) and 30 s to 27 s (adequate) [11]. Other performance criteria remained unchanged (Table 4).

D. Experimental Conditions

Data were collected in two separate sessions under different conditions (Table 5). Session 1 spanned two consecutive days of data collection. Session 2 lasted only one day but took place six weeks after the completion of Session 1 due to scheduling constraints. Since the main purpose of the study was to use SIMDUCE to evaluate an MR headset, only MR visuals were used in Session 1. The hover mode was disengaged during this session as this resulted in more direct control of vehicle attitude (i.e., bank angle instead of lateral velocity control with lateral right stick movements), which was thought to be beneficial in providing attitude VCRs. Full motion was used as the test requires the highest simulator fidelity available for a particular simulator. During Session 2, however, simulator motion was not available as the R-Cab was not installed on the motion system but in a fixed-base lab. To help mitigate resultant deficiencies, the vehicle hover control mode was engaged as it was thought that this would allow handling qualities to be closer to Level 1 as required by the SIMDUCE test procedure.

E. Pilots

A total of nine different pilots participated in this study, with eight pilots participating in Session 1 and three pilots participating in Session 2. All eight pilots in Session 1 were test pilots. Four were primarily fixed-wing pilots, while the other four only had rotorcraft experience. One of the four fixed-wing pilots had significant experience flying an eVTOL aircraft, while another had experience with the ‘Unified’ control scheme, which helped form the basis for the SVC concept used in this study [12].

Two of the pilots that participated in Session 2 were test pilots returning from Session 1. One had rotorcraft experience, while the other had fixed-wing and eVTOL experience. The third pilot in Session 2 was not a test pilot but had extensive experience flying the concept eVTOL vehicle used in this simulation (in addition to commercial fixed-wing experience).

Table 6 Visual Cue Ratings for all pilots and tasks.

		Attitude VCR			Translational Rate VCR		
Condition	Task	N	Mean	STD	N	Mean	STD
1	Pirouette	7	2.314	0.761	7	3.000	0.964
	Lateral Repo.	7	2.614	0.712	7	3.427	0.562
	Vertical Repo.	7	2.286	0.749	7	3.057	0.604
2	Pirouette	3	2.500	0.707	3	3.167	0.850
	Lateral Repo.	2	2.500	0.000	3	3.167	0.236
	Vertical Repo.	-	-	-	1	3.000	-
3	Pirouette	1	2.500	-	1	4.000	-
	Lateral Repo.	1	3.000	-	1	2.000	-
	Vertical Repo.	1	3.000	-	1	3.000	-

III. Results

This section provides the results of the two data-collection sessions (Table 5). Table 6 provides the means and standard deviations (STD) of the attitude and translational rate VCRs over all pilots who performed a certain task and condition. N is the number of data points. Note that some pilots did not perform all tasks resulting in an N lower than the total number of pilots.

Standard deviations of the VCRs in Session 1 generally remained below 0.75, except for the pirouette task which had a standard deviation of 0.761 for the attitude VCR and 0.964 for the translational rate VCR (Table 6). Per the SIMDUCE procedure, a standard deviation greater than 0.75 is considered unsatisfactory for a complete evaluation [1]. In Session 2, both MR and CV had the same attitude VCR for the pirouette task, although the CV had a higher translational rate VCR, albeit based on a single pilot rating. The CV had a higher attitude VCR but a lower translational rate VCR than MR for the lateral reposition and hold task. The pirouette task once again had a high standard deviation in the translational rate VCR (0.85, N=3), specifically in the MR configuration. Due to schedule constraints, only one pilot completed any of the three HQTEs with conventional visuals in Session 2. Similarly, only one pilot completed the vertical reposition and hold HQTE with MR but noted that it was difficult to provide an attitude VCR for that task (Table 6).

For Session 1, the UCE results for MR visuals are provided in Fig. 6. The UCE for all three tasks in this test was found to be Level 2, with the UCE for the pirouette and vertical reposition and hold HQTEs falling slightly closer to the Level 1 threshold (Fig. 6). Fig. 7 provides the UCE results for MR and CV as determined in Session 2. Note that some UCEs are based on only a single VCR (Table 6). Despite this limitation, both MR and CV were found to have a Level 2 UCE for the three tasks (Fig. 7).

Fig. 8 depicts the average and standard deviation of the collected HQRs for all experimental conditions in Table 5, per MIL-DTL-32742 [1]. In all three experimental conditions, the pilots rated the vehicle as having Level 2 handling qualities, with the exception of the vertical reposition and hold HQTE when using MR in hover mode, which was found to be Level 1 by the single participant. These HQRs across the three HQTEs were relatively consistent within each test condition. All six hover mode cases were rated between an HQR of 3.0 and 5.0, and the three HQRs when using MR in non-hover mode were slightly higher than their hover mode counterparts, averaging between 5.0 and 6.0 (Fig. 8).

Since five pilots had mainly fixed-wing aircraft experience and four pilots only had rotary wing experience, Fig. 9 breaks out the HQRs by aircraft experience to verify if this had an effect on the results. In Session 1, pilots with fixed-wing experience (N=4) rated the handling qualities between a 5.50 and 7.00 on average, giving the highest rating in the vertical reposition and hold HQTE and the lowest rating in the pirouette HQTE. Rotorcraft pilots (N=3) rated the handling qualities between a 4.33 and 6.33 on average, with the highest rating in the lateral reposition and hold and the lowest in the pirouette (Fig. 9). Note that due to various external factors, only three of the four total rotorcraft pilots were able to complete each of the three HQTEs. These data were also collected for Session 2 but are not presented here because of the small sample size.

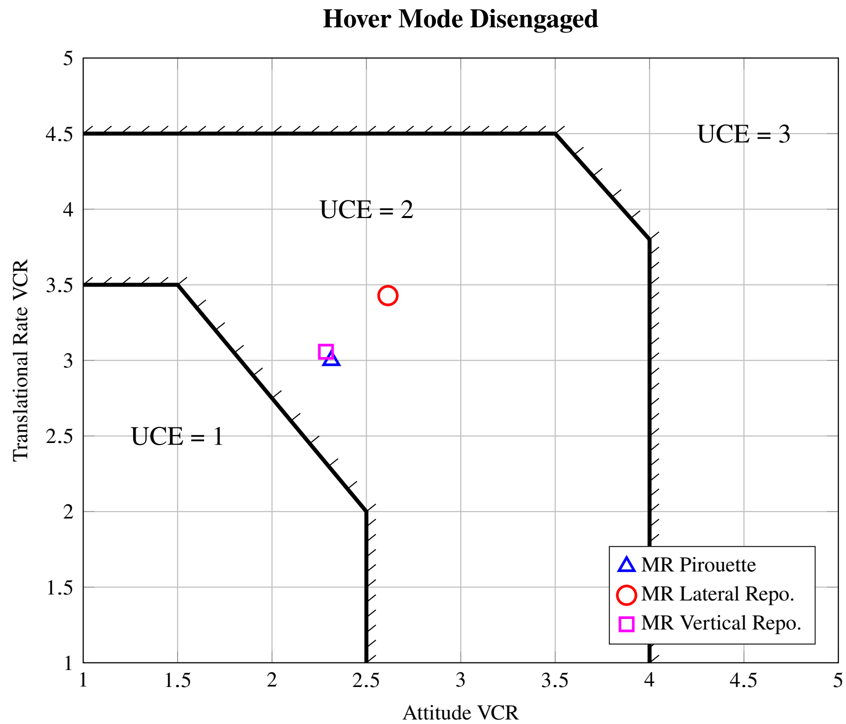

Fig. 6 UCE for MR and hover mode disengaged.

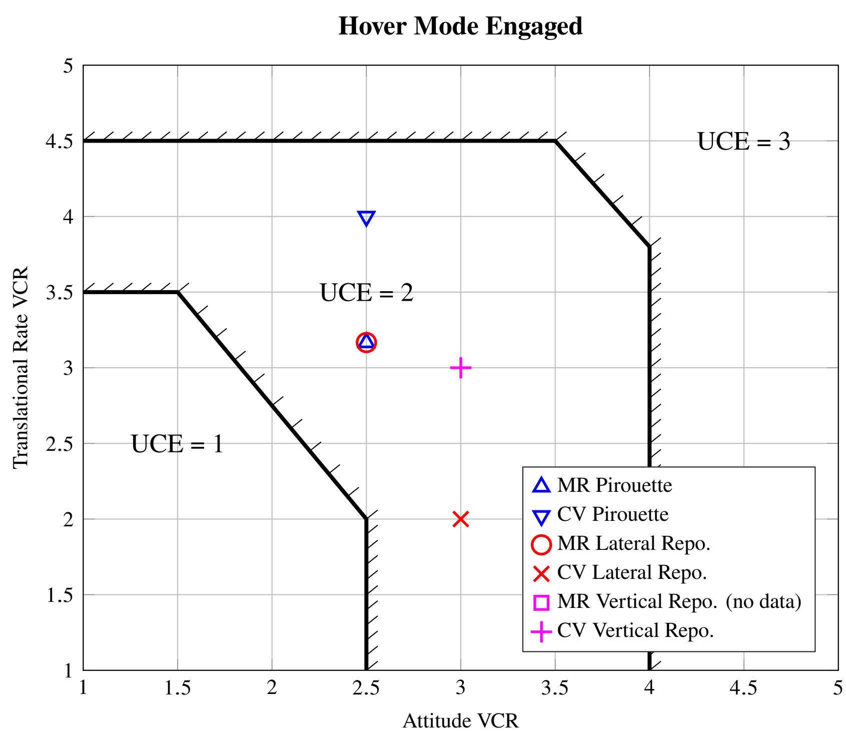

Fig. 7 UCE for MR and conventional visuals, and hover mode engaged.

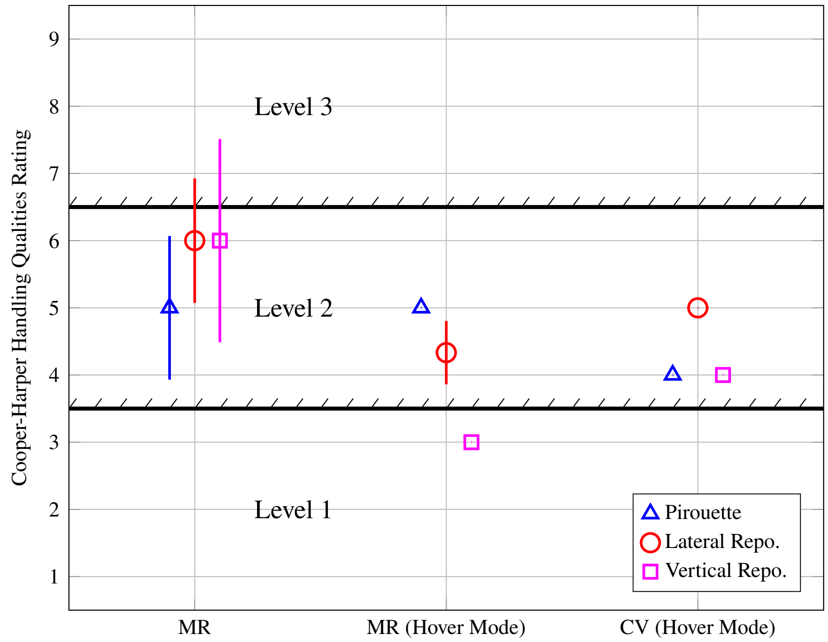

Fig. 8 Cooper-Harper handling qualities ratings.

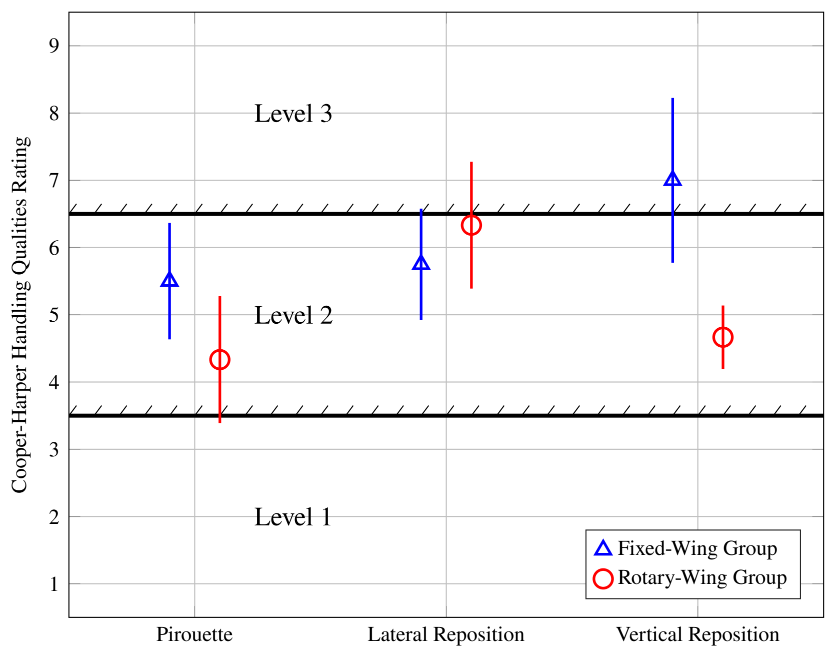

Fig. 9 Cooper-Harper handling qualities ratings by pilot experience.

IV. Discussion

A. SIMDUCE Evaluation

The main goal of this paper was to evaluate the UCE of an MR headset using the SIMDUCE paradigm. The UCE of the MR headset was determined to be Level 2 for all HQTEs. An opportunity arose to collect additional data in a second session under different experimental conditions, including conventional visuals. There was no significant difference in the SIMDUCE rating for any of the test conditions in this study. All HQTEs were given a UCE of Level 2 in both MR and CV and hover and non-hover modes (Fig. 6 and Fig. 7). Simulator motion also did not affect the UCEs.

Standard deviations of the VCRs were generally below 0.75, with the exception of the pirouette HQTE, which had a standard deviation higher than 0.75 in both test conditions with at least one participant (Table 6). Notably, unlike the other two tasks, the pirouette course was placed on a runway at the San Francisco International Airport, which lacked the rich visual references of nearby buildings present in the other tasks in downtown San Francisco. It is possible that the comparative lack of visual references outside the course may have affected the variation in pilot ratings, although this would require further experimentation to confirm.

Regardless, as desired during a SIMDUCE evaluation, the UCEs presented here were consistent with the corresponding HQ levels, which all fell between the upper edges of Level 1 and Level 2 on the Cooper-Harper scale. Hover mode seemed to improve the HQR of the vehicle slightly across all three tasks – most notably in the vertical reposition and hold HQTE – though the small sample size in Session 2 of this experiment makes it hard to claim any significance in this observation (Fig. 8).

When considering the HQRs based on pilot background, the lateral reposition and hold is the only HQTE in which fixed-wing pilots tended to rate the vehicle better in handling qualities than the rotorcraft pilots, by a margin of -0.58 points. On the other hand, during the vertical reposition and hold HQTE, fixed-wing pilots rated the handling qualities 2.33 points higher (i.e., worse) than rotorcraft pilots (Fig. 9). This trend seems to be consistent with the relative familiarity of the HQTE motion to that of a fixed-wing pilot (i.e. aircraft roll is more familiar than isolated heave), though the small sample size limits the significance of this observation.

The SIMDUCE paradigm was successfully applied to determine the UCE of an MR headset. Despite limitations to the actual results (as discussed in the following section), the SIMDUCE method seems an adequate starting point for the measure of the UCE of an MR system. There may, however, be certain properties of VR/MR visuals that warrant further consideration. In this study, for example, several pilots commented on the benefit of being able to expand their FOV by moving their head, whether by leaning to look past window edges or looking up or back through additional windows in the virtual vehicle cockpit while using the MR headset. In this case, this advantage was largely due to different geometries of the simulated LPC vehicle in MR and CV (i.e. smaller, limited windows in the physical R-Cab cockpit compared to the MR cockpit), but it highlights a major advantage of MR systems that SIMDUCE may fail to capture; that is, VR/MR setups are more flexible, and potentially provide more realism, in their representations of simulated aircraft than conventional simulators, especially for non-traditional aircraft geometries, such as an eVTOL. On the other hand, if these features of MR are indeed beneficial in completion of an HQTE, they should directly influence a pilot's ability to be aggressive and precise, making the SIMDUCE paradigm arguably satisfactory in its evaluation of MR systems. Unfortunately, the limited results of this paper make it difficult to do more than start a discussion on the matter.

B. Data Limitations

Data from Session 2 were most notably limited by the availability of participants and simulator time. Except for the pirouette HQTE in MR, every test condition in Session 2 of data collection had fewer than the three complete VCRs required by the SIMDUCE procedure. Furthermore, the standard deviation for at least one VCR exceeded 0.75 for both pirouette HQTEs in MR in Sessions 1 and 2 (Table 6). Therefore, only the results from Session 1 (N=7) for the lateral reposition and hold and the vertical reposition and hold HQTEs satisfied the SIMDUCE procedure requirements for a valid evaluation, at least from a sample size perspective [1].

Even when ignoring sample size limitations of the data collected in Session 2, it is not straightforward to directly compare the results of the MR and CV tests, making it hard to put the MR results into perspective. Due to software limitations of the available image generators, each visual system needed to use a different image generator with different versions of the visual database. Consequently, the visual references were not identical, with some buildings, for example, appearing in the CV visual database but not the MR database. The HQTE courses placed within these databases were identical, but discrepancies in the background visuals could have had a significant impact as well. Future experiments directly comparing the UCE for MR and CV should use identical image generator systems to allow for one-to-one comparison of the results.

The SIMDUCE protocol dictates that the test rotorcraft shall have response characteristics that rank no higher than a Rate Response-Type and shall be shown to have Level 1 handling qualities [1]. Since the LPC eVTOL aircraft used in this study was a conceptual NASA design, the actual HQ levels of the vehicle were unknown. In addition, not all controlled axes had a Rate Response-Type. In Session 1, with the hover mode disengaged, the HQ levels of the vehicle were determined to be Level 2 in the simulator environment (Fig. 8). Even though the handling qualities might be reduced because of the simulated environment, an attempt was made to improve handling qualities in Session 2 by running with the hover mode engaged. However, scheduling conflicts meant that this session was flown in a fixed-base simulator with an incomplete sample. Only a single participant completed the HQTEs with both MR and CV, and the remaining participants completed at most two HQTEs with MR, simply due to a lack of time. This limits the results of the SIMDUCE evaluation itself but doesn't affect the intent of this paper to present an initial discussion of the application of the SIMDUCE procedures to an MR headset, including key issues and lessons learned.

The aircraft and its control scheme used in this study were dictated by the availability of simulated vehicle models at the VMS. Most pilots were unfamiliar with the SVC concept of the simulated eVTOL vehicle, likely significantly impacting the results of this study. When providing comments on their VCR ratings, multiple pilots noted that their lack of familiarity with the control scheme detrimentally impacted their performance. These comments suggest that ratings may have been influenced, and potentially dictated, by participants' unfamiliarity with the controls, rather than their ability to be precise and aggressive alone. This limitation could potentially have been minimized by allowing pilots to get more familiar with the control scheme and performing more training runs for each HQTE.

Additionally, there were multiple comments expressing difficulty in providing a representative attitude VCR for tasks that did not require much attitude control, such as the vertical reposition and hold in hover mode. As a seemingly fair criticism, this observation potentially calls into question the arbitrariness of attitude VCRs given by other pilots when attitude control may not have been a factor.

C. Lessons Learned

Pilot familiarity, or lack thereof, with the simulated aircraft was found to have significant impacts on data collection and data accuracy. On multiple occasions, pilots failed to complete a given HQTE primarily due to negative habit transfer of the controls, thus decreasing the number of complete ratings on an already condensed schedule and minimal sample size. Moreover, for pilots that did complete the HQTEs, their lack of familiarity with the vehicle seems to have negatively impacted their subjective VCRs, as explained in the previous section where control deficiencies were often cited as justification for a poor VCR. The subjective nature of the VCR to begin with makes a pilot's comfort with the vehicle a substantial factor to consider. Therefore, steps should be taken to improve this comfort, either by using a simulated vehicle model with which pilots are more familiar, recruiting pilots whose prior experience more closely aligns with the simulated vehicle, or dedicating more time in the schedule for pilot training and familiarization with the vehicle prior to the start of data collection. Though the simulated vehicle was found to have a Level-2 handling qualities, this is just as likely a reflection of pilots' lack of familiarity with the control scheme and lack of training than it is the vehicle's actual handling qualities.

Additionally, greater effort should be made to isolate the rating of precision and aggression in the HQTE maneuvers in determining the VCRs, whether through follow-up questions or regular reminders of the guidelines, as pilots regularly reverted to rating visual cues instead of their ability to be precise and aggressive in the tasks. This approach may risk contaminating the pilot's ratings, however, so it should be done thoughtfully. Increasing the number, and variety, of the employed HQTEs may also help to refine the process for the pilot through repetition. Similarly, adding additional HQTEs that require more attitude control, for example, may facilitate pilots in providing better attitude VCRs, which was a noted limitation of this study.

Another important lesson gained from this study is the potential effect of visual complexity on consistent VCR results, as was possibly the case in the high standard deviations for the pirouette VCRs. It is advisable to place all task courses in the same location in a high-complexity area of the visual database to ensure that there are sufficient visual cues in the pilots' surroundings. This may be of additional importance in MR, where the pilot's FOV is possibly wider or less fixed than a conventional simulator. This recommendation also ensures that visual references external to the courses are consistent across all tasks. Repeating the experiment with this change could offer more data on whether the presence of multiple visual references in the visual database affected the standard deviation of the pirouette VCRs.

V. Conclusion

The SIMDUCE paradigm was successfully applied to determine the UCE of an MR headset in a motion-base simulator. Despite limitations making the interpretation of the UCE results of this study difficult, the SIMDUCE method seems an adequate starting point for the measure of the UCE of an MR headset to determine its validity in handling qualities evaluations. There may, however, be certain properties of MR headsets that are not captured by the SIMDUCE paradigm and warrant further consideration. To determine if VR/MR headsets are suitable for handling qualities evaluations, future work should focus on making a more thorough comparison between MR and CV with pilots adequately trained on the vehicle and its control scheme.

Acknowledgments

The authors would like to thank the simulation technicians and operators who helped with running the study. We would also like to thank the flight test pilots and engineers from the National Test Pilot School who participated. Finally, we would like to thank management of the Aerospace Simulation Research & Development Branch and the Aviation Systems Division at NASA Ames Research Center for supporting this study.

References

[1] Department of Defense, "Handling Qualities for Military Rotorcraft," Detail Specification MIL-DTL-32742, United States Department of Defense, March 2023. URL https://publishers.standardstech.com/content/military-dod-mil-dtl-32742.

[2] Nguyen, B., Sonnenfeld, N., Duruaku, F., and Jentsch, F., "Modern Training Practices: Studying the Use of Virtual Reality for Training Procedures in Flightcrew Training," Final report, U.S. Department of Transportation, Federal Aviation Administration, September 2023. URL https://rosap.ntl.bts.gov/view/dot/76580.

[3] 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/doi:10.1007/s13272-023-00688-5.

[4] European Union Aviation Safety Agency, "FSTD Special Conditions for the use of Head Mounted Displays (HMD) combined with a motion platform with reduced envelope," Information paper, European Union Aviation Safety Agency, 2023. URL https://www.easa.europa.eu/en/domains/aircrew-and-medical/flight-simulation-training-devices-fstd.

[5] Cooper, G. E., and Robert P. Harper, J., "The Use of Pilot Rating in the Evaluation of Aircraft Handling Qualities," Technical Note NASA TN D-5153, National Aeronautics and Space Administration, April 1969. URL https://ntrs.nasa.gov/citations/19690013177.

[6] 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.

[7] Theodore, C. R., Malpica, C. A., Blanken, C. L., Tischler, M. B., Berger, T., Lawrence, B., and Lindsey, J. E., "Effect of Control System Augmentation on Handling Qualities and Task Performance in Good and Degraded Visual Environments," *Journal of the American Helicopter Society*, Vol. 65, No. 4, 2020, pp. 1–14. https://doi.org/doi:10.4050/JAHS.65.042007.

[8] Malpica, C., Withrow-Maser, S., Aires, J., Schuet, S., Suh, P., Barnes, K., Hanson, C., Ruan, A., Altamirano, G., and Foster, J., "Handling Qualities of Multirotor RPM-Controlled Electric-Vertical Take-Off and Landing (eVTOL) Aircraft for Urban Air Mobility (UAM)," *Vertical Flight Society's 79th Annual Forum & Technology Display*, West Palm Beach, FL, 2023. https://doi.org/doi:10.4050/F-0079-2023-18079.

[9] Varjo, "Mixed Reality Headset for Professionals – Varjo XR-4 Series," https://varjo.com/products/xr-4/, 2025. Accessed: 5 August 2025.

[10] Life Performance Research, "LPVR-DUO: LPVR Middleware for In-Vehicle AR/VR visualization solution," https://www.lp-research.com/products/lpvr-virtual-reality/lpvr-duo/, 2025. Accessed: 5 August 2025.

[11] 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," Final Report DOT/FAA/TC-23/59, U.S. Department of Transportation, Federal Aviation Administration, Atlantic City, New Jersey, March 2024. URL https://rosap.ntl.bts.gov/view/dot/74025.

[12] 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*, AIAA, Las Vegas, Nevada, 2024. https://doi.org/10.2514/6.2024-4207.

