Visual Circuit Flying with Augmented Head-tracking on Limited Field of View Flight Training Devices

L. Le Ngoc[*] and R. S. Kalawsky[†]

Advanced VR Research Centre, Loughborough University, LE11 3TU, United Kingdom

The virtual reality technique of amplified head rotations was applied to a fixed-base, low-fidelity flight simulator enabling users to fly a visual flying circuit, a task previously severely restricted by limited field of views and fixed field of regard. An exploratory experiment with nine pilots was conducted to test this technique on a fixed-base simulator across three displays: single monitor, triple monitor and triple projector. Participants started airborne downwind in a visual circuit with the primary task of completing the circuit to a full stop landing while having a secondary task of spotting popup traffic in the vicinity simulated by blimps. Data was collected to study effects on flight performance, workload and simulator sickness. The results showed that there were very few significant difference between displays, in itself remarkable considering the difference in display size and field of views. Triple monitor was found to be the best compromise delivering flight performance and traffic detection scores just below triple projector but without some peculiar track deviations during flight and a less chance of simulator sickness. With participants quickly adapting to this technique and favorable feedback, these findings demonstrated the potential value of upgrading flight training devices and to improve their utility and pave the way for future research into this domain.

I.   Introduction

GENERAL aviation in the United States has the highest percentage (62%) of accidents occurring during the critical flight phases of takeoff, climb, approach, and landing, despite the relatively short duration of these phases compared to the entire profile of a normal flight.[1] This is mainly caused by the increased workload on flight crew and aircraft during these phases. For example, during both takeoff and landing, the flight crew must control the aircraft, change altitude and speed, communicate with air traffic control (ATC) and/or other aircraft, and maintain separation from obstacles and other aircraft. Aircraft systems are also stressed during takeoff and landing with changes to engine power settings, the possible operation of retractable landing gear, flaps, etc. While the aircraft is at low altitude during takeoff and landing, it is also most susceptible to hazards caused by wind and weather conditions. In Australia, midair collisions account for about 3% of fatal accidents involving general aviation aircraft, most (78%) of the midair collisions that have occurred since 1961 have occurred in or near the circuit area. This reflects the higher traffic density in this area. Statistics further showed that a high proportion of the collisions (35%) occurred on final approach or the base-to-final turn.[2]

Civil flight simulators are classified into Flight Simulation Training Devices (FSTDs) categories, as set to standards by various national aviation authorities,[3] ranging from instrument trainers with no visuals, PC based desktop flight training devices to large motion based full flight simulators airline crew certification. But there is no harmonization and hence no easy mapping between US and European standards for example.[4] This needs to be updated to solve regulatory changes, new aircraft types, new training methods and new technologies.[5] By mapping simulator requirements to support each clearly defined training task,

---

[*]PhD-Student, Advanced VR Research Centre, Loughborough University. Holywell Park, Loughborough, LE11 3TU, United Kingdom. AIAA Student Member. E-mail: L.Le-Ngoc@lboro.ac.uk

[†]Professor, Advanced VR Research Centre, Loughborough University. Holywell Park, Loughborough, LE11 3TU, United Kingdom. E-mail: R.S.Kalawsky@lboro.ac.uk

American Institute of Aeronautics and Astronautics

standardization and efficient procurement can be established,[3] a process which has been started, reducing the plethora of FSTD levels down to just seven, but not adopted worldwide yet.

By providing pilots with effective (simulation) training, future accidents can potentially be prevented from happening. However, flying schools and clubs do not currently have access to high-fidelity simulators with large, visual systems that provide a wide field of view (FOV) compared to commercial operations. However, low-fidelity simulators have proven to be capable of providing equivalent training benefits without additional cost if used appropriately.[6] On typical low-fidelity flight training devices, field of view restrictions limit the utility for training visual flying. Even experienced pilots find flying a precise visual pattern on a desktop simulator to be difficult, resulting in over- and under shoots during turns.[7] Smart integration of commercial-off-the-shelf technologies can provide technical solutions to bring affordable training solutions.[8]

One such technique/technology originating from virtual reality (VR) research is the implementation of amplified head-rotations[9] in order to overcome FOV and field of regard (FOR) restrictions. In this paper, we investigate usability and effects of amplified head rotations on visual flight maneuvers by comparing pilot performance, workload, simulator sickness human factors in a fixed-based, low-fidelity simulator equipped with single display, triple display and triple projector configurations. This ultimately paves the way for future transfer of training effectiveness studies.

II.   Non-constant head amplification

A.   Desktop applications

On desktop systems, the physical constraint posed by the display means users must still be able to view the display when rotating their head unlike helmet mounted displays with automatic head position view slaving. In the VR domain, the technique of amplified head rotations enables users through natural interaction the capability to increase utility and immersion in visual simulations.[10, 11] By exploiting the visual dominance effect between what the eyes perceive and the body sense of head position, pilots only need to make slight head rotations and the virtual scene is then panned at an amplified rate in that direction.[11] This effectively enables mapping of limited field of view systems with a fixed field of regard to true 360 degrees of coverage.

Amplified head rotations have been implemented on desktop systems[12] and fishtank VR setups[13] previously but this emerging capability has not been applied to flight simulators and scientifically reported. Also deployment on a multi-monitor desktop and large projection displays have not been fully documented.[10] Potential cybersickness and workload effects with this technique in an applied piloting task is also a relatively new area of research.[9]

B.   Amplification mapping

Static displays such as desktop systems restrict available user head rotation angles in pitch and yaw. Constant amplification is not suitable in this case: the maximum desired virtual head angle divided by the maximum real world yields a set gain, which although a user can reach the maximum desired virtual angle, this gain would be excessive near the centered head position causing jitter. A gain factor closer to unity for small movements is better here. Therefore, combining varying gains using control points in a custom profile is suggested, which is widely used in modern gaming applications.[14]

Table 1. Display configuration screen characteristics

Display	HFOV × VFOV (deg)	Physical size
Single monitor (SGL)	33 × 23	27"
Triple monitor (TRP)	100 × 23	3x 27" in flat landscape
Triple projector (PRO)	112 × 29	3.2m radius curved, 2m height

When different (physical) screen sizes and aspect ratios are used, separate profiles are required for each

American Institute of Aeronautics and Astronautics

amplified axis of head movement. The characteristics of the three display types we've mapped are listed in Table 1. Building upon prior work on single displays,[9] we've expanded mappings to cover the usage of triple monitors and even larger triple projectors based on trial-and-error fine-tunings in conjunction with empirical feedback from naive test volunteers. The amplification profiles for all three displays are shown in Figure 1.

In essence, a control point (x,y) coordinate of (10,2) translates $10°$ of real life head angle into $20°$ of virtual head angle. For example, with a known maximum value for virtual head angle required of $120°$ (allowing a pilot to visually look at his right past the wing) and the maximum viewable angle of a display to be $30°$ to the right as well, you would end up with the right end node for this particular yaw mapping to be (30,4). The remainder of the profile and its non-linearity is caused by the need for view stability around the center zone to avoid jitter and user comfort. Ramping up of the profile slopes is then needed in order to map usable display viewing angles with the desired virtual angles. What Figure 1 also shows is that for instance in case of mapping for the single monitor (SGL), the yaw profile has a steeper gradient and larger mapping values compared to pitch. This is due to the need of larger yaw viewing angles compared to pitch (pilot does not need to look at his feet or cockpit ceiling).
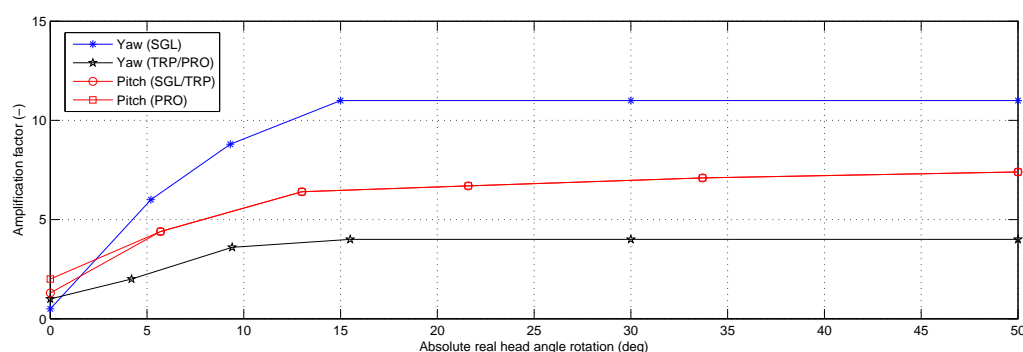

Figure 1. Amplification profile for single, triple monitor and triple projector displays

Pitch mapping for single monitor and triple monitors (TRP) was initially identical because the vertical viewing angles are the same on both configurations (TRP simply had 2 extra monitors to the side compared to SGL). Later, it was found out that there was a slight tweak necessary required at the first node-to-node gradient for increased comfort. The yaw profile for the triple monitor and triple projector (PRO) was not surprisingly similar as their respective horizontal FOVs were very close. What Figure 1 also theoretically suggests is that on the extreme edge of (50,7) for pitch (SGL/TRP) this would mean looking $50°$ to the right physically would translate into $350°$ of virtual angle. A simple limiter on maximum virtual angles simulating maximum human head rotation angles in place obviously eliminates this issue.

C.   Implications on adaptability

Studies have shown that this novel technique is both acceptable and useful: users preferred an average fixed amplification of 1.26x albeit while using a helmet mounted display with 60 degrees FOV citepJaekl2002. Subjects also found amplification natural: they noticed head movement attenuations significantly faster than amplifications.[15] Scene movement in the opposite direction of head rotations (i.e., amplifications) should be avoided though.[16] A recent study[11] evaluated the effects of different levels of amplification and the interaction between amplification and the display FOV. They investigated constant amplification factors up to 3x level. Subjects had a visual scanning and a counting task to perform. For visual scanning, FOV changes were immediately noticed whereas amplification was only detected by half of the participant. It also suggested that visual scanning tasks can be performed without a performance penalty. In the counting task, nobody was able to detect amplification being enabled but performance was affected negatively despite statistics showing no main effect.

Simulator sickness, also popularly known as cybersickness today, defines the possible symptoms associated with simulator and virtual environments.[17] This often results from exposure to simulated environments

on the ground such as flight simulators, driving simulators, and virtual environments.[18] Factors of interest for amplified head-rotations are when different FOVs are being tested as FOV is a major contributing factor to simulator sickness.[19] Wide field of views stimulate peripheral vision more, inducing greater vection. This increased stimulation potentially increases the chance of simulator sickness.[20] It must be noted that vection is an important visual cue in the aviation domain especially in manoeuvring tasks. Therefore, when considering single versus triple display setups, it must be kept in mind that the increased FOV might come with a sickness penalty if used for long durations even without view augmentation.

D.   Experimental task selection

To evaluate the utility and potential expansion of the visual training tasks amplified head rotations add to low-fidelity simulators, applicable tasks must be identified first which form the basis for exploratory human-in-the-loop trials. Visual flying tasks have been widely used in visual cueing, pilot performance and simulator fidelity studies.[21, 22] The air combat domain is predominantly showcased to benefit the most with head augmentation in the gaming industry.[14] One versus one Basic Fighter Maneuvering has been studied in the past where subjects dogfighted each other in aerial combat with the object to put the target aircraft within a 20 degrees frontal cone as quick as possible to simulate a win.[23] This experiment task would be very suitable due to its air combat origins, but it is not feasible to conduct due to lack of access to qualified fighter pilots as test subjects. Formation flying and teamwork in a mission scenario environment is also a popular performance measurement test in the military domain.[24] Even civilian formation flying is quite a complex task in itself and again the ability to recruit sufficiently qualified pilots to support such an experiment was prohibitive.

Reflecting on the general aviation mishaps near the airport area and the relevant flight phases/piloting tasks involved as discussed in the introduction, a distillation of this case was considered the most suitable to implement as a pilot study. As such, an experiment was designed to test a visual circuit flying scenario with traffic detection to replicate this. Visual circuit flying as a flying task in experiment has been done before[25] and since flying a visual pattern is a prerequisite during basic flight training, the pool of candidates as participants widens up to more obtainable numbers. Furthermore, traffic detection is a visual scanning task which has also been documented as a suitable task for amplified head-rotation applications.[11]

III.   Experiment

A.   Method

To study the effects of amplified head rotations in combination with larger display sizes and FOV, we compared its integration with a single monitor, triple monitor and triple projector setup (Table 1). The goal is to evaluate usability of augmentation and how it affects piloting performance, workload and simulator sickness.

1.   Participants

Nine male participants were recruited from the local flying club at East Midlands Airport (EGNX) for this experiment. All were licensed pilots or currently in flight training, with a mean age of 27 years and an average experience of 141 flight hours. Table 2 has the age and flying experience of individual pilots listed. None of the participants had any prior experience flying any of the experimental simulator configurations.

2.   Apparatus

The experiment was conducted in the Virtual Engineering Laboratory at the Advanced VR Research Centre (AVRRC) of Loughborough University. The laboratory consisted of an experiment area with a fixed-base simulator (Figure 2), control desk and a projection room area next door. For the single and triple monitor setup, the pilot sits in a (fighter type) cockpit in a darkened environment. Directly in front of the pilot are three 27″ 1080p LCD screens in landscape mode. Together, these form a single, large display area of (bezel-corrected) 6160 × 1080 pixels. The left and right displays were disabled for the single monitor configuration. All displays projected a virtual cockpit with an outside visual world altogether. Aircraft

American Institute of Aeronautics and Astronautics

Table 2. Participant characteristics

Pilot	Age	Hours	Aircraft types
1	22	365	C152/172, PA28/38
2	25	36	G115
3	20	65	C172, AT01, SR20
4	19	20	G115
5	22	90	C152/172, PA28/38, T67
6	58	575	PA28
7	19	51	C150, PA28/38
8	26	60	C152, PA28
9	25	12	Glider, C152, G115

control was available via a right-handed, mechanically, sprung loaded, passive center stick (elevator and aileron). The throttle quadrant was available on the left side panel with only the throttle levers being used. An infrared head-tracker (TrackIR 5) measured participant head position/orientation at 120Hz with a 9ms latency which controls the virtual pilot head camera in the simulation. The amplification profiles of Figure 1 were used allowing only pitch and yaw control of the virtual pilot head camera for viewing the virtual environment.

(a) Fixed-based Cockpit Simulator

(b) Immersive Projector Lab

Figure 2. AVRRC Research Flight Simulators

3. Aircraft characteristics and experiment conditions

X-Plane 9.70 by Laminar Research was used with the default CirrusJet aircraft model. The aircraft was preset in the landing configuration with flaps and gear down, trimmed for level flight at 1500 feet at 85 knots indicated airspeed. Participants did not have to operate the trim, rudders (yaw damper provided instead), brakes, flaps and gear. The aircraft would automatically deploy wheel brakes upon touchdown. Furthermore, there was no wind or turbulence simulated in this experiment.

4. Independent variables

The experiment had one within-subject variable, namely: simulator display configuration (DISP) with three levels: single monitor (SGL), triple monitor (TRP) and triple projector (PRO). The horizontal field of view (HFOV) and vertical field of view (VFOV) characteristics of each configuration was previously tabulated in Table 1.

5. Design

Participants were tasked to take control of an airborne aircraft starting downwind in a visual traffic circuit pattern, as shown in Figure 3. The primary task was therefore to complete the circuit pattern by flying

American Institute of Aeronautics and Astronautics

downwind and performing two turns (indicated by Turn 1 for downwind-to-base and Turn 2 for base-to-final) prior to a full-stop landing. To avoid participants memorizing timings to initiate the turns, the aircraft is randomly positioned at 0, 3 and 6 second time intervals on the downwind leg.
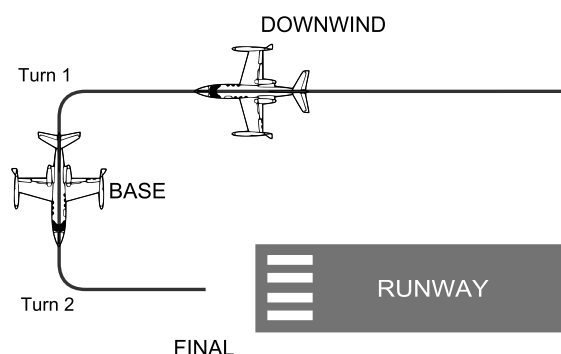

Figure 3. Downwind-base-final portion of visual flying circuit

In the terminal airport area, pilots need to maintain a visual scan for other traffic to maintain their situational awareness and prevent mid-air collisions. This action therefore has been assigned as a secondary task in this experiment, providing objective indication for workload. Traffic is simulated by means of blimps with a strobing anti-collision light appearing randomly in a designated area, dubbed spawn boxes, as illustrated in Figure 4. Each spawn box represents an airspace of 2.0 by 0.5 nm wide and covers 1000-2000 feet above ground level in the vicinity of the runway. A total of three blimps appeared in sequence throughout each experiment run, participants were only briefed that there will be more than one and that they had to press a button on the flightstick whenever they had detected one.

The blimp sequence was as follows: on downwind at the start of the run, the aircraft would trigger a timer after which the first blimp would spawn in area 1 at a randomly selected time interval. Once further downwind, the first blimp would disappear and another timer triggered for when Blimp 2 would spawn in it's box. A similar procedure applied to Blimp 3, however the final blimp had two box areas to randomly choose to spawn in. Timings for the blimps were selected as such that Blimp 1 would be detectable on the downwind leg only, Blimp 2 near the end of the downwind left through to base leg. Blimp 3 would only spawn whilst transiting from base leg to touchdown. Even though participants might have remembered the spawn sequence, the timings in combination with the spawn boxes still made it sufficiently challenging so participants had to actively visually search in order to detect a blimp.
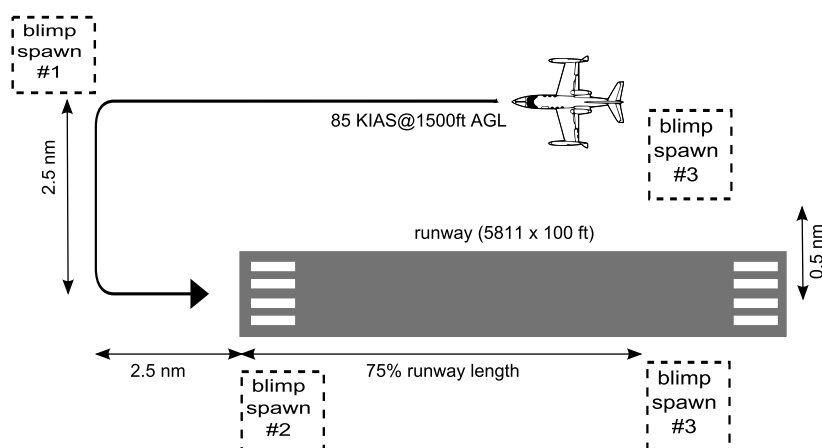

Figure 4. Experiment scenario with traffic spawn areas

American Institute of Aeronautics and Astronautics

6. Procedure

The experiment took each participant about two hours to complete and consisted of two phases. The first phase was considered to be a *training phase* in which the participants got an oral briefing and got accustomed to landing with the various simulator configurations in about 5 runs. This was to eliminate learning curve effects so they could converge to set a consistent performance for measurements later on. The second phase was the *measurement phase* consisting of 6 runs to be flown per display type. A single measurement run lasted around three minutes. After completing the runs on each display, the participant filled out a NASA TLX workload sheet and a SSQ form. The order of DISP that each participant flew was randomly allocated.

Because this study explores the performance of particular participants and compares them within their group and with themselves whilst the task (complexity) remains the same, NASA TLX is best suited in this manner,[26] it also being the industry standard workload evaluation tool in aerospace research. To exclude participants with high existing SSQ symptom levels, or having to dismiss participants who had a severe headache when they came to the experiment,[27] all participants were asked to be well rested before flying each DISP experiment to ensure their pre-experimental SSQ scores were all nil. This in effect meant the post-experimental scores became the absolute score.

There was a short break in between DISP runs and this in conjunction with the amount of runs/time spent per display had been deliberately chosen beforehand to prevent fatigue and boredom from occurring as well as full recovery from any prior SSQ effects of the previous DISP. Finally, participants filled out a general questionnaire at the end of the experiment indicating preferences amongst other feedback.

7. Dependent measures

PERFORMANCE.   To obtain flight performance data that were sensible as measurement variables, the entire flight was divided into separate flight segments. Each flight segment then had its own performance measures as following:

1. *Downwind*

   (a) *Altitude deviation (meters):* root mean square (RMS) of altitude deviation from 1500 feet during downwind leg

   (b) *Cross-track error (meters):* RMS lateral deviation from ground track 2.5nm parallel to runway centerline

   (c) V_{REF} *error (knots):* RMS deviation from approach reference indicated airspeed of 85 knots

2. *Turn 1 (Downwind-to-base)*

   (a) *Maximum bank angle (degrees):* root mean square (RMS) of maximum bank pilot used to perform the turn to base

   (b) *Turn starting distance from runway threshold (meters):* provides comparison of how far downwind pilots started the turn to base

   (c) V_{REF} *error (knots):* RMS deviation from approach reference indicated airspeed during Turn 1

3. *Base leg*

   (a) *Altitude deviation (meters):* root mean square (RMS) of altitude deviation from 1500 feet during base leg

   (b) V_{REF} *error (knots):* RMS deviation from approach reference indicated airspeed

   (c) *Heading error (degrees):* RMS of base leg magnetic heading deviation from ideal orthogonal runway heading

4. *Turn 2 (Base-to-final)*

   (a) *Maximum bank angle (degrees):* root mean square (RMS) of maximum bank pilot used to perform the turn to final

American Institute of Aeronautics and Astronautics

   (b) *Turn starting distance from runway threshold (meters):* provides comparison of how far downwind pilots started the turn to final

   (c) *V_{REF} error (knots):* RMS deviation from approach reference indicated airspeed during Turn 2

5. Initial line-up

   (a) *Runway alignment error (degrees):* heading deviation from runway heading upon roll-out from base-to-final turn

6. *Final approach*

   (a) *V_{REF} error (knots):* RMS deviation from approach reference indicated airspeed

   (b) *Glideslope deviation (deg):* RMS of vertical deviation from 3 degree glideslope

   (c) *Cross-track error:* RMS lateral deviation from (extended) runway centerline

7. *Touchdown location*

   (a) *Longitudinal (meters):* touchdown point measured along runway length from threshold

   (b) *Lateral (meters):* touchdown point distance from runway centerline

WORKLOAD. The workload was measured using objective and subjective data. Objective data was obtained from the secondary task performance of blimp spotting using two measures: blimp detection rate and detection time. Detection rate is the percentage amount of blimps spotted for each of the three blimps. Detection time is the elapsed time in seconds between a blimp spawn and button press. Subjective workload measures provided by the results of the administered NASA TLX forms.

SIMULATOR SICKNESS. Simulator sickness was measured by the results from the SSQ forms.

8. *Final questionnaire*

A questionnaire regarding preferences and opinion on the displays was given to the pilots after they have finished all experimental runs. Pilots were asked to indicate what cues they used during specific phases of the landing. They also had to rank the displays by providing a preference to use which display during a phase of the flare.

B. Hypotheses

The performance and workload were hypothesized to be the best for PRO and worst for SGL, as it was expected that larger display sizes and FOV would benefit flying and visual scanning accordingly. Simulator sickness, however, was hypothesized to be least prevalent on TRP, followed by PRO and the most symptomatic on SGL. TRP was expected to be better than PRO for SSQ as the larger display size of PRO would give stronger visual stimuli which in conjunction with the moving view during flying would potentially give more adverse symptoms. SGL was ranked the worst for SSQ since it was the most restrictive for viewing, compounded by larger gains for the amplification mapping, potentially being the most unnatural.

IV. Results and discussion

The participants generated a total of $(9 \times 18 =)162$ measurement runs. The performance dependent variables were analyzed using univariate Analysis of Variance (ANOVA) per flight segment. With the expectancy that multiple dependent variables per segment would be correlated, a Multivariate ANOVA (MANOVA) was conducted to protect against the Type I error rate.[28, 29] The assumptions required for ANOVA/MANOVA tests regarding normally distributed data could not always be satisfied. However, with homogeneity of variance assured and equal sample sizes, ANOVA is still robust enough and has been reported/used as such in numerous studies.[30] Results using Wilk's lambda was taken in case both Box's M test for homogeneity of variance-covariance and Levene's test for homogeneity of error variances were satisfied, else Pillai's trace test was used instead.[29]

American Institute of Aeronautics and Astronautics

A. Performance

1. Downwind

MANOVA (Pillai) reported a significant difference on downwind performance between DISP, $F(6,316) = 2.483, p = 0.023$. Follow-up univariate ANOVAs with Bonferroni correction accepting a reduced significant level ($p = 0.0167$) found that the downwind cross-track error was significant between DISP, $F(2,159) = 4.297, p = 0.015$. Both pattern altitude, $F(2,159) = 2.634, p = 0.023$, and Vref deviations, $F(2,159) = 0.840, p = 0.434$, were not significant. Figure 5(a) shows the boxplot for the downwind cross-track error across the three DISP levels, indicating the means and confidence level were higher for the triple projector compared to both monitor setups. Although the scenario is identical across all DISP with the aircraft trimmed for level flight, participants tended to drift to the right more on the triple projectors, they were not aware of this themselves even when they were 5 degrees off the original heading. This could be explained due to the physical commonality of SGL and TRP where for the downwind segment, the center screen sufficed to provide all the required visual cues. Flying on the triple projector would therefore be more different and the larger physical size providing stronger compelling visuals coupled with the head movement to cause this drift. However, a post-hoc Tukey test (with alpha correction $p = 0.0056$) revealed that the only significant difference yet was between SGL and PRO ($p = 0.017$). SGL versus TRP ($p = 0.833$) and TRP versus PRO ($p = 0.074$) were insignificant.
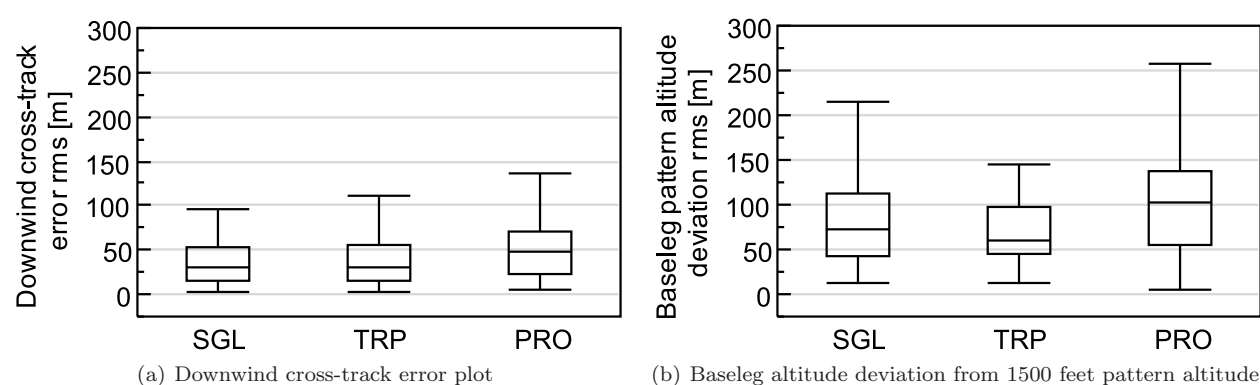

(a) Downwind cross-track error plot   (b) Baseleg altitude deviation from 1500 feet pattern altitude

Figure 5. Boxplots of cross-track and altitude deviations for two significant flight phases

2. Turn 1 (downwind-to-base)

Wilks' lambda reported no significant difference on Turn 1 between DISP, $F(6,316) = 1.610, p = 0.144$. Although the ground tracks seemed to indicate differences in Turn 1, the scatter and variance was sufficient to cause statistical insignificant conclusions.

3. Base leg

Wilk's lambda found a significant difference between DISP, $F(6,314) = 2.683, p = 0.015$. This was followed-up with ANOVAs (Bonferroni correction) at a reduced significant level ($p = 0.0167$). Baseleg heading, $F(2,159) = 3.229, p = 0.042$, and Vref deviations, $F(2,159) = 1.244, p = 0.291$, were not found to be significant. Base leg altitude deviation was found to be significant, $F(2,159) = 4.528, p = 0.012$, shown in Figure 5(b). Despite this finding, post-hoc Tukey test (alpha corrected $p = 0.0056$) did not detect significant pair-wise differences: SGL-TRP ($p = 0.490$), SGL-PRO ($p = 0.160$) and TRP-PRO ($p = 0.009$), probably again due to limited sample size. This can also be explained by individual piloting styles as some participants liked to lose more altitude during base to avoid having too much potential energy left to have to bleed off during final approach.

American Institute of Aeronautics and Astronautics

4. Turn 2 (base-to-final)

Wilks' lambda reported no significant difference on Turn 2 between DISP, $F(6, 314) = 0.799, p = 0.572$.

5. Initial line-up

One-way ANOVA showed a borderline significant difference for the runway alignment error, $F(2, 159) = 2.825, p = 0.062$. This is visualized in Figure 6, suggesting a trend of increasing positive angular errors from SGL to TRP to PRO. This meant that participants would tend to overshoot the runway extended centerline, being most prevalent on PRO.
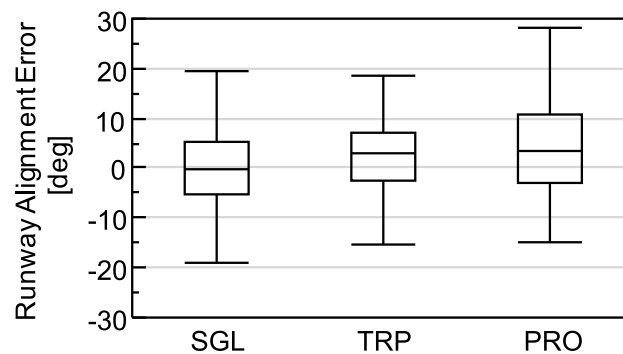

Figure 6. Boxplot of the runway alignment error during final approach

6. Final approach

MANOVA (Pillai) indicated significant difference between DISP during final approach, $F(6, 316) = 2.820, p = 0.011$. Following-up with univariate ANOVAs (Bonferroni corrected) did not yield further explanation of this fact. Glideslope deviation ($F(2, 159) = 1.302, p = 0.275$), approach cross-track error ($F(2, 159) = 2.983, p = 0.053$) and Vref deviation ($F(2, 159) = 2.896, p = 0.058$) were yet to be found no different between DISP.

7. Touchdown

MANOVA (Pillai) for touchdown location found no significant difference, $F(6, 318) = 1.829, p = 0.123$, all DISP yielded similar results upon landing. Just the task of a straight-in approach and landing is in itself a complex and challenging, with past studies favoring larger (horizontal) FOVs in particular for peripheral vision, albeit recent ones have proven that it is possible to achieve equivalent results across limited FOV displays for landing if the pilots are aware of the strategies available to them.[31]

B. Workload

1. NASA TLX

To compensate for different subjective rating ranges per participant, all resulting TLX workloads were converted into z-scores per participant[32] to allow proper comparisons between the participants. One-way ANOVA of the TLX workload z-scores did not result in any significant differences between DISP, $F(2, 24) = 1.322, p = 0.285$. ANOVA for any of the TLX workload demand subscales did not find any significant differences either between DISP. Thus subjectively according to participants, head-tracking augmentation could be implemented across these three DISP levels with no impact on the TLX workload nor individual workload demands.

2. Secondary task

Participants did occasionally fail to spot blimp(s) during a measurement run. Table 3 shows the count of blimps that were missed per DISP for each of the three blimps that would spawn. The percentages in

American Institute of Aeronautics and Astronautics

brackets is the corresponding detection rate overall for each blimp per DISP. Pearson's chi-squared test was then applied to check if each of the blimp spot counts were significantly different due to DISP. This was not the case with Blimp 1 ($\chi^2 = 2.971, p = 0.226$) nor Blimp 2 ($\chi^2 = 3.419, p = 0.181$) but Blimp 3 proved to be significant ($\chi^2 = 18.459, p < 0.01$). The lowest detection rate occurred on SGL whereas PRO resulted in the highest. This could be expected beforehand since the triple monitor and projector offered increasing larger viewing areas to perform the visual scan hence an easier time to detect the blimp.

Table 3. Number of missed blimp spottings and corresponding detection rates in percentages

DISP	Blimp 1	Blimp 2	Blimp 3
Single monitor	7 (96%)	2 (99%)	26 (84%)
Triple monitor	5 (97%)	5 (97%)	14 (91%)
Triple projector	2 (99%)	1 (99%)	6 (96%)

Statistical tests were conducted separately on each blimp detection time to see if DISP influenced the time it took to detect the blimps. Figure 7 plots the detection times for each blimp per DISP. Overall mean times considered, Blimp 1 was the quickest to be spotted followed by Blimp 3. Blimp 2 proved to take the longest time for participants to detect.
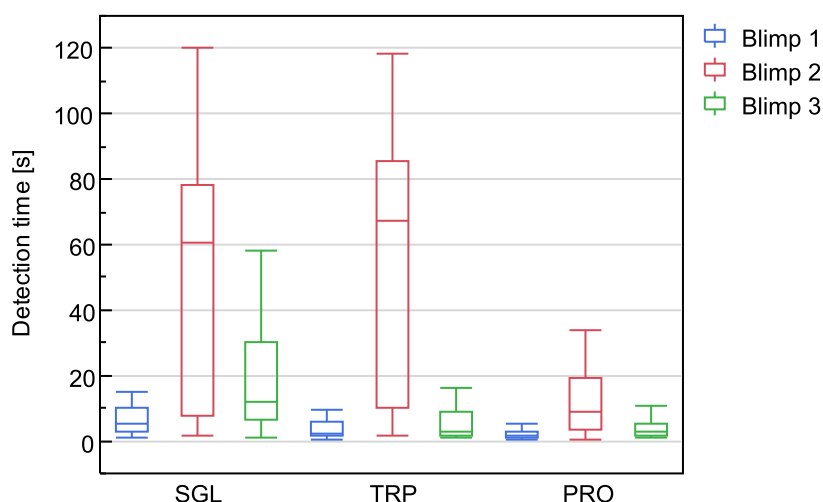

Figure 7. Blimp detection time

Although attributed to spawning within the central viewing area during the low workload downwind leg so it was relatively easy to spot across all DISP, ANOVA showed no significant difference yet in detection times for Blimp 1, $F(2, 145) = 2.324, p = 0.101$. Since the data for Blimp 2 times was unsuitable for ANOVA and could not be transformed accordingly, a non-parametric Kruskal-Wallis was used instead. This gave a significant result for Blimp 2, $H(2) = 20.053, p < 0.01$. Due to unequal groups in sample size, further post-hoc investigation was done via Tamhane's test (with significance levels again corrected per Bonferroni to $p = 0.0167$). This showed SGL and TRP to be inconclusive ($p = 0.553$) whereas SGL-PRO and TRP-PRO were both significantly different (both $p < 0.01$). Participants detected Blimp 2 the quickest on the triple projector DISP. ANOVA of Blimp 3 revealed a significant difference, $F(2, 113) = 5.105, p = 0.008$. Post-hoc (Tukey) tests (with corrected $p = 0.0167$) revealed SGL-TRP ($p = 0.013$) and SGL-PRO ($p = 0.014$) to be significantly different, this was not the case for TRP-PRO ($p = 0.990$). Besides Blimp 3 requiring searching a larger portion of the sky, this had to be done concurrently during the final approach which via this measure has proven to yield the highest workload. A further investigation post-experiment into the full flight data recordings confirmed this: in cases where pilots missed Blimp 3, they were indeed so preoccupied with landing that they were completely focussed on the runway straight ahead and did not do any visual scanning at all as the virtual camera did not move about!

American Institute of Aeronautics and Astronautics

C.   Simulator sickness

The results of the administered SSQ showed that some participants reported simulator sickness symptoms during this experiment. Table 4 provides the number and corresponding percentage of participants out of the total of nine that got symptoms. Oculomotor symptoms appeared to have been the least reported symptom of the three categories. Pearson's chi-squared test was then applied to check if each of the individual symptom categories were significantly different due to DISP. This was not the case with nausea $\chi^2 = 2.077, p = 0.354$, oculomotor $\chi^2 = 0.355, p = 0.837$, and disorientation $\chi^2 = 0.297, p = 0.862$.

Table 4.   Number and percentage of participants getting symptoms

DISP	Nausea	Oculomotor	Disorientation
Single monitor	3 (33%)	4 (44%)	4 (44%)
Triple monitor	5 (56%)	3 (33%)	5 (56%)
Triple projector	6 (67%)	3 (33%)	5 (56%)

The SSQ total score as well as the factors score are both shown in Figure 8. Statistical analysis (ANOVA) indicated no significant differences for the total SSQ score between DISP, $F(2, 24) = 0.342, p = 0.713$. However, the bar charts for the mean values suggested there was a trend of reduced simulator sickness going from single monitor on one end to triple projector on the other end. ANOVA of the SSQ factor scores did not reveal any significant differences between DISP. Again, the bar charts suggested a reduction in oculomotor factor from single monitor to triple projector. Furthermore, the indication of higher oculomotor factor means compared to nausea and disorientation is not surprising since this experiment focuses on the visual display system with head-tracking.[33]
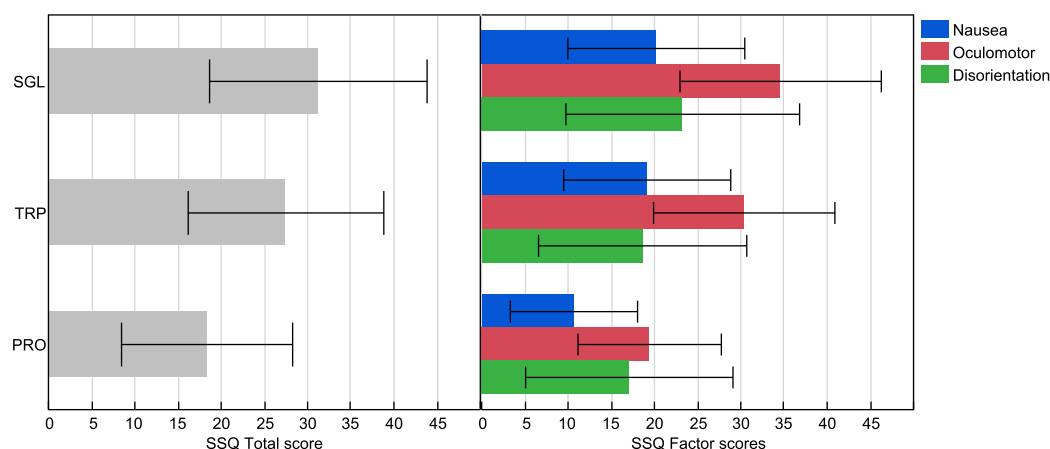

Figure 8.   Mean SSQ Total and Factor scores with standard error bars

D.   Correlation analysis

Due to the data non-normal distribution of the dependent performance measures, a non-parametric Spearman test was used for correlations, again for flight segments. However during applicable flight segments, the blimp detection times from the objective workload measures data were combined to cross-check workload indications.

The test showed that there was no correlation during the downwind leg between the performance dependent measures themselves and Blimp 1 or Blimp 2 detection time. For Turn 2, the turn distance from runway threshold had a significant positive correlation to the maximum bank angle ($r = 0.479, p < 0.01$) and reference airspeed deviation ($r = 0.173, p < 0.05$). This implied that if the downwind turn to base was started further away from runway threshold, pilots would use larger bank angles as well deviate more from

American Institute of Aeronautics and Astronautics

the reference airspeed. There was no correlation between Turn 2 maneuvering and detection time of Blimp 2.

On base leg, Blimp 2 and Blimp 3 detection times were added to the existing three performance measures. The test showed that base leg altitude deviation had a significant positive correlation to reference airspeed deviation ($r = 0.223, p < 0.01$) and Blimp 3 detection ($r = 0.383, p < 0.01$). This indicates that larger altitude deviations were coupled with larger airspeed deviations, which is fundamental in flight dynamics. Furthermore, the larger altitude deviations also resulted in a larger workload to correct for hence the longer it took to detect Blimp 3.

Turn 2, from base-to-final with Blimp 3 added, the test showed that maximum bank angle had a significant negative correlation to turn start distance from centerline ($r = -0.354, p < 0.01$), positive correlation to airspeed deviation ($r = 0.165, p < 0.01$) and Blimp 3 detection time ($r = 0.191, p < 0.01$). This is comprehendible as pilots who started their final line-up turn closer to the runway (extended) center line had less turning room to do so, hence leading to larger bank angles used, meaning larger airspeed deviations had to be corrected, in short a higher amount of workload leading to a longer time before spotting Blimp 3.

For final approach and touchdown, the three approach variables were combined with initial runway alignment error, touchdown location and Blimp 3 detection time. The test found a significant positive correlation between glide slope deviation and cross-track error ($r = 0.309, p < 0.01$). This means that pilots who were struggling with maintaining the glide slope tended to also have more issues with lining up with the runway. Furthermore, glide slope deviation had a significant correlation to both touchdown locations, for longitude ($r = 0.732, p < 0.01$) and latitude ($r = 0.370, p < 0.01$). This meant pilots who were strayed too far from the intended glide slope would land longer on the runway with a larger chance of being off-center. Further lateral touchdown location correlation was for cross-track error ($r = 0.562, p < 0.01$) and longitudinal touchdown location ($r = 0.155, p < 0.01$), meaning that the more misaligned the pilot was on the runway, the more off-center the resulting touchdown would be as well as tending to be further down the runway.

Finally, the test between the TLX subjective workload and SSQ scores did not reveal any correlations. So any perceived workload was not linked to any experienced simulator sickness symptom(s).

E. Final questionnaire

A Friedman's ANOVA test was conducted to determine whether participants had a differential rank ordered preference for the three levels of DISP. Results indicated that there was a significant rank preference, $\chi^2 = 9.250, p = 0.010$. The median for triple projector was being first choice, triple monitor as second choice and single monitor the least preferred. Post-hoc comparison of the preference rankings was conducted using a Wilcoxon signed-rank test to determine if these medians could be substantiated by significant pair-wise differences. A Bonferroni correction was applied to account for three tests,[29] thereby reducing the statistical significance level of acceptance to $p < 0.0167$. Results of this indicated that there were significantly more favorable rankings of triple monitor over single monitor ($Z = -2.640, p = 0.008$), borderline favoring of triple projector over single monitor ($Z = -2.309, p = 0.021$) but no significant preferences between triple monitor and triple projector ($Z = -302, p = 0.763$).

The pilot ratings of the aircraft dynamics and outside visuals are shown in frequency plots of Figure 9. All participants indicated that the aircraft dynamics and simulator apparatus were not detrimental in any way to carry out the experiment task, this was to justify the focus being on comparing the effects due to visual/head-tracking without adverse influence would the aircraft require too much human resources to just control. This is reflected by the Cooper-Harper ratings between 1-3 for all but one respondent. These ratings fall under the Level 1 US Military Specifications for Handling Qualities verifying that indeed the flying qualities in this setup were clearly adequate for the mission flight phase at hand.[34]

Participants were very pleased with outside visuals, with the majority scoring it as realistic or very realistic. Positive comments in particular were about the high resolution and detailed virtual airport environment such as individual modeling of runway lights. Half of the participants were initially skeptical regarding the usability of head-tracking augmentation but were pleasantly surprised at its utility and short learning time to adapt.
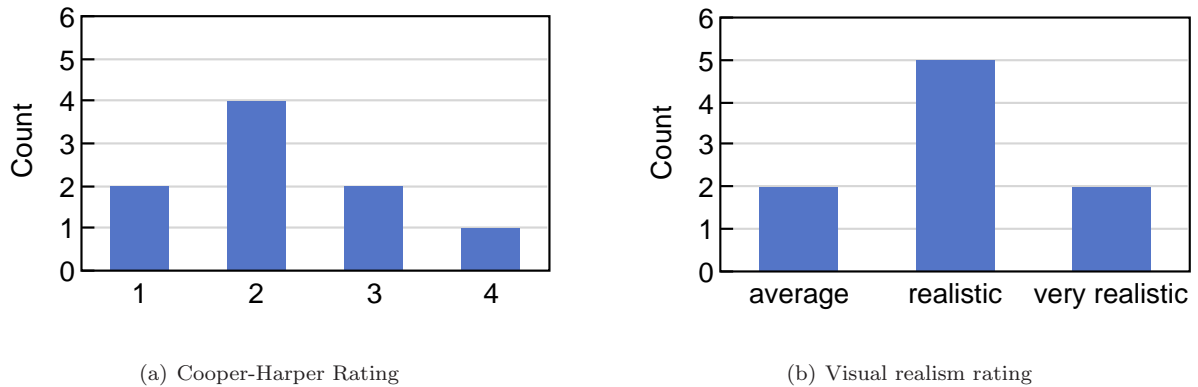
American Institute of Aeronautics and Astronautics

(a) Cooper-Harper Rating

(b) Visual realism rating

Figure 9. Pilot ratings of aircraft dynamics and outside visuals

V.   Conclusions and recommendations

Mapping profiles of head-tracking augmentation to a triple monitor and triple projector visual displays for a fixed-based flight simulator has been achieved, building on prior research with a single monitor. This paper has shown that amplified head rotations can be successfully applied to flying visual maneuvers even on various field of view displays by adding unlimited field of regard viewing capability.

The results of the experiment showed that there were very few statistically significant differences between the three displays in terms of flight performance for each flight segment. In this regard, it is remarkable that the smallest display (single monitor) with a very limited size and field of view was capable of near equivalent performance to the triple monitor and projector. Display differences that were found for the base leg altitude were attributed more to individual piloting technique rather than display. The triple projector did have a peculiar effect on participants, with them drifting right on the downwind start and a tendency to overshoot the runway centerline after the base-to-final turn.

In terms of workload, it is encouraging for any new technology implementation to being easily accepted and quickly learned by users. The subjective workload ratings indeed confirm this all participants rated all three displays equivalently. The objective secondary task of blimp detection, however, resulted in ranking the triple projector having the lowest workload yielding the highest detection rates and quickest detection times. Simulator sickness was surprisingly not significant yet across the displays, but the trend yielded an interesting finding: the single monitor had the least symptoms whereas the triple projector suffered the most sickness effects.

Overall, triple monitor edged out triple projector so this configuration could be recommended to achieve the best compromise between flight performance, workload and simulator sickness for flying a visual circuit. Triple projector lost out because it exhibited a peculiar drift during flying the circuit and was trending to be the worst cause of simulator sickness. For flight training and mission rehearsal, anomalous behaviors on flight performance and potential cause for simulator sickness is unacceptable whereas a slightly higher workload could be accommodated considering the added training value.

Developments like this are very attractive for integration into lower-cost, low fidelity flight simulators to provide cost-effective training solutions. This would translate into affordable, smart training access to a large number of users. It will also be flexible and capable of deploying on a large scale and into live/virtual/construct environments for further training tasks previously unattainable with lower fidelity simulators.

It is recommended for future research of amplified head rotations to study its effects on even larger displays and field of views with multiple visual channels would give insight on the upper end of simulator display

American Institute of Aeronautics and Astronautics

configurations, potential of retrofitting these and expanding their utility. It is also advised to incorporate head roll into the view augmentation as this becomes more natural to a user at the extreme edges of head yaw position. More complex flying tasks would also be interesting to document in future studies as well as tasks featuring non-aviation domains such as driving and virtual navigation.

References

[1]National Transportation Safety Board, "U.S. General Aviation, Calendar Year 2002. Annual Review of Aircraft Accident Data," Tech. rep., NTSB, Washington (DC), USA, 2006.

[2]Australian Transport Safety Bureau, "Review of Midair Collisions Involving General Aviation Aircraft in Australia between 1961 and 2003," Tech. rep., Australian Government, Canberra, Australia, 2004.

[3]ICAO, *Manual of Criteria for the Qualification of Flight Simulation Training Devices*, Vol. I, International Civil Aviation Organization, 3rd ed., 2009.

[4]Adams, R., "27 Into 7 Equals 9625 - The New Regulatory Matrix," *Civil Aviation Training Magazine*, , No. 1, 2009.

[5]Morrison, G., "Applying Simulation to First Officer Competencies," *2011 National Training Aircraft Symposium*, Daytona Beach, FL, USA, 2011.

[6]Wickens, C. D. and Hollands, J. G., *Engineering Psychology and Human Performance*, Prentice Hall, Upper Saddle River, NJ, USA, 2000.

[7]Bradley, D. R. and Abelson, S. B., "Desktop flight simulators: Simulation fidelity and pilot performance," *Behavior Research Methods, Instruments, & Computers*, Vol. 27, No. 2, Jun 1995, pp. 152–159.

[8]Alexander, A. L., Brunye, T., Sidman, J., and Weil, S. A., "From Gaming to Training: A Review of Studies on Fidelity, Immersion, Presence, and Buy-in and Their Effects on Transfer in PC-Based Simulations and Games," *DARWARS Training Impact Group*, No. 11, Aptima, Inc., Woburn, MA, USA, 2005, pp. 1–14.

[9]Le Ngoc, L. and Kalawsky, R. S., "Evaluating Usability of Amplified Head Rotations on Base-to-Final Turn for Flight Simulation Training Devices," *Proceedings of IEEE Virtual Reality 2013*, IEEE, 2013.

[10]Jay, C. and Hubbold, R., "Amplifying Head Movements with Head-Mounted Displays," *Presence: Teleoperators and Virtual Environments*, Vol. 12, No. 3, Jun 2003, pp. 268–276.

[11]Kopper, R., Stinson, C., and Bowman, D. A., "Towards an Understanding of the Effects of Amplified Head Rotations," *3rd IEEE VR Workshop on Perceptual Illusions in Virtual Environments*, 2011.

[12]Poupyrev, I., Weghorst, S., Otsuka, T., and Ichikawa, T., "Amplifying spatial rotations in 3D interfaces," *CHI '99 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '99, ACM, New York, NY, USA, 1999, pp. 256–257.

[13]Mulder, J. and van Liere, R., "Enhancing fish tank VR," *Proceedings IEEE Virtual Reality 2000*, IEEE Comput. Soc, 2000, pp. 91–98.

[14]Naturalpoint, Inc., "TrackIR," http://www.naturalpoint.com/trackir/, 2012.

[15]Bolte, B., Bruder, G., Steinicke, F., Hinrichs, K., and Lappe, M., "Augmentation techniques for efficient exploration in head-mounted display environments," *Proceedings of the 17th ACM Symposium on Virtual Reality Software and Technology*, VRST '10, ACM, New York, NY, USA, 2010, pp. 11–18.

[16]Jerald, J., Peck, T., Steinicke, F., and Whitton, M., "Sensitivity to scene motion for phases of head yaws," *ACM Symposium on Applied Perception in Graphics and Visualization*, ACM Press, New York, NY, USA, 2008, pp. 155–162.

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

[18]Draper, M. H., Viirre, E. S., Furness, T. A., and Gawron, V. J., "Effects of Image Scale and System Time Delay on Simulator Sickness within Head-Coupled Virtual Environments," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, Vol. 43, No. 1, Jan 2001, pp. 129–146.

[19]Mollenhauer, M. A., "Simulator Adaptation Syndrome Literature Review," Tech. rep., Realtime Technologies Inc., Royal Oak, MI, USA, 2004.

[20]Edwards, C. J., Creaser, J., Caird, J. K., Lamsdale, A., and Chisholm, S., "Older and younger driver performance at complex intersections: Implications for using perception-response time and driving simulation," *Proceedings of the Second International Driving Symposium on Human Factors in Driver Assessment Training and Vehicle Design*, Park City, UT, USA, 2003.

[21]Beall, A. C., *Visual control of the base-to-final turn in fixed-wing aircraft*, Ph.D. thesis, University of California, 1998.

[22]Bell, H. H. and Waag, W. L., "Evaluating the Effectiveness of Flight Simulators for Training Combat Skills: A Review," *The International Journal of Aviation Psychology*, Vol. 8, No. 3, 1998, pp. 223–242.

[23]Chien Wei, C., "Low Cost Virtual Cockpits for Air Combat Experimentation," *Interservice/Industry Training, Simulation, and Education Conference*, No. 1596, 2004, pp. 1–11.

[24]Estock, J. L., Alexander, A. L., Stelzer, E. M., and Baughman, K., "Impact of Visual Scene Field of View on F-16 Pilot Performance," *Human Factors and Ergonomics Society Annual Meeting Proceedings*, 2007, pp. 75–79.

[25]Covelli, J. M., Rolland, J. P., Proctor, M. D., Kincaid, J. P., and Hancock, P. A., "Field of View Effects on Pilot Performance in Flight," *The International Journal of Aviation Psychology*, Vol. 20, No. 2, Mar 2010, pp. 197–219.

[26]Rubio, S., Díaz, E., Martín, J., and Puente, J. M., "Evaluation of Subjective Mental Workload: A Comparison of SWAT , NASA-TLX , and Workload Profile Methods," *Applied Psychology: An International Review*, Vol. 53, No. 1, 2004, pp. 61–86.

[27]Häkkinen, J., Pölönen, M., Takatalo, J., and Nyman, G., "Simulator sickness in virtual display gaming," *Proceedings of the 8th conference on Human-computer interaction with mobile devices and services - MobileHCI '06*, ACM Press, New York, New York, USA, 2006, p. 227.

American Institute of Aeronautics and Astronautics

[28]Coolican, H., *Research Methods and Statistics in Psychology*, Hodder Education, London, 5th ed., 2009.

[29]Field, A., *Discovering Statistics Using SPSS*, SAGE Publications, 3rd ed., 2009.

[30]Borst, C., Mulder, M., and Van Paassen, M. M., "Design and Simulator Evaluation of an Ecological Synthetic Vision Display," *Journal of Guidance, Control, and Dynamics*, Vol. 33, No. 5, Sept. 2010, pp. 1577–1591.

[31]Le Ngoc, L., Borst, C., Mulder, M., and Van Paassen, M., "The Effect of Synthetic Vision Enhancements on Landing Flare Performance," *AIAA Guidance, Navigation and Control Conference 2010*, AIAA, 2010.

[32]Borst, C., *Ecological Approach to Pilot Terrain Awareness*, Ph.D. thesis, Delft University of Technology, 2009.

[33]Kennedy, R., Drexler, J., Compton, D., Stanney, K., Lanham, D., and Harm, D., "Configural scoring of simulator sickness, cybersickness and space adaptation syndrome: similarities and differences," *Virtual and Adaptive Environments Applications Implications and Human Performance Issues*, edited by L. Hettinger and M. Haas, chap. 12, CRC Press, 2003, pp. 247–278.

[34]Hodgkinson, J., *Aircraft Handling Qualities*, AIAA Education Series, American Institute of Aeronautics and Astronautics, 1999.