Loss of control is the largest contributor to the yearly aviation death toll, with energy mismanagement in low-energy conditions as one of its main causal factors. This has led to a large emphasis from both scientific and aviation safety communities on the prevention of aircraft upset conditions. Changes in aircraft configuration largely impact performance, and improved insight therein should allow pilots to better predict potentially dangerous situations, maintain suitable safety margins and more effectively react to unforeseen events. This paper presents the design and experimental evaluation of a Vertical Situation Display (VSD) with enhancements visualizing changes in the flight performance envelope. Sixteen pilots were tasked to fly approach and go-around scenarios with both a baseline and an ecological VSD, some of the scenarios containing flight control failures. Results show that the new display makes pilots maintain larger margins in velocity, thus spending less time below the advised minimum speed limit in final approach. However, these larger velocity margins also led to larger errors with respect to target velocities. Flight control failures were more often and more quickly discovered, and pilots reported feeling better able to predict dangerous situations. No significant differences in workload were recorded. These results conclude that the new VSD design enhances safety performance, but simultaneously raise the question whether the effect of enlarged safety margins is desirable if it causes a reduction in velocity tracking.

I. Introduction

The past 50 years have seen an exponential increase in flight safety. However, flying is still not free of risks, as is made clear by the news and yearly incident statistics. Both surprisingly and worryingly, the past 10 years have shown the same main cause of fatalities across all categories of civil aviation: Loss Of Control In-flight (LOC-I) [1, 2]. The NASA Langley Research Center reports that over three-quarters of LOC-I incidents are the result of flight outside the regular operational flight envelope, a condition better known as aircraft upset [3]. These upset are caused by various human, system or environmental factors, and often include an element of startle or surprise. Energy mismanagement (e.g. aerodynamic stall) is among the top causal factors in both commercial and general aviation, most commonly occurring in low-energy flight phases such as take-off and approach [4].

Research and industry are taking action at large to improve pilot training for LOC scenarios, such as the Manual on Aeroplane Upset Prevention and Recovery Training by the FAA, the project to create extended aircraft simulation envelopes by SUPRA, and over 130 companies offering upset prevention and recovery training worldwide [5-7]. These organizations agree that more focus should be placed on training pilots for situation awareness and manual skills in preventing upsets. One way to create such awareness is through Ecological Interface Design (EID) [8]. Founded by Vicente and Rasmussen [9], this framework aims to make complex interactions in the work environment accessible to users through visualizing operational boundaries. EID has successfully been adapted to improve pilot awareness in terrain avoidance [10], vertical combined traffic and terrain avoidance [11] and horizontal self-separation [12]. Furthermore, in these experiments EID has allowed pilots to better cope with unforeseen circumstances. Managing failures at low-energy flight phases is critical, as shown by the British Airways Flight 38 incident [13]. Following an engine failure in final approach the pilot of the Boeing-777 changed flap settings to reduce drag, thus extending the glide...
and preventing a crash with the ILS antenna. This raises the question: how can such interfaces be used to increase safety and performance for prevention of energy mismanagement in low-energy conditions, even after an unexpected event?

The objective of this research is to design and empirically evaluate an interface for increasing pilot performance and safety in an approach and go-around by visualizing the impact of configuration changes on the vertical flight performance. This display is an enhancement of the Intentional Vertical Situation Display (IVSD) by Comans [14], which in its turn was based on the Experimental VSD (EVSD) by Rijneveld et al. [11]. These VSDs yielded positive results in terms of pilot situation awareness and insight into maneuverability limits. The new version was tested by sixteen licensed pilots and compared to a baseline version without ecological cues. It was expected that by making aircraft safety margins more insightful, higher levels of performance and safety are facilitated at a lower workload. Moreover, if unforeseen circumstances do occur, EID-inspired displays should allow pilots to more timely diagnose the problem.

This article is structured as follows. Section II analyzes background on LOC incidents, the work domain, and earlier VSD versions. Section III then introduces the new interface enhancements. In Section IV the experiment design is outlined. The results obtained are presented in Section V after which they are discussed in Section VI. Finally, Section VII concludes this report and offers recommendations for further research.

II. Background

A. LOC Incident Analysis

Although the root cause of an upset might vary greatly, analysis of previous incidents show that a LOC incident usually develops as follows: the aircraft experiences a failure (system) or encounters an external hazard (environmental) to which the crew may not have appropriately responded (human). A resulting upset occurs, which might then lead to a loss of control [15]. Comparing causes between aircraft categories shows that Part 121 operations are more likely to suffer from system errors, whereas general aviation is more prone to human piloting errors [16-18]. Nonetheless, both scenarios would benefit from insight in aircraft capabilities combined with manual piloting skill to prevent or resolve upsets. Sorting upsets by flight phase shows that low-energy flight phases, such as take-off and landing, rank highest in terms of incident rates [16-18]. These phases often involve multiple procedures, limiting the pilots’ available mental resources. An upset at lower altitude allows for less time to recover and less potential energy to use in recovery. Additionally, these flight phases often involve changing the aircraft configuration, thus altering its performance characteristics. Visualizing these dynamic performance boundaries can make potentially dangerous situations more apparent.

B. Upset Prevention and Recovery Training

Unlike basic stall training, Upset Prevention and Recovery Training (UPRT) is not mandatory for pilots. In this sense UPRT can be compared to skid training for cars: a non-compulsory course that teaches drivers to keep control over their vehicle in adverse conditions. Various flight-safety organizations have advised structuring UPRT according to a specific multi-step process [5, 19-21]. First, the basic principles of aerodynamics relevant for LOC are taught in a classroom environment. These form the basis for both prevention techniques and recovery procedures, which are then practiced on Flight Simulator Training Devices (FSTDs). Unfortunately, current technology only allows limited simulation of LOC in FSTDs, as extended flight envelope models are still being developed [22-24]. Finally, on-aircraft training is used to practice recovery from full upset conditions, including effects of startle and disorientation. Two theories regarding aircraft performance and aerodynamics form the base of UPRT: the altitude envelope (Figure 1a) and the load-factor envelope (Figure 1b). Both describe aircraft maneuverability limitations, and thereby define operational constraints which pilots must respect.

1. The Altitude Envelope

The altitude envelope shows the range of true airspeeds which can be achieved at each altitude. At greater altitudes the lower air density allows an aircraft to obtain a range of higher true airspeeds. This is true until the maximum speed exceeds the maximum Mach number. Changing configuration from clean will change envelope limits. Extended flaps or gear decrease velocity limits, thus shift the envelope to the left. As airspeed and altitude are both measures of energy, the altitude-airspeed diagram represents a total energy diagram. When flying near the edges of the altitude envelope either the total energy contained by the aircraft must be adjusted or kinetic and potential energy must be
Fig. 1  The Altitude Envelope (a) and Load Factor Envelope (b), where $V_{MO}$ and $V_{DF}$ are the Maximum Operating and Design Velocities Respectively, $M_{MO}$ and $M_{DF}$ are the Maximum Operating and Design Mach Numbers and $V_S$ the Stall Speed. Adapted from [5].

exchanged. Exchanging energy is often faster than changing the total energy, making this the preferred option [5]. The most dangerous region of the altitude envelope is that for low altitude and low airspeed, as the combination indicates a low total energy state. Unintentionally approaching this region leaves pilots with reduced maneuverability, as there is a time delay for adding total energy through additional thrust.

2. The Load Factor Envelope

The load factor envelope, or V-n diagram, shows the interaction between load factor and equivalent airspeed. Use of equivalent airspeed allows using a single envelope rather than one for each altitude. The roof and floor of the V-n diagram in Figure 1b correspond to the common structural limits of +2.5 g and -1 g for general aviation aircraft, or +2 g and -1 g with flaps extended. These limits differ with aircraft type, and pilots should be aware of the maneuvering limits of the aircraft they are flying. Excursions of the left side of the V-n envelope indicate a deficient airspeed, which will stall the wing. For maneuvering in (near-)upset conditions, an aircraft will most often find itself close to the left-most border for positive load factors. Stalls for negative load factors are uncommon, as achieving a negative load factor requires the aircraft to make a diving maneuver, which will allow the aircraft to gain airspeed and divert from the envelope border. Note that this does not hold true for inverted flight conditions.

C. Review of Earlier VSD Enhancements

1. Visualizing the Flight Envelope

The basis for both reference VSD designs is the flight envelope, which shows maneuverability space in the vertical plane. The envelope encompasses all steady state velocities which can be achieved with a certain aircraft configuration. The current position within the flight envelope is shown by the aircraft velocity vector. The angle this vector makes with respect to the horizon corresponds to the flight path vector, its length indicates the velocity of the aircraft. Limiting factors for this envelope are the maximum and minimum thrust ($T_{max}$ and $T_{zero}$) and the maximum operating and stall speeds ($V_{MO}$ and $V_S$) These boundaries together form the performance envelope, shown in Figure 2. The steepest possible climb ($V_X$) is indicated by a green dot. In some aircraft, such as the Boeing 737 and 787, VSDs are used to display terrain height and send ground proximity warnings [25]. To integrate the performance envelope with a terrain database the envelope is expressed in distances by multiplying all velocities with a certain look-ahead time. For example, a look-ahead time of 60 seconds will result in an envelope of all locations that can be reached in one minute with a constant steady state velocity. If any wind is present, the envelope can be shifted as a whole to account for effects wind has on ground speed.

2. The Experimental VSD (EVSD)

Rijneveld et al. have added cues to the VSD for purposes of traffic and terrain avoidance [11]. Areas within the flight envelope which would result in a loss of separation with another aircraft are colored red, and the peak in nearby terrain is indicated by a brown line. The resulting display is shown in Figure 3. The upper axis shows the horizontal component of the velocity in Knots Indicated AirSpeed (KIAS), the right axis shows the vertical velocity component as the Rate Of
Climb (ROC). Distance is expressed in nautical miles on the bottom horizontal Along Track Distance (ATD) axis, the left vertical axis shows Altitude (ALT) in feet. As an additional ecological cue, the excess kinetic energy (more than is required for steepest climb) is expressed as a green bar on the velocity axis. This energy can be converted into a specific amount of potential energy, which is expressed as a green bar on the altitude axis. The total energy line within the flight envelope can be used to determine if the total amount of energy is increasing or decreasing.

Rijneveld et al. compared the EVSD to a VSD showing only climb and glide limitations in a pilot-in-the-loop experiment with 12 professional airline pilots. Although performance did not improve, pilots did report lower workload and increased situation awareness. Pilots noted that they did not often use the energy cues on velocity and altitude tapes. They also had longer reaction times using the EVSD but said to feel more confident in their decisions. The longer reaction time can be attributed to the greater amount of information needing to be processed, which also increases the certainty with which decisions can be made. It was recommended to explore integrating other types of information into the VSD such as intruder intent, aircraft configurations and malfunction scenarios.

3. The Intentional VSD (IVSD)

The IVSD by Comans [14] is an adaption of the EVSD where constraints are split into two types: causal and intentional. Causal constraints are determined by physical limitations to operation, such as the terrain or an intruder aircraft. The constraint around the intruder is approximated with a small cylinder to account for wake turbulence. Surrounding these are intentional constraints: the terrain clearance height and full intruder protected zone. Violating these boundaries will not directly cause an accident, but does put the aircraft in an increased state of risk. Both types of constraints are represented by filled polygons; causal constraints are more opaque than intentional constraints to differentiate their severity. Since all areas of the flight envelope that would result in any type of conflict are fully colored, the task of evading conflicts is now simplified to keeping the aircraft velocity vector outside of colored areas.
An experimental evaluation of the IVSD by 8 novices and 8 licensed pilots showed no significant effect on safety or performance compared to the EVSD. The IVSD did reduce the spread in performance, suggesting that adding intentional constraint information makes pilots more aware of their position within safety boundaries, thus leading to more fine-tuned control strategies. The extent to which this effect holds is unknown, but it raises interest in further research into adding additional information to the VSD.

Both the experiment by Rijneveld et al. and by Comans used scenarios flown exclusively in clean configuration, which means the flight envelope only varied in shape with altitude. Aircraft configuration, however, has a large influence on the flight envelope. Reflecting these effects requires the flight envelope to be made dynamic, which will be a central feature of the VSD enhancements presented in this research.

III. The Configuration VSD

This section will propose a new enhancement of the VSD to aid pilots in low-energy flight maneuvering. First the adjustments to the performance envelope are discussed in Section III.A, after which the integration into the VSD is explained in Section III.B. As this design will show the pilot information based on current aircraft configuration, it is named the Configuration VSD (CVSD).

A. Visualizing Configuration in the Performance Envelope

Changing aircraft configuration alters performance characteristics, which can be visualized using the flight envelope discussed in Section II.C. The flight envelope, however, is only able to reflect what velocities can be maintained with the current configuration. An ecological display shows the entire available work domain, so extra cues are added to give insight to what would happen when flap or gear state is altered. This envelope is shown in Figure 4.

The most prominent features of the visualization are the current velocity vector and current flight envelope. The velocity arrow must point inside the lines to be flying within flight envelope bounds, and if the arrow points close to the envelope edge it indicates a heightened risk of envelope excursions. Low velocities are dangerous for risk of aerodynamic stall, made apparent by a transparent orange area.

To keep the velocity arrow inside the envelope, both the vector and the envelope limits can be controlled. The velocity is controlled through changing thrust force and pitching moment, which in their turn depend on throttle and stick inputs. As pilots are assumed to know of these means-ends links and to limit the amount of cues presented simultaneously, no additional cues are dedicated to this interaction.

The envelope shape is prescribed by aircraft performance and limitations, which in their turn depend on configuration and aircraft specifications. Specifications, such as weight or maximum thrust, are mostly uncontrollable while airborne. The configuration, however, can be changed to alter the envelope shape. Means-ends links in the form of dotted lines are added to make this interaction explicit. Lines are grouped by color, and include congruent text indicating the configuration change corresponding to these limits.

Deploying flaps increases the lift and slightly increases drag, as well as decreasing stall and maximum operating speeds. This causes the flight envelope from Figure 4a to shift left towards lower speeds and slightly up towards higher rates of climb in Figure 4b. Lowering the gear greatly increases drag, causing the envelope to ‘drop’ in Figure 4c. Since drag increases with higher velocities, the effect is more significant on the far end of the envelope.
Additionally to adding configuration cues, an intentional cue is added regarding velocity. During approach the Final Approach Speed \((V_{\text{FAS}})\) is a safety constraint for pilots to abide to, which serves as a buffer in case any unexpected event occurs. Velocities below this threshold inside the flight envelope are indicated by a transparent orange area. The \(V_{\text{FAS}}\) in knots is typically determined by:

\[
V_{\text{FAS}} = 1.3 \cdot V_S + CF
\]

in which the Correction Factor (CF) depends on aircraft type and environmental conditions. For the Citation II, the model of which will be used in the experiment, this CF ranges from 0 to 20 kts.

B. Integration with VSD Design

As the display does not concern terrain avoidance the EVSD terrain peak line is removed for the CVSD. The total energy cues are also removed because their similar appearance to the newly added configuration cues might be confusing.

The CVSD includes all other features from the EVSD and IVSD. The performance envelope \(\odot\) is updated as explained in Section III.A. A magenta line \(\bigcirc\) shows the two-dimensional flight path, which includes waypoints indicating desired actions. The altitude and velocity goal for the next waypoint are indicated with magenta markers on the altitude \(\odot\) and velocity \(\bigcirc\) axes.

The envelope still shows the steepest climb by a green dot \(\bigcirc\), and the corresponding velocity is indicated in green on the velocity strip \(\odot\). The red bars for stall and maximum speeds \(\bigcirc\) are made dynamic to reflect speed limits for the current configuration. Finally a direct readout for the current flap angle and gear status is included \(\bigcirc\).

IV. Experiment Design

The CVSD was evaluated in a pilot-in-the-loop experiment. This experiment was designed to test the effectiveness of the CVSD in improving performance, detecting failures and adapting to unforeseen circumstances by comparing it to a Baseline VSD (BVSD) without ecological cues.

A. Participants

Sixteen licensed pilots took part in the experiment. All participants flew two similar sets of scenarios with both displays. Their task was to fly a scenario as best they could within safety limits. The order of the displays was varied to control for learning and fatigue effects; the first group of eight pilots started by using the BVSD (avg flight experience
2261 hours, $SD = 3198$), the second group of eight pilots started by using the CVSD (avg flight experience 2003 hours, $SD = 2588$). After completing a set of runs, participants took a break and flew a second set with the other of the two displays. In previous EID research, this method prevented display order from having significant effects on dependent measures \cite{26}.

Both groups comprised four PPL pilots, three CPL pilots and one military pilot. All types of licenses require training to perform basic low-energy flight maneuvers, so license type is assumed not to make a difference in performance. Participants had different levels of experience with glass cockpit displays, and therefore they were able to provide a wide range of feedback after the experiment.

### B. Independent Variables

The experiment was set up as a within-subjects repeated measures design, meaning that all participants flew using both the CVSD and BVSD displays. The former is described in Section III, the latter is exactly the same but without the velocity envelope and configuration cues.

The experiment comprises two sub-experiments: one concerning nominal operation and one concerning failure conditions. For both VSD variants a set of five runs was flown, four of which are nominal runs and one containing a flight control failure. The failure occurred once as the third and once as the fourth run in a set. As flight control failure, either the gear does not retract or the flaps get stuck at 15° during the go-around procedure. The order of these failures and combinations with displays were distributed evenly over the participants. Using a Latin-square design with 16 participants, each combination of display and failure occurred 8 times.

### C. Apparatus

The experiment was conducted in a fixed base flight simulator at Delft University of Technology. A non-linear six degree of freedom model of a Cessna 550 Citation II was loaded on the simulator. Throttle inputs were controlled by a thrust lever, control surface inputs were given through a hydraulic side stick with trim buttons.

The aircraft model was locked in the vertical plane, therefore yaw pedals and lateral side stick motions were frozen. A gear lever and flaps switch were used to change the configuration of the aircraft. International Standard Atmospheric conditions were used with zero wind, but all scenarios did include a mild turbulence using a Dryden model to increase the difficulty of the task.

### D. Experiment Displays

During each part of the experiment participants used a Primary Flight Display (PFD) and one of two VSD variants. No other flight displays were used.

1. **Primary Flight Display**

   A generic PFD based on a Garmin G1000 was used to give participants a familiar reference for their basic flight parameters (Figure 6). The PFD shows a pitch ladder and green flight path vector over a virtual horizon. An altitude tape is shown on the right, a velocity tape showing indicated airspeed on the left. Both include a magenta marker corresponding to the desired altitude and velocity at the next waypoint. The optimal climb speed $V_X$ is marked by an X on the velocity tape. The bank angle indicator at the top and compass at the bottom were frozen due to lateral inputs being disabled.

   Engine information is included on the far left side. Two bars indicate engine fan speeds with numerical readouts below. The bottom two numbers are the turbine speeds for both engines.

2. **Vertical Situation Display**

   Two versions of the VSD were used: the CVSD (Figure 5) and the BVSD (Figure 7). Although the BVSD does not include the flight envelope visualization the aircraft velocity vector was still present, allowing horizontal and vertical speeds to be read out on the KIAS and ROC tapes. The current flight configuration can be read from cues such as the numerical readout, red $V_S$ and $V_{MO}$ indicator bars and the three green indicator lights on the control panel.
Fig. 6  The PFD as Used in the Experiment.

Fig. 7  The baseline VSD as Used in the Experiment.

<table>
<thead>
<tr>
<th>Configuration</th>
<th>Clean</th>
<th>Flaps 15°</th>
<th>Flaps land</th>
<th>Go-Around procedure</th>
<th>Clean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Velocity</td>
<td>175 – 185 kts</td>
<td>170 kts</td>
<td>150 kts</td>
<td>119 kts</td>
<td>170 kts</td>
</tr>
<tr>
<td>Altitude</td>
<td>2000 ft</td>
<td></td>
<td></td>
<td></td>
<td>200 kts</td>
</tr>
</tbody>
</table>

Fig. 8  Schematic Representation of the Experiment Scenario.
E. Scenario

Participants were tasked to fly an approach and go-around procedure, as it allows combining multiple low-energy flight maneuvers in a single scenario. The scenario consisted of a standard non-precision approach using a 3° glideslope with a go-around at an altitude of 600 ft. The flight path and corresponding procedures were based on the Citation II Pilot Manual. An overview of the flight plan is shown in Figure 8.

Each run started in trimmed condition with an indicated airspeed between 175 and 185 knots at a distance of 1 to 3 nautical miles from the top of descent. From this point the participant had to follow the profile as closely as possible with intermittent velocity and configuration goals. All goals were allowed to be anticipated upon (such as configuring before a waypoint) except the go-around, which was not allowed to be initiated until after the go-around waypoint marker.

All participants reviewed the standard go-around procedure before the start of the experiment:
1) Gain velocity by giving full throttle
2) Simultaneously pull stick back to stop losing altitude
3) At $V_{REF} + 5$ kts, reduce drag by setting flaps to 15°
4) With positive ROC, retract gear
5) Clean up configuration to 0° flaps

Additionally, participants were told that they were free to deviate from standard procedure if they thought it would be in the interest of safety.

F. Procedure

The experiment started with a general briefing on the flight simulator, scenario and pilot objectives. Additionally, the briefing explained the information shown on their first VSD variant, which was then put into practice through multiple training runs. Finally, five measurement runs were made. Just before the measurement runs started, participants were alerted to the possibility of a flight failure as follows: “Unexpected events might take place. In such an event, the main goal will stay the same: to best execute the approach and go-around maneuver within safety limits.”

After a break, participants were briefed on the second VSD variant. This briefing and training runs were slightly shortened since the participant was already familiar with the flight controls, dynamics and procedures. The experiment is concluded by a questionnaire and debrief. The complete experiment procedure is shown in Table 1.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Experiment Procedure.</th>
</tr>
</thead>
<tbody>
<tr>
<td>15 min</td>
<td>Briefing</td>
</tr>
<tr>
<td>30 min</td>
<td>Training</td>
</tr>
<tr>
<td>20 min</td>
<td>Measurements</td>
</tr>
<tr>
<td>20 min</td>
<td>Break</td>
</tr>
<tr>
<td>5 min</td>
<td>Briefing</td>
</tr>
<tr>
<td>25 min</td>
<td>Training</td>
</tr>
<tr>
<td>20 min</td>
<td>Measurements</td>
</tr>
<tr>
<td>15 min</td>
<td>Questionnaire and Debrief</td>
</tr>
</tbody>
</table>

G. Dependent Variables

Measures for safety, performance and workload include both objective and subjective variables. From the simulation data, the Root Mean Squared Deviation (RMSD) of both altitude and velocity were used to assess participant performance. Safety is measured by the amount of envelope excursions, the minimum velocity obtained during a run and the cumulative time spent at velocities below $V_{FAS}$. Control activity is taken as an objective measure of workload. Participants' procedures were qualitatively analyzed to find differences in control strategies between displays. During scenarios in which a failure occurs, the time it took for the participant to diagnose the failure is measured. To obtain this information participants were requested to think aloud during the experiment, and timestamps were deduced from the audio recordings.

After each run participants were requested to submit a workload score on the Rating Scale Mental Effort (RSME), which was used as a subjective workload measure. A post-experiment questionnaire was used to collect further data on
safety and workload that participants experienced, as well as general comments on their preferences and suggestions.

H. Hypothesis

It is hypothesized that using the CVSD will increase performance, as the $V_{\text{FAS}}$ constraint assists pilots in velocity tracking when their workload is largest and the steepest climb indicator shows the quickest method to recover altitude at the start of go-around. Similarly, it is expected that the participants using the CVSD will fly more safely since velocity limits and $V_{\text{FAS}}$ boundary are made more explicit. Third, workload metrics are expected to be lower using the CVSD due to the mapping of higher level information in EID. Finally, as the CVSD includes ecological cues regarding aircraft configuration, it is hypothesized that detection and diagnosis of control failures will be faster than using the BVSD.

V. Results

This section presents the results of the experiment described in Section IV. The results are split into five parts corresponding to the categories of metrics used: performance, safety, workload, failure runs analysis and pilot feedback.

Each participant flew two runs which included a type of failure, one with each display variant. These failure runs are excluded from nominal performance and safety metrics, since a gear or flap failure impairs the performance characteristics of the aircraft. These runs are analyzed separately instead.

Runs which did not include a flight failure were effectively the same scenario with a random starting velocity and distance offset. These runs were aligned by disregarding data until 0.5 nautical miles before the top of descent. Test statistics were computed for each run and then averaged per pilot, resulting in fewer yet more reliable data. No runs were found eligible to be removed as outliers.

A. Performance

Generally altitude RMSD scores were similar between displays, and velocity RSMD scores showed better performance using the BVSD. Visualization of the data through box plots in Figure 9 shows a smaller spread in altitude error using the CVSD. The total range is however still similar.

No significant difference in altitude performance was found using Wilcoxon-Signed-Rank test ($Z = 0.362, p > 0.05$). Velocity RMSDs were compared using paired $t$-test, which showed a significant better velocity tracking performance using the BVSD ($t(15) = -2.19, p < 0.05$).

An improved velocity tracking performance using the BVSD is contrary to the hypothesis, which stated that this improvement would be found for the CVSD. Further analysis into the origin of this result is done by creating a running 95% confidence interval of the velocity profile along the length of the approach and go-around. Intervals for both displays are superimposed in Figure 10, showing similar trends at all velocity targets except at 4 NM. At this point the velocity goal is equal to $V_{\text{FAS}}$, which is explicitly shown by the intentional constraint on the CVSD. Participants using the CVSD more actively prevented their velocity from violating this intentional constraint, often keeping some margin to account for the randomness of turbulence. The variance in velocity during go-around was also decreased, but this
effect is not reflected in performance metrics as no velocity targets are present in this section of the experiment scenario.

To further analyze strategies participants used, the two performance metrics are compared to each other. Pearson’s correlation coefficient shows that for both displays a relationship between the two performance metrics exists (r = -0.637, p < 0.01 and r = -0.715, p < 0.01 for BVSD and CVSD respectively). This is visualized by plotting both performance metrics against one another in Figure 11.

Trend lines for BVSD and CVSD are of comparable slope, and show that participants must chose between optimizing altitude or velocity tracking performance. Furthermore, Figure 11 can be used to confirm that pilot license seems to have no effect on performance, as no type of pilot is consistently above or below the trend line. This is confirmed by multivariate ANOVA tests, varying license type for altitude RMSD (F(2,15) = 0.376, p = 0.694) and license type for velocity RMSD (F(2,15) = 1.066, p = 0.373). Similarly, MANOVA tests showed display order not have a significant effect on performance metrics.

B. Safety

One of the goals of the CVSD is to make pilots more aware of causal and intentional velocity limits. No envelope excursions occurred during the experiment, so only intrusions of the intentional constraint are analyzed. This is done by looking at intrusion depth and total intrusion time, which translates to the minimum obtained velocity and the total time spent below $V_{FAS}$. These metrics are represented by box plots in Figure 12a and 12b respectively.

The large difference in minimum velocity is confirmed by a $t$-test ($t(15) = -2.71$, p < 0.05). This is in agreement with the results found in Section V.A, in which performance metrics were in favor of the BVSD. Since the target velocity at go-around and the velocity safety limit coincide, the BVSD median minimum velocity being approximately $V_{FAS}$ means that half the minimum velocities violated the intentional constraint. The CVSD minimum speeds are generally above the $V_{FAS}$ mark, trading performance achieved by the BVSD for better performance regarding safety.

Since minimum velocities for the CVSD generally did not drop below $V_{FAS}$, it is no surprise that most participants averaged 0 seconds below $V_{FAS}$ in this condition, as can be seen in Figure 14. Eleven out of sixteen participants didn’t exceed $V_{FAS}$ in any of their four nominal runs, bringing the median and bottom quartile both to zero. The single CVSD outlier whith a mean time of 8.7 seconds often pitched up too quickly during go-around in order to follow the reference height, thus neglecting their velocity for a brief moment. As data for time below $V_{FAS}$ did not pass the Shapiro-Wilk test, a Wilcoxon-Signed-Rank test was used to confirm significance ($Z = -2.29$, p < 0.05). As both safety metrics are in favor of the CVSD, the hypothesis that CVSD will allow pilots to fly more safely is considered confirmed.

C. Workload

Workload was analyzed objectively through control input variation and subjectively through self-reported RSME ratings. Control inputs used were side stick deflection and thrust setting. Using the control rates rather than deflections
Fig. 11  Velocity versus Altitude RMSD’s for Both Displays, Including Regression Lines. Participants Visually Distinguishable by Pilot License Type.

Fig. 12  Minimum Velocity (a) and Time Spent Below $V_{FAS}$ (b).
eliminates the effect a different trim position might have, thus allows comparing results of participants using varying amounts of trim.

At least one large peak in both control inputs is expected each run at the start of the go-around. Since this input is present for all runs it introduces the same bias in all results, and thus does not prove a problem for this analysis. Boxplots with deviation of control input rates are shown in Figure 13.

A decrease in rate variability is visible, especially for elevator inputs in Figure 13a. This might be explained by participants being able to see that their aircraft is not flying close to any safety limits, thus relaxing the need to correct higher frequency errors introduced turbulence.

Wilcoxon-Signed-Rank tests found no difference for standard deviation in elevator input rates (Z = -0.958, \( p > 0.05 \)) or throttle deflection rates (Z = -0.675, \( p > 0.05 \)). These results are replicated by self-reported workload ratings, showing no improvement in reported subjective workload scores (t(15) = -0.798, \( p > 0.05 \)). These metrics provide evidence that there is no difference in workload between displays.

### D. Failure Run Analysis

It was expected that flight control failures would always be discovered, but this did not end up being the case. Failures were diagnosed in half of the BVSD failure runs and three-quarters of the CVSD failure runs, shown in Table 2.

Possible reasons for missed detection are discussed in Section VI. Although diagnosis rate is in favor of the CVSD, not enough data were generated to confirm this with a reliable test statistic. Five participants did not notice the flight control failure with the BVSD but did with the CVSD, whereas the reverse result occurred only once. For all failures successfully diagnosed, the time it took for the participant to do so is shown in Figure 14. Times are measured from the moment a rejected input is given until the start of an audible diagnosis by the participant. No boxplots are drawn because of the limited amount of data.

Diagnosis times for the CVSD are generally shorter than for the BVSD. This was expected, as the CVSD flight envelope visualization offers an additional cue for detection and diagnosis. Several failures were diagnosed within seconds of being introduced, but most required some additional event to take place. Diagnosis times for both gear and flap failures that initially went undetected seem grouped at certain time intervals. This grouping is correlated with specific moments in the go-around occurring which facilitate diagnosis.
For gear failures, this specific moment was either the participant going through a self-directed post-go-around checklist when using the BVSD, or when extending the flaps to 0 with the CVSD. The latter often led to immediate detection by the participant, as having the gear deployed without flaps results in a downward pointing shape of the performance envelope which is easily distinguishable (Figure 15).

CVSD flap failures were often noticed shortly after the rejected input had been given. Using the BVSD generally caused these failures only to be noticed by the end of the go-around after the speed goal changed to 200 knots. This corresponds to $V_{MO}$ for the Citation II with flaps 15°, and pilots noticed something was wrong as their velocity indicator on the VSD horizontal speed tape approached the red area.

Although no reliable statistical analysis is possible, trends in data show the CVSD having a positive effect on both diagnosis rate and diagnosis time of flight control failures. Not all failures were discovered however, for which potential reasons are discussed in Section VI.

After detection and diagnosis, participants generally slightly adjusted their control strategy for that run. Flying an aircraft with the gear deployed adds drag, which participants accounted for by setting a higher throttle. Having flaps at 15° adds both drag and lift, but more noticeably reduces $V_{MO}$ to 200 knots. Participants originally gave more throttle, but later reduced to keep their velocity around 190 knots, again sacrificing performance for safety. These adjustments were similar for both displays after the failure was diagnosed.

Failure runs have also been analyzed using the performance, safety and workload metrics, and showed similar trends to the nominal conditions. The CVSD has not lead to different control strategies such as changes in configuration schedule.
E. Pilot Feedback

Two forms of feedback were collected from participants at the end of the experiment. Firstly, a questionnaire asked participants to compare both displays, and participants were requested to elaborate on their answers. Secondly, participants were asked for feedback on the displays, simulator or any other aspect of the research in an open format.

1. Questionnaire

Each of the ten questions in the questionnaire presented participants with a scale showing the BVSD on the left, CVSD on the right and a small mark in the middle for neutral/equal. Participants ticked a location on the scale to indicate for which display they found the statement to be most true, as well as the weight they gave to their opinion. This was used to construct a box plot for each question, which are shown in Figure 16.

Participants mostly indicated that the CVSD allowed them to better handle failures, fly safer and better predict dangerous situations. Even pilots who did not notice either of the two failures reasoned that additional information on the CVSD must make failure detection and diagnosis easier, although they submitted less positive scores than those who successfully diagnosed the flight control failures. Two out of sixteen pilots disliked the CVSD, and noted that the flight envelope visualization sometimes caused cluttering of the display.

Generally, pilots indicated experiencing the CVSD ecological cues as useful additions to the VSD. They say to have experienced lower workload, although this claim is not supported by objective nor subjective metrics analyzed in Section V.C. Additionally, multiple participants said the CVSD was more difficult to work with (Question 1), yet still rated their workload in favor of the CVSD. When asked about this seeming inconsistency, a participant explained: “The CVSD wasn’t easy to learn and requires more of my attention, but it offers me peace of mind knowing I’m not near the limits of my aircraft.” This is confirmed by the answers to question 8 on ‘most concentration to use’.

Further correlations in the questionnaire results were found by means of a principle component analysis. Three significant categories (with eigenvalues greater than 1) were created, cumulatively capturing 80% of variance in the answers given. The categories are marked with numbers 1 to 3 in Figure 16 based on overlapping topics of the questions:

1) Workload
2) Performance
3) Situation Awareness

Whereas questions on failure detection and diagnosis seem to belong under the ‘Situation Awareness’ topic, a larger correlation was found with the question on achieving altitude/velocity goals, placing this question in the ‘Performance’ category. This can be explained by the nature of the display and scenario design: multiple flight control failures have happened during the experiment, which impeded aircraft performance. Flight envelope visualization gave insight in performance limits after a failure, thus timely detection and diagnosis allowed pilots to consciously account for these adjusted performance limits in controlling the aircraft.
Questions regarding failure detection and diagnosis were heavily in favor of the CVSD, a result which was expected due to the increased amount of cues a pilot can observe when a failure occurs. Interestingly, the CVSD scores higher on failure detection than on diagnosis. Various reasons along the same lines were given, best articulated by a participant as: “After a few runs you know how the envelope is supposed to move, thus when it doesn’t [move as expected] you know to start a search for the underlying reason”.

Interestingly, personal preference correlates with both workload and situation awareness, although correlations are less strong than commonly found within categories (with correlation weights of 0.71 and 0.63 respectively). This result gives insight in what factors might effect pilot preference in working with one display over another, with workload as one of the highest priorities.

2. Further Pilot Comments

Since the pilot comments had an open format, topics varied among pilots. Still various themes could be distinguished, within which one or two common opinions were held by groups of participants.

Pilots often mentioned the usefulness of the green dot indicating the steepest climb on the CVSD. Reasons given include its clear purpose, good visibility and ease of use by simply moving the velocity vector there. Especially at the workload-intensive time during go-around when the steepest climb is commonly used, these aspects are amplified in their importance and thus perceived usefulness.

Several suggestions were done to expand on display features, such as having the aircraft icon rotate according to the current aircraft pitch angle and providing a cue to indicate the current trend of the velocity vector. The most requested display feature was some form of direct feedback for the effect of control inputs on the velocity vector. For a successor to the CVSD, some form of acceleration vector might be considered to more closely link inputs to the effect they have, thus giving operators more insight in system dynamics.

Critique of the display mostly focused on display elements which, in their current state, are not easy to use. Since most cues are centered around the left of the VSD, the ROC tape on the right is difficult to include in a natural-feeling crosscheck. Additionally, the scale of the vertical velocity tape is too small for some readouts, and participants rather used the numerical velocity readout on the PFD. Two participants disliked the CVSD in general, as they said it makes the situation unnecessarily complex.

VI. Discussion

This research aimed to investigate the effects of an ecological VSD enhancement on the ability of pilots to perform a safe and accurate approach and go-around maneuver with potential flight control failures. An analysis of flight paths and control inputs shows that pilots were better able to abide by safety constraints when using the ecological CVSD then when using the more basic BVSD. However, this did come at the price of reduced velocity tracking performance, indicating that the CVSD caused pilots to place a higher emphasis on maintaining safety margins than on optimizing their performance. The desirability of this effect can be questioned when it would become too large, as the marginal gain of larger safety buffers might not outweigh significant decreases in performance.

A turbulence of medium intensity was simulated throughout all of the experiment runs, which caused small random movements of the velocity vector. This uncertainty might be a reason for pilots to maintain an additional velocity margin to the CVSD $V_{FAS}$ boundary, thus decreasing CVSD performance metrics. Previous research by Comans has shown EID to cause a reduced spread in flight path traces [14], an effect which can also be seen in the velocity traces for the CVSD (Figure 10). Reducing the turbulence might cause pilots to reduce the velocity buffer size they maintain. This would result in flight consistently close to but rarely under $V_{FAS}$, thus optimizing both safety and performance.

Objective workload measures do not indicate any significant difference between displays, and neither do subjective post-run workload scores. However, comparing the two displays in the post-experiment questionnaire, pilots did indicate experiencing a workload reduction when using the CVSD. This does come at the cost of more concentration required to interpret the cues it gives, especially in early phases of using the display. Pilots believed that more training and experience in using the CVSD would allow them to retain their workload reduction without the concentration penalty. In conclusion, the hypothesis stating that the CVSD leads to a workload reduction cannot be accepted, as the results are conflicting and inconclusive.

The CVSD did increase flight control failure detection rates and reduce diagnosis times. Multiple failures went undetected, and a reduction in the amount of data available rendered producing a reliable statistic impossible. Visual analysis of the data supports the hypothesis, but more tests will be necessary for a full confirmation.
Although pilot license did not effect performance metrics, recent experience with FSTDs might have affected metrics for failure diagnosis. Three participants had frequent simulator experiences during the past six months: two during CPL training and one as FSTD instructor. These participants consistently diagnosed failures within 10 seconds regardless of flight display, and achieved all diagnosis times below 10 seconds for the BVSD. This would indicate that recent FSTD training outweighs the effect EID might have on handling certain types of unforeseen circumstances. Whereas the CVSD might not play a significant role for well-trained pilots, its use as diagnosis aid for participants who have not had recent simulator training is strengthened when these data points are accounted for.

It was assumed pilots would always detect failures after they occurred, and the question would be how quickly they would be able to diagnose them. However, the experiment has shown that with neither VSD all failures were detected. Reasons for unnoticed failures are suspected to lie in their non-critical consequences and the limited amount of cues they can be detected by. Participants were only presented with visual cues during the experiment, whereas something as an extended gear would usually also have audible and haptic cues. Including these in the simulation is expected to reduce detection times considerably, whereas diagnostic times are expected still to be in favor of the CVSD. This is to be confirmed by future research, investing more effort into increasing simulation fidelity for better failure simulation.

Throughout the experiment, a few issues arose regarding the simulator and simulation that might have affected the results. When using the simulator for multiple experimental trials on a single day, the hydraulic side-stick could overheat which caused the simulator to stop operating. This has forced to restart a handful of scenarios, and on two occasions forced to take a second, small recess during the experiment. The effect of the extra experience, rest and/or fatigue which some participants accumulated this way is unclear. However, the effect is brought to a minimum when analyzing the results by averaging metrics for all runs a participant flew. No individual runs qualified to be removed as outliers. Some participants commented on the highly sensitive pitch response of the simulation model. This may be due to poor tuning of the aircraft model, but is more likely related to the experience these participants had flying mostly larger aircraft. Participants experienced flying the Cessna Citation II had no objections on simulation fidelity.

**VII. Conclusion & Recommendations**

In this research, two versions of a VSD were evaluated by 16 pilots to test their effect on the performance of an approach and go-around maneuver. Some scenarios included a gear or flap failure which pilots had to diagnose.

The CVSD showed pilots their flight envelope, which was variable with configuration and included a $V_{FAS}$ indication. Compared to a baseline VSD, the CVSD increased safety at the cost of tracking performance, as pilots avoided violating safety constraints by maintaining an additional velocity buffer. Participants said the CVSD reduced their workload, although this statement is not backed by objective metrics. Flight control failures were more often detected and diagnosed more quickly. However, not all failures were detected and pilots said the CVSD required more concentration, indicating more training might be necessary for the display to be optimally used.

For future research, it is recommended to look into redesigning parts of the CVSD, as well as testing its use in other scenarios. Various small yet valuable suggestions were made by pilots participating in the experiment, such as pitching the ownship visualization and including a velocity trend vector, but also more prominent aspects such as the effect of a constant wind should be considered. This would allow for a wider variety of more realistic scenarios to be flown and tested.

Failures tested in this experiment were limited in variety and cues. Future experiments could aim to add others failures such as windshear, microburst or engine failures. Previous versions of the VSD were used as tools for traffic and terrain awareness, incorporating these in scenario design will lead to a more varied experiment and potentially to a more versatile display.

Finally, it is recommended to further explore the effect turbulence has on strategies employed by pilots. If introducing random variance to a controlled system with an ecological display causes pilots to change their control strategy, this may have unintentional consequences. Understanding these consequences is paramount to adapting ecological displays for use in the real, turbulent outside world.

**References**


