Evaluating Stick Stiffness and Position Guidance for Feedback on Flight Envelope Protection

Van Baelen, Dirk; van Paassen, M.M.; Ellerbroek, J.; Abbink, D.A.; Mulder, Max

DOI
10.2514/6.2021-1013

Publication date
2021

Document Version
Final published version

Published in
Proceedings of the AIAA Scitech 2021 Forum: 11–15 & 19–21 January 2021Virtual/online event.

Citation (APA)
Van Baelen, D., van Paassen, M. M., Ellerbroek, J., Abbink, D. A., & Mulder, M. (2021). Evaluating Stick Stiffness and Position Guidance for Feedback on Flight Envelope Protection. In *Proceedings of the AIAA Scitech 2021 Forum: 11–15 & 19–21 January 2021Virtual/online event.* Article AIAA 2021-1013 American Institute of Aeronautics and Astronautics Inc. (AIAA). https://doi.org/10.2514/6.2021-1013

AIAA SciTech Forum
11–15 & 19–21 January 2021, VIRTUAL EVENT
AIAA Scitech 2021 Forum

10.2514/6.2021-1013

Evaluating Stick Stiffness and Position Guidance for Feedback on Flight Envelope Protection

Dirk Van Baelen*, M.M. (René) van Paassen†, Joost Ellerbroek‡, David A. Abbink§ and Max Mulder¶
Delft University of Technology, Kluyverweg 1, 2629 HS Delft, The Netherlands

Modern aircraft use a variety of fly-by-wire control devices and combine these with a flight envelope protection system to limit pilot control inputs when approaching the aircraft limits. The current research project aims to increase pilot awareness of such a protection system through the use of force feedback on the control device, i.e., haptics. A previous design used asymmetric vibrations to *cue* the pilot on the flight envelope. The evaluation showed no improvement in metrics at the first emergency encounter, yet did show a potential training benefit. Therefore, a new haptic feedback concept was designed with the specific aim to *guide* the pilot when approaching a limit and provide support from the first time use. This paper evaluates these haptic feedback designs with 36 active PPL/LAPL pilots who flew a challenging vertical profile and encountered a windshear in a fixed-base simulator. The pilots were divided in three groups who received either cueing, guidance, or no haptic feedback. It was hypothesized that: (i) cueing haptic feedback provides a faster learning rate compared to no-haptics, and (ii) guidance haptic feedback results in best performance from the first run yet worse metrics when no feedback is provided. Comparing the results of the cueing and no-haptic feedback groups confirmed the first hypothesis. Results also showed that the guidance haptic feedback resulted in improved metrics at the first run, and the worsening of metrics when no longer provided.

Nomenclature

Symbols

a	Acceleration, m/ s^2
b	Damping, Nms/rad
D	Drag, N
g	Gravitational acceleration, m/ s^2
K	Gain, -
k	Spring, N/rad
m	Mass, kg
n	Load factor, g
q	Pitch rate ($\dot{\theta}$), rad/s
T	Thrust, N
t	Time, s
V	Velocity, m/s
W	Weight, N
x	Distance from starting position, m
α	Angle of attack, rad
β	Side slip angle, rad
γ	Flight path angle, rad
δ	Control device deflection, rad
θ	Pitch angle, rad
ϕ	Bank angle, rad

Subscripts

lat	Lateral value
lon	Longitudinal value
max	Maximum value
MO	Maximum operational value
min	Minimum value
nom	Nominal value
prot	Protected region value
stall	Value when stall occurs

---

[*] PhD student, Delft University of Technology - Control & Simulation, d.vanbaelen@tudelft.nl. Student Member AIAA
[†] Associate Professor, Delft University of Technology - Control & Simulation, m.m.vanpaassen@tudelft.nl
[‡] Assistant Professor, Delft University of Technology - Control & Simulation, j.ellerbroek@tudelft.nl
[§] Professor, Delft University of Technology - Cognitive Robotics, d.a.abbink@tudelft.nl
[¶] Professor, Delft University of Technology - Control & Simulation, m.mulder@tudelft.nl. Associate Fellow AIAA

I. Introduction

Both international aviation safety boards, such as the European Union Aviation Safety Agency EASA, and airline associations, for example the International Air Transport Association IATA, identify loss of control in flight as one of the key risk areas resulting in most fatalities within aviation. [1, 2] A safety issue contributing to such a loss of control is identified as the inadequate monitoring of the main flight parameters and automation modes. To ensure and improve current safety levels, these loss of control events should be prevented.

Improving the information presented to pilots is expected to help reducing the loss of control occurrences. This can be achieved by augmenting the visual displays on the flight deck with information on the limits of the aircraft, i.e., the flight envelope. Research showed that this can improve safety by reducing the risk of violations of those limits. [3] Once the limits are exceeded, for example in a stall, the information on the Primary Flight Display (PFD) can be augmented with recovery guidance which delivers recovery performance improvements as shown in three simulator evaluations. [4]

Apart from the visual channel, pilots can also perceive information through the sense of touch. An example is the haptic interface, which provides force feedback through the control device. This form of information can have a significant positive effect when a pilot is guided along the approach path. [5, 6] Additionally it can be used to show a set of predicted controllability limits, which was shown to be used by pilots in an experiment. [7] Research indicates also that haptic feedback can be used to show pilots information on the Flight Envelope Protection (FEP). [8]

The latter experiment had two groups where the first group started with haptic feedback which was ‘cueing’ the pilot on the flight envelope limits, and the second group had no haptic assistance. After a break the groups switched: only the second group received haptic feedback. The initial hypothesis for this experiment was that haptic feedback would support performance, and that performance would reduce after reverting to a condition without haptic support. Contrary to this, however, it was found that haptic feedback mainly contributed to pilot learning, and performance persisted after haptic support was removed. In addition, haptic support did not improve performance during the first run, which indicates that when implemented on an aircraft, it might not provide pilots with support the very first time they encounter a new situation. As the haptic feedback system aimed to support pilots also in new, unforeseen circumstances, a new iteration of the haptic feedback is required.

Actively supporting the pilot has been found to help at the first encounter, yet is subject to reversion to base performance when the support is removed. In a skill acquisition task where a slider had to be moved left and right, four groups of participants received feedback on their performance in a training phase at different times: after each run, or an average score after every five, ten or fifteen runs. [9] Their results showed that increasing the amount of feedback increases performance. Immediately after the training phase, another set of measurements was performed where no feedback was provided. There, the group with the most amount of feedback in the training performed worst, although not significantly different from the other groups. Another measurement was performed two days after the initial training, which showed again a tendency for decreasing performance with increasing feedback during training. This phenomenon is called the “guidance hypothesis”: a dependency on the feedback develops while learning the task; disabling this feedback then results in worse performance, due to required re-adaption. This phenomenon was also reported in a similar, vertical task. [10]

Within the field of haptic feedback, different applications have been recently designed to support the human operator in a task, and to provide support from the first encounter. Examples of this are a support for an abstract control task ([11]), a lane keeping assist in the automotive domain ([12, 13]), and an obstacle-avoidance system for UAV tele-operation. [14] These examples used active haptic feedback, for example an increased stiffness or actively moving control device, to guide the operator to complete the task. Transferring these active haptic feedback principles to the aircraft flight envelope protection system might provide a feedback system which supports pilots from the first run and solve the issue with our previous ‘cueing’ system. [8] Nevertheless, such implementations of haptic support have been found to be also hindered by to the guidance hypothesis described before, and it should be investigated whether this is also true in our particular application.

The aim of this paper is to present a new haptic feedback for FEP design which is more actively ‘guiding’ the pilot, and to compare the results of this guidance haptic feedback system, as well as the existing ‘cueing’ haptic feedback system, to the results of a group of pilots who did not receive any haptic feedback at all. It is hypothesised that the group without haptic support requires more time to learn the task when compared to the results of the ‘cueing’ group, and that the guidance haptic feedback design is able to support pilots from the very first run, however, with possible reversion in performance when the haptic assistance is removed.

This paper first discusses the different haptic designs used in Section II. Section III presents the experiment where the participants were required to operate an aircraft at the limits. In Section IV and V, results of the experiment are described and discussed. Finally, the conclusions are stated in Section VI.

II. Feedback Design

The haptic feedback design is based on a control structure similar to an Airbus A320. Full details are given in our earlier work, see Ref. [15], only the relevant elements for understanding the current experiment will be explained in this section. Two designs are elaborated which use haptic feedback to communicate the flight envelope protection limits by changing the feel on the control device.

Note that the designs shown here do not include a breakout force, i.e., a minimal force required to move the side stick, which is present on an actual A320 aircraft. The two haptic feedback designs to communicate the flight envelope limits are discussed below, respectively a cueing and guidance haptic support system. But first some basic knowledge on the A320 control structure is presented.

A. Airbus A320 Control Structure

Modern-day Airbus aircraft, like the A320 and the A330, all employ a Fly-By-Wire (FBW) system. This means that there is no mechanical connection between the control surfaces and the control device. The latter acts as an interface for the pilot to provide inputs to the Flight Control Computers (FCCs) which then command the control surfaces with hydraulic actuators. This allows a Flight Envelope Protection (FEP) system to be used, which can check and, if necessary, limit pilot inputs, to ensure that no flight envelope limits are violated.

Longitudinal control in a FBW Airbus, with all sensors functional (a mode designated as the normal law control law), is provided using C^*-control, which is a combination of both pitch rate (q) and load factor (n). [16–19] On top of this control law, a hard envelope limit is employed which protects the pilot from exceeding limits on angle of attack (α), load factor (n), and maximum velocity (V_{MO}). This protection is depicted in Fig. 1, where the nominal flight envelope is the extreme limit which can not be exceeded, the safe flight envelope is the point where protections start acting. The envelope is constructed by the maximum (n_{max}) and minimum (n_{min}) load factor, their protection limits ($n_{\mathrm{max}_{\mathrm{prot}}}$ and $n_{\mathrm{min}_{\mathrm{prot}}}$, respectively), the maximum operation velocity (V_{MO}, and protection $V_{\mathrm{MO}_{\mathrm{prot}}}$), and minimum velocity (V_{stall}, and protection V_{prot}).

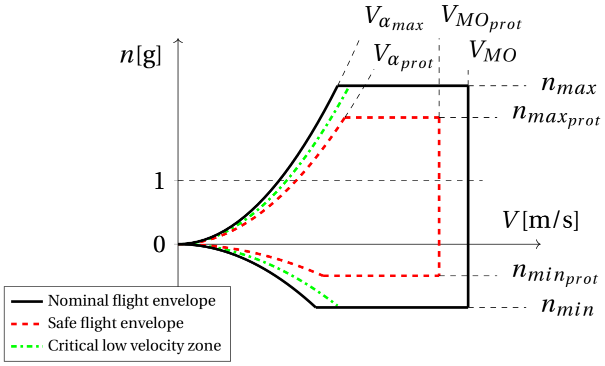


Figure 1 Flight envelope, velocity (V) versus load factor (n)

When multiple FCCs fail, or when a sensor failure occurs, the control is reverted to a degraded control law. In this research, we will consider a control law close to the Airbus alternate law without reduced protections, where the same protections apply as before, only the angle of attack protection is lost. Hence, in alternate law the aircraft can be stalled, and it allows the pilot to give more extreme control actions.

Lateral control in normal law is a bank (ϕ) rate command from −33° till +33° of bank. Beyond these limits, positive roll stability is achieved such that the aircraft rolls back to the protection value (ϕ_{prot}) of ±33°. The maximum achievable bank, with full lateral side stick deflection is ±66° of bank. In alternate law, lateral control reduces to a pure rate command, irrespective of the actual bank angle. More details on the control laws and degraded control laws can be found in Ref. [15].

Given that for both longitudinal and lateral control, a degradation of the control law results in a different effect for a given control input, a clear indication of both the limits and the active protections of the flight envelope is required. Nevertheless, accidents have occurred where pilots were not aware of what control law was active, and what protections were still active. [20] As such, a clear and intuitive way of presenting this information can be found in haptic feedback and a new design is proposed in the following.

B. Cueing haptic feedback design

In the cueing haptic feedback design, the pilot is cued about the flight envelope limits using forcing functions (forces on the control stick) which are asymmetric in both time *and* amplitude. To visualize the feel, the amount of force required to displace the side stick to a certain deflection is combined in the haptic profile as given in Fig. 2. This figure shows the nominal feel on an Airbus side stick with a neutral point, the point at which no force on the side stick is required, and a linearly increasing force with a certain spring coefficient with an increased stiffness at 6°. [21] Such a haptic profile provides the pilot with information on the input magnitude: larger inputs require larger forces.

Previous research showed that such an asymmetric vibration can be used to both cue the pilot about an imminent limit, as well as indicate a required control action to move away from that limit. [22] Such a forcing function is vertically shifting the default haptic profile (Fig. 2). It is assumed that the forcing functions are short in time and/or amplitude such that the input to the aircraft is minimal. The feedback design uses three cues to communicate the flight envelope limits to the pilot:

1) When the aircraft state leaves the safe flight envelope, i.e., crosses the red line on Fig. 1: a sawtooth-shaped forcing function of 1s with an amplitude of 0.282Nm and frequency of 2Hz is activated.
2) As long at the aircraft state remains outside the safe flight envelope: one sawtooth-shaped 'tick' is provided every second, where the intensity of the tick is linearly increasing with the magnitude of the safe flight envelope excursion, up to a maximum of twice the default magnitude.
3) When the velocity drops below $\left(V_{\text{stall}} + V_{\text{prot}}\right)/2$, i.e., left of the green line on Fig. 1, a stick shaker signal defined by a sinusoid with amplitude of 0.426Nm and frequency of 20Hz is activated.

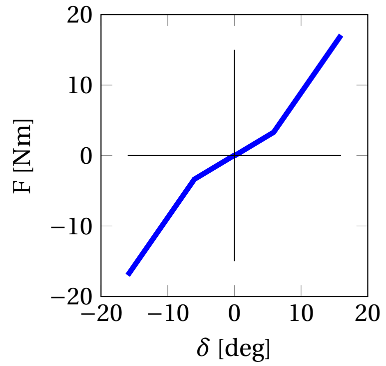


Figure 2 Default haptic profile, force required on the side stick for a given deflection, in the cueing design

The *direction* of the sawtooth-shaped forcing functions is used to suggest a control direction to move away from the limit. As such, the cue is forward/push for high angles of attack and high load factors, the direction is opposite for other conditions. More details and an example can be found in Ref. [8].

C. Guidance haptic feedback design

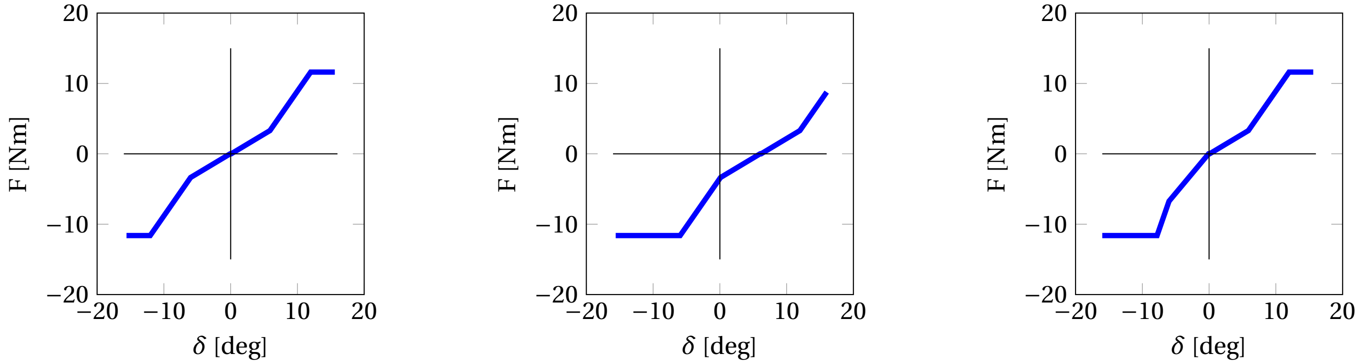
The guidance haptic feedback design informs the pilot on the limits of the flight envelope using two changes to the haptic profile: i) a change in spring coefficient and ii) a displacement of the neutral point position. To guarantee that the pilot has the final authority of the side stick, the maximum amount of force required to displace the stick to the maximum position is limited to 11.6Nm which results in a default haptic profile for the guidance design as shown on Fig. 3a. This maximum value is chosen based on the forces exerted by pilots on the stick in the experiment discussed in the previous experiment, see Ref. [8]. A summary of all tuning parameters can be found in Table 1. The two cues introduced above are elaborated next.

1. Stiffness feedback

Increased manipulator stiffness has been investigated in previous research for indicating an undesired control deflection when a pilot-induced oscillation is imminent ([23, 24]), signaling a lagging adaptive controller ([25]), or indicating a limit on the main rotor setting of a helicopter. [26] In our scenario, an undesired control deflection is defined as an input which brings the aircraft closer to the limits of the flight envelope, which can be e.g., a control deflection in

Table 1 Summary of all haptic feedback tuning parameters

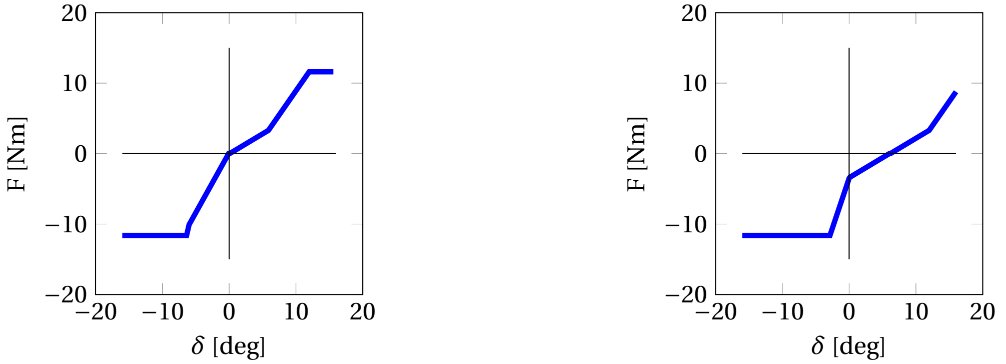
Property	Value
δ_{max}	16°
$\tau_{\text{overspeed}}$	5s
Δt_{α}	3s
F_{max}	11.6Nm



(a) Default feeling in guidance design

(b) Default guidance design, Fig. 3a with shifted neutral point

(c) Guidance design with increased stiffness (severity 0.5)



(d) Guidance design with increased stiffness (severity 1.0)

(e) Guidance design with increased stiffness (severity 0.5), Fig. 3c and shifted neutral point

Figure 3 Haptic profiles in the guidance design, dashed lines indicate default properties

one specific direction. As such, our haptic feedback system will increase the spring coefficient in the direction of the unwanted deflection, leaving the other direction unchanged as shown on Fig. 3c.

The amount of stiffness change is determined by the magnitude of the safe flight envelope excursion, similarly to the amount of stiffness change in previous research to indicate a criticality. [25] Starting at the edge of the safe flight envelope until the flight envelope limit (respectively, the red-dashed line and black line on Fig. 1), the stiffness is gradually increased. Using a generic symbol ν for the different limits of the flight envelope (maximum velocity, max/minimum load factor, maximum angle of attack), the default stiffness of the unwanted direction is multiplied with a factor K_k, determined by the gain K_ν and the severity of the violation:

$$K_k = \begin{cases} 1 & \text{if } \nu < \nu_{\text{prot}} \\ 1 + K_\nu & \text{if } \nu > \nu_{\text{nom}} \\ 1 + K_\nu \frac{\nu - \nu_{\text{prot}}}{\nu_{\text{nom}} - \nu_{\text{prot}}} & \text{else} \end{cases} \tag{1}$$

The severity is defined as the ratio of the violation of the safe flight envelope, $\nu - \nu_{\text{prot}}$, where ν_{prot} is the value at the edge of the safe flight envelope, and the distance between the safe and nominal flight envelope, $\nu_{\text{nom}} - \nu_{\text{prot}}$, where ν_{nom} is the value at the edge of the nominal flight envelope. To illustrate this, the haptic profile with a stiffness change for a severity of 0.5 is shown on Fig. 3c. Increasing the severity to 1 results in a haptic profile shown on Fig. 3d which requires even more force for a backwards stick deflection. In this experiment, K_ν is set to 2 for all limits.

2. *Neutral point feedback*

A shift in the neutral point can be used to indicate a required deflection to follow a certain flight path ([6]) or, in automotive applications, to follow the road ahead. [27] If a positive/push deflection is required, this would result in a haptic profile as shown Fig. 3b. In our scenario, the aircraft is nearing its limit and the required deflection to return to the safe flight envelope can be indicated through the side stick. Since the aircraft dynamics at the different edges of the flight envelope are not equal (i.e., high velocity, angle of attack, and load factor), for each of these limits a required side stick deflection is determined as follows:

Velocity protection ($V > V_{\text{prot}}$) When an overspeed occurs, the speed has to be reduced actively by the pilot by either reducing the throttle, or by pitching up such that kinetic energy is rapidly exchanged for potential energy. The Airbus control law will implement a forced nose-up command (see Subsection II.A), which could be translated to a change in neutral point. Nevertheless, the actual implementation of this signal is not known for this research and is approximated as described below. The main reason for this cue is to inform the pilot that maintaining the stick at zero deflection does *not* solve the flight envelope violation, and action needs to be taken. Note that here our research deviates from the A320 FEP: the nose-up command is not activated when crossing V_{MO}, it is already activated when crossing $V_{\text{MO}_{\text{prot}}}$.

For this research, the nose-up command, and therefore the magnitude of the neutral point shift, is governed by the change in load factor required to bring the positive acceleration to zero. It is determined by starting from the longitudinal equations of motion ([28]), where we assume engine thrust to be parallel to the aircraft longitudinal body axis:

$$T\cos(\alpha) - D - W\sin(\gamma) = m\frac{dV}{dt} \tag{2}$$

The pilot can manipulate the aircraft flight path (γ), through moving the stick. Here, the neutral point is shifted to obtain a flight path angle such that there is no positive acceleration, $\frac{dV}{dt} = 0$. If the aircraft is accelerating before the activation of the neutral point shift, the left part of Equation 2 is not zero and can be rewritten to obtain a steady flight path:

$$\gamma_{\text{steady}} = \arcsin\left(\frac{T\cos(\alpha) - D}{W}\right) \tag{3}$$

Thrust and drag cannot be measured directly, their effects can be measured through accelerometers, mounted on the aircraft body, which therefore must first be rotated to the velocity reference frame:

$$\begin{aligned} T\cos(\alpha) - D &= ma_{x_a} + W\sin(\gamma) \\ &= m\left(a_{x_b}\cos(\beta)\cos(\alpha) + a_{y_b}\sin(\beta) + a_{z_b}\cos(\beta)\sin(\alpha)\right) + W\sin(\gamma) \end{aligned} \tag{4}$$

Combining Equation 3 with Equation 4 then yields the required change in flight path angle for zero acceleration ($\gamma_{\text{steady}} - \gamma$), all expressed in measurable quantities.

As discussed above, the side stick gives load factor commands for high velocities and therefore also a relation between the change in flight path angle and load factor is required. Load factor is governed by the time derivative of the flight path angle, therefore a tuning factor ($\tau_{\text{overspeed}}$) is chosen which is a measure of the recovery speed:

$$n_{\text{req}} = \frac{V}{g}\tan(\dot{\gamma}) = \frac{V}{g}\cdot\tan\left(\frac{\gamma_{\text{steady}} - \gamma}{\tau_{\text{overspeed}}}\right) \tag{5}$$

Angle of attack protection ($\alpha > \alpha_{\text{prot}}$) When the angle of attack is above the maximum value, the required change to bring it back to the protection value should be translated to the side stick. The required change in load factor can be obtained by starting from the effect of pitch rate on load factor:

$$n = V \cdot q \tag{6}$$

Furthermore, the required pitch rate can be approximated by a required change in angle of attack over a certain time, assuming that for short periods of time the change in pitch is dominated by a change in angle of attack. As a desired angle of attack is available (α_{prot}), and by choosing a time, the required change in load factor is determined by:

$$n = V \cdot q \approx V \cdot \frac{\alpha - \alpha_{\text{prot}}}{\Delta t_{\alpha}} \tag{7}$$

This results in one tuning parameter (Δt_{α}) which can be used to indicate how responsive the side stick will move for a given required change in angle of attack. In the current setup, this tuning parameter is set to 3s.

Load factor protection ($n > n_{\text{prot, pos}}$ or $n < n_{\text{prot, neg}}$) When a load factor outside the safe flight envelope occurs, a required change in control inputs can readily be obtained since side stick inputs are proportional to a change in load factor. The required load factor in case of positive load factors is $n_{\text{prot, pos}} = 2.0$g, in case of negative load factors $n_{\text{prot, neg}} = -0.5$g, resulting in a required stick deflection:

$$\delta_n = \begin{cases} \left(n - n_{\text{prot, pos}}\right) \cdot \frac{n_{max,pos}}{\delta_{\text{max}}} & \text{if } n > n_{\text{prot, pos}} \\ \left(n_{\text{prot, neg}} - n\right) \cdot \frac{n_{max,neg}}{\delta_{\text{max}}} & \text{if } n < n_{\text{prot, neg}} \\ 0 & \text{else} \end{cases} \tag{8}$$

When this haptic feedback system is implemented, it presents the pilot with continuous feedback which uses the stiffness to indicate an undesired deflection, and a shift in neutral point to show the required deflection to return to the safe flight envelope. The stiffness change and neutral point shift can occur simultaneously, for example in Fig. 3e where a positive neutral shift is combined with an increased stiffness for backwards deflections. The combination of these two cues might result in unacceptable high forces required to move the side stick. This is prevented with the implementation of the maximum force, resulting in a flat slope on the haptic profile. The remainder of this paper discusses the results of an experiment to evaluate both cueing methods.

III. Method

To evaluate the haptic interface designs, an experiment was performed which uses the same setup as used in a previous experiment which investigated the 'cueing haptic feedback', see Ref. [8].

A. Independent Variables

The experiment had a between-participants design, with one independent variable. The participants were divided in three groups: the cueing group, the guidance group, and the manual (no-haptics) group. Each group (12 participants per group) performed two blocks of four runs each, elaborated below, and summarized in Table 2.

In the first block, participants were presented with one of the three haptic support conditions. Literature found that an increasing amount of feedback in this initial stage, results in worse performance when that feedback is removed: the "guidance hypothesis". [29] To investigate the consequences of removing the feedback in our application, all participants performed a second block in the manual, no-haptics condition.

The results of the cueing group (12 participants) were obtained from our previous experiment, see Ref. [8], which had the exact same experimental setup. Twenty four new participants were invited and numbered in sequence of experiment participation. Even-numbered participants were placed in the guidance group group, all odd-numbered participants are part of the manual group. A total of 36 participants results from combining the previous and present experiment groups.

Table 2 Experimental design

Block	1				2			
Run	1	2	3	4	5	6	7	8
Cueing group	Cueing haptics				No haptics			
Guidance group	Guidance haptics				No haptics			
Manual group	No haptics				No haptics			

B. Participants and Instructions

For this experiment, data from 36 pilots (1 female, 35 male) with a current Private Pilot License (PPL) or Light Aircraft Pilot License (LAPL) license were used. As these pilots are not Airbus pilots, they were reminded that the aircraft model used has a mass of 64,000kg and had to be handled with more care than a general aviation aircraft. The experience of the three different groups can be found in Tables 3, 4 and 5. A visual comparison of the flight hours per group is shown Table 4. A Kruskal-Wallis rank sum test did not show statistical significant differences in experience between groups ($\chi^2 = 3.17$, $p > 0.2$).

Table 3 Participants in the manual group

Participant	Age	Flight hours	License
M1	65	400	PPL
M2	52	1,500	CPL / IR / FI
M3	66	1,860	PPL / IR
M4	20	150	PPL
M5	62	430	PPL
M6	57	180	LAPL
M7	49	420	CPL
M8	20	82	PPL
M9	62	175	PPL
M10	50	200	PPL
M11	20	55	PPL
M12	23	65	PPL
Mean	45.5	459.8	-
Std.Dev.	19.1	590.4	-

FI	Flight Instructor
IR	Instrument Rating

Table 4 Participants in the cueing group

Participant	Age	Flight hours	License
T1	34	116	PPL
T2	25	80	LAPL
T3	49	205	PPL / E-IR
T4	40	630	PPL
T5	48	350	PPL
T6	24	500	PPL
T7	66	900	PPL
T8	46	250	LAPL
T9	48	500	PPL
T10	50	400	PPL
T11	33	240	PPL
T12	53	600	PPL / E-IR
Mean	43	397.6	-
Std.Dev.	12.3	240.2	-

E-IR	Enroute-Instrument Rating

Table 5 Participants in the guidance group

Participant	Age	Flight hours	License
G1	67	475	PPL
G2	57	300	PPL / IR
G3	26	100	PPL
G4	30	78	PPL
G5	44	170	PPL
G6	50	80	PPL
G7	43	150	PPL
G8	47	500	CPL
G9	71	300	PPL
G10	52	250	PPL
G11	50	200	PPL
G12	60	200	PPL
Mean	49.8	233.6	-
Std.Dev.	13.3	140.5	-

IR	Instrument Rating

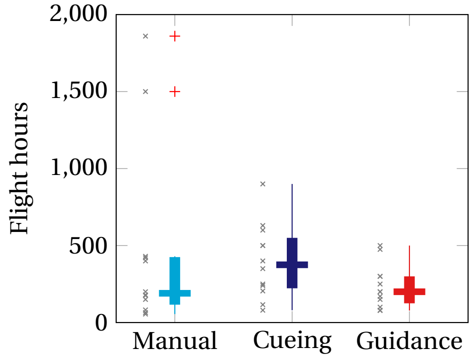


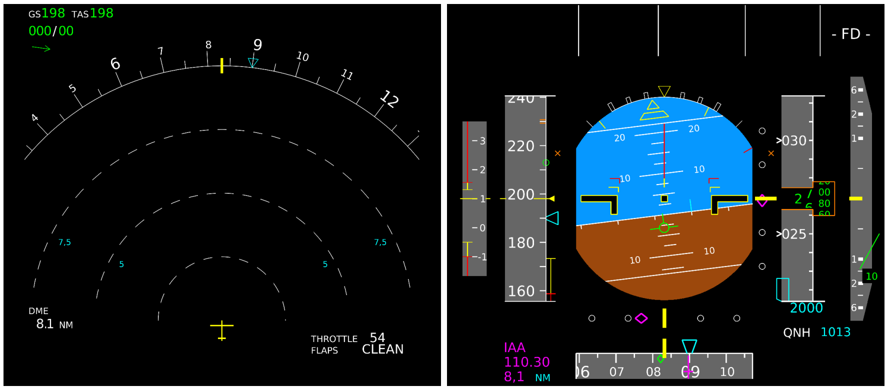
Figure 4 Flight hours per group

Participants were instructed to always remain within the nominal limits of the flight envelope (black line on Fig. 1) which are shown on the PFD using the red indications proposed in Ref. [30]. Additionally, it was mentioned that a simulation run would stop when the aicraft reached an altitude of 50ft above ground level, irrespective of any other event/performance.

C. Experimental Setup

The experiment was performed in the Human Machine Interaction (HMI) research simulator of Delft University of Technology. It is a fixed-base simulator, with a near 180° outside field-of-view, used in the first officer position of which an inside-view is shown in Fig. 5. Since the pilot was sitting in the first officer position, the display to his front-left was the Navigation Display (ND) showing a top-down overview of the situation, shown in Fig. 6a, combined with a basic engine N1-indication and slats/flaps indication. The display right in front of the pilot was the PFD showing the critical flight states, shown in Fig. 6b, which included display indications used to show *why* and *when* the haptic feedback is active. [30] Next to the visual information, auditory warnings were presented when the aircraft angle of attack was above the maximum value, and when the velocity was above the maximum velocity.

Figure 5 Inside view of the HMI flight deck



(a) ND **(b) PFD**

Figure 6 Flight deck display setup used in the experiment

A custom-made, hydraulically driven side stick with programmable dynamic properties is located at the right-hand side and was configured to Airbus side stick properties. [21] To the left, a throttle quadrant is present which was used to control the throttle and high lift device settings. Centrally placed, a Boeing 737 Mode Control Panel (Airbus terminology: Flight Control Unit (FCU)) enabled the interface with the heading, velocity and altitude references on the displays. Outside visuals were generated using FlightGear* and showed the airport infrastructure, terrain and important buildings at the airport. A proprietary A320-like flight dynamics model, including control laws from the German Aerospace Center (DLR), was used as the simulated aircraft. [31]

---

*Open source flight simulator available at `http://flightgear.org`

The nominal, no-haptics control device settings for this experiment, including mass (m), spring coefficient (k), damping coefficient (b) and maximum deflection ($\delta_{\max}$), for both side stick axes are given in Table 6.

The 'cueing haptic feedback' used: (i) a sawtooth shape of intensity 0.282Nm, duration 1s and frequency 2Hz when exiting the safe flight envelope, (ii) a sawtooth shape with varying intensity proportional to the relative distance of the protection and flight envelope limit, duration 0.5s and frequency 2Hz when remaining outside the safe flight envelope and (iii) a stick shaker for low velocities as a sinusoid with frequency of 20Hz and magnitude 0.426Nm. Further details can be found in Ref. [8].

'Guidance haptic feedback' used an increase in spring coefficient to maximal twice the nominal stiffness, and neutral point shifts to provide recommended side stick deflections as discussed in the design section.

Table 6 Control device in the experiment

Property	Value
m	0.2 kg m^2
$k_{\text{lon}_{\text{nom}}}$	36.3 Nm/rad
b_{lon}	0.4 Nm s/rad
$\delta_{\text{lon}_{\max}}$	0.279 rad
$k_{\text{lat}_{\text{nom}}}$	21.8 Nm/rad
b_{lat}	0.4 Nm s/rad
$\delta_{\text{lat}_{\max}}$	0.314 rad

D. Experiment Scenarios

The haptic feedback system was designed to communicate the 'proximity of the flight envelope limits' to the pilot and therefore required an evaluation at these limits. In analogy to our previous experiment, the scenarios presented a stringent flight path, as discussed below, followed by the emergency scenario encountered during each flight.

1. Flight path

Each run was started when the aircraft was flying 140kts (72.0mps) at 2,500ft (762m) with slats and flaps set for approach (Airbus setting 3), overhead the threshold of runway 23 of Zoersel (Belgium) and aligned with the respective runway. This location was chosen as it has no special terrain features close-by, and the runway was not visible from the starting position as illustrated on Fig. 8a. Additionally, the auto-throttle was set to 140kts and activated, reducing the variability of the initial aircraft state when the event was triggered and should provide more consistent results. From this position, pilots were presented with visual markers (squares of 60m by 60m) on the outside visual display to help them fly a flight profile consisting of six 'hills', for which an example path is presented in Fig. 7 and visualized on the outside visual as shown in Fig. 8.

A hill was 2.27NM (4,200m) long and had one of three possible amplitudes: 150ft (45.72m), 300ft (91.44m) or 500ft (152.4m). Combining six hills yielded a saw-tooth trajectory. The flight path started with a horizontal segment of 0.41NM (750m) and one hill of the smallest amplitude as run-in. This was followed by a randomized order of hills such that each amplitude of hill occurred twice in the flight path. Each flight ended with a horizontal segment of 0.54NM (1,000m) as run-out. This setup of hills was chosen as it was expected that it allowed the results to be evaluated for each hill separately. Eight different realizations of the randomization were obtained to present pilots with variability in the scenarios. The resulting trajectories are all shown in Appendix A.

2. Emergency scenario

As the pilots of our initial experiments did express the potential added value of the haptic feedback system in a windshear event, this event is re-used for this research. [32] A wind shear is a meteorological phenomenon where wind velocities are locally rapidly changing, and can be caused by a large cylinder of air suddenly "droping" towards the earth. During such a wind shear event, downdrafts can push the aircraft dangerously close to the ground. [33] To recover from this event, the pilot has to move as close to the stall limit as possible to prevent further height loss and maximise aircraft performance. [19]

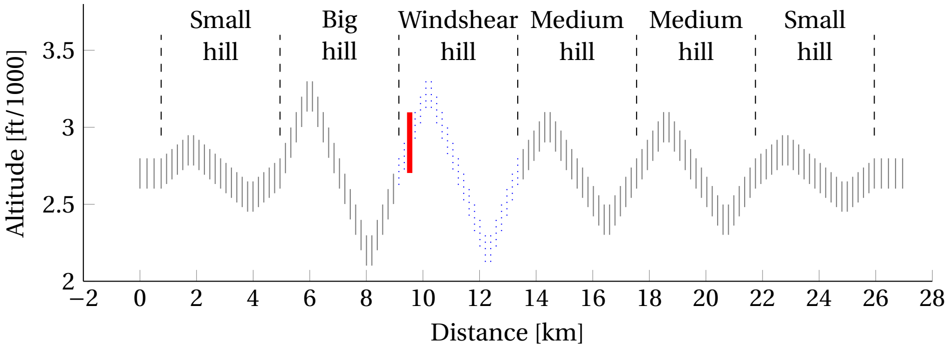


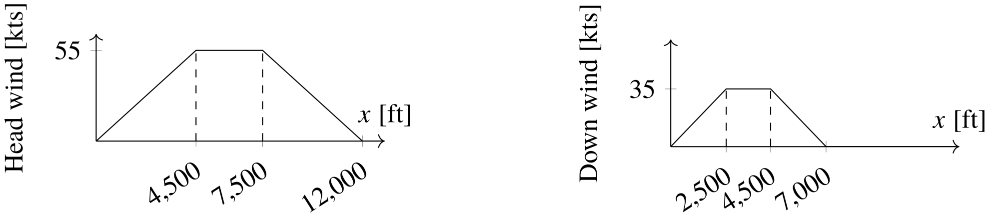
Figure 7 Flight path side-view, solid black vertical lines indicate "fly-through gates" shown on the outside visual; the thick red line indicates the trigger point of the windshear (not shown on the outside visual); the dotted blue lines indicate the windshear section used in our evaluation

(a) View at start of run

(b) Perspective view on flight path (viewing angle is for illustrative purpose only, never encountered during flight)

Figure 8 Example of the outside visual flight path visualization

The windshear in each run was *always* started when the aircraft moved through the visual marker of the windshear hill at an altitude of 2900ft (883.92m). Each flight path contained two hills with the largest amplitude, only one of them was selected at random to contain the windshear trigger point. The windshear itself was modeled by both a head-on and top-down component as shown in Fig. 9. [33] Once the windshear was initiated, the visual and aural warning trigger, and the pilot had to apply the windshear recovery procedure as stipulated in Fig. 31a, which was based on the Airbus Flight Crew Operating Manual. [19]



(a) Headwind component

(b) Downwind component

Figure 9 Windshear component distribution

When providing only windshear as the emergency scenario during each run, pilots might anticipate this event, even in the first run. To prevent this, two more checklists for an emergency were presented to the pilots beforehand: the actions required for a single engine stall (Fig. 31b), and for a sudden center of gravity shift (Fig. 31c). Therefore, pilots were expecting one of these three emergency scenarios, but were unaware of what scenario was actually triggered. Note that the checklists presented in Fig. 31a and 31b are heavily modified from the FCOM, and the checklist for the sudden center of gravity shift is non-existing in the FCOM.

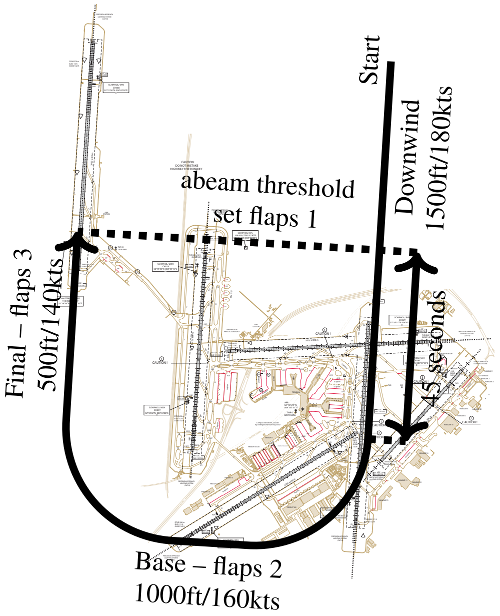


Figure 10 Traffic pattern flown to runway 36L at Schiphol (Schiphol layout from AIP [34])

E. Experiment Design

To allow pilots to become sufficiently familiar with the simulator and the haptics (if applicable), a familiarization phase was performed, followed by measurement runs.

1. Familiarization

After a briefing on the simulator safety procedures, all pilots were explained the controls and displays by presenting the flight envelope (an image similar to Fig. 1), and the PFD (Fig. 6b) to the pilot. In this setup, no aircraft model was used, yet the flight envelope state was changed directly (hence changing the velocity and load factor) and all visual and auditory cues were elaborated. After that, pilots in the cueing or guidance groups *felt* the design rationale behind the haptic feedback design using an image similar to Fig. 3.

Then the aircraft model was introduced to the pilot by flying a traffic pattern twice to a final approach at Schiphol (EHAM) as shown on Fig. 10, without the haptic feedback, hence focusing on familiarization with the model. Pilots were instructed to follow the instructions as indicated. Some pilots encountered a stall and/or an overspeed condition during these first runs. If the pilots did not hit one or both limits, they were asked to deliberately explore those boundaries to ensure that all pilots encountered them before the measurement runs.

2. Measurements

The measurement phase contained eight realizations of the flight path presented above. They were flown in a randomized fashion, distributed over all participants using a Latin-square distribution. Each group performed two blocks of four runs, with a break in between, with haptic feedback as shown in Table 2.

After each run, pilots were asked to indicate their workload using a Rating Scale Mental Effort (RSME) rating [35], and complete a post-run situation awareness questionnaire, to indicate how helpful the visual, auditory and haptic (if supplied) elements are. They also provided a misery scale rating to measure and account for possible effects of motion sickness. [36] Once this was completed, pilots were informed on how much time they spent inside the flight envelope, which they had to maximise.

After each block of four runs, pilots were asked to complete a questionnaire with a modified Cooper-Harper rating scale ([37]), and a Van der Laan-rating scale. [38] After the experiment was completed, pilots were asked to complete a post-experiment questionnaire, which contained a number of questions with Likert-scales on how they experienced the haptic feedback system.

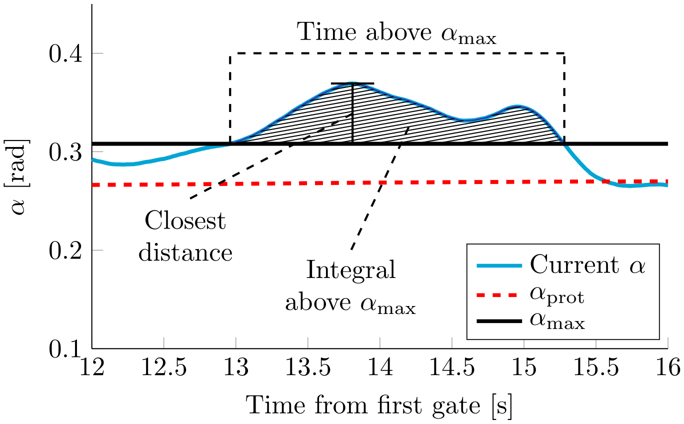


Figure 11 Time trace of velocity with safety metrics indicated

F. Dependent Measures

The dependent measures are split into objective and subjective measures.

1. Objective measures

The objective measures are retrieved from the windshear recovery procedure, and focus on performance and safety. To illustrate why these metrics were chosen, a time excerpt of a windshear recovery is shown in Fig. 11. Another example is further elaborated in the results section, for now it is sufficient to understand that this shows a participant aiming for the best performance of the aircraft, flying close to the maximum angle of attack.

Looking at the example, one can argue that a safe flight is performed when the aircraft state is within the flight envelope limits, indicated with the solid black line representing α_{max}. Although participants are instructed to stay within the limits at all times, at certain moments in time the pilot could control the airplane beyond these limits. A first performance metric was therefore the *time spent outside the angle of attack limits*.

Participants can push the aircraft by flying very close to its limits, even above the limits, or they can choose to remain well away from the limits. A straightforward metric to determine this safety definition was the *maximum angle of attack obtained relative to the flight envelope limit*: it can indicate how close to the limits the participant dares to control the airplane.

Time by itself only informs about the length of the limit violations, it does not take into account the closest distance: two different limit violations might be of equal time, yet one just slightly over the limit while another one is in a deep stall. As such, a safety metric combining both the time and the magnitude of the violation was the *integral of the angle of attack over the flight envelope limit*.

One additional performance metric on the overall windshear recovery procedure was used: the *total amount of altitude lost during the recovery*. Although not communicated to the participants, the maximum altitude lost from the windshear initiation to the end of the windshear recovery is considered here as an indication of how much of the available aircraft performance is utilized by the participants. Best performance is achieved when this amount of altitude lost is minimum.

Previous research showed that the level of risk humans experience is mostly kept the same when support increases, i.e., risk homeostasis. [39] This was found in an automotive studied where supplying haptic feedback resulted in participants driving at higher velocities. [40] For the current experiment, risk homeostasis was also expected, and can be defined by improved performance, combined with objective safety metrics closer to the maximum value, as pilots obtain a better awareness of the risk involved when supplied with haptic feedback.

2. *Subjective measures*

Subjective measures are obtained by asking the pilot for an opinion, or experience. The categories and measures are:

- Workload: after each run, the pilot was asked to provide a RSME rating [35]
- Situation awareness questions: after each run, the pilot was asked to answer two questions on a linear scale (0–100) ranging 'Never' left (0), and 'Always' right (100):
  1) Did you have the feeling you were in control of the situation?
  2) Did you have the feeling you missed critical information?
- Usefulness:
  1) Pilots were asked after each run to rate the usefulness of all display and haptic elements on a Likert scale
  2) After each block, pilots were asked to provide a modified Cooper-Harper rating
  3) After each block, pilots were asked to fill a Van Der Laan-questionnaire
- Pilot experience: after the experiment, the pilot was asked to fill in a questionnaire regarding the experience with the haptic feedback system.

The questionnaire presented to the pilots after the experiment used five point Likert-scales where all points are labeled. A different set of questions was presented to the participants in the manual, no-haptics group because they did not experience any haptic feedback at all.

G. Hypotheses

In the experimental evaluation, the manual group served as a baseline to compare pilot behaviour during windshear recovery. The expected behaviour of the other two groups is summarized in Fig. 12 and explained in the following.

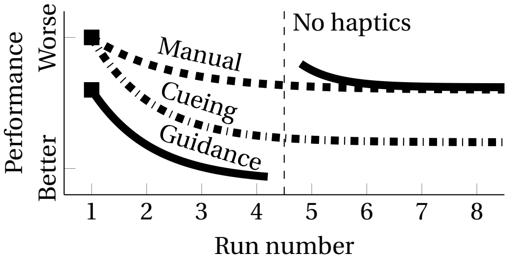


Figure 12 Schematic representation of the expected results

We expected the *cueing group* to perform initially at the same performance level, yet have a faster learning rate over the first four runs, have an improved performance level at Run 4, and keep performance equal when no haptic feedback is provided in the final four runs, i.e., no after-effects. In terms of dependent measures, this means no change in performance/safety between the manual and cueing group at Run 1. At Run 4, the cueing group has an improved performance between-groups compared to the manual group, and within-group compared to Run 1. Comparing the metrics of Runs 4 and 8 within the cueing group, should give no differences to indicate no after-effects.

The *guidance group* was expected to have an improved performance from the first run as long as haptic feedback is provided, but when this haptic guidance is not provided, (Run 5), we expected the performance to suddenly worsen following the "guidance hypothesis". [29] In terms of dependent measures, this would translate to improved performance and safety margins at Run 1 when between-groups comparing the manual and guidance groups. At Run 4, the guidance group has an improved performance between-groups compared to the manual group. After-effects were expected to show up when comparing performance and safety margins of Runs 4 and 5 within the guidance group.

We expected pilots to perceive the *cueing haptic feedback* as a useful source of information, yet the information still needs to be interpreted. Therefore, the subjective workload ratings at Run 1 of the cueing group were expected to not differ from the manual group, yet indicate this group to have an improved situation awareness. At Run 4, the workload of the cueing group is expected to be lower due to familiarization.

The *guidance haptic feedback* was expected to be supporting pilots from the first run, yet it can be less clear in the reason why it provides a cue. It was expected that the subjective workload ratings at Run 1 of the guidance group is lower compared to the manual group, yet deteriorate when no haptic feedback is supplied anymore (at Run 5). Subjective situation awareness ratings of the manual and guidance groups were expected to be similar.

For both haptic designs, we expected the Modified Cooper-Harper ratings and Van Der Laan-ratings to improve. The remainder of this paper looks into the results of this experiment and discusses whether our hypotheses can be supported.

IV. Results

Before the metrics are discussed, Subsection IV.A shows one example case where a pilot used the guidance haptic feedback system and two other noteworthy events which happened during the experiment. Next, the objective and subjective measures are presented in, respectively, Subsection IV.B and IV.C. Answers to the debriefing questionnaires are presented in Subsection IV.D. For reference, all flown trajectories included in the analysis are shown in Appendix A.

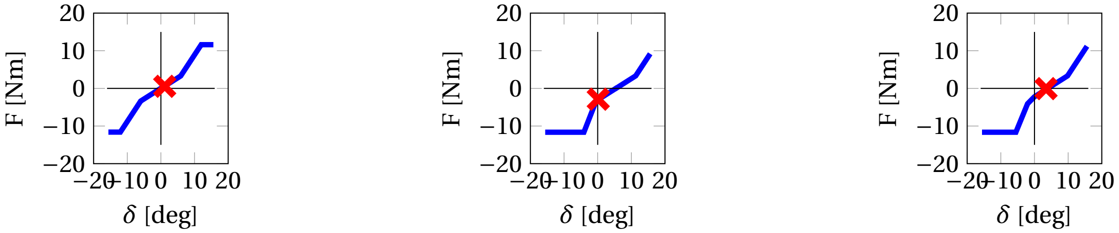
When presenting data using box plots, medians are indicated using a horizontal thick line, outliers are indicated using plus-signs; all individual data points are presented next to the boxes using crosses. Furthermore, statistical analyses are performed in R ([41]) and results are only reported if p-values of 0.05 or lower are found. Results are compared both within- and between-groups: within-group the differences of Runs 1, 4 and 8 are examined, the between-group comparison investigates the difference between groups at Runs 1, 4 or 8. Tests are performed using a Kruskal-Wallis rank sum test which indicates whether there is a statistically significant difference. If a difference is found, a post-hoc test is performed using a pairwise Wilcox test where p-values are adjusted using the method proposed by Benjamini and Hochberg. [42]

A. Time trace examples

This section discusses three time traces from the experiment: the first shows a participant using the guidance haptic feedback, the second shows the only crash which occurred, and finally a design flaw of the guidance haptic feedback is illustrated.

1. Example use of guidance haptic feedback

This is an example where Guidance Participant 2 during his first run made use of the guidance haptic feedback system proposed in Subsection II.C. Three haptic profiles are given in Fig. 13 which correspond to the frames indicates on the time traces for pitch angle, angle of attack and control device deflections, respectively, Fig. 14a, 14b, and 14c. The haptic feedback can be seen in the haptic profiles, and on the control device deflection plot: the neutral point (δ_{np}) is the stick shift by the haptic feedback, the actual control device deflection (δ_{cd}) is the sum of haptic feedback and the human operator.

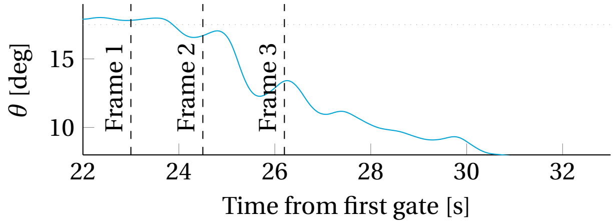


(a) Frame 1 **(b) Frame 2** **(c) Frame 3**

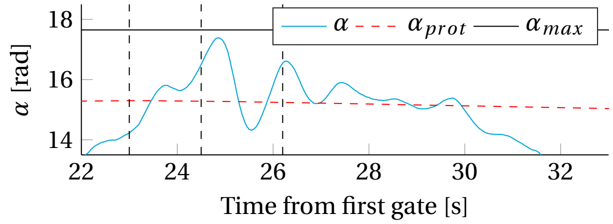
Figure 13 Haptic profiles for frames indicated in Fig. 14, cross indicates the current state

The windshear recovery procedure requires the pilot to use all of the available performance of the aircraft, which can be achieved by operating the aircraft near the maximum angle of attack. Initially, the pilot has to obtain a pitch angle of 17.5°, which is achieved at Frame 1 as can be seen on Fig. 14a. Here, the current state is still within the safe flight envelope and the corresponding haptic profile shows the nominal stick feeling on Fig. 13a. Next, Frame 2 shows the participant exerting back pressure on the stick to maintain pitch and to avoid the aircraft from descending, despite the haptic feedback indicating that a pitch down input is required to return to the safe flight envelope (shifted neutral point, Fig. 14c, and increased stiffness, Fig. 13b). Subsequently, the participant notices that a sustained back pressure brings the aircraft too close to the stall and starts following the haptic feedback cues to operate the aircraft near its limits. This can be seen by the matching of the neutral point and the actual control device position on Fig. 14c, for which one haptic profile is given in Fig. 13c.

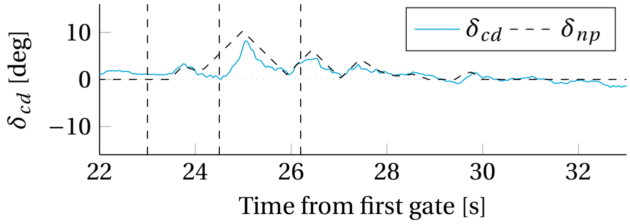
This example shows that this pilot used the haptic feedback, even in the first run where participants were expected not to be fully familiar with the aircraft model, task, or emergencies procedures. The metrics presented next can be used to further investigate whether this is a one-off example, or participants can indeed use the haptic feedback effectively from the first encounter.



(a) Pitch angle



(b) Angle of attack



(c) Control device deflection, i.e., input from the side stick to the FCC

Figure 14 Time traces of Guidance Participant 2 using the haptic feedback during the recovery in Run 1

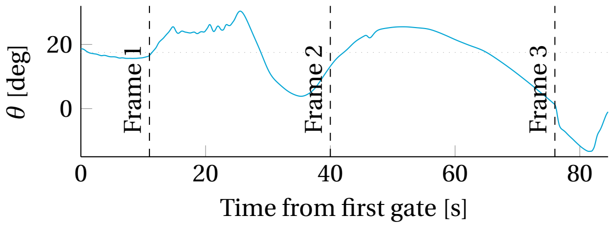
2. Only crash of the experiment

This example shows the first run of Partipant 7 in the manual group, where Fig. 16 shows the haptic profiles for the time frames shown on the time traces in Fig. 15. As this was the first run, this was also the first time the participant encountered the flight path and windshear. After the windshear warning, the participant aimed for a pitch angle of 17.5°, as can be seen on Fig. 15a before Frame 1. At Frame 1, the participant notices that the aircraft is not climbing anymore (Fig. 15d), and as indicated in the checklist, increases nose-up input (Frame 1 at Fig. 15c, Fig. 16a).

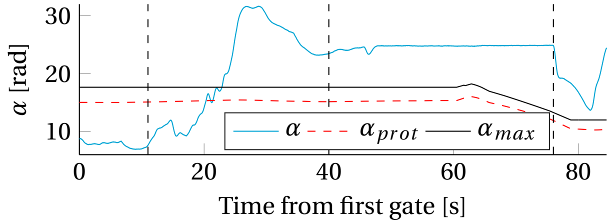
After this, the windfield suddenly pushes the angle of attack above the maximum, as shown on Fig. 15b, and on the aural stall warning, the particpant starts applying *more* back pressure on the side stick. The input, shown on Fig. 15c, one snapshot at Frame 2 in Fig. 16b, shows a negative (pull) input of more then $50s$ during which sustained visual and aural stall warnings are provided, yet no haptic feedback as the participant is part of the manual group. About $60s$ after the windshear trigger, the participant retracts flaps which reduces in a reduced maximum angle of attack enlarging the problem.

Near the end of the flight, at Frame 3, the participant starts using a positive input and starts to solve the angle of attack excursion. Nevertheless, the action is too late and not sufficient altitude is left for the recovery. The flight is, as indicated in the briefing, stopped 50ft above ground level.

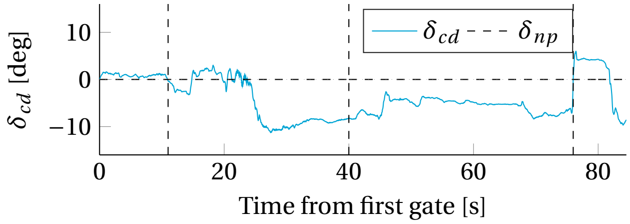
After this run, the participant filled the required questionnaires and he was told that he had a sustained stall, reducing his time inside the flight envelope. Additionally, he was reminded that one of the windshear recovery items indicates *not to change configuration*, i.e., do not change flaps. During the next run, the participant was able to recover from the windshear and complete the next flights. Although this crash is a one-off example, the only crash throughout the entire experiment campaign did occur when no haptic feedback of any form was present.



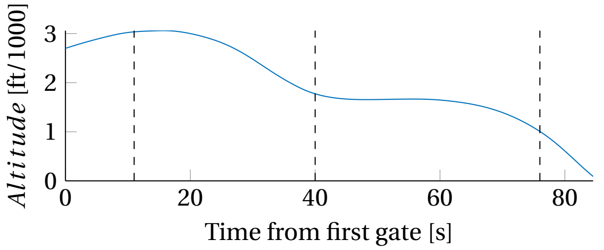
(a) Pitch angle



(b) Angle of attack

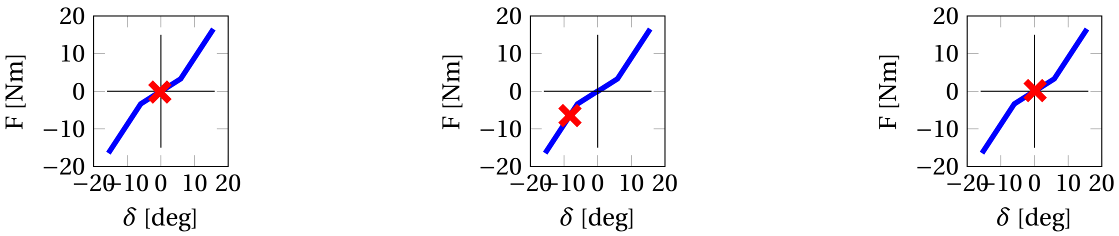


(c) Control device deflection, i.e., input from the side stick to the FCC



(d) Altitude

Figure 15 Time traces of the crash which occurred during Run 1 of Manual Participant 7

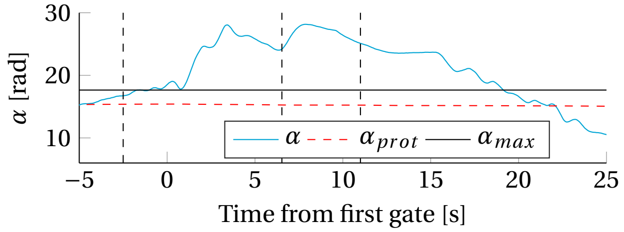


(a) Frame 1 **(b) Frame 2** **(c) Frame 3**

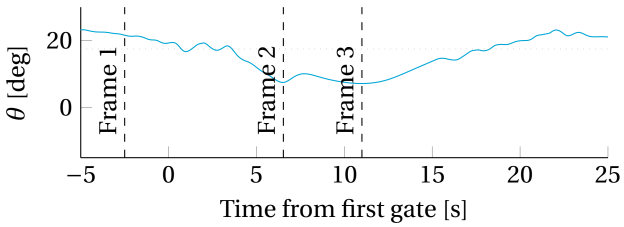
Figure 16 Haptic profiles for frames indicated in Fig. 15, cross indicates the current state

3. Guidance haptic feedback design issue

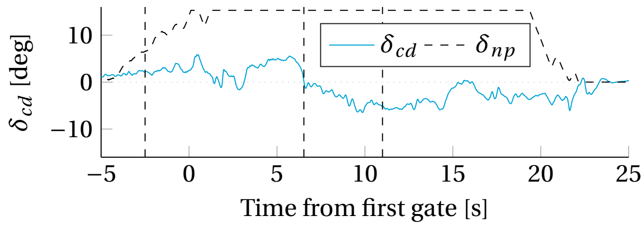
The third example shown is the first run of Guidance Participant 5 with time traces in Fig. 17 and corresponding haptic profiles in Fig. 18. It is a show-case of a flaw in the current design of the guidance haptic feedback: a possible haptic ‘lock-in’ where eventually all haptic feedback is lost. The origin of the flaw can be traced back to the method used to guarantee that the pilot always has final authority over the automation on the side stick: with a changing neutral point and an increasing stiffness, a certain deflection of the side stick might require a force level which is not reasonable anymore to be achieved. Therefore, as mentioned in the design section, the maximum amount of force required to move the side stick is limited to 11.6Nm.



(a) Angle of attack

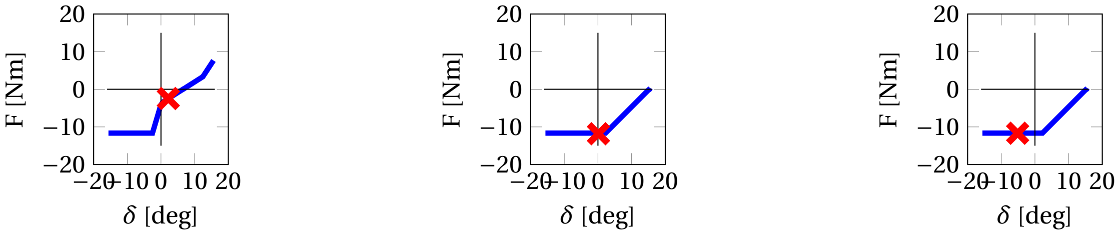


(b) Pitch angle



(c) Control device deflection, i.e., input from the side stick to the FCC

Figure 17 Time traces of the design flaw which occurred during Run 1 of Guidance Participant 5



(a) Frame 1

(b) Frame 2

(c) Frame 3

Figure 18 Haptic profiles for frames indicated in Fig. 17, cross indicates the current state

Looking at the angle of attack before the windshear trigger, indicated as negative times on Fig. 17a, at Frame 1 the state of the aircraft was already near the limits and the participant was informed as such using a shift in neutral point and increased stiffness for negative inputs as shown on Fig. 18a. At that moment, the participant was trying to follow the tunnel-in-the-sky presented on the outside visual and needed all available performance to do so as he was slightly below the tunnel. On the windshear trigger, the condition worsened and the haptic feedback provided a full stick forward input which is maintained throughout the time trace as can be seen on Fig. 17c.

Starting from Frame 2, the participant reached a backwards pressure of the limiting 11.6Nm, resulting in a flat haptic profile as in Fig. 18b. The participant maintained the backwards input, nevertheless, one level of force on the side stick is required for all negative deflections. As a result, the participant was inputting significant pitch-up commands for a significant time, with one snapshot in Frame 3 (Fig. 18c), possibly without noticing the magnitude of the input due to the lack of a force gradient: the participant was ‘locked-in’. Additionally, in this situation, whatever the neutral point shift or stiffness changes, the participant is not able to perceive this feedback.

One possible solution to avoid such a haptic ‘lock-in’ is to not use a flat haptic profile, yet implement a very shallow slope. By using a minimal slope, the participant is able to distinguish between different magnitudes of input, and the neutral point shift by the haptic feedback can still be observed. In the runs after this, the participant was able to apply the recovery procedure without entering a haptical lock-in.

B. Objective measures

Objective measures are directly retrieved from the simulation data and are discussed in the following.

1. Altitude lost during recovery

The amount of altitude lost during the windshear recovery is shown in Fig. 19.For the manual and cueing groups, a learning process is present: from Run 1 onward, performance increases and less altitude is lost during the recovery. Comparing Runs 1, 4 and 8 within one group shows a statistical significant difference for the manual group ($p < 0.05$, $\chi^2 = 6.87$) where the post-hoc indicates that only Run 1 and Run 8 contain a difference ($p < 0.05$). This indicates that the group improved performance from the start till end, yet at Run 4, the participants were not fully learned yet. For the cueing group, differences are present ($p < 0.005$, $\chi^2 = 12.87$) between Run 1 and Run 4 ($p < 0.005$), and between Run 1 and Run 8 ($p < 0.005$). This confirms that after Run 1, the participants quickly learned how to handle the windshear recovery, yet were not able to reach the final performance at the first run. Within the guidance group, no large differences are observed, which is confirmed by no significant differences between Runs 1, 4 and 5.

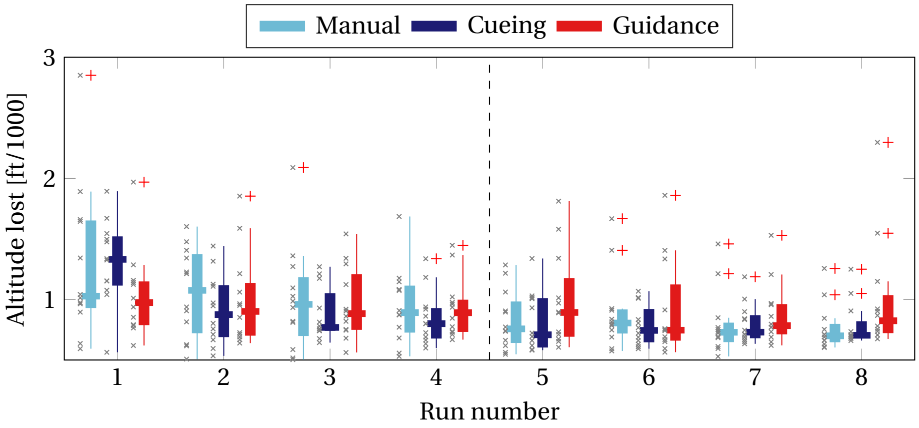


Figure 19 Altitude lost during windshear recovery

At first glance, the guidance group seems to have a lower median and lower spread compared to the other groups. At Run 4, the differences do not seem large. Nonetheless, no statistical significant results are found by the Kruskal-Wallis test. When no haptic feedback is supplied, performance of the guidance group seems to have a larger spread for worse performance. For Run 8, there is a ‘near statistical significant’ result of the Kruskal-Wallis test ($p = 0.06$, $\chi^2 = 5.64$), and the post-hoc test points to a difference between the guidance and cueing group ($p = 0.09$), as well as between the guidance and manual group ($p = 0.09$), indicating that the observation can be right, yet not supported by clear statistical significance.

2. Time above maximum angle of attack

The duration of the flight spent above the maximum angle of attack, Fig. 20 – a metric that was also communicated to the participants received after each run which they had to optimize – clearly shows the learning effect: starting from Run 1, this metric reduces, and thus participants also spent less time with the stall warning active. Within each group, this learning effect is confirmed with statistical tests: the manual group ($p < 0.005$, $\chi^2 = 12.60$) has significant differences between all runs (Run 1 and 4 $p < 0.05$, Run 1 and 8 $p < 0.005$, Run 4 and 8 $p < 0.05$) indicating that they improved over the course of the entire experiment. The cueing group has significant differences ($p < 0.001$, $\chi^2 = 23.50$) between Run 1 and 4 ($p < 0.001$), as well as between Run 1 and 8 ($p < 0.001$), indicating that they improved performance from Run 1 to 4, yet kept there performance level afterwards. The guidance group has only statistical differences ($p < 0.05$, $\chi^2 = 7.11$) between Runs 1 and 4 ($p < 0.05$) indicating an improvement over the first block of four runs, yet the lack of significant differences between Runs 1 and 5 might indicate the slight deterioration of the metric which is also visible on Fig. 20.

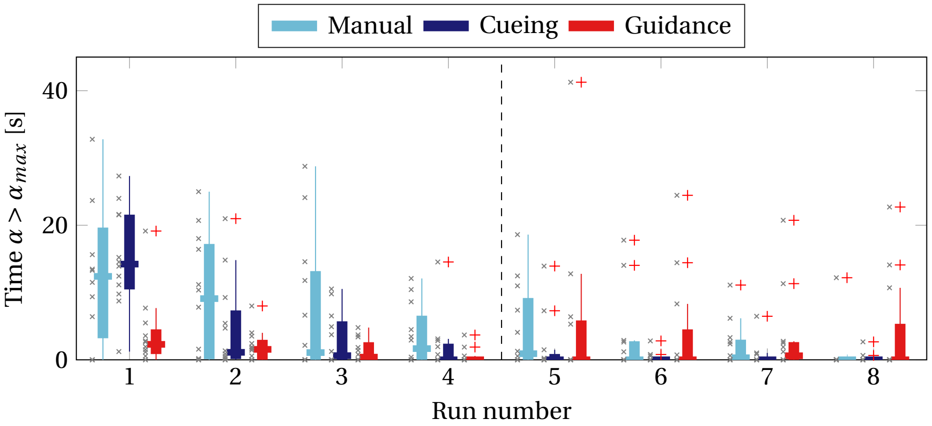


Figure 20 Time with angle of attack above maximum value during windshear recovery (Manual Participant 7, $61.7s$, not shown)

The time above the maximum angle of attack shows that at Run 1 the guidance group has a better performance compared to the two other groups. This is confirmed by a significant result of the Kruskal-Wallis test ($p < 0.01$, $\chi^2 = 9.92$), and post-hoc analysis indicates differences between the guidance and cueing group ($p < 0.005$), and 'near significant' difference between the guidance and manual group ($p = 0.10$). At Run 4, the boxplot of the manual group shows more participants still encounter a stall warning, yet this is not supported by statistical evidence. At Run 8, the guidance group appears to have three participants encountering a stall warning, compared to one in each of the other groups, but without statistical significance.

Another between-groups observations can be made regarding the spread of the data on the first block of four runs: each run of the manual group has the largest spread, and each of the guidance group has consistently the lowest spread. This might indicate that the guidance feedback is more stringent in its communication of the flight envelope limits, compared to the cueing feedback, and especially compared to the manual group.

3. Highest angle of attack

Fig. 21 shows how the largest angle of attack obtained throughout the recovery relates to the maximum angle of attack, i.e., it indicates how close pilots operate the aircraft near the limits, and if a maximum angle of attack violation is made, its magnitude. Again, a learning effect is present: the metric improves over the runs. Within the manual group, the plot shows that the learning effect seems to be rather slow, and this is confirmed by the statistical analysis: there are statistically significant difference ($p < 0.01$, $\chi^2 = 9.28$), more specifically, the post-hoc test reveals significant differences between Runs 1 and 8 ($p < 0.05$), and Runs 4 and 8 ($p < 0.05$), not between Runs 1 and 4.

Combining the lack of statistical difference between Run 1 and 4, and the large spread of the data indicates that participants in the manual group need more time to learn how to properly control the angle of attack. The values obtained for the cueing group show the worst results on the plot but improves over the four runs: the results have significant differences between runs ($p < 0.001$, $\chi^2 = 19.18$) and post-hoc indicates differences between Runs 1 and 4 ($p < 0.001$) and Runs 1 and 8 ($p < 0.001$), indicating that indeed participants made a significant difference in the first four runs and maintained this control strategy during the subsequent runs. For the guidance group, the Kruskal-Wallis test indicates 'near statistical' differences ($p = 0.076$, $\chi^2 = 5.15$) between runs, yet the post-hoc test only indicates 'near statistical' differences between Runs 1 and 4 ($p = 0.06$) and Runs 1 and 8 ($p = 0.07$). The plot does show a decrease in median during the first four runs, and an increase in spread when no haptic feedback is provided anymore.

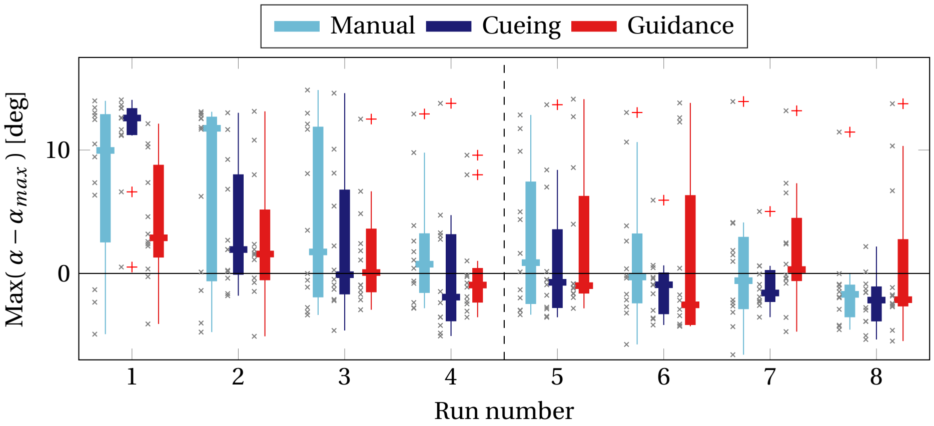


Figure 21 Highest angle of attack obtained during the windshear recovery, relative to the maximum angle of attack (positive values result in a stall warning)

At Run 1, the guidance group seems to have the lowest, yet still positive, safety margin compared to both other groups. This is confirmed partly as significant differences between groups ($p < 0.01$, $\chi^2 = 9.61$) are present, yet only between the cueing and guidance groups ($p < 0.005$) indicating that the participants in the guidance group have consistently lower maximum angle of attack violations compared to the cueing group. No significant difference is found between the manual and guidance groups probably due to the spread of the data. One key difference between groups which is not captured by the statistical test is where the median is located with respect to zero, in other words, whether a median angle of attack above the maximum is achieved. At Run 1, the guidance group clearly has the median closest to zero. Furthermore, it takes the manual group until Run 6 to achieve a median below the maximum value, whereas both the cueing and guidance group achieve this at Run 3.

4. Integral above maximum angle of attack

Combining time and distance above the maximum angle of attack results in the integral as shown in Fig. 22. The figure shows that from Run 1, the guidance group has the least amount of angle of attack above the maximum value. Additionally, the learning effect of both other groups is clearly visible, and the cueing group appears to have a smaller spread compare to the manual group. All groups though, seem to be able to reduce the amount of angle of attack above the maximum to an reduced level by Run 4.

Within the manual group, the learning effect is confirmed by statistical test: a significant difference is found ($p < 0.005$, $\chi^2 = 12.14$) and the post-hoc test indicates statistical significant differences between all runs (Runs 1 and 4 $p < 0.5$, Runs 1 and 8 $p < 0.01$, Runs 4 and 8 $p < 0.05$). The cueing group has differences ($p < 0.001$, $\chi^2 = 22.86$) between Runs 1 and 4 ($p < 0.001$), and between Runs 1 and 8 ($p < 0.001$), again indicating the initial learning effect and no after-effects in this metric. Statistical difference are present in the guidance group group ($p < 0.05$, $\chi^2 = 6.26$), with the post-hoc pointing to a difference between Runs 1 and 4 ($p < 0.05$), no difference between Run 4 and 5. This again indicates that participants perform better over the first four runs, but the next run has no statistical evidence. The figure does show that in the second block of four runs, without haptic feedback, the spread of the guidance group is slightly higher.

At the first run, we saw that the guidance group has the lowest median. The statistical test indeed confirms differences ($p < 0.01$, $\chi^2 = 9.76$), yet the post-hoc only confirms differences between cueing and guidance groups ($p < 0.005$). Differences between manual and guidance groups are almost significant ($p = 0.10$). At Run 4 and 8, no statistical significant differences are found.

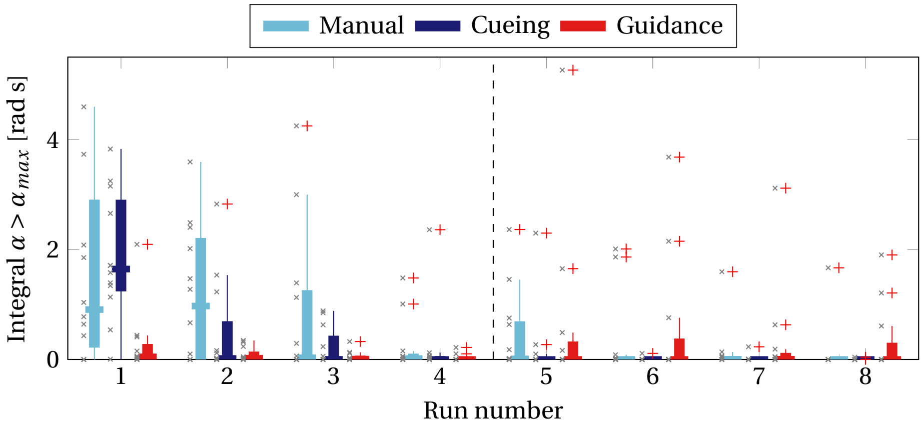


Figure 22 Integral above angle of attack limit during windshear (Manual Participant 7, 8.45rad s, not shown)

C. Subjective measures

The first subjective measures are obtained after each run: a measure for situation awareness by asking the pilot whether (s)he was in control and whether (s)he was missing information, and a measure of workload. After each block of four runs, another questionnaire asked the participants to provide a Van Der Laan and MCH rating.

1. Was the pilot in control?

The results of the first measure can be found in Figs. 23 and 24, where no clear differences between groups can be observed. All groups seem to show an improving trend from Run 1 to Run 8.

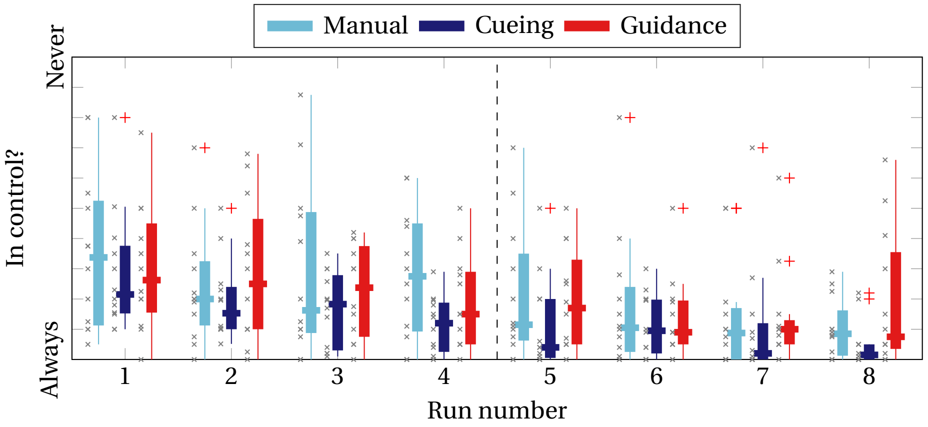


Figure 23 Was the pilot feeling in control?

Statistical tests show significant differences between the runs for the manual group ($p < 0.05$, $\chi^2 = 8.86$), post-hoc indicates a significant difference between Runs 1 and 8 ($p < 0.05$), and 'near significance' between Runs 4 and 8 ($p = 0.06$). This confirms the observation of improving over all runs and indicates a gradual change of the runs. The differences within the cueing group are also significant ($p < 0.001$, $\chi^2 = 14.34$), and post-hoc indicates differences between Runs 1 and 4 ($p < 0.05$), between Runs 1 and 8 ($p < 0.005$), and not between Runs 4 and 8. Again, this can indicate that the metric improved over the first block, yet did not significantly improve after that.

2. Was the pilot missing information?

Results for the amount of information missed by the pilots did not result in any statistical significant results. Fig. 24 shows that the cueing group has an increased spread from Run 4 to Run 5. At the latter run, no haptic feedback was provided, possibly pointing out that the haptic feedback was providing the pilot with extra information. Additionally, the spread of the guidance group seems to be larger over most runs compared to the other groups, especially for the last three runs, where the spread is larger compared to the other groups.

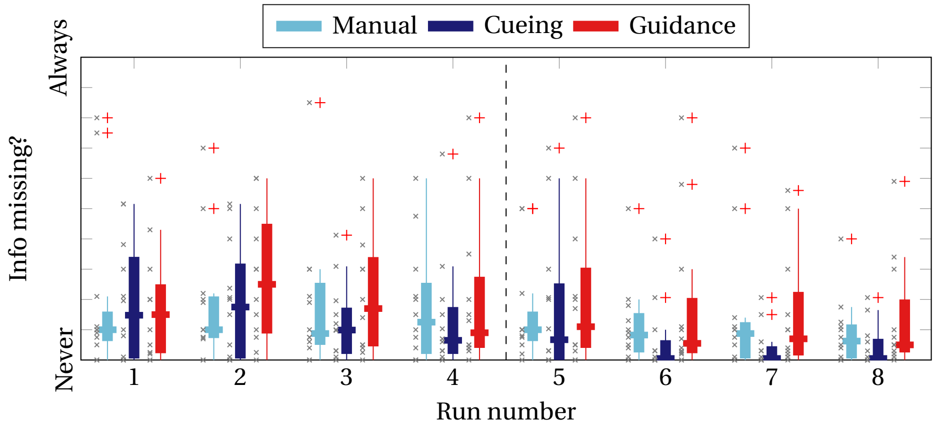


Figure 24 Was the pilot missing information?

3. RSME

The mental load required for each of the runs was measured using the RSME and is shown in Fig. 25. Within groups, a general decreasing trend, can be observed. Statistical analysis for the cueing group shows differences ($p < 0.05$, $\chi^2 = 7.48$) and the post-hoc indicates significant differences between Run 1 and 8 ($p < 0.05$), confirming the general decreasing trend. Two between-group observations can be made: the median of the manual group for each run, except Run 7, is the highest, and the median of the cueing group, except Run 1, is the lowest. These last observations are not statistically significant.

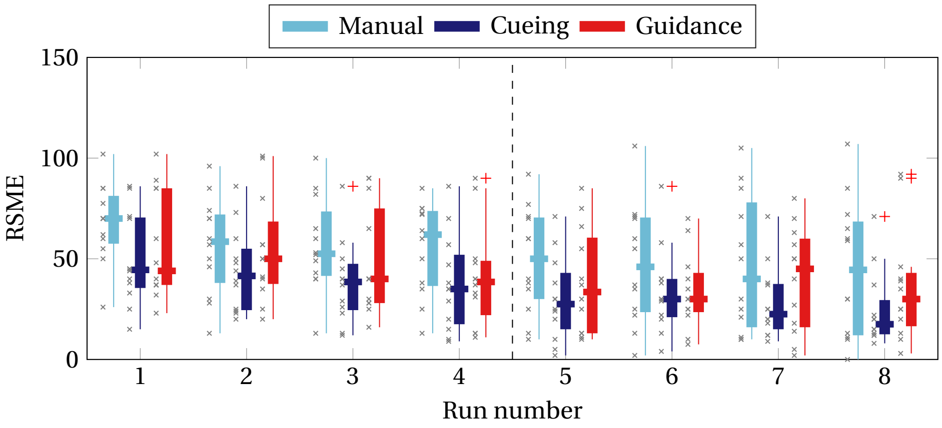
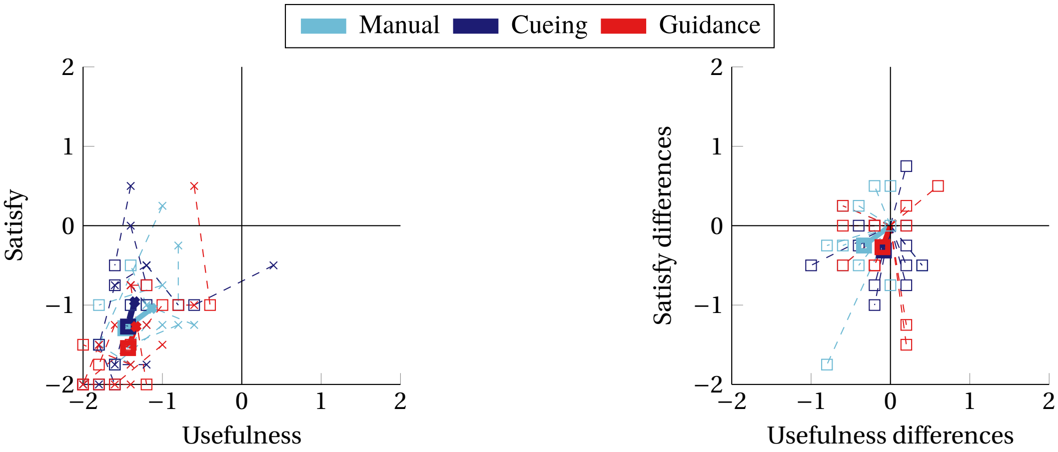


Figure 25 Rating Scale Mental Effort (RSME)

4. Van Der Laan ratings

The Van Der Laan uses nine questions to score the system on usefulness and satisfaction, a perfectly useful and satisfying system would score minus two on both scales. As suggested by Van der Laan, both the absolute values and the difference from Block 1 to Block 2 are shown in respectively Fig. 26a and Fig. 26b. The absolute values show that the systems are in general well received and that the averaged-per-group results, indicated by thick indications, are situated close together, i.e., there is little effect of the haptic feedback support on the initial rating. No statistical significant difference is found between blocks, or between groups within one block.

Looking at the change in rating after Block 2, the manual group seems to have an improvement in both usefulness



(a) Raw data **(b) Change in rating from Block 1 to Block 2**

Figure 26 Van Der Laan-ratings; crosses represent the score after Block 1, squares after Block 2, bold indicate group means

and satisfaction, both other groups change mainly in satisfaction, to a lesser extent in usefulness. This corresponds to the verbal comments pilots gave after the second block: *"The visuals need time to learn to understand and use them."*, implying that the system becomes more useful after more training. The groups provided with haptic feedback made such comments less frequently. Nevertheless, this difference is again not statistically significant.

5. MCH ratings

Results for the MCH rating, shown in Fig. 27, are obtained using a decision-tree format and the main decision points are indicated with horizontal dashed lines. Note that lower scores are better, and going from score three to two requires that 'The information facilitates efficient decision making.' Two ratings do stand out: one participant rated the system twice as ten, where he indicated that all the information was too cluttered (especially the visual display), and warnings should only be provided when close to the limits. Other ratings provided by the participants in the guidance group did not change between blocks. Ratings for the two other groups have reduced spread and shift towards an improved rating: more participants rated the system one or two after Block 2. This indicates that the systems can provide efficient decision making, yet after training. Nonetheless, no statistical differences are found between blocks, nor between groups within a block.

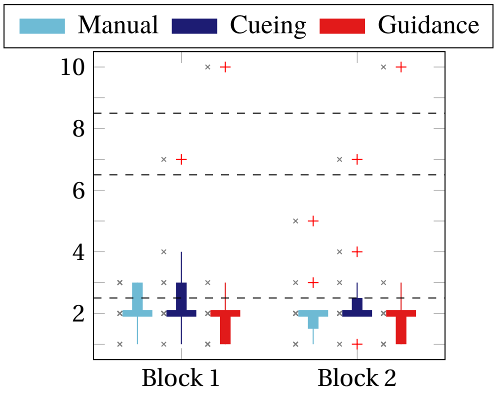


Figure 27 Modified Cooper-Harper rating

D. Debriefing questionnaire

A final questionnaire is used to structure the debriefing session and involved 19 questions when participants were in either the cueing or guidance group, 14 when in the manual group as questions concerning haptics were excluded. The first question simply queried the pilots of the cueing and guidance groups whether they prefer flying the aircraft with, or without haptic feedback. The results in Fig. 28 show that in the guidance group all but one pilot prefered to fly with the haptic feedback, in the cueing group opinions are divided evenly. The difference between groups is confirmed with statistical significance ($p < 0.05$, $\chi^2 = 42$).

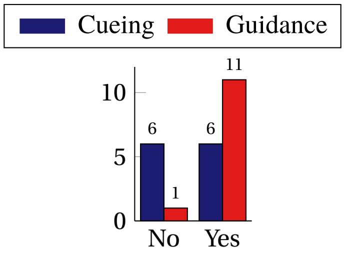


Figure 28 Do you prefer to fly with the haptic feedback system?

Unless mentioned otherwise, debriefing questions presented in this paper used a five-point Likert scale, the results are shown in Fig. 29. Fig. 29a shows that about half of the participants in the cueing group agrees that the supplied haptic feedback is distracting, whereas the guidance group was more inclined to disagree with this statement. Concerning the visual system, Fig. 29b, pilots in general disagree, i.e., the visual system is not distracting. For both the haptic and visual interface, all groups are undecided on whether much training is required (Fig. 29d and 29e). The majority of the pilots who received haptic feedback did agree that the visual and haptic feedback did not give conflicting signals, attesting to the design work reported in Ref. [30]. Nevertheless, a difference in groups might be present on whether the pilots were fighting the haptic system: Fig. 29f shows that for the majority of the cueing group this was not an issue, whereas the majority of the guidance group reported to be ‘fighting’ the haptic system.

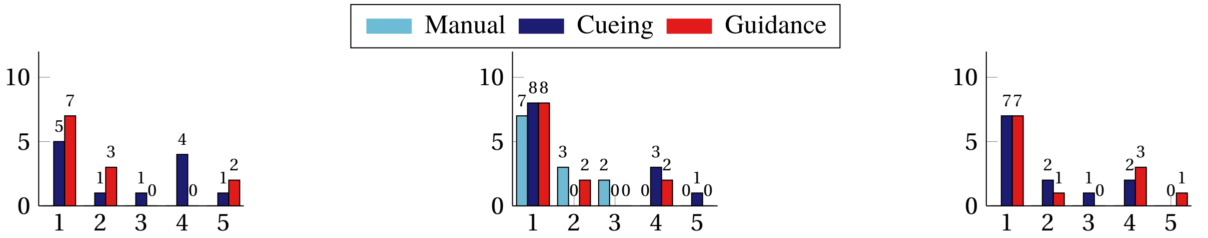
Concerning the visual system, Fig. 29b, pilots in general disagree, i.e., the visual system is not distracting. For both the haptic and visual interface, all groups are undecided on whether much training is required (Fig. 29d and 29e). The majority of the pilots who received haptic feedback did agree that the visual and haptic feedback did not give conflicting signals, attesting to the design work reported in Ref. [30]. Nevertheless, a difference in groups might be present on whether the pilots were fighting the haptic system: Fig. 29f shows that for the majority of the cueing group this was not an issue, whereas the majority of the guidance group reported to be ‘fighting’ the haptic system.

Looking at the subjective effect on workload, Fig. 29g, the manual group indicates a decrease, the cueing group indicates an increase, and the guidance group is undecided. This is the second question where a statistical difference is found between groups ($p < 0.05$, $\chi^2 = 6.15$), and the post-hoc test indicates that there is a significant difference between manual and cueing groups, confirming the difference between these two groups.

Participants are almost unanimous on the possible learning effect: Fig. 29h indicates that their understanding of the interface increased throughout the experiment. Finally, when one of the systems (visual, cueing haptics, of guidance haptics) is implemented, participants expect the knowledge on the aircraft performance boundaries to increase (Fig. 29i), it prevents critical situations to occur (Fig. 29j), and if a critical situation occurs, it can help in resolving it (Fig. 29k). In conclusion, pilots expect the visual and haptic systems to increase safety (Fig. 29l).

The questions at the end of the debriefing questionnaire allowed participants to elaborate on any of the above questions, comment on the reality of the simulator, and possible other comments. The manual group indicated that the display increased awareness, yet might could lead to pilots focussing on the instruments, not looking outside. Participants in the cueing group used the haptics as an alerting cue, and the visual display as a measure of criticality. The guidance haptic feedback made flying at the limits harder, yet provided intuitive feedback.

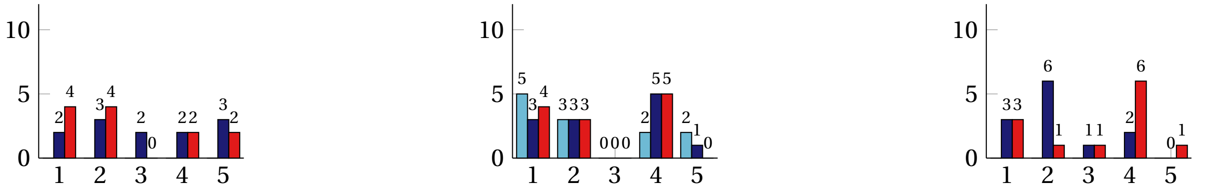
Both groups who received haptic feedback indicated that more training might be required. All groups indicated that their performance might have improved solely due to familiarity with the simulator. Verbatim comments can be found in Appendix C.



(a) The haptic interface is distracting.

(b) The visual interface is distracting.

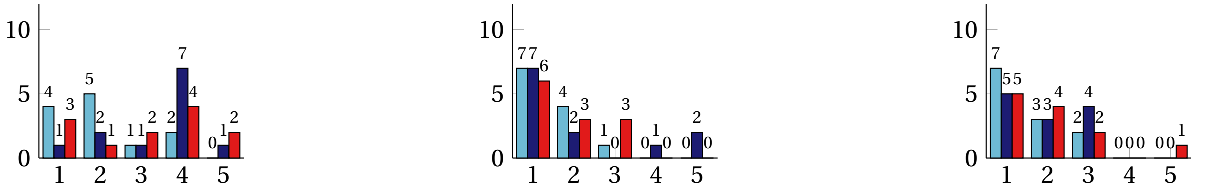
(c) The visuals and haptics gave conflicting signals.



(d) The haptic interface requires a lot of training.

(e) The visual interface requires a lot of training.

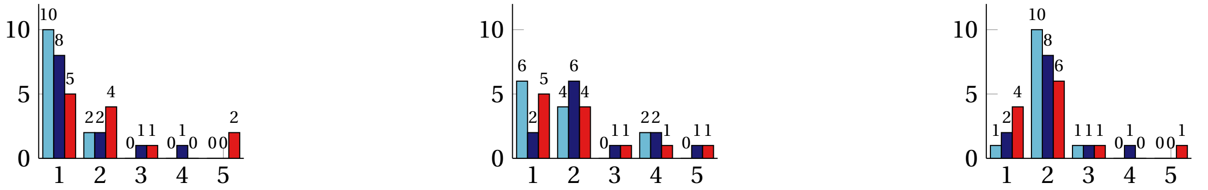
(f) I was fighting the haptic interface.



(g) The interface affected my workload.[1]

(h) During the experiment, my understanding of the interface increased after each flight.[2]

(i) Using the interface, my knowledge on the edges of the aircraft performance changed.[3]



(j) The interface helps in preventing critical situations.[2]

(k) If a critical situation occurs, the interface helps in resolving it.[2]

(l) When implementing this system on an aircraft, what would be the effect on safety?[4]

Figure 29 Debriefing Likert-scale questions. Possible answers (unless specified otherwise) were: 1) Disagree 2) Slightly disagree 3) Disagree nor agree 4) Slightly agree 5) Agree

[1] Possible answers: 1) Decreased 2) Marginal decrease 3) Did not change 4) Marginal increase 5) Increased
[2] Possible answers: 1) Agree 2) Slightly agree 3) Disagree nor agree 4) Slightly disagree 5) Disagree
[3] Possible answers: 1) Increased 2) Marginal increase 3) Did not change 4) Marginal decrease 5) Decreased
[4] Possible answers: 1) Much safer 2) Safer 3) Safer nor unsafer 4) Unsafer 5) Much unsafer

V. Discussion

With the results of the experiment presented before, two designs of haptic feedback were evaluated for their effect on training and their general acceptance by pilots. First of all, the debriefing questionnaires showed that pilots appreciated the haptic feedback and expect the system to support in knowledge of the flight envelope limits, prevent critical situations, and help in resolving them. However, each group voiced a concern for over-reliance on the haptic and/or visual system.

The design principle with haptic feedback is to keep the pilot in the loop, thereby making the pilot more aware of the situation and possibly less likely to be surprised to an automation failure compared to supervising an autopilot. Nonetheless, a next step to investigate this can be the inclusion of haptic feedback failures in the experiment design. In the following paragraphs, both the cueing and guidance haptic feedback design is discussed separately to evaluate the hypotheses presented before.

A. Cueing haptic feedback

Using the results of the manual group as a baseline, the cueing group was expected to obtain similar performance levels and safety margins at the first run. All metrics showed the expected result at Run 1, both as observed from the graph, and confirmed by lack of significant differences between the results of the manual and cueing groups.

Next, the cueing group was expected to learn faster, which would result in differences between both groups at Run 4. Although not directly visible in the altitude lost or integral of angle of attack over the maximum, looking at the time where the maximum angle of attack is exceeded, the cueing group is indeed closer to zero compared to the manual group (although not statistically significant). Additionally, the median of the highest angle of attack achieved during the recovery for the cueing group is below the maximum, whereas this is not the case for the manual group.

Furthermore, the cueing group shows a quicker improvement of the safety margins over Runs 2, 3, and 4. When removing the haptic feedback from the cueing group, no after-effects are expected, i.e., no changes in metrics from Run 4 to Run 8. This latter is present in all objective metrics used, hence the pilots used the haptic feedback to learn quicker how to stay within the angle of attack limits, and they are still able to do so when no haptic feedback is provided anymore.

These observations are in line with the results found in our previous experiment, see Ref. [8]. That experiment contained a group which did not receive haptic feedback in Block 1, followed by the cueing haptic feedback in Block 2. Since the second block contained haptic feedback, the data from that group can not be used as a comparison for training effects. The first block did, however, show also that participants who did not receive haptic feedback had a much harder time keeping the aircraft within its limits. Participants in the earlier experiment had even more difficulty compared to the current manual group in this experiment.

The cueing haptic feedback was expected to provide pilots with an increased situation awareness at the cost of an increased workload and without the benefit of immediate improvements at the first run. Both the information missing as the feeling of being in control for the manual and cueing group did not differ, providing no evidence that an increased situation awareness is present at Run 1.

The RSME, however, did show a (non-statistical significant) difference: the median of the cueing group is lower compared to the manual group, actually indicating less workload in contradiction to what was expected. The workload for cueing group was expected to decrease to Run 4 due to familiarization, which is partly confirmed by the RSME data: the statistical analysis only showed a difference between Runs 1 and 8, nonetheless, a decreasing trend in median RSME is visible supporting the expected decrease in workload due to familiarization.

The subjective situation awareness scales did not provide any evidence that a difference between groups is present at Run 1, 4, or 8. It did show an improving trend over the runs for all groups. This is to be expected due to familiarization with the task and model. Interestingly, going from Run 4 to Run 5, i.e., disabling the cueing haptic feedback, increased the spread for information missing. This can indicate that the haptic feedback is providing pilots information which they are consciously integrating in their control loop.

Modified Cooper-Harper ratings did not indicate any differences compared to the manual group. The change in Van Der Laan ratings between blocks showed that the cueing group did not have an improvement in usefulness, whereas the manual group did. This might indicate that the haptic feedback made the pilots understand the visuals more easily, yet differences are very small. For these two system evaluation ratings, differences are considered too small to draw any conclusions from them.

In contrast to the observations on workload before, the debriefing questionnaire showed that participants who received the cueing feedback indicated a marginal increase in workload. This might be attributed to the fact that the haptic feedback is slightly distracting, as the pilots indicated as well. Because of the contradicting answers in the RSME and the debriefing no clear conclusions can be draw for a change in workload when using the cueing haptic feedback.

In conclusion, the cueing haptic feedback system is able to provide the pilots with support in the learning process without after-effects, and without having large effects on workload or situation awareness. It does lack the ability to support the pilot at the very first encounter with a certain limit. Therefore, it might be well-suited to provide such feedback to pilots in simulators to support the learning process, without the disadvantage of dependency on such a system when transferring to the real aircraft.

B. Guidance haptic feedback

The guidance group was expected to have an improved performance and safety margin compared to the other groups already at Run 1. For all metrics, there is a significant improvement in relation to the cueing group, and only ‘ near statistical significance’ differences compared to the manual group. Nevertheless, the plots of all metrics show that the guidance group obtained a lower median *and* a lower spread. This means that the participants in the guidance group did not only do better, they were also more consistent across all participants. In other words, when using this haptic feedback system, a safer operation from the first time can be expected.

At Run 4, performance and safety margins are not statistically different from the manual group, yet the time of angle of attack above the maximum only shows two outliers above zero, and the highest angle of attack is clearly below the maximum value. This indicates that better metrics are achieved compared to the manual group. They are, however, of similar value as the cueing group. An improvement for the guidance group in time and integral of angle of attack above the maximum is found from Run 1 to Run 4, indicating that a minor learning process is still present.

When the haptic feedback is disabled, we expected the performance and safety margins to deteriorate. This did not largely happen in the amount of altitude lost, yet in the other safety margin metrics an increase in spread of the data can be seen. In other words, when no haptic feedback is supplied, three participants (or 25%) did exceed the angle of attack limits, giving some evidence for the “guidance hypothesis”. This improved spread also contributes to the lack of significant changes between Run 1 and Run 5 for the guidance group.

The guidance haptic feedback was expected to actively support the pilot from the first run and thereby reducing the workload, at a cost of decreased situation awareness. The lack of changes in information missing and feeling of being in control at Run 1 gives no evidence that a decrease in situation awareness is present. The workload ratings at Run 1 show a (non-statistical significant) lower median rating compared to the manual group, indicating a reduced workload. This is a similar observation as with the cueing group which can indicate that for this setup/model/task, a multi-modal system, including haptic and visual feedback, is preferred over visual only.

Looking at all runs, the workload rating of the guidance group is consistently between the ratings of the cueing and the manual group, indicating that the guidance haptic feedback provides a reduced workload compared to no haptic. Nevertheless, in the debriefing questionnaire, not all pilots were convinced that a lower workload is achieved, and it shows that a number of them reported fighting the haptic feedback system at times. As the RSME and questionnaire are again contradicting, no clear conclusions can be draw with respect to workload.

Although the subjective situation awareness did not differ at Run 1, the larger spread on the amount of information missing for the guidance group can indicate that the haptic feedback is supporting the pilots, yet it is not transparent in why it is doing so. Additionally, during the final four runs, the spread of the data is higher compared to the manual group, indicating that the pilots who do not receive guidance feedback anymore have difficulty adjusting to the new situation.

As before, modified Cooper-Harper ratings did not change, and no change in usefulness was present for the Van Der Laan ratings of the guidance group. The latter might indicate that the visuals were better understood because of the presence of haptics, i.e., multi-modal information. Nonetheless, differences for both ratings are considered too small to have any meaningful indication.

Compared to no haptic feedback, the guidance haptic feedback system is able to provide pilots with support during operation of the aircraft, without having large effects on workload or situation awareness. It is subjected to after-effect, as the spread of both the objective metrics and the amount of information missing increases when haptic feedback is no longer supplied. As such, it might be well-suited to provide such feedback to pilots during continuous operation of the aircraft. The after-effects have to be mitigated, for which more research is needed. Possibly a visual/aural indication can be added such that pilots are aware when the system is disabled and do not rely on it.

The design of the guidance haptic feedback was setup to make sure the pilot is in final control and can always over-ride the haptic inputs. To guarantee this with an increasing stiffness, a maximum force required to move the side stick was implemented after which a zero-slope force was present. The last example presented in the results showed one design flaw: when the participant exerts a force on the side stick of this maximum value, the feeling can be ‘lock-in‘ on the flat slope. In this situation, the stiffness and neutral point can still change, yet the participant will not notice this as (s)he is already pulling at the maximum force level. Future designs are recommended to avoid such a condition by, for example, implementing a shallow force gradient beyond the maximum force such that at least the effect of the neutral point shift can be felt.

VI. Conclusion

This paper compared two designs of a haptic feedback system, using force feedback on the control device, to a situation without haptic feedback. These systems are expected to inform pilots on the flight envelope limits and protections in a modern fly-by-wire aircraft. One design used forcing functions which are asymmetric in time and amplitude, and warn the pilots of an approaching limit, as well as of the direction of the corrective action. In other words, this design is haptically *cueing* the pilot on the limits.

The second design actively changed the side stick mechanical properties to indicate a required direction to remain clear of the limits, as well as increased the stiffness for an input which brings the aircraft closer its limits. Therefore, this second design is haptically *guiding* the pilot near the limits.

Three groups – one with the cueing, one with the guiding, and one without haptic feedback – of each 12 PPL/LAPL pilots were used to evaluate the systems by flying 8 runs. Each run contained a trajectory shown on the outside visual and during the run a windshear was encountered. The first 4 runs were flown with haptic feedback for two out of three groups, the last four runs were flown without haptic feedback.

The results of the pilots flying with the cueing haptic feedback, compare to those who flew without haptic feedback, showed that the cueing feedback is able to provide support in the learning process. It can do so without suffering after-effects, i.e., when transitioning to a condition without haptic assistance. Additionally, it does not have large effects on workload or situation awareness. This system does not support the pilot at the first encounter with a certain limit.

Results show that the guidance haptic feedback system is able to provide the pilots with support during operation of the aircraft, also without having large effects on workload or situation awareness. It is sensitive to an after-effect, and creates a reliance on the system, as the spread of both the objective metrics and the amount of information missing increases when no haptic feedback is supplied.

Acknowledgments

We would like to thank all 24 pilots from this experiment campaign, as well as all pilots from the previous campaign. They all freed-up three hours in their schedule (excluding travel) to participate in this experiment.

References

[1] EASA, "EASA Annual Safety Review," Tech. rep., European Aviation Environmental Report, 2019. URL `https://www.easa.europa.eu/eaer/`, Available at `https://www.easa.europa.eu/eaer/`.

[2] IATA, "IATA Safety Report 2017," Tech. rep., International Air Transport Association, 2018. URL `https://aviation-safety.net/airlinesafety/industry/reports/IATA-safety-report-2017.pdf`, Available at `https://aviation-safety.net/airlinesafety/industry/reports/IATA-safety-report-2017.pdf`.

[3] Ackerman, K. A., Talleur, D. A., Carbonari, R. S., Xargay, E., Seefeldt, B. D., Kirlik, A., Hovakimyan, N., and Trujillo, A. C., "Automation Situation Awareness Display for a Flight Envelope Protection System," *Journal of Guidance, Control, and Dynamics*, Vol. 40, No. 4, 2017, pp. 964–980. doi:10.2514/1.G000338.

[4] Schuet, S., Lombaerts, T. J. J., Stepanyan, V., Kaneshige, J., Hardy, G., Shish, K. H., Robinson, P., Etherington, T., Kramer, L., Evans, E., Daniels, T., Young, S. D., and Rodzon, D., "Piloted Simulation Study Findings On Stall Recovery Guidance," *Proceedings of the AIAA Modeling and Simulation Technologies Conference, San Diego (CA)*, 2019, p. 23. doi:10.2514/6.2019-0981, AIAA-2019-0981.

[5] de Stigter, S., Mulder, M., and van Paassen, M. M., "Design and Evaluation of a Haptic Flight Director," *Journal of Guidance, Control, and Dynamics*, Vol. 30, No. 1, 2007, pp. 35–46. doi:10.2514/1.20593.

[6] Beeftink, D. G., Borst, C., Van Baelen, D., van Paassen, M. M., and Mulder, M., "Haptic Support for Aircraft Approaches with a Perspective Flight-path Display," *IEEE International Conference on Systems, Man, and Cybernetics, Miyazaki*, 2018, pp. 3016 – 3021. doi:10.1109/SMC.2018.00512.

[7] Stepanyan, V., Krishnakumar, K., Dorais, G., Reardon, S., Barlow, J., Lampton, A., and Hardy, G., "Loss-of-Control Mitigation via Predictive Cuing," *Journal of Guidance, Control, and Dynamics*, Vol. 40, No. 4, 2017, pp. 831–846. doi:10.2514/1.G001731.

[8] Van Baelen, D., Ellerbroek, J., van Paassen, M. M., Abbink, D. A., and Mulder, M., "Using Asymmetric Vibrations for Flight Envelope Protection," *Proceedings of the AIAA Modelling and Simulation Technology conference, Orlando (FL)*, 2020, p. 27. doi:10.2514/6.2020-0409, AIAA-2020-0409.

[9] Schmidt, R. A., Young, D. E., Swinnen, S., and Shapiro, D. C., "Summary Knowledge of Results for Skill Acquisition: Support for the Guidance Hypothesis," *Journal of Experimental Psychology Learning Memory and Cognition*, Vol. 15, No. 2, 1989, pp. 352 – 359. doi:10.1037/0278-7393.15.2.352.

[10] Winstein, C. J., and Schmidt, R. A., "Reduced Frequency of Knowledge of Results Enhances Motor Skill Learning," *Journal of Experimental Psychology Learning Memory and Cognition*, Vol. 16, No. 4, 1990, pp. 677 – 691. doi:10.1037/0278-7393.16.4.677.

[11] Kuiper, R. J., Heck, D. J. F., Kuling, I. A., and Abbink, D. A., "Evaluation of Haptic and Visual Cues for Repulsive or Attractive Guidance in Nonholonomic Steering Tasks," *IEEE Transactions on Human-Machine Systems*, Vol. 46, No. 5, 2016, pp. 672–683. doi:10.1109/THMS.2016.2561625.

[12] Petermeijer, S. M., Abbink, D. A., and De Winter, J. C. F., "Should Drivers Be Operating Within an Automation-Free Bandwidth? Evaluating Haptic Steering Support Systems With Different Levels of Authority," *Human Factors*, Vol. 57, No. 1, 2015, pp. 5–20. doi:10.1177/0018720814563602.

[13] Wang, Z., Kaizuka, T., and Nakano, K., "Effect of Haptic Guidance Steering on Lane Following Performance by Taking Account of Driver Reliance on the Assistance System," *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Miyazaki (JP)*, 2018, pp. 2717–2723. doi:10.1109/SMC.2018.00464.

[14] Lam, T. M., Boschloo, H. W., Mulder, M., and Van Paassen, M. M., "Artificial Force Field for Haptic Feedback in UAV Teleoperation," *IEEE Transactions on Systems, Man, and Cybernetics Part A:Systems and Humans*, Vol. 39, No. 6, 2009, pp. 1316–1330. doi:10.1109/TSMCA.2009.2028239.

[15] Van Baelen, D., Ellerbroek, J., van Paassen, M. M., and Mulder, M., "Design of a Haptic Feedback System for Flight Envelope Protection," *Journal of Guidance, Control, and Dynamics*, Vol. 43, No. 4, 2020, pp. 700–714. doi:10.2514/1.G004596.

[16] Chatrenet, D., "Les qualités de Vol des Avions de Transport Civil à Commandes de Vol Électriques," *Active Control Technology: Applications and Lesson Learned*, AGARD, Neuilly-Sur-Seine, France, 1995, p. 5.

[17] Favre, C., "Fly-by-wire for Commercial Aircraft: the Airbus Experience," *International Journal of Control*, Vol. 59, No. 1, 1994, pp. 139–157. doi:10.1080/00207179408923072.

[18] Niedermeier, D., and Lambregts, A. A., "Fly-By-Wire Augmented Manual Control - Basic Design Considerations," *28th International Congress of the Aeronautical Sciences, Brisbane*, Vol. 100, 2012, pp. 1 – 14.

[19] Airbus, "A319/A320/A321 Flight Crew Operating Manual," Tech. Rep. 36, Airbus, 2003.

[20] Bureau d'Enquêtes et d'Analyses pour la sécurité de l'aviation Civile, "Final Report On the accident on 1st June 2009 to the Airbus A330-203 registered F-GZCP operated by Air France flight AF 447 Rio de Janeiro - Paris (English courtesy translation)," Tech. Rep. June 2009, Bureau d'Enquêtes et d'Analyses pour la sécurité de l'aviation Civile, 2012.

[21] Corps, S. G., "Airbus A320 Side Stick and Fly By Wire - An Update," *Proceedings of the Aerospace Technology Conference and Exposition, Long Beach (CA)*, 1986, p. 13. doi:10.4271/861801.

[22] Van Baelen, D., Ellerbroek, J., van Paassen, M. M., Abbink, D. A., and Mulder, M., "Just Feeling the Force: Just Noticeable Difference for Asymmetric Vibrations," *Accepted to Proceedings of the 2020 IEEE International Conference on Human-Machine Systems, Rome (Italy)*, 2020, p. 6.

[23] Jeram, G. J., and Prasad, J. V. R., "Tactile Avoidance Cueing for Pilot Induced Oscillation," *Proceedings of the AIAA Atmospheric FLight Mechanics Conference, Austin (TX)*, 2003, pp. 1–9. doi:10.2514/6.2003-5311.

[24] Klyde, D. H., and McRuer, D. T., "Smart-Cue and Smart-Gain Concepts Development to Alleviate Loss of Control," *Journal of Guidance, Control, and Dynamics*, Vol. 32, No. 5, 2009, pp. 1409–1417. doi:10.2514/1.43156.

[25] Klyde, D. H., Richards, N., and Cogan, B., "Mitigating Unfavorable Pilot Interactions with Adaptive Controllers in the Presence of Failures/Damage," *Proceedings of the AIAA Atmospheric Flight Mechanics Conference, Portland (OR)*, 2011, p. 17. doi:10.2514/6.2011-6538.

[26] Whalley, M. S., and Achache, M., "Joint U.S./France Investigation of Helicopter Flight Envelope Limit Cueing," *Proceedings of the 52th American Helicopter Society Forum, Washington D.C. (DC)*, 1996, p. 30.

[27] Vidulich, M. A., Tsang, P. S., and Flach, J. (eds.), *Advances in Aviation Psychology, Volume 2*, 2^{nd} ed., Taylor & Francis, London, 2017, Chap. 12, pp. 237 – 254. doi:10.4324/9781315565712.

[28] Ruijgrok, G. J. J., *Elements of Airplane Performance*, 2nd ed., VSSD, Delft, 2009, Chaps. 1,3, pp. 1–23,51–62. URL `http://www.vssd.nl/hlf/ae02.htm`.

[29] Schmidt, R. A., and Bjork, R. A., "New Conceptualizations of Practice: Common Principles in Three Paradigms Suggest New Concepts for Training," *Psychological Science*, Vol. 3, No. 4, 1992, pp. 207 – 217. doi:10.1111/j.1467-9280.1992.tb00029.x.

[30] de Rooij, G., Van Baelen, D., Borst, C., van Paassen, M. M., and Mulder, M., "Supplementing Haptic Feedback through the Visual Display of Flight Envelope Boundaries," *Proceedings of the AIAA Guidance, Navigation, and Control Conference, Orlando (FL)*, 2020, p. 25. doi:10.2514/6.2020-0373, AIAA-2020-0373.

[31] Lombaerts, T. J. J., Looye, G., Seefried, A., Neves, M., and Bellmann, T., "Development and Concept Demonstration of a Physics Based Adaptive Flight Envelope Protection Algorithm," *International Federation of Automatic Control*, Vol. 49, No. 5, 2016, pp. 248 – 253. doi:10.1016/j.ifacol.2016.07.121.

[32] Van Baelen, D., Ellerbroek, J., van Paassen, M. M., and Mulder, M., "Evaluation of a Haptic Feedback System for Flight Envelope Protection," *Proceedings of the AIAA Guidance, Navigation and Control Conference, San Diego (CA)*, 2019, p. 25. doi:10.2514/6.2019-0367, AIAA-2019-0367.

[33] FAA, "Windshear Training Aid - Volume 1," Tech. rep., Federal Aviation Administration, 1990.

[34] Netherlands AIP, "AD 2.EHM-ADC Aerodrome Chart," Online, Feb. 2018.

[35] Zijlstra, F. R. H., "Efficiency in work behaviour: A design approach for modern tools," Ph.D. thesis, Delft University of Technology, Nov. 1993.

[36] Wertheim, A. H., Bos, J. E., and Bles, W., "Contributions of roll and pitch to sea sickness," *Brain Research Bulletin*, Vol. 47, No. 5, 1998, pp. 517–524. doi:10.1016/S0361-9230(98)00098-7.

[37] Roscoe, A. H., and Ellis, G. A., "A subjective rating scale for assessing pilot workload in flight: a decade of practical use," Tech. Rep. 900019, Royal Aerospace Establishment, 1990.

[38] Van Der Laan, J. D., Heino, A., and De Waard, D., "A simple procedure for the assessment of acceptance of advanced transport telematics," *Transportation Research Part C: Emerging Technologies*, Vol. 5, No. 1, 1997, pp. 1–10. doi: 10.1016/S0968-090X(96)00025-3.

[39] Wilde, G. J. S., "Risk homeostasis theory and traffic accidents : propositions, deductions and discussion of dissension in recent reactions," *Ergonomics*, Vol. 31, No. 4, 1988, pp. 441–468. doi:10.1080/00140138808966691.

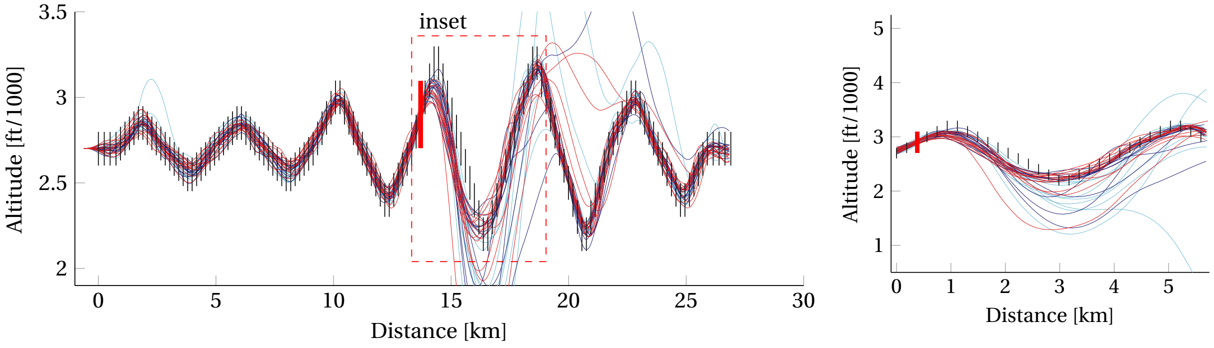
[40] Melman, T., de Winter, J. C. F., and Abbink, D. A., "Does haptic steering guidance instigate speeding? A driving simulator study into causes and remedies," *Accident Analysis and Prevention*, Vol. 98, No. January, 2017, pp. 372–387. doi:10.1016/j.aap.2016.10.016.

[41] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2017. URL `https://www.R-project.org/`.

[42] Benjamini, Y., and Hochberg, Y., "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing," *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 57, No. 1, 1995, pp. 289–300.

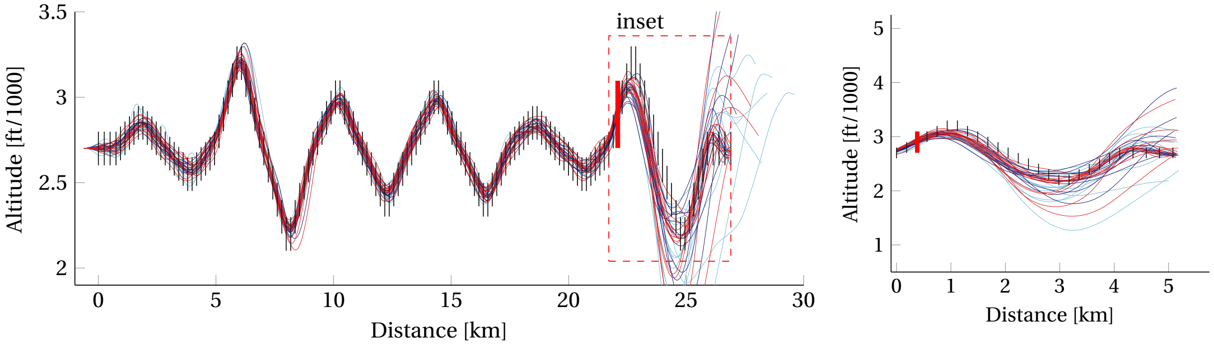
Appendix A. Full experiment data

This appendix contains the flights of all participants, grouped per realization of the random-ordered flight paths. For each realization, Fig. 30 shows both the entire flight of each pilot and a zoomed inset which focuses on the start of the windshear recovery procedure. The metrics used in the paper are calculated on this specific part of the flights. Light blue lines represent the group without haptic feedback, dark blue lines represent the group with cueing haptic feedback, and red lines represent the group with guidance haptic feedback.



(a) Flight path realization one.

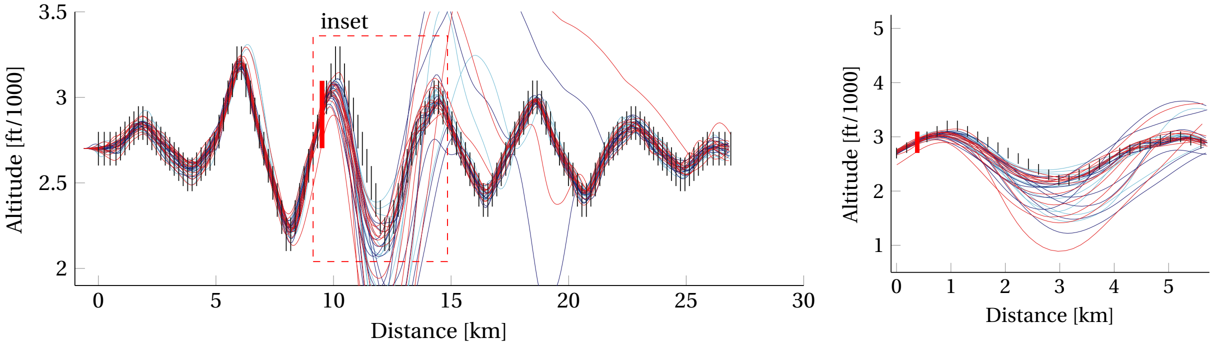
(b) Inset for realization one.



(c) Flight path realization two.

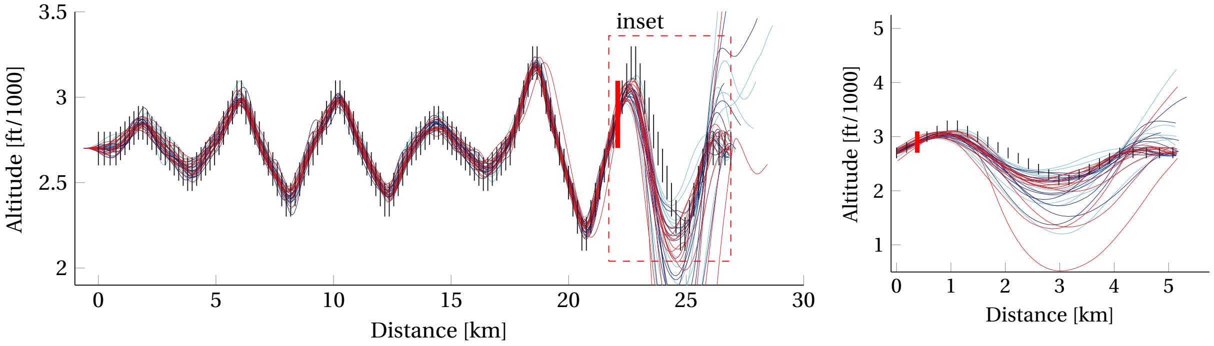
(d) Inset for realization two.

Figure 30 All flown flight paths by all pilots, vertical lines indicate the gates on the outside visual, vertical thick red line indicates windshear trigger.



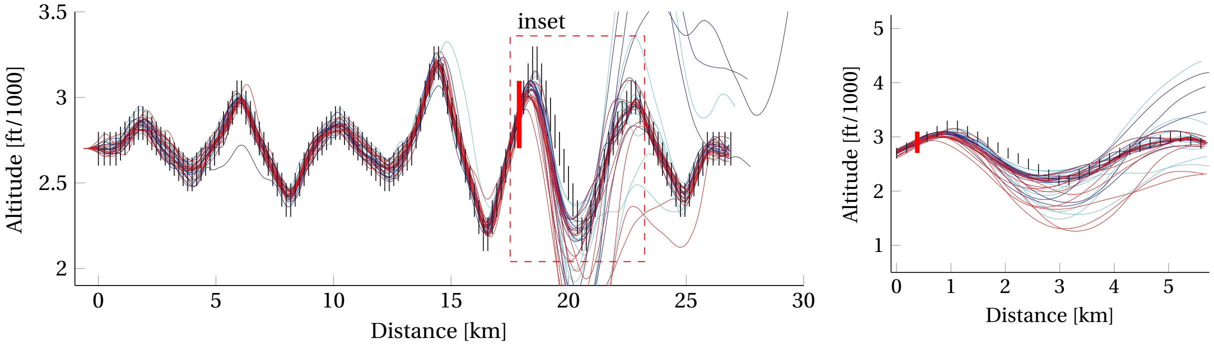
(e) Flight path realization three.

(f) Inset for realization three.



(g) Flight path realization four.

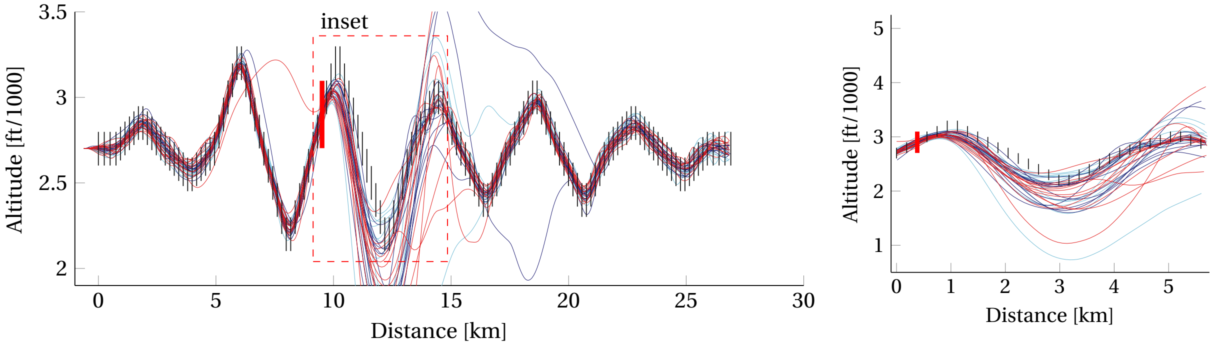
(h) Inset for realization four.



(i) Flight path realization five.

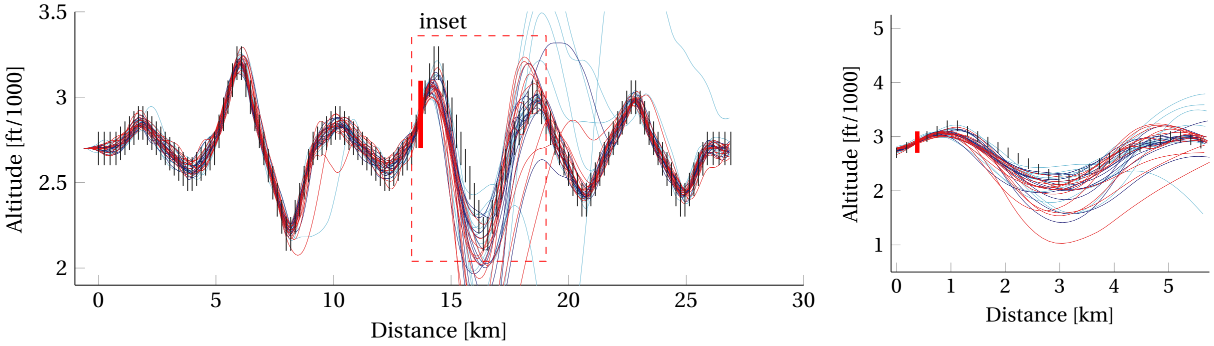
(j) Inset for realization five.

Figure 30 (continued)



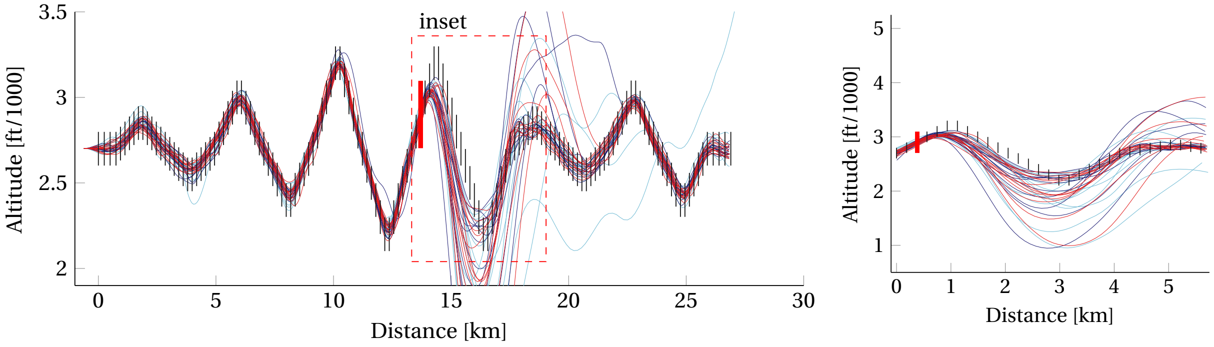
(k) Flight path realization six.

(l) Inset for realization six.



(m) Flight path realization seven.

(n) Inset for realization seven.



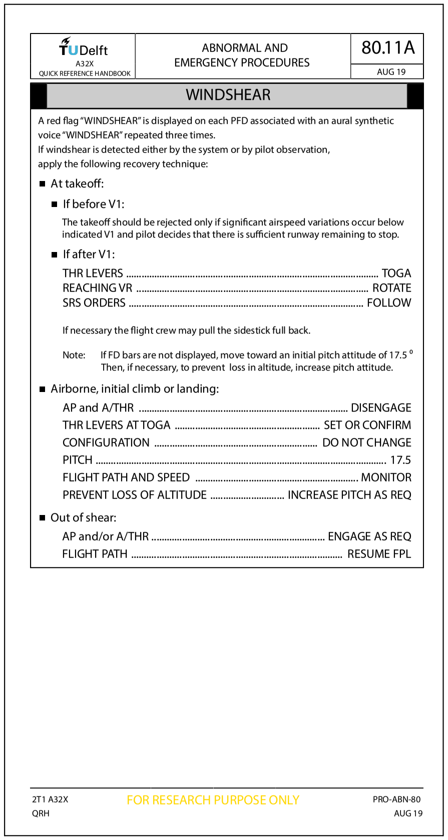
(o) Flight path realization eight.

(p) Inset for realization eight.

Figure 30 (continued)

Appendix B. Emergency checklists

This appendix shows the checklists used in the experiment. When an emergency situation presented itself, pilots were expected to follow the checklist shown in Fig. 31 to mitigate it. All checklists were shown and explained to the pilots before the start of the measurements.



TUDelft
A32X
QUICK REFERENCE HANDBOOK

ABNORMAL AND
EMERGENCY PROCEDURES

80.11A
AUG 19

WINDSHEAR

A red flag "WINDSHEAR" is displayed on each PFD associated with an aural synthetic voice "WINDSHEAR" repeated three times.
If windshear is detected either by the system or by pilot observation,
apply the following recovery technique:

- At takeoff:
  - If before V1:

    The takeoff should be rejected only if significant airspeed variations occur below indicated V1 and pilot decides that there is sufficient runway remaining to stop.
  - If after V1:

    THR LEVERS .... TOGA
    REACHING VR .... ROTATE
    SRS ORDERS .... FOLLOW

    If necessary the flight crew may pull the sidestick full back.

    Note: If FD bars are not displayed, move toward an initial pitch attitude of 17.5 °
    Then, if necessary, to prevent loss in altitude, increase pitch attitude.
- Airborne, initial climb or landing:

  AP and A/THR .... DISENGAGE
  THR LEVERS AT TOGA .... SET OR CONFIRM
  CONFIGURATION .... DO NOT CHANGE
  PITCH 17.5
  FLIGHT PATH AND SPEED .... MONITOR
  PREVENT LOSS OF ALTITUDE .... INCREASE PITCH AS REQ
- Out of shear:

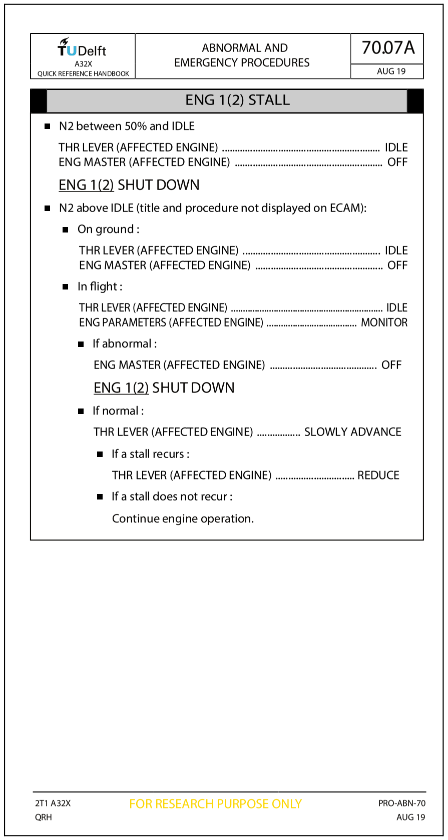
  AP and/or A/THR .... ENGAGE AS REQ
  FLIGHT PATH .... RESUME FPL

2T1 A32X
QRH

FOR RESEARCH PURPOSE ONLY

PRO-ABN-80
AUG 19

(a) Windshear recovery



TUDelft
A32X
QUICK REFERENCE HANDBOOK

ABNORMAL AND
EMERGENCY PROCEDURES

70.07A
AUG 19

ENG 1(2) STALL

- N2 between 50% and IDLE

  THR LEVER (AFFECTED ENGINE) .... IDLE
  ENG MASTER (AFFECTED ENGINE) .... OFF

  ENG 1(2) SHUT DOWN
- N2 above IDLE (title and procedure not displayed on ECAM):
  - On ground :

    THR LEVER (AFFECTED ENGINE) .... IDLE
    ENG MASTER (AFFECTED ENGINE) .... OFF
  - In flight :

    THR LEVER (AFFECTED ENGINE) .... IDLE
    ENG PARAMETERS (AFFECTED ENGINE) .... MONITOR
    - If abnormal :

      ENG MASTER (AFFECTED ENGINE) .... OFF

      ENG 1(2) SHUT DOWN
    - If normal :

      THR LEVER (AFFECTED ENGINE) .... SLOWLY ADVANCE
      - If a stall recurs :

        THR LEVER (AFFECTED ENGINE) .... REDUCE
      - If a stall does not recur :

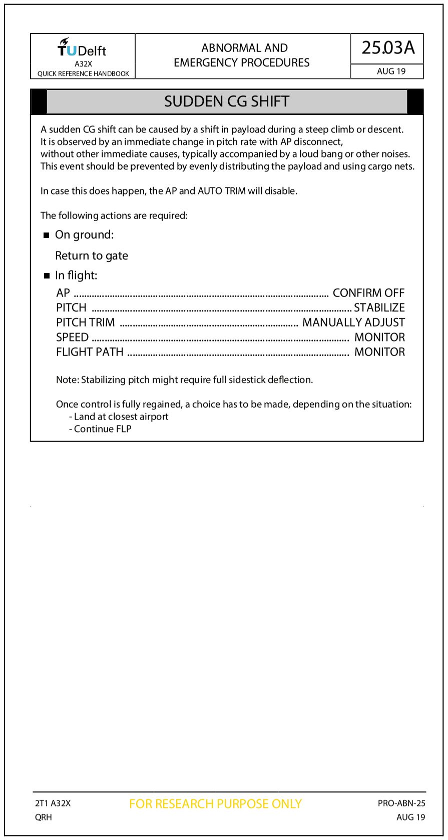
        Continue engine operation.

2T1 A32X
QRH

FOR RESEARCH PURPOSE ONLY

PRO-ABN-70
AUG 19

(b) Single engine stall



TUDelft
A32X
QUICK REFERENCE HANDBOOK

ABNORMAL AND
EMERGENCY PROCEDURES

25.03A
AUG 19

SUDDEN CG SHIFT

A sudden CG shift can be caused by a shift in payload during a steep climb or descent.
It is observed by an immediate change in pitch rate with AP disconnect,
without other immediate causes, typically accompanied by a loud bang or other noises.
This event should be prevented by evenly distributing the payload and using cargo nets.

In case this does happen, the AP and AUTO TRIM will disable.

The following actions are required:

- On ground:

  Return to gate
- In flight:

  AP .... CONFIRM OFF
  PITCH .... STABILIZE
  PITCH TRIM .... MANUALLY ADJUST
  SPEED .... MONITOR
  FLIGHT PATH .... MONITOR

  Note: Stabilizing pitch might require full sidestick deflection.

  Once control is fully regained, a choice has to be made, depending on the situation:
  - Land at closest airport
  - Continue FLP

2T1 A32X
QRH

FOR RESEARCH PURPOSE ONLY

PRO-ABN-25
AUG 19

(c) Sudden CG shift

Figure 31 Checklists for the emergency scenario's (provided to pilots on A4 paper size)

Appendix C. Experiment debriefing statements

The questions at the end of the debriefing questionnaire provided the participants to elaborate on any of the above questions, comment on the reality of the simulator, and possible other comments. Additionally, expected negative effects and reported changes in behavior are reported below. Not every individual comments is shown, we tried to make sure that all comments made within one group are captured by one of the comments below.

Comments from the manual group include:

- *Better anticipation of required aircraft control input based on visuals.*
- *More general awareness.*
- *More relaxed and more time to see ahead.*
- *A negative consequence could be that one starts flying by following only the visual signals.*
- *Display can distract from the actual flying.*
- *Resolving a critical situation could be further assisted with audible commands like "climb", "descend", etc.*
- *Besides this info, throttle awareness is the next most important item needing assistance.*
- *Beginning was quite a start-up, but I got used to the flying of the airplane.*

Participants in the cueing group provided the following comments:

- *I was not looking at instruments anymore, I reacted naturally to the haptics, and looked outside to monitor the behavior of the airplane.*
- *I made slightly stronger inputs.*
- *The haptic feedback could lead to too much stimulus together with sound/visual.*
- *I used the haptics to double check whether my input was correct, if so I made a stronger input.*
- *The haptic feedback indicates a problem, the visuals gives the possible space left (criticality).*
- *I followed the indicated direction for the suggested correction.*
- *Increased awareness of critical situation in angle of attack and speed.*
- *I made more subtle movements and anticipated more with the haptic feedback.*
- *When using all warning systems, extra training is recommended to process everything correctly.*
- *I think my performance improved solely due to experience with the sim.*
- *Audio warnings are best for VFR, with minimal visual information, and only shaker for stall.*
- *Sound can be added to support the protections.*

Finally, the comments made in the guidance group included the following:

- *You need to work harder to fly intentionally on the limit.*
- *Steers you easily towards the correct actions.*
- *Haptics makes you do a faster response, and you than check visuals for clues.*
- *I was struggling with conflicting visual and haptic inputs.*
- *More attention to critical situation, where you can follow your feeling.*
- *The haptic feedback might give a false sense of safety.*
- *I was distracted to what the reason was of the force? What is going on?*
- *Angle of attack should appear more prominently on the visuals.*
- *Training effect of each session has effect on my relation with haptic feedback.*
- *If possible, make the haptic settings adjustable to personal preferences.*