

Evaluation of an Approach Stabilization Advisory system in a B737 full flight simulator

*Frédéric Dehais¹, Fabrice Cauchard¹, Frank Rister², Yujia Cao³, Ivan Lacko³,
Frantisek Mikulu³, Felix Helmke⁴ & Jan-Patrick Osterloh⁵*
¹ISAE, DMIA, Université de Toulouse, France
²Truestream aerospace human factors, Hamburg, Germany
³Honeywell International s.r.o., Brno, Czech Republic
⁴Lufthansa Flight Training, Berlin, Germany
⁵OFFIS Institute for Information Technology, Oldenburg, Germany

Abstract

Unstabilized approach has been identified to be a major causal factor of approach-and-landing accidents (e.g. off-runway touchdowns, tail-strikes etc.). In the D3CoS project, we conducted experiments in order to analyze pilots' workload during approaches. Therefore 15 type-rated, commercial pilots flew 4 different approaches each in a B737 full flight simulator. Geometry characteristics, winds and weather conditions were manipulated in order to induce unstabilized approaches. The pilot flying's eye gaze, heart rate and subjective data (NASA-TLX) were collected. Flight data were also recorded and aggregated with an algorithm to provide a stabilization performance indicator. Flight data analysis suggests that the scenarios were able to induce unstabilized approaches. Moreover, our results showed that only half of the unstabilized approaches were subjectively perceived as critical by the participants. Interestingly enough, a scenario at Dalaman airport was very efficient to induce unstabilized approach and elicited higher physiological responses, as well as higher Nasa TLX scores. The next step is to implement an Approach Stabilization Advisory System (AStA) that monitors aircraft performance/configuration and pilot's behavior/cognitive state. When AStA detects a potential occurrence of an unstabilized approach, it suggests corrective actions to re-stabilize the approach or to go-around. AStA will be tested in the next experimental campaign of D3CoS.

Introduction

Unstabilized approaches have been identified to be a major causal factor of approach-and-landing accidents (e.g. runway overruns, tail-strikes etc.). Poor pilot performance in aircraft handling, system control or crew resource management during approach and landing reveal that, from the years 2001 to 2010, 49% of all fatal accidents worldwide occurred during the initial approach, final approach, or in the landing phase (Boeing, 2010). To approach and land safely, pilots are required to follow approach profiles fulfilling predetermined stabilization criteria based on

flight parameters defined by the authorities, such as vertical speed, airspeed, or landing configuration in relation to the height above ground (ICAO, 2006; Airbus, 2006, 2009). If the criteria for stabilized approach are not met at the stabilization height (e.g. 1000ft), a go-around is mandatory. However, continuation of an unstabilized approach has been found to be a causal factor in 40% of all approach-and-landing accidents and in 75% of runway excursions or overruns (Flight Safety Foundation, 2009). This covers 19 percent of all aviation accidents since 1988! Combined with a system philosophy based on a master (human)-slave (machine) relation, today's flight deck automation has a significant, negative impact on this demanding flight phase. Mode confusion, or improper system state awareness, significantly contribute to approach destabilization (Sarter & Woods, 1995). The aerodynamic characteristics of modern aircraft wings aggravate the competing physical interplay of altitude loss, deceleration, (vertical) speed restrictions and airplane configuration. This always puts some sacrificing factors into the pilot's decision-making process and lead to high workload situation. A way to avoid this drawback is to develop an assistant system to enhance situation awareness and to facilitate decision-making during approach and landing. However, the design of such systems is critical (Ayaz et al., 2011) so as to provide assistance only when necessary (Scerbo, 1996; Hancock & Verwey, 1997). Therefore as key issue is to develop a reliable method to derive the operator's mental workload (Parasuraman, 2003).

In an attempt to implement an assisting tool to facilitate landing, a first step is to conduct experiments to analyse pilot's behaviour and especially pilot's workload during both stabilized and unstabilized approach (Rister et al., 2012). In aeronautics, the impact of workload on flight performance is a well-known issue (Roscoe, 1978; Sloan & Cooper, 1986; Stokes & Kite, 1994, Causse, Dehais, Faaland & Cauchard, 2012). Various subjective methods are commonly used to assess this concept from a subjective point of view such as NASA TLX (Hart & Staveland, 1998). Moreover, numerous studies with psychophysiological measurements have been conducted to derive level of stress and workload from measurements of the autonomic nervous system (ANS) activity. Indeed, Veltman & Gaillard (1996) showed that heart rate (HR) and blood pressure were both affected by the levels of task difficulty of segments during a simulated flight scenario. Lee and Liu (2003) showed that HR variability (Δ HR) evolved significantly according to the flight phase in a Boeing 747-400 flight simulator. In their study, the Δ HR was greatest during landing (18.8 bpm), followed by take-off (14.2 bpm), approach (10.6 bpm), and cruise (7.1 bpm) phase. In addition, Lee and Liu found that the Δ HR was significantly related to mental workload (as assessed by NASA TLX). Similarly, De Rivecourt et al. (2008) and Wilson (2002) showed that HR increased in response to the evolution of the mental demand.

In this study we aimed at analysing psychophysiological, behavioural and subjective responses when pilots faced unstabilized approaches. All these measurements were compared to normative (procedural) data that have been modelled with a formal procedure editor (Rister et al., 2012). This first of three experimental cycles took 15 pilots into approaches of 4 different airports on a Boeing 737 full flight simulator of Lufthansa Flight Training.

Material and method

Participants

Fifteen healthy pilots (all men; mean age = 46.9 years, SD = 17.7; flight experience: 12868 hours, range = 9814 hours).

Flight simulator

The experiments were conducted on a Lufthansa Boeing 737-300 full-flight-simulator of the CAE 600 series. It has a hydrostatic motion system with six degrees of freedom (6DOF), a Rockwell Suprawide Vision System and is certified as Level D/Zero Flight Time – this means, that the simulator reflects the aircraft so realistic, that operator training can be accomplished without the necessity to do further training on the real aircraft before the trainee pilot is allowed to fly with passengers on board.

Scenario description

Scenarios of the experiments include ILS straight-in approaches as well as circling/visual offset approaches at 4 different airports, namely Frankfurt (FRA), Hannover (HAJ), Amsterdam (AMS) and Dalaman (DME). FRA was used for collecting baseline data and for getting the subject accustomed to the experimental situation. It was flown once by each subject. The other 3 airports were expected to induce unstabilized approaches, due to their approach geometry characteristics, induced tailwinds and other weather conditions.

Procedure

A subject pilot arrived at Lufthansa one hour ahead of the dedicated simulator time slot. Before entering the simulator, the experimenter introduced the eye tracker and the ECG sensors to the subject. The instructor pilot then briefed the subject on what is flown (airfields, approach-types) and finally, the subject filled in a demographics questionnaire and performed a computer-based test (to account for contributing human factors, e.g. time of day, biorhythm, rest period, cognitive aging). The subject then received NASA-TLX Workload Scale Instructions from the experimenter.

After entering the simulator, the experimenter mounted the ECG sensors and eye tracker. The subject was asked to adopt a relaxed position and close his/her eyes while a 5-minute ECG baseline was collected. Then the eye tracker was calibrated using a few IR-labels located on the cockpit panels. It was planned for each subject to fly 7 scenarios in sequence. Before each flight, the subject received all required documents (approach chart, clearances, weather info etc.) and completed all necessary preparations (set-ups, briefs etc.). After each scenario, the subject rated his/her task load via the NASA-TLX. At the end of the experiment, subject was asked to complete a debriefing questionnaire. The duration of one experiment was about three hours.

Psychophysiological measurement

An electrocardiogram (ProComp Infinity system, Thought Technology) was used to collect (sampling rate = 2048 Hz) the participants' cardiac activity during the entirety of each flight. Three electrodes connected to an extender cable were applied to the participant's chest using a Uni-Gel to enhance the quality of the signal. The BioGraph Infiniti© software was used to export and filter the heart rate (HR) derived from the inter-beat interval. Due to a commonly observed difference in HR baseline values among participants, HR values were then standardized to allow a between-participant comparison. The mean HR of the working period (whole flight duration or a given flight phase) was divided by the mean HR calculated during rest. This data reduction provided the mean HR response during flying periods. Mean HR was computed for each flight phase and for the whole flight duration.

Results

Statistical Analysis

All behavioural data were analyzed with Statistica 7.1 (© StatSoft). The Kolmogorov-Smirnov goodness-of-fit test showed that data distribution was normal, therefore, parametric repeated-measures ANOVA and Fisher's LSD (Least Significant Difference) were used to examine the effects of the landing scenario type on HR and Nasa TLX scores.

Behavioural results

Altogether, 15 subject pilots flew 94 approaches within 32 simulator hours. According to the debriefing questionnaires, the subjects stated that they were not always stabilized in 10 out of the 94 approaches. We pre-analyzed 87 flight parameters that were recorded from the simulations, out of which 58 were used as indicators for the detection of unstabilized approaches. According to the data viewed to date, 26 of the 94 approaches were not stabilized at all stages, or stabilized very late (close to the runway), or even stayed unstabilized until touchdown or go-around initiation. Six approaches out of the 26 remained unstabilized past the 1000ft gate with 3 or more parameters exceeding limits. Only two pilots subsequently performed a go-around. The 4 remaining pilots landed unstabilized with high descent rates, low power settings and high landing speeds.

Subjective measurements: self reports and Nasa TLX score

Thirteen approaches were reported to be unstabilized by the participants during the debriefing: 10 occurred during the Dalaman landing, and one for each of the three other scenarios.

Mean Nasa TLX score is plotted in Figure 1 as a function of scenario type. The ANOVA on Nasa TLX mean score revealed a main effect of scenario type ($p < .0001$). Post-hoc pairwise comparisons showed that Dalaman scenario lead to significantly higher mean Nasa TLX score than all the other scenarios: Amsterdam ($p < .003$), Frankfurt ($p < .001$), and Hannover ($p < .0001$).

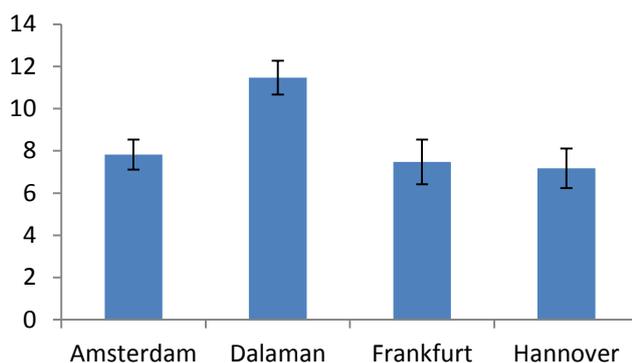


Figure 1. Mean Nasa TLX scores for each scenario. Bars represent standard errors.

Physiology

Mean HR response is plotted in Figure 2 as a function of scenario type. The ANOVA on HR showed a main effect of scenario type ($p < .001$). Post-hoc pair wise comparisons showed that Dalaman scenario elicited statistically higher HR than the other scenarios: Amsterdam ($p < .004$), Frankfurt ($p < .05$), Hannover ($p < .037$).

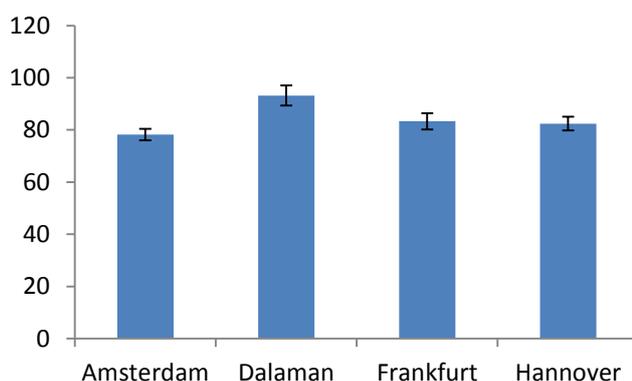


Figure 2. Mean HR for each scenario. Bars represent standard errors.

Discussion

The objective of this study was to collect behavioural, subjective and physiological data during unstabilized approach. In this perspective, different experimental scenarios were designed to induce such critical landings. Taken together our results suggest that this problem is an important issue for flight safety. First, the analysis of the flight parameters revealed that 27,6% of the approaches were objectively categorized as unstabilized. Interestingly enough, 10 of these approaches lead the pilots to hazardous behaviour such as perseveration (Dehais, Causse, Vachon & Tremblay, 2012) or violation (Reason, 95). Secondly, it is worth noticing that pilots sometimes failed to

perceive the degradation of their flight profile.. This was particularly true during the Dalaman scenario. Indeed, this latter scenario was particularly difficult to handle as it was a non-standard approach geometry with a high/steep descent gradient and especially involving time-constraint decision-making (i.e. a choice between a straight-in or circle approach): when visual with the runway (breaking clouds), the pilot needs to decide whether (s)he is in a position to safely land straight-in, or rather join the visual traffic pattern to increase track mileage to touchdown and get a better stabilization subsequently. This subjective feedback is consistent with the Nasa TLX scores and the physiological measurements. Indeed, our measurements showed that the Dalaman scenario elicited statistically higher mean Nasa TLX score and mean HR responses than all the other scenarios. These results suggest that this scenario, and by extension unstabilized approaches, may lead to higher workload, psychological stress and the mobilization of mental energy to deal with the situation (Causse, Sénard, Démonet, & Pastor, 2010).

Perspectives

These experiments delivered precious data to support our hypotheses that approach destabilization can be made predictable and that machine agents have the potential to counteract unstabilized approach in impending phases. During these experiments, the participants were also equipped with an eye tracker. These data are currently processed as shown in Figure 3. We therefore intend to correlate these psycho-physiological deteriorations (e.g slow scan/too rapid scan) to the pilot-machine interactions and detect deviations from normative activities.

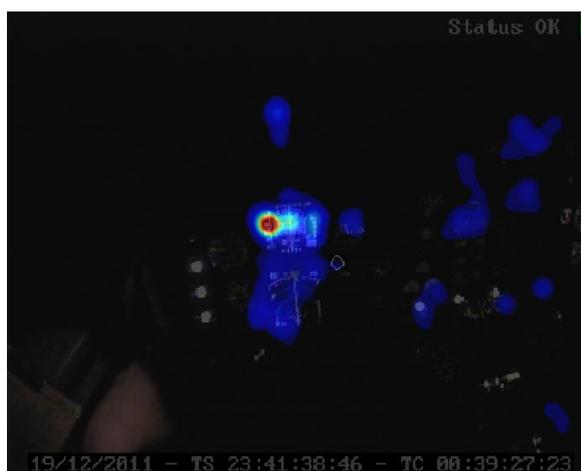


Figure 3. Amsterdam landing: heatmap. Note that the pilots is mostly focused on the speed indicator, neglecting critical parameters such as altitude, visual cues, autopilot mode indicator...

The next step is to implement an Approach Stabilization Advisory System (AStA) that monitors aircraft performance/configuration and pilot's behaviour/cognitive state. When AStA detects the potential occurrence of an unstabilized approach, it suggests corrective actions to re-stabilize the approach or to go-around. AStA will be tested in

the next experimental campaign. In this currently ongoing campaign, entire flight profiles are planned to be flown by the same subjects, and a first version of AStA demonstrator is being connected to a 737NG flight simulator at Lufthansa. A formal task model, specified with a tool called PED, is used to simulate and predict workload, allowing to fine tune the workload of the scenarios for the campaign. Also, workload peaks will, in conjunction with cycle two findings, allow for cooperative system design, e.g. testing task-reallocation functions and shared authority amongst human and machine agents during the third round of experiments in 2013.

Acknowledgement

This research has been performed with support from the EU ARTEMIS JU project D3CoS (<http://www.d3cos.eu/>) SP-8 - GA No.: 269336. Any opinions, recommendations, findings, or conclusions expressed herein are those of the authors and do not necessarily reflect the views of the ARTEMIS JU and/or the EC.

Bibliography

- Airbus (2006). Flight Operations Briefing Notes – Approach Techniques: Flying Stabilized Approaches. Blagnac, France: Airbus Customer Services.
- Airbus (2009). A320 Flight Crew Training Manual. Toulouse, France: Airbus Industries.
- Ayaz, H., Shewokis, P. A., Bunce, S., Izzetoglu, K., Willems, B., & Onaral, B. (2012). Optical brain monitoring for operator training and mental workload assessment. *NeuroImage*, 59, 36-47.
- Boeing (2010). Statistical Summary of Commercial Jet Airplane Accidents. Retained from the internet : <http://www.boeing.com/news/techissues/pdf/statsum.pdf>.
- Causse, M, Sénard, JM, Démonet, JF, & Pastor, J. (2010). Monitoring cognitive and emotional processes through pupil and cardiac response during dynamic versus logical task. *Applied psychophysiology and biofeedback*, 35(2), 115-123.
- Causse, M., Dehais, F., Faaland, P.O., & Cauchard, F. (2012) An analysis of mental workload and stress in pilots during actual flight using heart rate. ICRAT'12 conference
- Dehais, F., Causse, Vachon, F., & Tremblay, S. (2012) Cognitive conflict in human-automation interactions: a psychophysiological study. *Applied Ergonomics*, 43, 588–595
- De Rivecourt, M., Kuperus, M.N., Post, W.J., & Mulder, L.J. (2008). Cardiovascular and eye activity measures as indices for momentary changes in mental effort during simulated flight. *Ergonomics*, 51, 1295–1319
- Flight Safety Foundation. (2009). Reducing the Risk of Runway Excursion. Alexandria, VA: Flight Safety Foundation.
- Hancock, P., Verwey, W. (1997). Fatigue, workload and adaptive driver systems. *Accident Analysis and Prevention*, 29, 495–506.

- Hart, S.G., & Staveland, L.E. (1988). Development of a Multi-dimensional Workload Rating Scale: Results of Empirical and Theoretical Research. In Hancock, P.A., & Meshkati N.(Eds.): *Human Mental Workload*. Amsterdam: Elsevier.
- ICAO (2006). Procedures for Air Navigation Services – Aircraft Operations (PANS OPS). In: ICAO Document 8168, Part IX, Ch. 1 – Stabilized Approach – Parameters. Montreal: International Civil Aviation Organization.
- Lee, Y., & Liu, B. (2003). Inflight workload assessment: Comparison of subjective and physiological measurements. *Aviation, Space and Environmental Medicine*, 74, 1078-1084.
- Parasuraman, R. (2003). Neuroergonomics: research and practice. *Theoretical Issues in Ergonomics Science*, 4, 5–20.
- Reason, J. (1995). A systems approach to organizational error. *Ergonomics*, 38, 1708-1721