Master internship report

Measuring pilot-affected safety of flight

東京大学
THE UNIVERSITY OF TOKYO

Marijn Nijenhuis
Mechanical Engineering
University of Twente
October 2013

Measuring pilot-affected safety of flight

Report of master internship on pilot-affected safety of flight in the Department of Aeronautics and Astronautics at The University of Tokyo.

Marijn Nijenhuis
October 2013

Supervision by J.O. Entzinger, The University of Tokyo
Examination by R.G.K.M. Aarts, University of Twente

東京大学
THE UNIVERSITY OF TOKYO

UNIVERSITY OF TWENTE.

Acknowledgements

The research project that this assignment contributes to is financially supported by the Japan Aerospace Exploration Agency (JAXA). My internship was partially financed by a monthly stipend from the Twente Mobility Fund.

I would like to thank Dr. Kohei Funabiki from JAXA for providing the eye-mark camera, Professor Dr. Shinji Suzuki for the opportunity to do research at The University of Tokyo, and Mr. Daiichiro Matsunaga for his assistance with the operation of the simulator equipment. My grateful thanks go out to Ex-JAL Captain Mr. Tsuneharu Uemura for kindly and diligently sharing his flight experience and providing the expert's point of view.

Dr. Jorg Entzinger has been the daily supervisor of my internship and life in Japan; his efforts were crucial in making my stay a very pleasant and fascinating one, for which I am grateful. I would like to sincerely thank him for the many interesting discussions, extensive review and new perspective I obtained, both professionally and personally.

I owe thanks to Mr. Morinobu Chikama and people at the Suzuki-Tsuchiya laboratory for helping me out with countless translations and making me feel very welcome and comfortable.

This has been a truly remarkable experience.

Contents

Chapter 1

Introduction

This report is the result of my internship at The University of Tokyo. As part of my master programme I have been given the opportunity to spend the alloted internship period abroad as "international internship trainee" to partake in a research project in Japan.

To me personally, facing some career decisions soon, it provided an insider's look into the academic world of research. To the people at the Department of Aeronautics and Astronautics my exploratory assignment should give some leads to further the initial phase of the project.

This chapter will continue with an explanation of the research subject. The next one (chapter 2) will provide some background information. After that, the analysis techniques are detailed (chapter 3), to be used in the experiments (chapter 4).

1.1 Project on safety of curved approaches

In line with their policy of stimulating research on aeronautics in Japan, The Japan Aerospace Exploration Agency (JAXA) has agreed to contribute to a project proposed by researcher dr. Entzinger from the School of Engineering of The University of Tokyo. The proposal outlines a nearly 3-year project that goes by the title "Research on the pilot's supervision, decision making and intervention requirements for automatic curved approach systems". It intends to look at the role of human factors in aviation and, specifically, in flights that exhibit a curved approach path.

When an aircraft is descending and approaching the destination airport, there is a strict set of rules that prescribe the correct procedure for doing so. Especially in bad weather conditions — when the pilot has to rely on his instruments without visual cues from outside — the aircraft has to adhere to a specified flight path. Commonly, this path consists of a series of connected straight lines (legs) that lead an aircraft through the airspace onto the runway. Such a 'straight' approach route aligns with the extended runway centerline a certain distance x before the landing point.

Technological progress, particularly concerning the navigation equipment, has now given aircraft the ability to accurately track curved flight paths. In the recently introduced 'required navigation performance' (RNP) specification for approach procedures, these paths have been included as special RNP Authorization Required (RNP AR) approaches, to guarantee a certain level of operational reliability. These new 'curved' approach routes greatly reduce the distance x before the landing point. Appendix A has more details about instrument approaches.

RNP is relatively new and in the process of adoption throughout the aviation industry, because it is said to lead to more efficient procedures and hence expedite air traffic [4]. The implementation comprises the introduction of new procedures and certifications for aircraft, aircrew and airport (personnel). Aeronautical research related to this development so far has focussed on implementation specifics, such as the approach path design and automation equipment. Little consideration has been given to human factors aspects, such as the pilot's situational awareness and supervision of cockpit automation.

Despite the large contribution of automation in aircraft control, a good understanding of the human factors concerning the pilot is essential for safe flight operation. The pilot, still shouldering the final responsibility in his new supervising role, has to be able to intervene and assume control to correct errors. To do so, he needs the skill to discern problems in the automated flight control and take over flight operation, unto a safe conclusion.

The notion that this is more difficult for the new curved approach than for the conventional straight approach underlies the research project. To articulate this idea, research objectives have been formulated accordingly:

- To understand differences in mental modes and cognitive processes between curved and straight-in approaches;
- To find out how to best support (through training or interfaces) the pilot in supervising automation and decision making.

Correspondingly, the overarching hypothesis is that *curved approach paths are less safe than straight approach paths*, due to the assumed relationship between increased (procedural and technological) complexity and reduced (flight) safety.

1.2 Internship assignment

My internship assignment began in January 2013, only half a year after the start of the project. Consequently, the intended effect of my assignment is not so much the acquisition of final and concluding experimental results, but rather the exploration of the different methods of getting them. My investigation of the feasibility of different methods should provide some insight in what course to pursue during the later research phases.

1.2.1 Hypotheses

Given the research hypothesis, it is easy to see that a tempting and decisive experiment to conduct at a later stage in the course of the project would be an actual flight in a commercial airliner on one of the new curved approach paths, whilst assessing and comparing the safety with that of a flight on a straight-in approach.

Obviously, such an experiment should be predicated on a good understanding and a tested measure of flight safety to be able to compare different flights. This means that a good starting point is the development of such a measure and its application to conventional approaches first. So, by formulating substitute hypotheses, the ideal and expensive experiment can be scaled down to obtain a more feasible version, which a) uses readily available facilities at the university and b) still provides the necessary insight on the usability of analysis techniques for the time being. Figure 1.1 shows how the experiment conditions are toned down, subject to new hypotheses, and put the university's fixed-base flight simulator and experienced ex-JAL airline captain to good use in this project.

The scope of my internship assignment will be defined and limited by these new hypotheses:

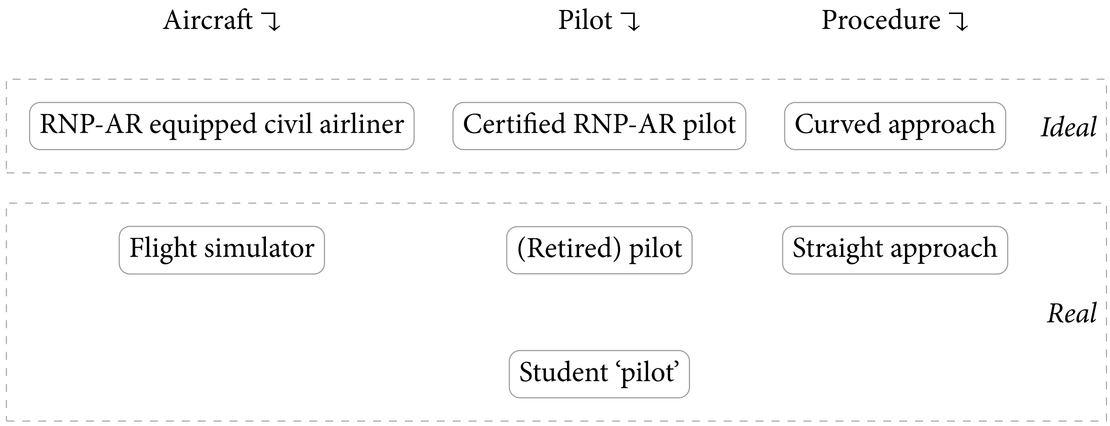


Figure 1.1: For the initial experiments the ideal setup is somewhat moderated.

- *approaches in IMC are less safe than in VMC* (appendix A) — this takes into account the importance of information availability;
- *the short final of the approach is less safe than the long final* — this takes into account the importance of time and timing;
- *a novice pilot flies less safe than an expert pilot* — this takes into account the importance of training.

This choice makes for a manageable and time-limited assignment within the larger project, while contributing to the main research objectives with a proof of concept.

1.2.2 Methods

The way this intership can contribute to the research project is by investigating fruitful methods that are applicable to later and more complex experiments, too. Specifically, the issue of measuring and comparing flight safety has to be tackled.

It is worthwile noting that, as reported by the International Civil Aviation Organization in 2013, the global commercial aviation accident rate is very low and has never been lower [1]. This means that unsafe events have become increasingly more rare. In order to keep pushing safety to a higher level, more and more extreme situations will have to be analyzed for obtaining significant results.

To address the issue of measuring and comparing flight safety, figure 1.2 visualizes the relationship between pilot, aircraft and relevant signals (feedback loops and disturbances have been omitted), as commonplace in typical control theoretic considerations. By tapping into this 'loop', various kinds of information about the flight operation can be obtained. The idea is that since flight performance and, hence, safety originate from this chain — though commonly observed in a rather statistical long-term manner, it has to be retrievable in a more direct manner as well.

> As a sidenote; theoretically, in case that the dynamics of the elements in the loop of figure 1.2 were known, it would not matter where one took measurements. However,

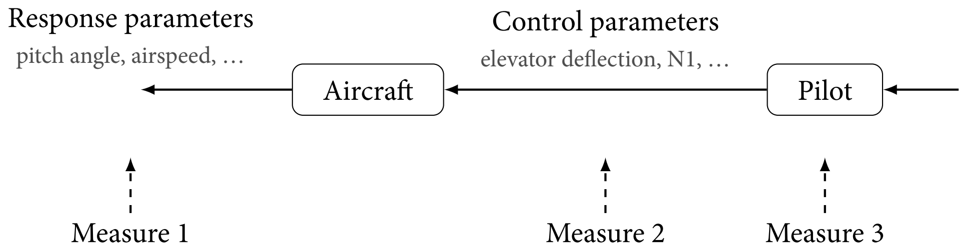


Figure 1.2: Different measures on which to base the concept of safety.

some inputs and especially the pilot dynamics remain largely unknown, making it useful to look at different signals.

Measures 1 and 2 are based on data that is typically provided by flight recorders: the first one is based on the aircraft response and the second one on aircraft controls that the pilot actuates with his actions. A third option is extracting data from the human pilot.

These measures provide starting points for relating available signals to the concept of flight safety. To this end, background information as a result of a literature investigation will be discussed, followed by a more concrete formulation of these measures.

Chapter 2

Background information

To make the research hypotheses more readily testable in experiments, this chapter will elaborate on approach procedures (section 2.1) and the relation between human performance and mental workload (section 2.2). This then leads to an expansion of the main safety hypothesis (section 2.3).

2.1 Approach procedures

Some important flight concepts related to the approach phase and relevant to this research are presented here. Appendix A details additional information on this subject.

2.1.1 Approach phase of flight

When entering the airspace of the destination airport for a descent, the approach phase of the flight will begin. There is a set of rules that specify how to safely land the aircraft, depending on weather conditions, available equipment, crew training and other traffic [4].

2.1.2 Meteorological conditions and flight rules

Weather conditions can preclude flight based on outside visual references. In that case, a pilot has to rely on instrument readings and stricter procedures, which is considered to be more challenging.

For flight purposes the weather conditions are referred to as either *visual* meteorological conditions (VMC) or *instrument* meteorological conditions (IMC), based on visibility, distance from clouds and cloud ceilings [5]. (VMC means "better" weather than IMC.)

Only in VMC ("good" weather), flight is allowed under visual flight rules (VFR), whereas degradation to IMC ("bad" weather) requires operation conducted under instrument flight rules (IFR) because of increased safety risks. These stricter IFR are also allowed to be used when conditions clear up and become VMC again, but for obvious safety reasons, this is not the case vice versa. The IFR are established by aviation authorities such as the (American) Federal Aviation Authority (FAA).

Under IFR there are special instrument approach procedures (IAPs) for the "orderly transfer of an aircraft [...] from the beginning of the initial approach to a landing [...]" [5].

2.1.3 RNP

Up to now it has been customary to define different kind of approaches (i.e. different IAPs) in terms of the navigational equipment that they are based on. The ground-based instrument landing system (ILS) is commonly used by commercial airliners and provides accurate horizontal and vertical guidance to a specific runway [4, 5].

For the benefits of shorter flight times, reduced fuel consumption, and increased traffic capacity [4, 47], 'required navigation performance' (RNP) approaches are being developed. The flight paths of the conventional procedures are usually dictated by the location of ground facilities, whereas these new approaches take full advantage of a GNSS (such as GPS) together with systems that guarantee on-board accuracy and integrity monitoring [45]. That way, aircraft can be positioned more accurately, requiring less separation between obstacles and other air traffic, even allowing curved flight paths. In essence, ground stations can only provide straight-in guidance, whereas a GNSS system — fully air-based — has flexiblity regarding the flight path. The adoption of RNP should lead to the "cleanest straight line [and] constant radius turn route to everywhere […]" [47].

Figure 2.1 shows the flight path (blue solid line) and corresponding ground track (red dotted line) of the last phase of a non-RNP, conventional approach and landing at Tokyo Haneda Airport. Notice that the path consists of two straight segments: the final segment is aligned with the runway and has to be straight for the ILS to work (i.e. provide guidance onto the runway).

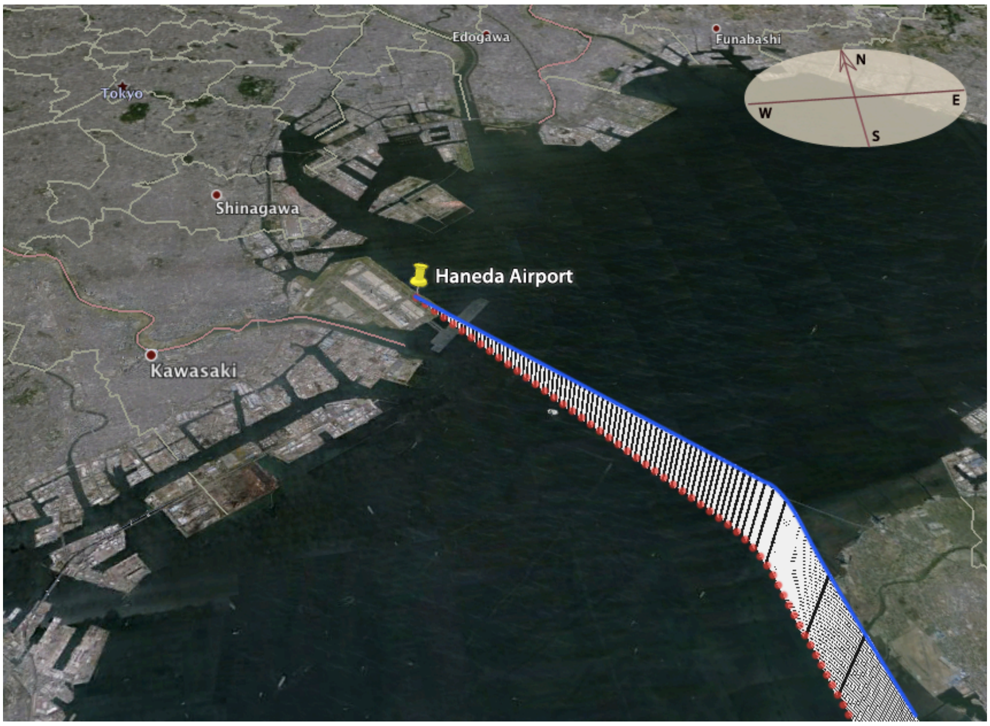


Figure 2.1: Last phase of conventional approach and landing at Haneda Airport, Tokyo Bay. Blue: flight path, red: ground track.

2.1.4 Stabilized approach

According to the Flight Safety Foundation, the concept of stabilized approach contributes significantly to reducing aviation accidents: "unstabilized approaches [...] were a causal factor in 66 of percent of 76 approach-and-landing accidents and serious accidents worldwide in 1984 through 1997" [16]. They describe such an approach as flying either too high/low or too fast/slow. More thoroughly, an approach is defined to be stabilized when at a certain minimum height the flight parameters do not exceed established criteria, published in the airline's standard operating procedures (SOPs). The three essential parameters that have to be stabilized are aircraft track, flight path angle and airspeed.

The FAA clearly explains why approaches should be stabilized, from a pilot's point of view [4], especially in IMC, where the pilot relies on instrument information to properly maneuver the aircraft or monitor autopilot performance. Even before reaching the decision point of the approach, the pilot already has to form a decision concerning the probable success of the approach.

> [The pilot's] decision-making process requires [him] to be able to determine displacements from the course or glidepath centerline, to mentally project the aircraft's three-dimensional flight path by referring to flight instruments, and then apply control inputs as necessary to achieve and maintain the desired approach path. This process is simplified by maintaining a constant approach speed, descent rate, vertical flight path, and configuration during the final stages of an approach. This is referred to as the stabilized approach concept [4].

2.2 Mental workload

A literature investigation has shown that human performance and mental workload are related (although not in a very straightforward manner). This makes the workload concept interesting in this research for assessing the flyability of new kinds of approach procedures.

Stemming from a lack of insight in the cognitive processes that govern the relation between operator performance, workload, resources, effort and task load is the absence of a generally-accepted definition of these terms. Therefore, some reasonable notions of workload and performance from literature will be echoed here. Our vision and definition, used throughout this research, will be outlined next.

2.2.1 Descriptions in literature

Mental workload is an expression of the task demands placed on an operator. These demands are influenced by the goals that are to be achieved, the time available for accomplishing these goals and the operator's skill [13]. Kramer [28] defines mental workload as the costs that a human operator receives while performing tasks, or in more detail:

> Although there is no universally accepted definition of mental workload, the recent consensus suggests that mental workload can be conceptualized as the interaction between the structure of systems and tasks on the one hand, and the capabilities, motivation, and state of the human operator on the other. [28]

In discussions about the physiological reaction to mental workload, authors often bring up the concept of mental effort. Vicente et al. [53] regard mental effort as the “operator’s reaction to the input load, [...] determined by the internal goals and criteria [the subjects] choose to adopt”. Aasman et al. [2] have a similar view, refering to effort as the “the willingness to spend capacity in order to cope with the demands of the task”.

The distinction between mental workload and effort is expressed quite clearly by Colombi et al. [13], who regard workload as a measure of the demanded mental effort for a task relative to the mental resources the operator possesses. It is useful to distinguish between the two concepts of mental workload and effort, as research suggests that the latter can be correlated with the intensity of physiological reactions [2, 11, 52, 53].

“The relationship between workload and performance has often been represented as an inverted U, with poor performance resulting when workload is quite low [...] and when it is quite high” [22]. This means that an optimal level of workload exists. Low workload levels lead to poor performance due to lack of stimulation, whereas high levels provide excessive demands. “Additionally, when operating [...] systems that require an excessive number or difficulty of tasks to be performed within limited time, the human operator will become overloaded and will be incapable of performing all of the tasks, potentially foregoing tasks that are critical for safe system operation” [13].

Ayaz et al. [6] stress the “key feature of the concept of mental workload [...] that it can be dissociated from performance output” and that, consequently, “it is important to assess mental workload independently of performance measures”. Moreover, since a workload increase can be predictive of a looming performance breakdown, it should be evaluated in advance. They refer to neuroergonomic approaches that measure brain activity for a reliable assessment of human mental workload in complex work environments.

2.2.2 Our view

We consider the concepts of *performance* (P) and *workload* (W) to be the result of the interplay of the factors

- *Available mental resources* (R): the mental resources that an operator has available for a task. These determine whether an operator is actually capable of performing a given task. The available resources are determined by operator experience, skill and other tasks he is performing at the moment.
- *Task load* (T): the task load that is imposed on an operator. It is an objective load, independent of operator skill and equal for all persons.
- *Mental effort willing to expend* (E): the mental effort that an operator is willing to invest in a task. This is dependent on fatigue, circadian rhythm, time of day, sleep deprivation, time on task, drugs, heat, noise [2], etc.

For the visualization of these factors some conventions will be adopted:

- The amount of available mental resources for a task is indicated by the height of the solid line (R).

- The magnitude or difficulty of the task load that is imposed on the operator is indicated by the grey bar (T).
- The amount of mental effort an operator is willing to expend in performing a task is indicated by the height of the red bar (E).

A safety margin is introduced to take the phenomenon of performance degradation over time into account: we propose that only a part of all available mental resources R can be used continuously or, in other words, that an operator can only temporarily use all available resources. Only a certain fraction α be used for a prolonged period of time without performance issues. This amount αR is indicated by a dashed line.

Workload definition

We consider the ratio between mental effort willing to invest and available mental resources to be the operator workload, or

$$\text{workload} := \frac{E}{R}. \tag{2.1}$$

Regarding R as a property of a person that is more or less fixed, at least for some amount of time, the effort E is crucial to workload assessment. Due to the psychological or mental nature of factors that influence E (i.e they are related to mood and willingness), this variable can differ considerably, even for one test subject.

Operator workload larger than 1 will render the operator incapable of functioning. Regarding the safety margin introduced earlier, there is a certain workload level α above which performance can only be maintained for a limited amount of time.

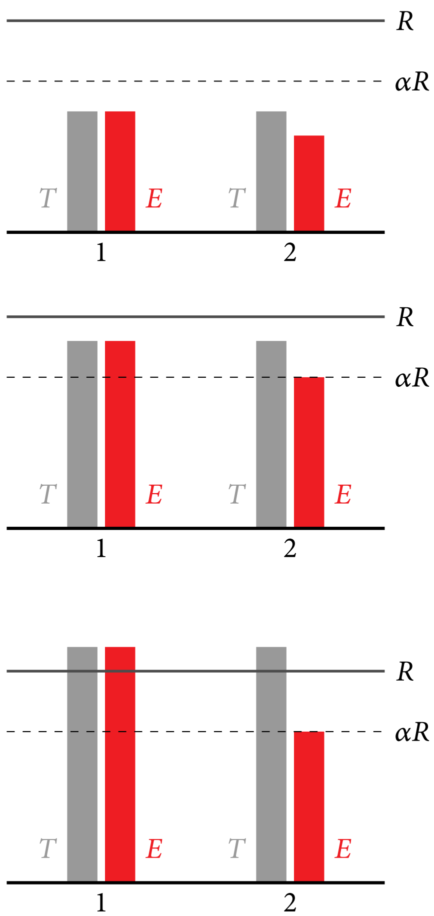
Performance definition

When an operator receives a task T, the performance is dictated by resources R and effort E. For adequate performance, both need to be sufficiently high: i.e. an operator needs to be both capable ($R \geq T$) and willing ($E \geq T$). If performance falls short (i.e. it is less than 1), the level is determined by R/T or E/T, whichever is lowest. This leads to a definition of

$$\text{performance} := \min\left(1, \frac{R}{T}, \frac{E}{T}\right). \tag{2.2}$$

Figure 2.2 lists three task load cases with different levels of effort an operator might invest and the effects on performance and workload.

To incorporate some kind of time-dimension in this representation, it is suggested to think of the horizontal baseline in figure 2.2 as time axis; the temporal course of the task load and invested effort can then be visualized intuitively by regarding them as continuous function of time, i.e. $T = T(t)$ and $E = E(t)$.



When $T \leq \alpha R$, workload is not at risk of becoming too high. Performance is dictated by E.

1. If the operator 'decides' to invest effort and match the task load ($E = T$), performance will be adequate.
2. If not ($E < T$), performance will be $E/T < 1$.

When $\alpha R < T \leq R$, the safety margin is compromised.

1. If the operator invests effort and matches the task load ($E = T$), the performance is adequate, though possibly degrading over time, since the workload is larger than α.
2. A coping strategy could be adopted: E is reduced, $E \leq \alpha R$, which compromises performance ($E/T < 1$) but limits workload to a sustainable level ($E/R < \alpha$).

When $T > R$, the task load exceeds the available resources and performance cannot be adequate, limited by $R/T < 1$.

1. Even though a motivated operator might be willing to invest the required effort ($E = T > R$), he will not function properly due to overload.
2. A coping strategy could be adopted by reducing the invested amount of effort E: this creates a tolerable workload level. Performance is then given by $\min(E/T, R/T) < 1$.

Figure 2.2: Visualization of three different task load scenarios.

2.2.3 Justification

An evaluation of pilot workload during the critical phases of flight is justified given concerns expressed by experts and pilots themselves [8, 14], warning against flight skill degradation among pilots due to use of automation in the cockpit. Casner [12] reports that pilots have been found to unwarrantably misplace trust in GPS, thinking it enhanced their situational awareness when in fact it did not. Also, the majority (88%) of recent fatal aviation accidents have been reported to be attributable to pilot error [20], making a tolerable level of mental workload for pilots an important aspect of safe flight.

2.2.4 Other research

In order to maintain and increase safety in aviation, NASA and FAA initiated a project to provide awareness of human performance issues related to performance-based navigation (PBN) [7].

Given that the role of the cockpit operator shifts from pilot to observer and that the procedures become trajectory-based with high reliance on accuracy, the observability of approaches — closely related to operator workload — is studied.

The introduction of NextGen cockpit technologies, which are expected to significantly increase the pilot's responsibilities, triggered studies of their effect on pilot workload [11].

2.2.5 Measuring workload

According to (referees of) Carlin [11], it is possible to assess workload as an index, which should be

- sensitive (to changes in task demands),
- diagnosing (the cause of workload variation),
- selective (by excluding non-contributing factors),
- unobtrusive (in that the index computation does not affect workload itself) and
- reliable.

The means of assessing workload can be categorized as subjective, performance and physiological measurements [24]. Note that Colombi remarks that truly objective measures are yet to be developed, all measures currently being to a greater or lesser extent only a proxy of actual mental workload [13].

Subjective index

Methods such as the NASA Task Load Index, Bedford Workload Rating Scale, or Subjective Workload Assessment Technique provide workload measures [11, 13, 18]. The subjective ratings of perceived workload do impose a risk on the validity of the measurement and can be intrusive [50].

Performance index

A performance measurement, such as the Multiple Resource Theory, infers workload from operator performance degradation while conducting a task [13, 18, 24]. It is based on the workload–performance relation described earlier and, hence, subject to the same pitfalls.

Physiological index

"Neurophysiological and psychophysiological variables are known to respond to cognitive demand in a relatively predictable manner" [6]. For instance, the electroencephalography (EEG) measure is known to be associated with task difficulty. Heart rate variability (HRV) has been shown to be related to mental effort [53] and indicative of different phases of flight (requiring varying levels of mental effort), even when subjective measurements cannot [11]. Ayaz et al. [6] advocate the use of functional near infrared (fNIR) spectroscopy for operator cognitive state monitoring, as it is less invasive and more practical than EEG. In an air traffic control experiment they have shown that "fNIR measures are sensitive to mental task load and practice level".

A lot of research has focussed on relating eye metrics to cognitive processes. A good overview of the suitability of pupil dilation measurements for mental workload indicators is given by Kramer [28]. The author concludes that pupil diameter can indicate mental activities, but notes that two human pupil reflexes (the "near reflex" and "light reflex") can easily overshadow the effect, making it difficult to produce reliable measurements outside a lab with carefully controlled experiment conditions. Various, somewhat loose descriptions of the mental activity effect on pupil diameter have been summarized by Boff and Lincoln [9] and, in a more recent 2009 literature overview, by Wang [54].

Predictive simulations

From mathematical psychology there are attempts to model the cognitive processes of a human operator, resulting in simulations that can predict operator workload. Although these attempts look promising, creating a model for a task as complex as flying is currently very time-consuming. More information about this kind of pilot modelling can be found in appendix B.

2.3 Hypothesis tree

The discussion about mental effort and workload (section 2.2.2) can be combined with the three hypotheses that govern this assignment (section 1.2.1). This way the 'hypothesis tree' in figure 2.3 is obtained. It is an analysis and reasoning of the consequences of reduced flight safety, focussing on the pilot's cognitive processes that could play a role.

The issue of reduced flight safety is present in all three hypotheses and summarized by 'Safety reduction'. Arguing that this is related to operator performance ('Performance reduction'), the link with the discussion of this chapter on mental workload is made.

At various places in this tree, we propose that information can be extracted, in line with the overview of the three safety measures described earlier (figure 1.2 on page 12):

1. Flight performance (based on e.g. flight recorder data) can be monitored and scored, hence the *performance index*.
2. Mental workload indicators are said to be available, referred to — for now — by *more corrective control*.
3. Mental effort has been shown to lead to *physiological responses*.

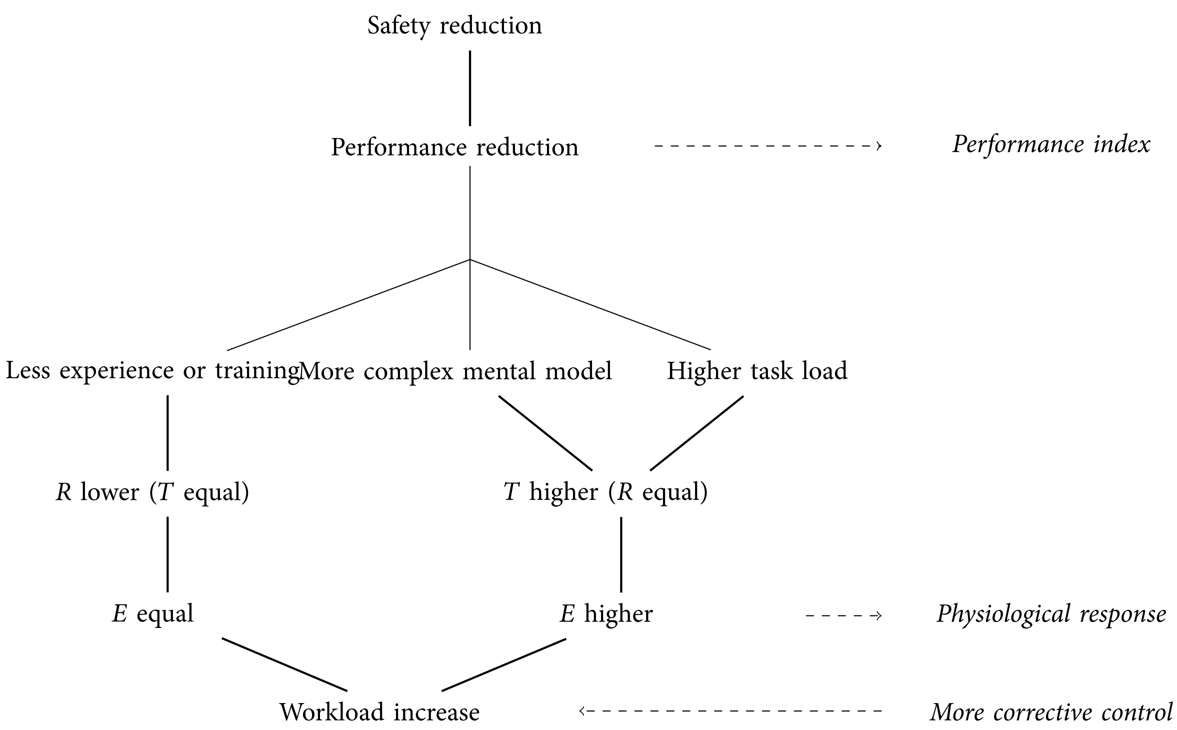


Figure 2.3: Hypothesis tree.

Chapter 3

Analysis techniques

Having discussed some background information, the three safety measures of chapter 1 (figure 1.2 page 12), repeated here, can be formalized.

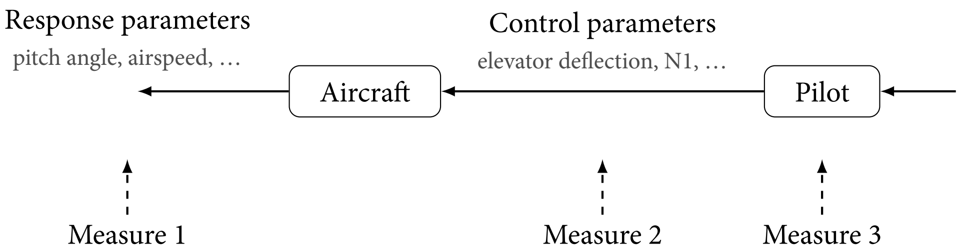


Figure 1.2: Different measures on which to base the concept of safety.

3.1 Measure 1: Time to crash

The aircraft response and performance is characterized by many parameters, which hinders a simple comparison. Given a specific definition of safety, these can be reduced to a more convenient one-dimensional measure.

3.1.1 Concept

After reviewing common workload assessment techniques (see subsection 2.2.5 for more details), Gawron [18] lists some shared issues and proposes an alternate approach: "the purpose of workload measure [is] to identify potentially dangerous situations". Leaning on the idea that they are caused by poor design, procedures, training, etc., the author concludes that "the most objective measure of danger in a situation is time until aircraft is destroyed if no control action is taken".

A similar approach was taken by the Air Force Test Center of the United States Air Force: they rely on a "Time Safety Margin" with a set of criteria for planning aircraft and dive maneuvers [23].

Defining high safety at some point during a flight to mean that it would take a long time before the aircraft crashed if at that point the pilot would no longer control the aircraft, the value of this time margin during the flight can be interpretated as measure of safety. It is based on the instantaneous performance and the idea that mistakes turn into an accident if the recovery time is low.

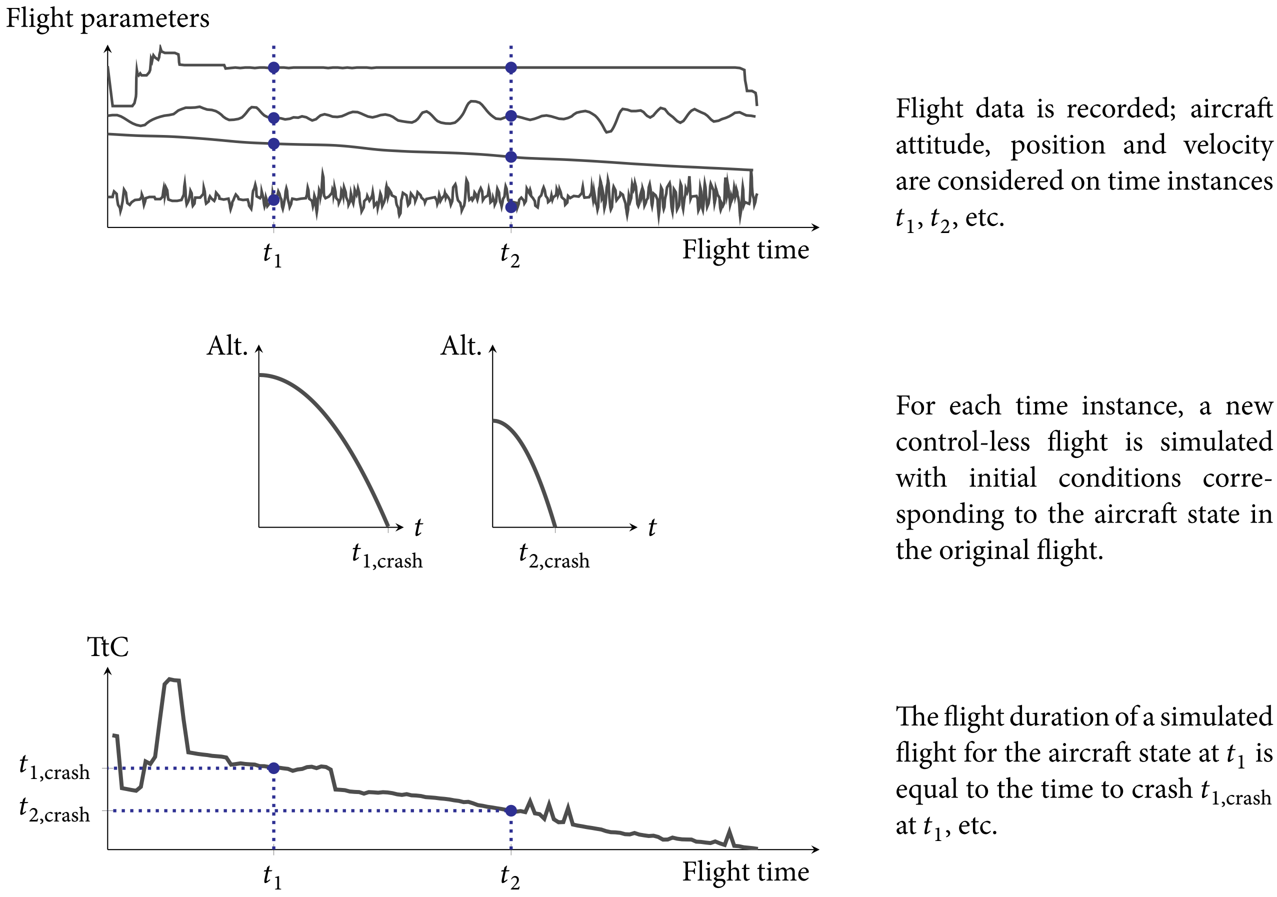


Figure 3.1: Overview of time to crash computations.

3.1.2 Computation

A time margin that indicates the time until impact if no control action is taken, can be computed from aircraft performance data (‘response parameters’ in figure 1.2). Figure 3.1 shows schematically how flight simulation software is used to produce these so-called time to crash (TtC) graphs. In order to have a reasonable time resolution, the TtC is calculated for every second of a typical 140-seconds-flight, meaning that 140 control-less flights are simulated.

The absence of pilot action is taken into account by starting a simulation given a certain aircraft state but setting (and keeping) the elevator, aileron and rudder pressure to their trimmed values at that state (as if the pilot suddenly took his hands of the yoke). The thrust value is not reset, because the thrust lever maintains the last setting when the pilots lets go of it.

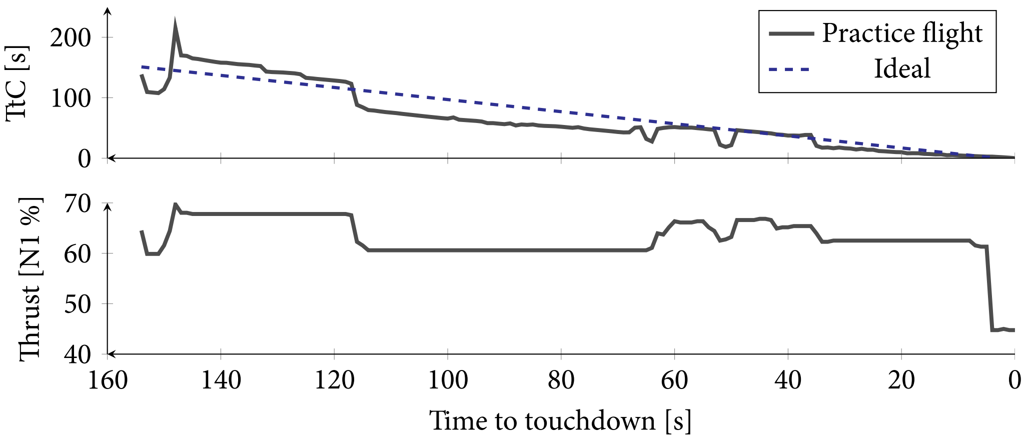


Figure 3.2: Indication of apparent relation between TtC and engine thrust setting.

3.1.3 Implementation

Stabilized approach

In terms of TtC time margin, the safest flight would be the one where $\int_t \mathrm{TtC}(t)\mathrm{d}t$ is highest. Within the framework of stabilized approaches though (section 2.1.4), a flight can be graded on its resemblance to the ideally stabilized one. Declaring this ideal approach to indicate a perfect performance, deviations from this flight should be regarded as performance drops.

Since the time margin is a function (that has a convenient safety interpretation) of the aircraft performance parameters, these deviations from the ideally stabilized approach should show up in TtC graphs and would indicate safety reductions.

Penalty function

While the TtC deviation from the ideal one, $\Delta\mathrm{TtC}(t)$, shows the time-progression, a scalar function $f(\Delta\mathrm{TtC})$ facilitates comparison of different flights. This penalty function should extract the features from $\Delta\mathrm{TtC}$ that indicate reduced safety.

During the flight training we have experienced that pitch and thrust control are essential to good flight performance. This becomes clear when e.g. an unexperienced pilot tries to fly with correctly trimmed pitch and thrust set to the ideal value. In this case, the pilot might be able to land safey, whereas manual control over thrust and unstabilized pitch would most surely lead to a crash. In line with this, inspection of flight recorder data and the corresponding TtC graphs has taught us that at least the thrust setting is visibly correlated with TtC. An example of this relation is given in figure 3.2.

Prolonged deviations

Predicated on the notion that a safe pilot should know when his instantaneous performance is inadequate, it makes sense to penalize longer-lasting deviations from the ideal TtC value. After all, for a safe flight a skilled pilot would correct such deviations promptly. We therefore propose that

$$\frac{1}{t_d}\int_0^{t_d} \Delta\mathrm{TtC}(t)\mathrm{d}t, \tag{3.1}$$

with flight duration t_d, be a contribution to the penalty function function f.

Temporary deviations

Based on the idea that quick and proper corrections by the pilot are safer and less pronounced in ΔTtC than slower inadequate corrections, it makes sense to analyze the smaller, shorter and temporary deviations from the ideal TtC. To this end, we propose a wavelet transform of ΔTtC. Wavelet analyis is briefly discussed in appendix .... Suffice it for now to say that it is suited for detecting the location and magnitude of discontinuities in a signal. By means of a heuristic approach we have found that the level 1 detail coefficients of the wavelet transform of ΔTtC with a Daubechies 2 wavelet adequately detect the smaller features in the signal. This is illustrated in figure 3.3: when the TtC is smooth, the detail coefficients $D(k), k = 1 \ldots N$ are practically zero, whereas a bump at k_1 shows up quite reliably as $|D(k_1)| > 0$. The larger the deviation is, the larger the corresponding wavelet coefficient becomes. Utilizing this behavior, the sum $\sum_k |D(k)|/N$ is included in the penalty function f as well.

Scaling

Choosing the penalty function f to be a linear combination of the temporary and prolonged deviations[1],

$$f(\Delta\mathrm{TtC}) = C_1\frac{1}{t_d}\int_0^{t_d} \Delta\mathrm{TtC}(t)\mathrm{d}t + C_2\frac{1}{N}\sum_k |D(k)|, \tag{3.2}$$

the weights C_1 and C_2 still need to be decided on, to ensure that both contributions have equal effect. These should be the same for all flights that are analyzed to make sure that findings cannot be attributed to flight-dependent choices of these weights. We therefore propose that C_1 is the inverse of the average of the numbers $\int_t \Delta\mathrm{TtC}/t_d dt$ for all analyzed flights and C_2 the inverse of the average of the numbers $\sum_k |D(k)|/N$ for all analyzed flights. Consequently, the penalty function values are dimensionless.

3.2 Measure 2: Control frequency and power

Information in the control parameters (figure 1.2) can be given a safety interpretation, as suggested by Gawron [18]: the number of control inputs per unit of time may be related to operator workload.

From our students' flight training we have learned that pitch control has a large role in flying straight courses. Although other controls are constantly being monitored, pressure on the rudder and ailerons or changes to the thrust settings are only applied occassionally. In line with this,

[1] Note that in the actual implementation of f the prolonged deviations are not calculated as a continuous integral but as a summation of the sampled ΔTtC values.

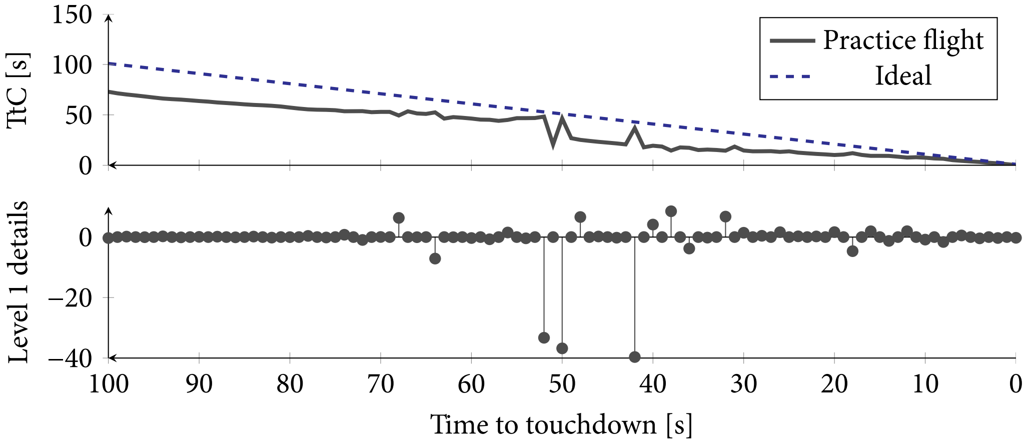


Figure 3.3: Illustration of detail coefficients of wavelet transform of ΔTtC.

inspection of the training results has taught that the elevator deflection, as imposed by the pilot, is the control input that contains the most information on control actions inflicted on the airplane.

Though a count of the number of control inputs (Gawron [18]) suggests that the control signal is binary, the elevator deflection obviously is not. By extending this interpretation for continuous signals, the frequency of the elevator deflection becomes the workload indicator.

To deal with the fact that this signal will in general consist of more than one harmonic component — it has a frequency spectrum (which makes comparison harder), the signal's power is calculated to take into account all harmonic components.

Although we arrive at the signal's power from a frequency consideration, it can be calculated directly, according to Parseval's theorem, from the time-signal $u(k), k = 1 \ldots N$ by computing $||u||_2^2/N$, where $||\cdot||_2$ indicates the ℓ^2-norm. Note that the signal's power can be computed for different time-intervals; in this report it will be mentioned specifically whether the reported power values are based on the entire flight duration, a specific part of it, or instantaneous values ($N = 1$).

3.3 Measure 3: Eye metrics

Eye metrics constitute this third measure. The use of this group of physiological indicators for mental workload and effort (section 2.2.5) is justified by their relation with operator performance (section 2.2).

3.3.1 Literature

Despite some unsuccesful attempts [9, 28], there have been promising experiments that relate pupil size to some notion of mental workload. Van Orden et al. [51] use a windowed relative pupil diameter for indicating workload in a visuospatial task. Klingner et al. [27] conduct simple tests, not taking the light reflex into account, for obtaining the task-evoked pupillary response (TEPR), which is "small […], involuntary, and reliably associated with a broad set of cognitive processes that are characterized as cognitive load". Similarly, Palinko et al. [39] introduce the mean pupil diameter change rate as a promising measure (after providing a decent literature overview). In later research, Palinko and Kun [38] build on this by adding light measurements and calculating a predictor for the eye's dilation and contraction responses to lighting changes, which they say leads to a method that could "track cognitive load-induced changes in pupil diameter". With the same goal, Marshall [35] has developed an index of cognitive activity (ICA) that she states discards the light reflex in the pupil size signal using a wavelet transformation. Schwalm [46] was able to experimentally confirm in his PhD-thesis a correlation between this index and driving performance. Stevens et al. [48], though, were unable to relate the ICA to a notion of workload.

Inconclusive as some aspects of the pupil diameter analysis may be, blink rate has been agreed upon to correlate strongly with the visual demands of a task [10, 41, 52, 55].

Moreover, Marshall [36] claims to have a set of seven eye metrics (including pupil diameter and blink frequency) that succesfully identify an operator's cognitive state in realtime, insensitive to changes in lighting and not requiring averaging over multiple trials.

3.3.2 Pupil diameter

In experiments in a (simulated) aircraft cockpit, perfect control over the intensity of the light that enters the pilot's eyes is unattainable. This means that pupil diameter measurements need to be compensated for the light reflex of the pupil post-hoc. The only method current available that is said to be capable of doing this without additionally measuring the light intensity, is the index of cognitive activity.

Details about the mathematics of the wavelet transformation in general can be found in [37, 42]. The filter that can be applied to clean the eye-camera signal for pupil diameter analysis, is described in section 3.3.4.

3.3.3 Blink frequency

The pupil diameter signal contains information about the blinking of the eye, since blinking obviously precludes an eye camera from taking diameter readings. We have developed a filtering process for inferring the blink rate from the time-history of the pupil diameter, which can be obtained with a head-mounted camera system (more details about the equipment in section 4.1.1 on page 31).

3.3.4 Blink filtering

Note that accurate kinematic descriptions of the pupil's reaction to cognitive activity are as of yet unavailable in literature; this makes filtering, when it has the purpose of obtaining a signal that is as good a representation of the pupil diameter as possible, more challenging. For the purpose of calculating a reliable blink rate, the filtering process relies less on knowledge of pupil reflexes.

Figure 3.4 shows the different steps in the filtering process, applied to illustrative data captured during flight. These steps are the result of a somewhat heuristic approach. The assumptions that govern the reasoning behind each step become increasingly more uncertain, this way making sure that the most effective and correct processing leads.

1. *Original signal*: the unprocessed signal contains incorrect diameter values, as suggested by the unlikely spikes.

2. *Detected by camera*: by design, when the camera processor cannot calculate the pupil size from the recorded images, it reports these measurements as 'unreadable': they get a specific, high numeric value. This is very likely caused by the eye lid obscuring the pupil from the camera during blinking, but could also be the result of noise. Simply marking all of these high values as moments of blinking in this case affects about 7% of the data. The beginning of each blink is indicated by a blue dot.

   The plot still shows some high and low spikes, especially in the neighborhood of already detected blinks. These are likely the result of the eyelid partially covering the eye and pupil at the beginning or end of a blink, leading to false readings.

3. *Extreme diameters*: from physiology [49] we know that the mean pupil size can vary greatly among persons. Therefore the recorded signal is inspected visually and some loose upper and lower thresholds are applied. These thresholds should only intersect the 'nearly vertical' spikes.

   Since this step shows a reduction in the number of blinks, it mostly affects those cases in which one true blink with a certain duration was reported by the camera as two shorter blinks, joined by some false threshold-exceeding reading(s).

4. *Velocity-based, next to blinks*: Stilma and Voorn [49] say that the human pupil is anatomically capable of dilating to a maximum width of 8 mm and contracting to a minimum width of 1.5 mm, depending on lighting conditions. The contraction reflex, in reaction to light, is the quickest, reducing the pupil to its minimum size in about 1 second. Based on this information we assume that the maximum pupil diameter velocity that is anatomically possible is 6.5 mm/s. Note that the described reflexes concern the reaction to light; we expect that the reaction to cognitive processes will not be quicker than the light reaction that has developed evolutionarily to protect the eye's retina.

   By discarding measurements that differ more than 6.5 mm/s from each other, within k samples that are only adjacent to already-detected blinks of at least length n, the number of blinks is not affected — as it should, but some false readings due to partial blinks are likely to have now been detected as blink moments. (For the data in figure 3.4, $k = 3, n = 2$ appear to give good results.)

5. *Singletons, first time*: inspection of the filtered data up to this stage shows that some 'singletons' have been created: one-sample measurements that are preceded and followed by blinks. These single values do not contain useful information and are more likely to be falsely separating one actual blink into two shorter ones. Removing these reduces the number of blinks present. The effect of this step in figure 3.4 is best visible by comparing the blue dots of the fourth and fifth plot.

We propose that these five steps are a fairly solid method for detecting blinks in the pupil size signal. It could still contain measurements that do not accurately represent the pupil diameter at that time, but for the purpose of calculating blink rate, these steps should suffice.

It is important to note that the filtering steps are predicated on the eye-camera processor marking some of the measurements as 'unreadable'. Though this is likely the result from mostly a blinking eye lid, other disturbances which are not related to blinking could also lead to measurements being marked this way.

Three additional steps are presented that can be applied, to get a cleaner pupil diameter signal; i.e. probably not yielding better blink detection but identifying more incorrect measurements.

6. *Velocity-based, anywhere*: the maximum pupil diameter velocity limit can be imposed on the entire signal, not just limited to those measurements adjacent to blinks. The incorrect samples that are identified in this step are not directly attributable to a specific cause (as was the case in the fourth step). Compared with the previous step, the signal does show to lose some of the spikes.
7. *Pre- and post-blink difference*: inspection of the data shows that the diameter just before and just after a blink is not always the same. Assuming that this should be the case, it can be corrected by calculating whether the difference would become smaller in case the blink were extended on either side. This means that we assume that directly adjacent to blink, there are still some incorrect diameter measurements, which should be eliminated. This step does that by trying to make the pre- and post-blink difference smaller if on either sides of already-reported blinks either 1, 2 … n additional measurements would be marked as incorrect. (For the data in figure 3.4, $n = 3$ appears to give good results.)
8. *Singletons, second time*: see step five.

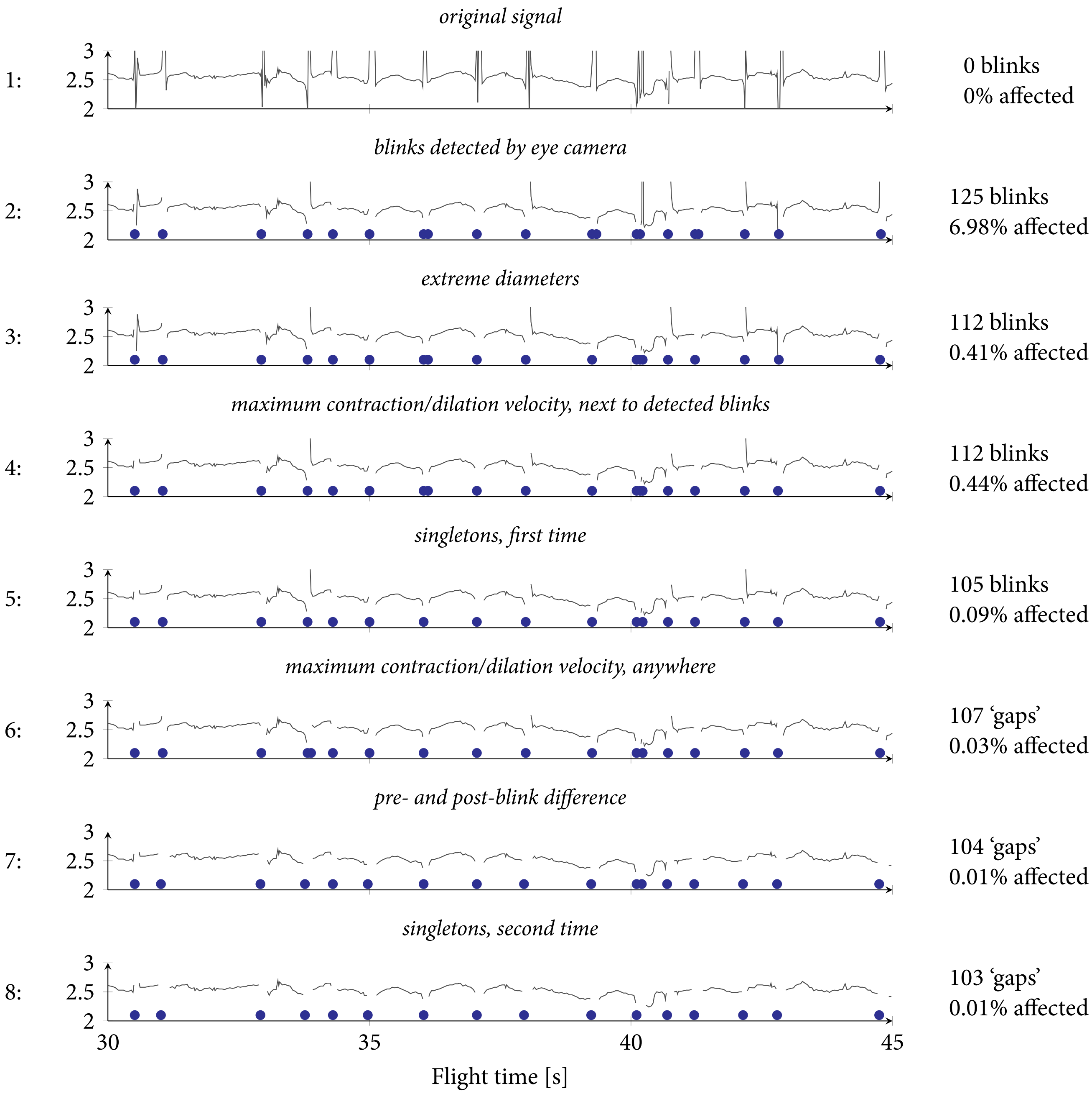


Figure 3.4: Blink detection and pupil diameter filtering. Number of blinks are reported for the entire 140-second flight. Vertical: pupil diameter [mm].

Chapter 4

Experiments

Ideally, all three measures that were described (time to crash, control power and eye metrics) would have been employed. The physiological measure of safety, consisting of the eye metrics of section 3.3, however, will remain inconclusive for now, due to difficulties regarding the data analysis. These are detailed in chapter 6.

This chapter will discuss the three experiments by using the remaining analysis techniques (time to crash and control power).

4.1 Short — long final

4.1.1 Materials and methods

Subjects

Simulated flight data from three professional pilots — named PP1, PP2 and PP3 — and two student pilots — named SP1 and SP2 — is available. All have agreed to the anonymous analysis of their data.

The professional pilots are (current or retired) captains from All Nippon Airways or Japan Airlines. Among the group of student pilots that received some kind of basic flight training in the simulator, SP1 and SP2 were chosen because their training was by far the most extensive. Their flights were conducted at the end of a three-month twice-weekly training, when the instructing pilot judged their performance to be adequate and, in this specific condition and flight procedure, on par with that of professional pilots.

Instruments

Flights are conducted in the fixed-base no-force feedback simulator of a Boeing 747-400, represented in figure 4.1a. The aircraft's dynamic response is calculated by software developed in-house. The graphics are provided by Microsoft Flight Simulator. Cockpit sound is simulated; the pilot can adjust the volume to a comfortable level. A flight director for ILS approaches is available. Next to the runway a simulated PAPI visual aid is present. Different weather conditions can be created by setting a wind turbulence and cloud ceiling level. Flight data is recorded at 20 Hz for all controls and aircraft response parameters.

Camera systems from NAC Image Technology provide eye measurement data. These are head-mounted systems that record images from both eyes and the visual field of the person wearing the system. The images are processed in realtime at a sample frequency of 60 Hz, to give pupil diameter and coordinates of the gaze point. The NAC EMR-8 (pupil diameter resolution 0.1 mm)

(a) Flight simulator (b) NAC Eyemark EMR-8 in use (c) NAC Eyemark EMR-9

Figure 4.1: Instruments.

was mostly used, although for a few flights the NAC EMR-9 (professionally calibrated, pupil diameter resolution 0.01 mm) was available. These systems are depicted in figure 4.1.

Flight procedure

The simulation starts with the aircraft headed for a landing on runway 34R of Haneda Airport in Tokyo. The initial altitude is 1800 feet AGL and initial indicated airspeed is 150 knots. The flight director indicates that the aircraft is on the ILS glideslope and localizer. The elevator is nearly fully trimmed (the aileron and rudder are not equiped with trim tabs).

Each pilot performs two of the described flights in VMC with 'light' turbulence, followed by a short rest with a duration to the pilot's liking. Then, two flights are conducted in IMC with the same turbulence setting. The visibility is such that the approach lights of the runway just become visible at 500 feet AGL when the aircraft is inbound on the ILS flight path.

The pilots are asked to perform the task to the best of their ability. Prior to the first recorded flight, the pilots have all had the opportunity to practise with at least one extra flight in VMC. Two pilots have performed the experiment, consisting of four flights, a second time on a different day.

Data analysis

The intention of comparing two phases of the approach is to see whether safety varies during the approach. Since the long and short final segments of the approach are not well-defined in terms of a flight parameter such as altitude, we choose the end of the long final and beginning of short final to coincide with the time at which the pilot in IMC will transition to using visual references for the landing; this transition marks a significant change in flying strategy. This moment occurs some seconds after the runway has first come into sight in IMC. With the IMC weather conditions as described, the runway becomes visible at about 500 feet AGL for an aircraft on a stabilized track and, in general, 40 to 35 seconds before touchdown. We have decided to call the last 35 seconds of the flight the short final.

Table 4.1: Time to crash penalty function values [-] for long and short final, per pilot and weather condition.

	SP1, day 1		SP1, day 2		SP2		PP1, day 1		PP1, day 2		PP2		PP3	
	long	short	long	short	long	short	long	short	long	short	long	short	long	short
VMC	1.34	1.06	1.97	1.18	1.89	1.47	2.48	0.92	1.59	1.77	2.58	1.37	2.22	1.58
	1.45	2.41	2.31	3.24	2.01	0.92	1.57	1.01	1.73	1.65	2.06	1.33	1.30	1.13
IMC	2.71	3.54	2.71	3.50	2.08	2.80	2.05	1.53	1.99	2.32	2.39	1.33	3.41	1.47
	2.81	1.89	3.22	2.46	1.43	2.30	1.59	1.15	1.36	2.31	1.56	1.67	2.08	1.32

Table 4.2: Prolonged deviations [s] contribution to time to crash penalty function for long and short final.

	SP1, day 1		SP1, day 2		SP2		PP1, day 1		PP1, day 2		PP2		PP3	
	long	short	long	short	long	short	long	short	long	short	long	short	long	short
VMC	8.05	1.68	7.01	3.36	8.22	3.70	9.52	4.81	11.88	6.48	14.08	5.71	10.84	5.19
	4.88	3.73	11.46	2.39	12.69	5.01	9.50	2.77	10.08	4.72	14.17	7.06	7.58	4.05
IMC	11.63	5.67	11.53	7.73	7.94	1.80	10.86	2.80	10.86	4.31	17.28	5.16	19.46	8.06
	11.84	6.17	10.73	6.94	9.20	3.82	9.18	1.44	9.59	3.33	8.91	2.14	14.29	3.73

4.1.2 Measure 1

The penalty function, equation (3.2) on page 25, has been applied to the short final (last 35 flight seconds) and long final (until the last 35 flight seconds) segment of all 14 VMC and 14 IMC approaches. This is tabularized in table 4.1. Based on the data of all pilots in both weather conditions, there is no statistically significant difference in time to crash penalty between the long and short final segment of the approach.

Discussion

To investigate the effect of the two contributions (the 'prolonged' and 'temporary' deviations) of the penalty functions to the short and long final approach segment, they have been studied independently.

Table 4.2 shows these prolonged deviations, i.e. $\int_t \Delta \mathrm{TtC}(t)\mathrm{d}t$, without the scaling weight C_1, calculated for the short and long final segments. These numbers consistently show that the prolonged deviations are larger on the long final than the short final, for all pilots and both weather conditions. From a one-way ANOVA it follows that this is a statistically significant result with $F(1, 54) = 89.62$, $p < 0.01$.

Table 4.3 shows the temporary deviations, i.e. $\sum_k |D(k)|$, without the scaling weight C_2, calculated for the short and long final segments. From comparison with table 4.2, it follows that the scaling weights are indeed necessary to get equal contributions to the penalty function. A one-way ANOVA shows that the temporary deviations are statistically significantly larger on the short final than the long final, for all pilots and both weather conditions, with $F(1, 54) = 7.33$, $p < 0.01$.

Table 4.3: Temporary deviations contribution [-] to time to crash penalty function for long and short final.

	SP1, day 1		SP1, day 2		SP2		PP1, day 1		PP1, day 2		PP2		PP3	
	long	short	long	short	long	short	long	short	long	short	long	short	long	short
VMC	0.34	0.66	0.90	0.61	0.74	0.81	1.08	0.30	0.21	0.80	0.77	0.56	0.77	0.76
	0.69	1.51	0.79	2.25	0.46	0.27	0.39	0.53	0.47	0.85	0.37	0.41	0.35	0.52
IMC	1.08	2.21	1.09	2.00	0.91	1.96	0.64	0.93	0.59	1.40	0.36	0.58	0.95	0.44
	1.13	0.91	1.54	1.28	0.31	1.42	0.43	0.75	0.23	1.47	0.44	1.08	0.38	0.69

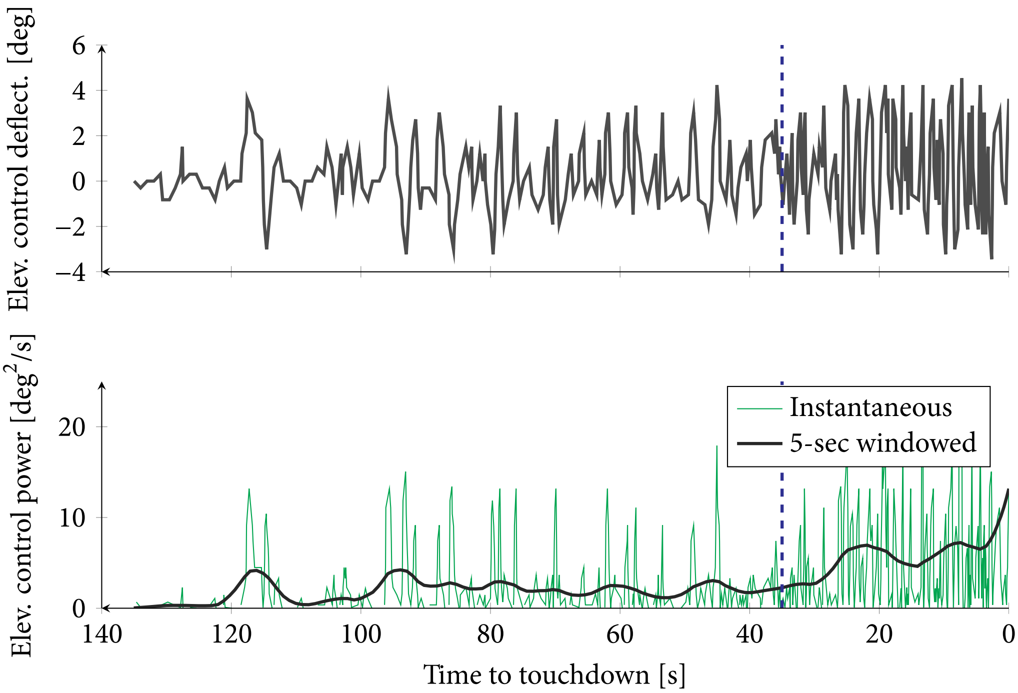


Figure 4.2: Power and magnitude of elevator deflection for PP1, day one, IMC flight 2.

4.1.3 Measure 2

Figure 4.2 shows the typical elevator deflection and power distribution. The instantaneous elevator control power (green) has been averaged with a 5-second window (black) to better see the trend. A separation line (blue dashed) indicates the end of the long final segment (until 35 seconds to touchdown) and beginning of the short final segment (the last 35 flight seconds). Clearly, the average power during the short final is higher. The values in table 4.4 represent the elevator control power averaged on the long and short final segment for all flights.

It can be seen that for nearly all flights the short final contains more elevator power than the long final segment. A one-way ANOVA shows that this is a statistically significant result ($F(1, 54) = 16.98$, $p < 0.01$).

Table 4.4: Elevator control power [deg²/s] on long and short final per pilot and weather condition.

	SP1, day 1		SP1, day 2		SP2		PP1, day 1		PP1, day 2		PP2		PP3	
	long	short	long	short	long	short	long	short	long	short	long	short	long	short
VMC	5.71	9.05	5.97	9.86	1.66	2.92	1.11	2.39	0.35	1.56	1.69	2.00	2.12	9.02
	5.25	12.43	6.61	7.80	1.96	1.53	0.70	2.12	0.62	1.21	1.19	6.17	2.87	5.23
IMC	8.19	17.54	4.06	12.57	3.22	10.53	1.38	3.48	0.93	1.94	1.66	7.53	0.93	2.12
	5.00	17.09	6.14	16.87	3.34	11.14	1.90	5.76	0.41	2.84	2.06	7.39	1.67	8.24

Table 4.5: Time to crash penalty function values [-] for VMC and IMC, per pilot.

	SP1, day 1		SP1, day 2		SP2		PP1, day 1		PP1, day 2		PP2		PP3	
	VMC	IMC	VMC	IMC	VMC	IMC	VMC	IMC	VMC	IMC	VMC	IMC	VMC	IMC
Flight 1	1.26	2.96	1.74	2.95	1.77	2.29	2.04	1.90	1.65	2.09	2.23	2.10	2.05	2.87
Flight 2	1.74	2.55	2.59	3.01	1.70	1.69	1.41	1.46	1.71	1.64	1.86	1.60	1.26	1.87

4.2 IMC — VMC

For investigating differences between IMC and VMC flights, the materials and methods are identical to the first experiment (section 4.1.1).

4.2.1 Measure 1

The penalty function that operates on the time to crash deviation, ΔTtC, has been used for studying the differences between VCM and ICM flights. Table 4.5 shows these function values. Though not consistently the case for every pilot, it appears that the penalty in IMC is higher than in VMC. A one-way ANOVA demonstrates that this result is statistically significant, with $F(1, 26) = 5.74$, $p < 0.05$.

4.2.2 Measure 2

Table 4.6 shows the elevator control power values averaged per flight. Considering all pilots, there is no statistically significant difference between the elevator control power in VMC (3.36 ± 2.50)[1] and IMC (4.23 ± 2.94).

The differences between pilots are large, though it is mainly SP1 who deviates from the others by using more power. Furthermore, PP3 appears to deviate by using less power in IMC, though not significantly so. A one-way ANOVA shows that SP2 uses significantly less elevator control power for landing the aircraft in VMC with $F(1, 2) = 711.96$, $p < 0.1$.

Discussion

Elevator control power averaged over the full flight duration for all pilots did not show significant differences between the VMC and IMC case. Though the effect appears to exist, in this analysis it

[1] The notation $\mu \pm \sigma$ indicates that the mean of the corresponding dataset is μ and standard deviation σ.

Table 4.6: Elevator control power [deg^2/s] for entire flight, in VMC and IMC per pilot.

	SP1, day 1		SP1, day 2		SP2		PP1, day 1		PP1, day 2		PP2		PP3	
	VMC	IMC	VMC	IMC	VMC	IMC	VMC	IMC	VMC	IMC	VMC	IMC	VMC	IMC
Flight 1	6.58	10.25	6.98	5.97	1.98	4.77	1.44	1.85	0.67	1.15	1.77	2.93	3.89	1.19
Flight 2	7.14	7.69	6.91	8.54	1.85	4.95	1.07	2.75	0.78	0.94	2.46	3.24	3.49	3.04

Table 4.7: Power distribution during flightsas per Eq. (4.1), in VMC and IMC per pilot.

	SP1, day 1		SP1, day 2		SP2		PP1, day 1		PP1, day 2		PP2		PP3	
	VMC	IMC	VMC	IMC	VMC	IMC	VMC	IMC	VMC	IMC	VMC	IMC	VMC	IMC
Flight 1	1.58	2.14	1.65	3.09	1.76	3.28	2.15	2.52	4.42	2.09	1.18	4.54	4.26	2.29
Flight 2	2.37	3.42	1.18	2.75	0.78	3.33	3.04	3.03	1.94	7.01	5.20	3.58	1.82	4.92

seems to have been obscured by large differences among pilots. This is best visible from table 4.4 by comparing the elevator control power that pilot SP1 uses with that of the others for succesfully landing an aircraft. Also, a day-to-day difference can be observed, e.g. in table 4.6 for pilot PP1.

To investigate this effect of pilot style on the VMC — IMC case, the control power results from the first hypothesis (short versus long final) are used: elevator control power is significantly different during approach phases, as depicted in figure 4.2. An increase in frequency and power appears to be present in the last part of the flight. The differences in control style between pilots are eliminated by introducing the dimensionless number

$$\frac{\text{average power on [35,0] seconds to touchdown}}{\text{average power on [140,35] seconds to touchdown}}, \tag{4.1}$$

calculated for all VMC and IMC flights. These numbers are shown in table 4.7. A one-way ANOVA shows that this ratio has a significantly higher average for all IMC flights (3.43 ± 1.32) than for the VMC flights (2.38 ± 1.35) with $F(1, 26) = 4.30$, $p < 0.05$.

4.3 Novice — expert

The pilot data of the first two experiments is not well-suited for a comparison of novice and expert pilots, since the student and professional pilots that contributed have performed equally on the task at hand. In fact, the student pilots were selected on their performance by the instructing professional pilot, with the purpose of extending the available dataset, and not for introducing novice - expert differences in the data. Variation among pilots that was detected has been attributed to personal control style differences.

For a full-fledged experiment, sufficient data is lacking. To give some future leads, just some training data of student pilots — named SP1 and SP4 — at our laboratory will be discussed.

4.3.1 Measure 1

Student pilot 1, who was capable of flying the approaches in the former experiments after training, received about three months of flight experience. Over the course of this twice-weekly training

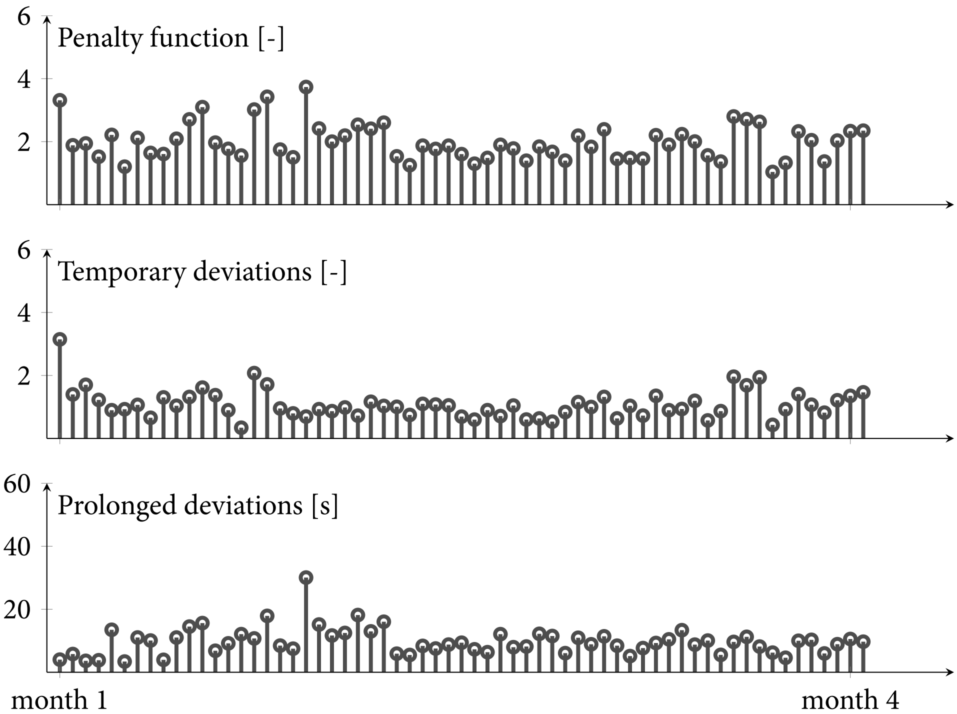


Figure 4.3: Time to crash penalty function values during three-month training, pilot SP1.

period, SP1 flew many approaches, as described in the first paragraph of 4.1.1. Gradually, the performance saw increases and the training became more challenging. Figure 4.3 shows the time to crash analyses of all succesful Haneda Airport landings over the course of the training period. Most interesting is the weighted penalty function (which takes into account both temporary and prolonged deviations): it clearly does not show a decreasing trend, but rather a constant one, with quite some variation, especially in the early phase. This is likely the result of the flight conditions becoming increasingly more challenging, combined with the development of some flight skill.

Student pilot 4, still in the initial phase of flight training, is just capable of succesfully flying the approach in very mild conditions. Over the course of one week, the pilot performed the same procedure (VMC without turbulence) five times. SP4 indicated to better understand and control the thrust setting of the aircraft during this period. Corresponding time to crash graphs are depicted in figure 4.4: clearly, the first flight in week 1 (above) deviates more from the ideal approach than the last flight of week 2 (below). The corresponding penalty functions, in figure 4.5, also show a decreasing penalty for the same procedure.

4.3.2 Measure 2

Looking at the elevator control power for SP1, figure 4.6, trends are obfuscated by the various kinds of procedures that were conducted during training. It does appear that the pilot started using more power nearing the end of the training. Also, the variability in the ratio between average

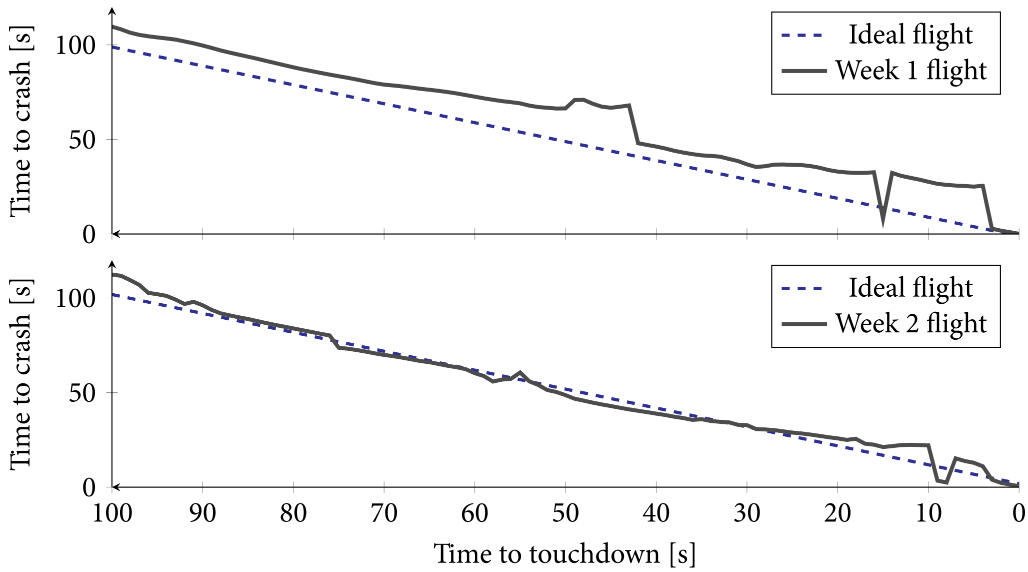


Figure 4.4: Time to crash in early week of training, pilot SP4.

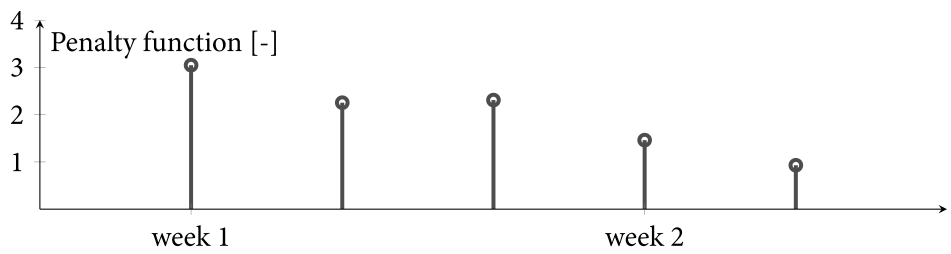


Figure 4.5: Time to crash penalty function values in early week of training, pilot SP4.

power on short and long final seems to be reducing.

Over the one-week period in which SP4 conducted the same flight, the difference between average power used on short and long final seems to have been reduced, as depicted in figure 4.7. This could be related to improved thrust control (reported at the time to crash analysis, measure 1), requiring less compensation from the elevator.

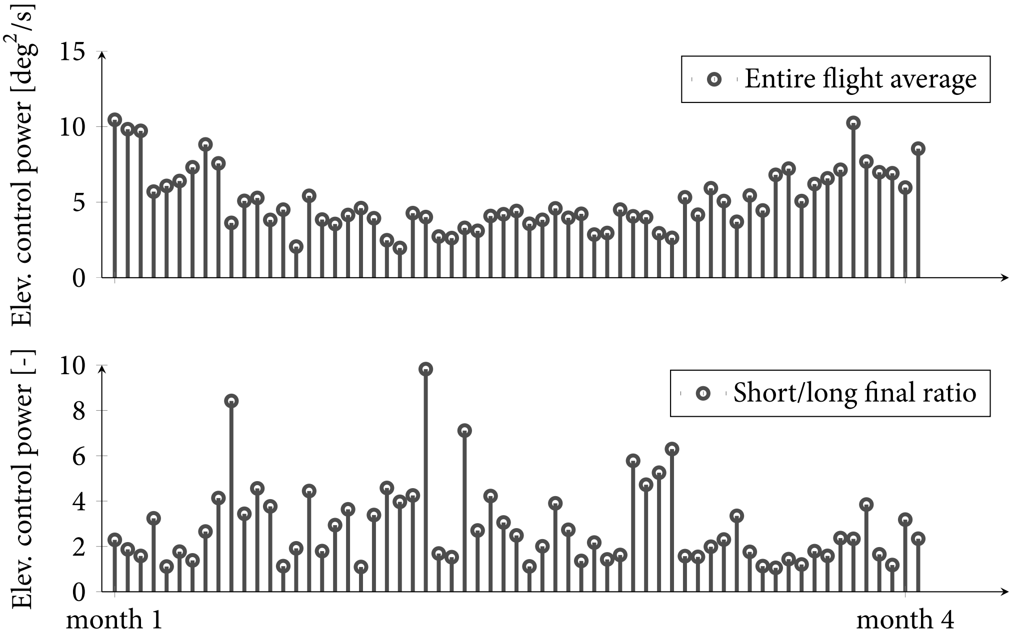


Figure 4.6: Elevator control power during three-months training, pilot SP1.

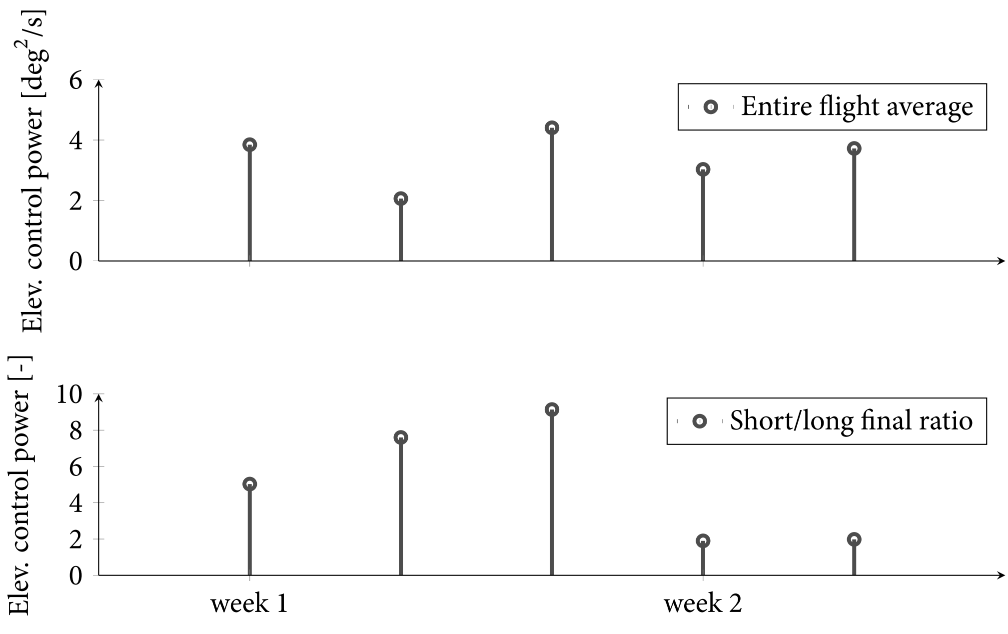


Figure 4.7: Elevator control power in early week of training, pilot SP4.

Chapter 5

Conclusion

Within a research project on the safety of new curved approach procedures, this internship is an investigation of methods to discern various levels of safety in flight, based on three cases: the long versus the short final approach, flight under visual versus instrument meteorological conditions, and the expert versus the novice pilot.

Since the scope of the term flight safety has been limited to include only those effects that the human pilot can influence, it makes sense to study potential failure modes of the pilot: mental workload and effort have been examined, and their usefulness for assessing flight safety through physiological indicators has been discussed.

Measures of safety based on the 'time to crash' (TtC) and elevator control power have been used to analyze experimental results. The former identifies deviations in flight performance from the ideally stablized approach, whereas the latter draws on the relationship between control actions, mental processes and risk. A first step towards using eye metrics as indicators for mental activity has been taken, in terms of a procedure for filtering artefacts in pupil diameter measurements due to blinking.

Student pilots and professional pilots have flown approach procedures in a fixed-base Boeing 747-400 flight simulator, with the purpose of obtaining experimental data. Analysis has shown that both (constituents of) the TtC measure and the power of the aircraft's elevator deflection signal (when corrected for personal control style) are indicative of the meteorological condition and approach phase.

In more detail, the TtC penalty function and (dimensionless) elevator control power have larger values for flight under IMC than VMC. Seeing as the TtC penalty function and elevator control power have been carefully selected to constitute our definition of flight safety, this result means that flight under IMC is less safe than VMC.

In the short final versus long final case, the elevator control power is equally distinguishing: larger values on the short final indicate that this phase is less safe than the long final of the approach. This does not hold for the TtC penalty function, though: of the two parts that make up this function, only the 'prolonged deviations' are larger on the long final, whereas the 'temporary deviations' are larger on the short final. This indicates that stabilizing the aircraft on the long final takes time, and that close proximity to the runway triggers small and quick corrections from the pilot.

For the novice versus expert pilot case, the TtC and control power measures have been applied to some flight training data. It followed that these measures are related to the pilot's performance level. Accordingly, they form valuable tools for assessing a pilot's flying qualities.

Chapter 6

Discussion and recommendation

In this chapter I intend to discuss challenges that I faced during the internship assignment. These challenges are relevant especially when they are caused by the limited scope of my assignment, seeing as it is part of a larger ongoing research project; either my successor or at least the remaining researchers on the project would then know in what areas I faced these limitations, and hopefully find ways to surpass them, for future gain.

The consequence of being in the earlier phase of a project and having limited time and resources was having to 'tone down' the experiment conditions, by formulating substitute research hypotheses. This meant that only a proxy of the ideal procedure, equipment and test persons could be used, while operating under the assumption that the reasons for doing so are justified.

This particular topic of research brings challenges in that the safety of aviation is a key point: the already very high level of safety in commercial flight hinders further improvements, since 'unsafety' is no longer omnipresent but only manifested as a coincidence of extreme conditions. Such rare-case events are harder to simulate experimentally with sufficient statistical signficance. The situation is remedied by good definitions and an understanding of safety and the events of interest.

More to the point, in this research the concept of safety is being related to notions of mental workload, effort, etc. (see research objectives in section 1.1); while this definitely seems a good approach to better understand the pilot, it is as of yet by no means an extensively tried and tested method. In fact, at the same time that we are using physiological indicators that presumably indicate mental effort, research into these indicators and the underlying cognitive processes is still ongoing, and not nearly matured.

In other words, while there are very promising reports of experiments that succesfully measure some kind of mental workload, a ready-to-use and proven toolkit for assessing an operator's mental state does not exist currently. Moreover, even if the mental state can be measured reliably, the relation with operator-affected safety will most likely remain assumed, due to the aforementioned high-standard nature of aviation safety.

This means that this research would benefit from developments in understanding cognitive processes. To contribute, we presented our own philosophy on some of the concepts, based on a literature investigation that did not show a lot of consensus, and in an effort to unify some ideas.

More substantively, I would want to know more about the so-called multi-dimensionality of mental workload: can an operator have high workload in some areas, but still have resources left for a different kind of task? Does e.g. a visuospatial task leave room for a memory-related task, or does the one impair the other? What about the resource- and data-driven task differences that are reported on? How do these apply to flying, or to specific tasks that comprise flight? Could flight under VMC be a predominantly resource-driven task and IMC a data-driven one? Assuming that the presence of operator workload is the absolute goal to point out, do e.g. different sections of the approach phase draw on separate mental resources, establish different kinds of mental workload/effort and hence require different measurement techniques (physiological indicators)?

Relating mental processes to safety in aviation is hard and, if possible, will probably occur via the operator's task performance. On this subject, the inverted-U relation and coping strategy at overload are relevant in the case of pilots. While flight training will teach pilots a certain degree of control over an aircraft, it also makes pilots develop sensitivity to only the relevant signals and cues in the cockpit. That is to say, a cockpit continuously shows an enormous amount of information, which the pilot cannot all process and, hence, has to filter. Unlike pilots in the air force, whose skill is mainly performing exceptionally demanding and flight-technical tasks, the commercial pilots in this research carry the passenger's safety responsibility. The pilot's ability to remain calm and clear-minded in every situation has to do with the ability to cope with stress, a higher task load, and filtering the most critical alerts and tasks. Knowing how this coping strategy works is very relevant. This is akin to asking how 'far' in the inverted U a pilot is, what factors influence α, and what the value of β is for a flight task $T = \beta R$ (section 2.2.2).

With regard to the hypothesis tree (section 2.3) and the mental processes, I would want to see some more thoughts about how mental effort works: the current reasoning leads to the idea that mental effort becomes higher when the task load on an operator is increased, but remains the same when an operator is subjected to the same task load of a different kind (i.e. a task the operator is not properly trained to perform). And given some kind of relation between mental effort and physiological response, only the former situation would be measurable.

The tree also shows the association of workload with the amount of control actions: while this seems rather intuitive, the sole literature source used was Gawron [18] and some other interesting research and insights might exist. Also, the causality of this relation is relevant: more corrective control can be said to imply a mental workload increase, but an increase in mental workload does not necessarily imply more corrective control, as exemplified by the chess player who thinks but does not move. Besides, the multi-faceted aspect of mental workload could play a part here, too.

While the time to crash quantity can be interpreted intuitively without bringing up an operator's cognitive processes, the sensitivity of the measure deserves some attention; i.e. at what flight performance level can it no longer discern a good flight from an even better one? (Do professional pilots at some point achieve equal TtC, while still not flying equally safe? Or vice-versa, when a group of professional pilots can be considered to fly equally skilled, what is the variation in TtC?) Which flight parameters affect TtC the most (a correlation between TtC and thrust setting has been pointed out, and it is likely to assume that pitch and sinkrate have an effect, too)?

To make TtC a good proxy for flight safety, a penalty function has been introduced that draws on the idea of an ideal flight in terms of the stabilized approach. It would be interesting to conduct extra validation of the particular penalty function by having a professional pilot judge flights and their corresponding ΔTtC penalties. I would want to know e.g. whether the 'prolonged deviations' should be penalized more as the aircraft approaches the runway ('mistakes close to ground are more dangerous'). Since the penalty function is based on mathematical transformations of ΔTtC that have a fairly intuitive interpretation, it can be readily tuned in case of discrepancies between the calculated penalty and the expert's judgment.

In that case, it should also be investigated whether the current TtC calculation procedure is adequate: presently, 'crash flights' are simulated for every flight second to make up one flight's TtC graph. (Should this frequency be increased to capture the 'temporary deviations' more accurately?) These crash flights are based on the pilot suddenly giving up all aircraft controls: hence rudder, elevator and aileron deflections are reset to their trimmed values, and the thrust control maintains the last setting. It should be noted that this particular choice actually defines time to crash and this quantity should therefore be interpreted accordingly.

A similar validation is useful for the elevator control power measure, especially since it has been shown to vary considerably among pilots. To obtain more information from this signal, analysis of the frequency content — and not just the power value — is relevant, as inspection of flight training results has hinted that the frequency content changes over the course of the weeks of flight training.

Currently, the elevator deflection signal has been analyzed, since it is crucial for a succesful landing and thought to contain the most relevant control information. While a pilot can fly the designated procedure of this experiment without rudder control, corrections with ailerons too are necessary for a safe conduct. Analysis of the aileron deflection time-history could therefore contain interesting safety- and performance-related clues as well.

The use of eye metrics as physiological indicators for mental activity has been announced and prepared in the sections of this report leading up to the experiments; there, however, they turned out to be inconclusive and as of yet incapable of giving any interpretation to experimental data. This should mainly be attributed to the lack of a proper description of the actual pupil response to cognitive activity.

In essence, to retrieve a representation of an operator's cognitive load from merely eye camera images, the pupil diameter will first have to be estimated. To do so, a few disturbances should be eliminated, such as

1. movement of eye camera relative to subject's head (the camera is mounted on a cap),
2. reflections due to external light sources,
3. reflections due to contact lenses and
4. distortion of recorded image due to eye movement (gaze angle).

Admittedly, these aspects are generally expected to have been taken care of by the manufacturer of the eye mark product, though the first three are fairly easy to check and monitor even during an experiment. The fourth should either be confirmed to have no effect on the particular experiment (i.e. eye movement is very little), or be compensated for by processing software from the manufacturer or even a self-developed calibration interface (e.g. Pomplun and Sunkara [43] used a neural network approach for correcting for gaze angle). It can be useful in later data analysis stages to have the original camera-recorded images available for an, albeit time-consuming, post-hoc data integrity check.

Having obtained a reliable pupil diameter estimate, factors that influence a human's pupil should be considered. The most important ones are

- light (when light intensity increases, pupil constricts to protect retina),

- accommodation (when focussing on near object, pupil constricts to adjust the depth of field) and
- cognitive activity.

It is said that the light reflex easily dominates the other two (i.e. it causes the largest diameter change). To cope, it seems that there are three options:

1. separate the effects of the different reflexes based on knowledge of their distinguishing features;
2. predict the effect of the light reflex based on luminance measurements, and then filter;
3. make the experimental conditions so that light intensity is constant.

The second and third option do not really qualify for keeping the experiments practical (requiring either an extra device for measuring the light incident on the pilot's eyes, or an impossible cockpit, whose lights do not change). The remaining first option is predicated on research of the dynamics of the different reflexes — precisely this has been stated in the beginning of this eye metric discussion as the primary reason for the lacking experimental results; Palinko and Kun [38] describe a predictor of the pupillary light reflex but a good model of the accommodation and cognitive activity response is not readily available. Boff and Lincoln [9] have an overview of some of the (earlier) research of the pupil's reaction to cognitive activity, which shows that the results are mostly limited to statements of the time between the onset of mental activity and the observation in the pupil diameter.

The 'index of cognitive activity' (ICA) of Marshall [34] is claimed to be a method that can separate the effects of changing lighting conditions from the desired indicator of mental effort or workload. However, unfortunately, it did not lead to meaningful results in this assignment. The problems can be attributed to the ICA method itself or to the data it has been applied to.

I like to think that the data supplied (i.e. the pupil diameter measurements) is of quite a good quality, since for some experiments the eye mark camera had even been calibrated by representatives from the manufacturer, who monitored the data acquisition continuously. The specifications of the equipment (resolution, sample frequency) are nearly identical to what was used by Marshall in the ICA patent. In the later processing stage where blinks have to be filtered, there did appear to be some kind of difficulty: occasionally it looked like eye movements are correlated with pupil diameter, and therefore with blink frequency (depending on blink filtering parameters). This could be one error source. It might be fixed with even better calibration but is possibly already mitigated by using data from both eyes cleverly (the artefacts do not necessarily show up on both eye signals at the same time).

Related to this problem is the issue of the ICA method itself: the ICA appears to be strongly dependent on the blink filter applied. This makes the previous issue more pressing. It also raises questions about the basis for the ICA method (which is not clearly stated in the patent): what is the underlying knowledge of the pupil's response to mental activity? How thorough is the validation of the index? While Marshall uses (mostly?) self-reported data as validation, other researchers have used the method with some results [46]. Also, Marshall's pursuit to estimate an operator's cognitive state from eye metrics seems ongoing (through a commercial software package), purportedly with success, and unlikely to be based entirely on false pretenses. It is not unreasonable to expect the vendor of the software to have some evidence of its effectivity.

The implementation of the ICA in this assignment often resulted in an index that resembled white noise. I suspect that this is not related to the specific calculation, since the wavelet transform — the only step in the algorithm that is a bit involving – itself has been used successfully elsewhere in this report (in the penalty function for the TtC).

A literature investigation did turn up one note about constraints of pupillometry [9], apparently related to the accommodation reflex, described earlier:

> Lengthy stimulus presentation at close distance (e.g., <3 m) will tend to result in constriction of the pupil. This effect is called the "near-vision pupil reflex" and is quite variable among subjects [...] This is the type of effect that can impede the use of pupillometry in a monitoring or vigilance situation. [9]

In summary, eye metrics could not be used for a few reasons. Pupil size is said to be correlated with cognitive activity, though an exact description or explanation is missing. Furthermore, it is known that light has a large effect on the pupil size, and blinking obstructs clear pupil measurements. Not knowing what to look for precisely, one has to rely on others for an appropriate cognitive activity index, and on a self-made blink filtering procedure. To succcessfully use pupil measurements in the future, the procedure for filtering unwanted reflexes has to be validated first of all. Then, the quality (calibration, etc.) of the measurements should be ensured. After that, blink filtering improvements could be carried out (if necessary, by manual inspection of the video frames).

Appendix A

Instrument approach procedures

The approach phase of flight is critical for a safe landing. It requires good judgement from the pilot as there is little margin for error. This appendix gives an overview of flight approaches in order to better see this research project in the aviation context. The information is based on FAA handbooks [4, 5].

A.1 Exiting the en route structure

Upon nearing the destination and finishing the en route phase of flight, the pilots most likely intend on landing the aircraft at the destined airport. The procedure that follows will usually fall in one of the following categories:

- Complete instrumental approach procedure;
- Approach with radar vector support;
- Visual approach;
- Holding pattern.

In the case of instrument flight rules (IFR), as opposed to visual flight rules (VFR) suited only for visual meteorological conditions (VMC), the regulations governing flight are stricter, allowing safe procedures in the more challenging instrument meteorological conditions (IMC). These procedures are published by the country's national aviation authority (NAA), such as the Federal Aviation Administration (FAA) in the United States and the Ministry of Infrastructure and the Environment (IenM) in the Netherlands.

The United States Terminal Procedures Publication (TPP) contains Standard Terminal Arrival Routes (STARs) that specify the routes transitioning the aircraft from the en route structure to a fix from which the approach can be conducted. The pilot has to adhere to the STAR rules, unless otherwise instructed by air traffic control (ATC). After completing the STAR, the pilot is supposed to proceed to the destination according to the instrument approach procedure (IAP) selected for the flight.

Depending on the available ATC service, the IAP is either executed completely or at some point superseded by instructions from ATC.

A.1.1 Fully flown instrumental approach procedure

In case of minimal ATC assistance, the published IAP is flown as a full approach. The pilot then conducts his own navigation. This situation is encountered in the event of a communications failure and — far more commonly — in areas without radar coverage.

If the destined airport does not have radar, nor an operating control tower, pilots are expected tot self-announce their arrival on the common traffic advisory frequency (CTAF) to inform other aircraft in the vicinity of their flight inbound on an instrument approach. This way the required separation between IFR and VFR traffic, which would otherwise have been provided by ATC, can be maintained.

Section A.2 details common types of IAPs.

A.1.2 Approach with radar vector support

In high traffic airspace, radar assistance from ATC — the so-called radar vectors — is usually available. It serves as course guidance to the pilot in the form of headings and altitudes, and as a means of expediting traffic to ATC. The radar controller assists the pilot in establishing an inbound course for the final (part of the) approach. The pilot is then expected to secure a visual reference to the runway and finish the approach and landing.

A radar approach often shortens flight time and distance, saving fuel and reducing emissions. Pilots indicate that they generally prefer this kind of procedure.

A.1.3 Visual approach

Another method to expedite the traffic flow near the airport is the visual approach. It can either be requested by the pilot or assigned by the controller. ATC can give clearance for a visual approach when the pilot confirms to have the runway or an aircraft to follow in sight. It allows an aircraft set out on an IFR flight to overrule the IAP and proceed visually if the weather conditions permit. When cleared, the pilot assumes the responsibility for separation and wake turbulence avoidance.

A.2 Types of IAPs

All instrument approach procedures have in common that they define a lateral and vertical trajectory to be flown under IFR, down to certain minimums where the pilot has to obtain visual references. The choice for a certain IAP is based on numerous aspects. The approach planning starts preflight and is finished inflight when the latest ground conditions have been considered.

A.2.1 Performance considerations

The aviation authority specifies aircraft performance limitations, which the pilots have to make sure to comply with. These are primarily intended to ensure safe flights. While the in-depth performance planning is normally carried out by air carriers prior to the aircraft's departure, the air crew conducts a review of performance considerations before the approach. Their source is the aircraft flight manual (AFM) or the pilot's operating handbook (POH). Typically, the air crew considers whether the aircraft, given its current weight, is capable of

- landing within a specified distance,
- maintaining a specified climbing gradient from the missed approach point (MAP) with one engine shut down and

- performing a go-around with a specified climbing gradient in the landing configuration with all engines operating.

Aircraft are grouped in approach categories based on a certain reference speed. These determine the allowed speed ranges for performing approach maneuvers. While an aircraft belongs to one category, pilots have to observe whether minimums of a different category apply in the given situation (e.g. in case of inoperative flaps, icing conditions or weight excesses).

A.2.2 Navigational equipment

The IAP charts associated with a particular airport distinguish different approaches primarily based on the navigation equipment (NAVAID) that is involved in the procedure. The technological progress of navigation systems over the last four decades has spawned many different approach procedures.

They are frequently divided into non-precision and precision approaches. Although different definitions of the precision approach exist, the important characteristic is that precision approaches also provide vertical guidance besides lateral guidance.

The instrument landing system (ILS) ground-based navigation system is most commonly used in a precision approach. It has greatly reduced the risk of controlled flight into terrain (CFIT) during final approach. Limitations of ILS are the sensitivity to signal obstructions (due to e.g. large buildings) and reflections (due to e.g. uneven terrain), the cost (every runway requires a new installation) and the approach trajectory (only straight lines).

Non-ILS approaches have seen large improvements over the years, trying to overcome the ILS limitations whilst offering proper navigational guidance. In short, the non-precision approaches (NPAs) in 1970 have evolved to stabilized (constant descent angle) NPAs in the 1980s, and to precision-like approaches in the 1990s and onward.

Traditional NPAs

In the 1970s the non-directional beacon (NDB), very-high frequency (VHF) omni-directional radio (VOR) and localizer (LOC) accounted for the non-precision approaches. They are all ground-based systems whose location affect the flight pattern and, hence, the approach difficulty. Ideally, they are positioned close to the runway, but the surrounding terrain occasionally does not allow this. The addition of distance measuring equipment (DME), co-located with these systems, gives the pilot more information on the aircraft position.

RNAV approaches

The development of accurate inertial reference systems (IRS) and the introduction of the ‘glass cockpit’ in the 1980s have lead to the area navigation (RNAV) system. It incorporates information from conventional ground beacons, the IRS and a navigation database to indicate the aircraft’s position on a display. It allows the pilot to navigate on any course within the coverage of station-referenced signals. The trajectory is indicated by waypoints, which are defined by reference to ground NAVAIDS or geographic latitude and longitude.

PBN and RPN

The availability of GPS in the 1990s has increased the accuracy of the RNAV systems. Further improvements were introduced by the wide area augmentation system (WAAS) that enhances position integrity information from satellites with additional ground stations.

Performance-based navigation (PBN) is a relatively new concept, currently being implemented by the International Civil Aviation Organization (ICAO). It allows for the specification of *navigation performance* (in terms accuracy, integrity, availablity, functionality, etc) requirements instead of *sensor* requirements for procedures. When procedures are not based on specific sensors but on the required performance instead, technological development of navigation equipment does no longer make the associated procedures obsolete. This makes maintenance of routes and procedures easier in case e.g. a VOR ground station is moved or global navigation satellite system (GNSS) equipment evolves.

Confusingly, RNAV is being redefined to be one of the navigation techniques of PBN — it now carries certain performance requirements. Currently, the other PBN navigation technique is RNP.

A required navigation performance (RNP) specification for a navigation system consists of additional requirements (on top of RNAV specs) for *on-board accuracy* and *integrity monitoring*. One key feature of approach procedures that can be executed as a result of RNP navigation systems (the so-called RNP approaches) is the curved flight track. They are intended to improve the use of airspace and create IFR approaches at airports where IFR flight was not possible before.

The term RNP is also used with a number to state specific performance requirements for operation in a certain airspace. (RNP 0.3 means that the system must be able to calculate its position to within a 0.3 nm radius circle.)

Among RNP procedures there is a categorization, one of the categories being the Special Aircraft and Aircrew Authorization Required (SAAAR) or also called Authorization Required (AR). All RNP procedures in the US belong to this category. Operators must meet special RNP requirements as outlined in FAA AC 90-101.

Appendix B

Pilot modelling

B.1 Conventional

Models of human operators have been developed in different fields for a variety of applications, such as control, physiology and cognition [56]. During and after World War II control engineers wanted to mathematically capture human behaviour and performance for military tasks, such as anti-aircraft gun tracking [26]. In aviation, McRuer in the 1950s set out to express the dynamics of flight in terms of control engineering transfer functions (instead of the traditional partial differential equations) in order to describe aircraft handling qualities. This required him to explore the modelling of the dynamic response of the human controller, thus sparking human performance modelling in aviation [17].

In [15, 56] an overview of the control-theoretic approach to pilot modelling is given. The introduction of soft-computing techniques has advanced the models: in addition to representing human behaviour, they allow for capturing knowledge, too. For instance, fuzzy logic was used to extract control rules in linguistic terms from experimental data [19, 57].

B.2 Cognitive architectures

B.2.1 Limitations of conventional methods

Limitations of the control-theoretic models are being addressed with the help from cognitive psychology [21]. According to Zemla et al. [58], in order to evaluate new technological equipment, human-in-the-loop (HITL) simulations are commonly employed. These experiments are generally labour intensive, expensive for even the slightest new changes and not always accountable for behavior on large scale. Current simulations are said to be limited in not dynamically simulating human pilot behavior:

> Responses to ground controllers are predetermined, meaning that the planes in these simulations always react to air traffic controllers without error and in zero time. Furthermore, off-nominal situations are neither detected nor corrected in these simulations due to the lack of consideration for pilot cognition. While such omissions are not uncommon in the early stages of research on an problem, they expose a serious gap in our ability to accurately predict the outcome of changes [to flight equipment designs]. [58]

Cognitive models can overcome these concerns since they have the benefit of being fast and inexpensive, and "integrating key components of human cognition and behavior […], such as pilot errors and response times" [58].

B.2.2 Cognitive science

The different approach that the cognitive sciences take is illustrated by Kurup and Lebiere [29], who compare the problems in robotics research to those in cognitive sciences:

> In contrast [to traditional robotics research], AI [artifical intelligence] and Cognitive Science are concerned with problems that are, in many ways, diametrically opposed to robotics, namely, open-ended tasks that span longer intervals of time in discrete domains, are knowledge-intensive and make assumptions about the existence of modules and tools that simplify interaction with the environment. Examples of such tasks include high-level planning and scheduling problems, language understanding, instruction following, diagnosis, and domain-independent execution monitoring and recovery. In order to focus on such tasks, AI and Cognitive Science make assumptions about the existence of low-level perceptual and motor modules that can provide the required information from the world and act on it. In addition, these assumptions also allow an AI or Cognitive Science agent to represent perception and motor-action in an abstract, modular way without having to worry about the specifics of the components (like the particular camera or actuators) used to build the agent. Robots, by virtue of having to exist in and interact with the real world do not have this luxury. Robots need functional vision and motor-action capabilities before they can be used to perform higher-order tasks. [29]

With regard to the increasing level of automation and, consequently, the changing role of the pilot, Holt et al. [25] state that automation in the cockpit, despite having been introduced in part to reduce errors in aviation, has created new sources and types of error. While in the past research on human-machine systems focused primarily on "physical and ergonomic characteristics of the interaction for optimizing construction and force feedback properties", now it is directed towards estimating the consequences of automation in high risk environments [32]. Detailed models of technical systems nowadays need to take cognition, memory span and mental representations into account [32]. Holt et al. [25] argue that cognitive modeling would be suitable for this purpose:

> One approach to studying error has been to classify or functionally group automation-related errors […]. However, this approach does not allow researchers to pinpoint the causes of errors. Further, this approach does not describe the process of pilot-automation interaction that results in the errors. This makes it impossible to know how to design interventions such as training or the redesign of instruments, displays, or software. An alternative to the taxonomic approach is cognitive modeling. Detailed cognitive modeling of the processes involved in human-automation systems should give a more complete and systematic picture of automation errors, their detection and possible mitigation. [25]

B.2.3 Cognitive architecture

Cognitive architectures give a "theory of human cognition and perceptual-motor actions that provides a rigorous framework for building, running and testing computational models of behavior" [44]. These can lead to a more psychologically probable representation of behaviour, since they make "a priori assumptions that constrain the representations and processes available for use in a model on the basis of the theories underlying the architecture. By virtue of these constraints, architectures are a distinct subset of the total set of possible human representation systems" [21]. As Zemla et al. [58] put it, cognitive architectures can simulate human cognition through interaction of lower-level psychological processes (e.g. memory retrieval and visual attention) that have been "well vetted in the psychological literature to ensure predictive accuracy" [58].

Langley et al. [30] describe a cognitive architecture as "those aspects that are constant over time and across different application domains". Typically, these aspects are said to include

- short- and long-term memories, storing content about the agent's beliefs, goals and knowlegde;
- representation of elements in these memories and their organization into larger-scale mental structures;
- functional processes operating on these structures (e.g. performance mechanisms using them and learning mechanisms altering them).

Alternative architectures differ in the specific assumptions that they make about these aspects [30].

B.2.4 UTC

Historically, Allen Newell (1973) hypothesized that building models of individual cognition phenomena had not lead to a model of general cognition. Rather, it had caused fragmentation in the field: although many different micro-models existed, they only addressed one issue at a time and were incompatible with each other. In reaction, with his proposal of the development of Unified Theories of Cognition, Newell sparked research of capturing the "task-invariant structures that underlie human cognition" in the form of cognitive architectures [29]. He argued for a unification of many findings into a single theoretical framework (followed by tests and refinement of that theory), hence capturing intelligent behavior at the systems level [30]. In this respect, cognitive architectures can be considered the antipole of expert systems, which provide skilled behavior in narrowly defined contexts.

B.2.5 ACT-R

A widely-used cognitive architecture is ACT-R (adaptive control of thought – rational), originally developed by Anderson at Carnegie Mellon University in 1993 and continuously developed ever since. It has seen many applications, such as in cognitive psychology, education, control, aviation and neuropsychology [3, 30, 33]. This can in part be attributed to the fact that ACT-R is distributed as open-source software, available on all major operating systems. Stimulated by

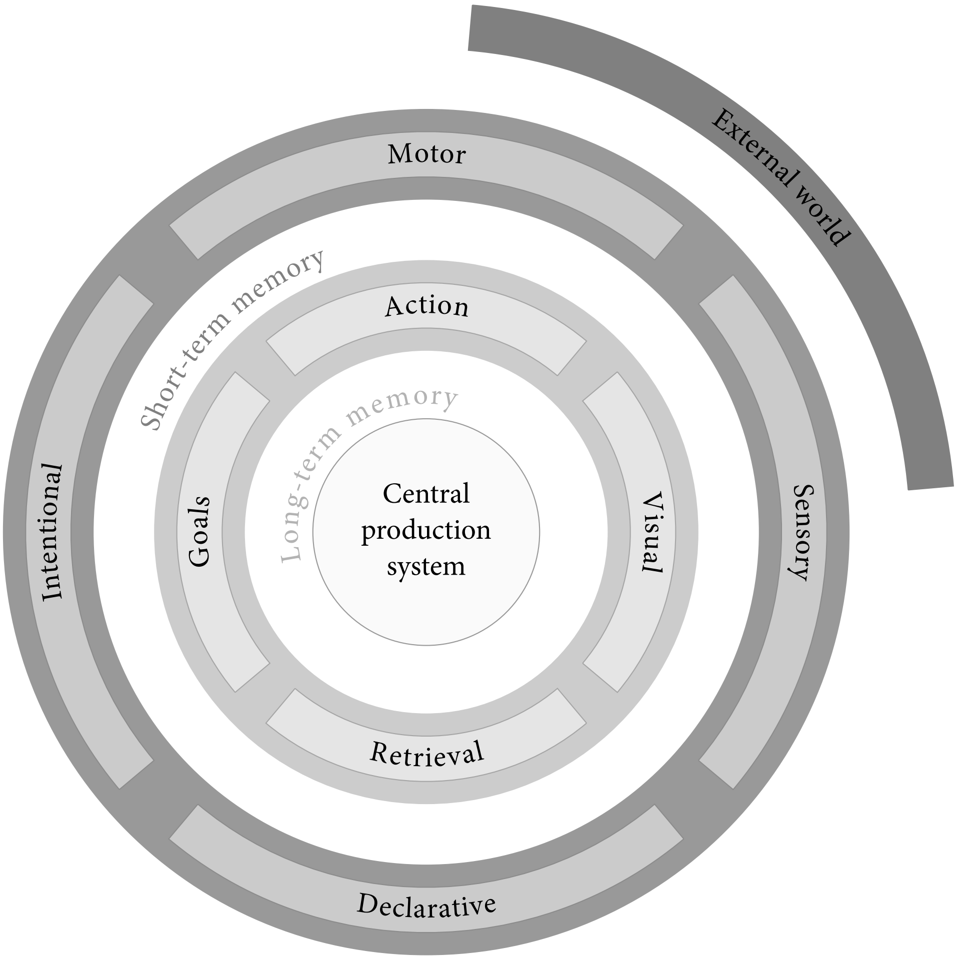


Figure B.1: Overview of ACT-R structure.

the adoption of his framework, Anderson more recently (2007) investigated the neural plausibility of ACT-R, trying to map the different components to areas of the brain with brain imaging techniques [30].

B.2.6 Overview

Figure B.1 shows a schematic overview of the structure of the ACT-R framework, in which cognitive models can be created. The *modules* are used for processing different kinds of information:

- The *sensory* module: visual processing;
- The *motor* module: action;
- The *intentional* module: goals;
- The *declarative* module: long-term knowledge [33].

ACT-R models commonly employ these four modules, but are not limited to these — researchers have extended the feature set of ACT-R []. Each module connects to the *central production system* through its *buffer*, which holds a single declarative unit of knowledge, called a *chunk*, at any time [31]. The buffers constitute the short-term memory. The central production system consists of *productions*: rules that coordinate the processing of modules and together form the long-term memory [33], representing procedural knowledge (memory of skill) [31]. A single production can be thought of as one simple step of cognition. The central production system can only access the modules through (two-way connection with) the module's buffer; this reflects the fact that humans do not have all available input information ready for processing at a given time (e.g. retrieving a memory might momentarily decrease visual vigilance) [33]. Sequentiality in processing information arises by allowing

- only one chunk to be in a buffer at a given time and
- only one production to fire per execution cycle.

B.2.7 Applications

Salvucci et al. found that a cognitive architecture could succesfully predict driver behaviour aspects of low-level control (such as steering) and higher-level cognitive tasks (such as task management and decision making) [44].

Also, NASA invested in developing cognitive architectures with the initiation of their HPM (human performance modeling) project. This is a 6-year research activity with teams from academia and industry that focussed on the advancement of cognitive modelling of human operators in aviation problems, in order to improve safety [17].

Not nearly as large-scale, but more illustrative of the power of a cognitive architecture, Peebles [40] has looked at modelling the cognitive and perceptual operations involved in interpretating graphs. The video "ACT-R graph comprehension model" on YouTube shows the model in operation: `http://www.youtube.com/watch?v=z2kAwrOrjIM`

Appendix C

Flight training notes

This appendix lists some of the notes that I took during the flight training from the (retired) professional pilot in our lab. Although these are mostly observations without coherent interrelation, they could be useful for someone taking the training in the future. Also, since one of the research objectives is understanding mental modes of certain flight procedures, the student pilot's perception and perspective can be interesting.

Scanning

Using a 'scanning technique' when flying the aircraft is the recommended way to monitor all systems. By getting used to continuously looking at the different instruments in a fixed order and for equal duration, the pilot is least likely to miss vital information that has appeared somewhere in the cockpit. This is a bit counter-intuitive as soon as some some major issue that deserves full attention pops up. E.g. when approaching the runway on the short final and getting ready for the flare and touchdown, it is still advised to check airspeed, sinkrate and thrust setting up until the very last moment in order to accurately time the flare, even though one might be inclined to only look at the runway itself. (I have noticed that after this flight training I apply the same scanning procedure to my driving; having established the habit of alternately focussing my attention on the revolution counter, speedometer, navigation system, rearview mirror and front view, it has become easier to manage more stressful situations in traffic, and stay calm.)

Corrective control

The type of control actions that are performed on the final approach procedure seems to be mostly corrective control/disturbance rejection. When the ILS glideslope and localizer have been intercepted, flying the approach has become a 'game' of constantly making sure that a few crucial flight parameters remain within bounds, and quickly spotting and correcting deviations.

When the weather permits, this can largely be done based on visual cues from outside. In that situation, flying still involves flight skill in the purest sense in my opinion of the word: just as a fighter pilot decades ago in a automationless cockpit would, or a pilot in a glider aircraft would, flying is about understanding the dynamics of the machine, having a feel for its controls and seeing where the pilot's actions take the aircraft in the outside world.

In case of bad weather and absence of outside visual clues, the dynamics of the aircraft have not changed, but those of the flying task at hand seem to have: instead of pointing the aircraft nose at the desired touchdown point, the pilot now has to keep the aircraft on the localizer centerline and glide-path center. He does so by monitoring his displays and mainly applying aileron or rudder pressure when the horizontal situation indicator (HSI) is off, and by applying elevator pressure or changing the thrust setting when the glideslope indicator is off. Doing this accurately requires getting a feel for the amount of correction that is needed given a certain indicated deviation on the display. Admittedly, it could be my lack of experience, but still this task seems to be considerably different from the one that does have outside visual cues. (This is one of the reasons that the VMC — IMC case was studied in this report.) A truly experienced pilot would probably still have a sense of what the entire aircraft is doing (his mental model is more comprehensive), but it is fair to say that it has become harder to visualize the aircraft state. I feel that flying in low visibility involves a different kind of skill. (Probably a good way to get better at it, is — apart from attempting many IMC approaches — to actively compare in good visibility the changes in visual scene with the changes in HSI and glideslope indicator, and try to mentally unite both representations.)

`Quick' and `slow' loop

When preparing for flight in the simulator, trying to learn from previous attempts, I discovered that I needed some kind of strategy to apply to the task of stabilizing the aircraft on a certain sinkrate, airspeed and heading.

It helped me to simplify the relation between the different controls and think in terms of a quick and slow control loop:

- *Pitch angle* responds very quickly to a change in elevator pressure, followed by a slower response of the *sinkrate*;
- *Engine thrust* responds quickly to a different lever setting, followed by a slower response of the *airspeed*;
- *Bank angle* responds very quickly to a change in aileron pressure, followed by a slower response of the aircraft's *heading*.

So, while e.g. airspeed is not only dependent on the engine thrust, I found out that to some extent it helps to think that this actually *is* the case. Then, when flying, there are the three separate tasks of maintaining (after all, corrective control) the correct sinkrate, airspeed and heading. The pilot can do so by instead looking very actively for the correct pitch, thrust and bank. When one of these 'quick' parameters stray from the ideal ones, they can be corrected very directly; as soon as this takes some more time, the 'slower' parameters will start to deviate as well. This little strategy helped me especially in the IMC flights.

Preparation

When calculating ideal values of some flight parameters a priori, it is much easier to perform the corrective control actions for flying the approach in a stabilized manner. For instance, if the pilot knows beforehand what values the pitch angle and thrust setting should approximately have for the current wind conditions, aircraft weight, etc. it is easier to stabilize the aircraft on the desired path with the ideal airspeed. If not, the pilot has to find the correct pitch to ensure a correct sinkrate while adjusting the thrust to actually provide the aircraft with just the right amount of power. When knowing what parameter values to aim for, the searching/hunting for the correct ones is reduced considerably.

From a safety perspective, this is also recommended, since major deviations would quickly alert the pilot that something is amiss: e.g. if the required thrust for maintaining the glidepath turns out to be much larger than calculated, the pilots would check for wrong flap settings, etc. right away.

Concentration and attention

The instructing pilot used 15 minutes as the maximum amount of time during which a human can concentrate to full extent. During training, therefore, we took breaks after this period. In cockpits in actual aircraft, the instructor would therefore recommend to take over manual control from the autopilot on the descent at about 5000 feet: doing so earlier would lead to manual flying for more than 15 minutes straight; doing so a lot later might not give the pilot enough time to get a proper feel for the aircraft (after possibly having rested and not flown for hours).

Simulator versus aircraft

The professional pilot judged the simulator in our lab to be a fairly accurate representation of the real Boeing 747-400, especially for the elevator/pitch and thrust behavior. Since force-feedback is lacking, the trimming behavior of the controls was said to be not very realistic and in fact a bit more difficult and less intuitive than in reality. Trimming the elevator pressure in the simulator almost requires some kind of mnemonic to remember which side of the button to use, whereas in reality this is supposed to happen nearly automatically. (The simulator is also different in that there is no prospect of life-threatening situations, nor the presence of passengers…)

Technique

During the approach it is recommended to keep only one hand on the control column; the other should rest on the thrust lever. By doing this already during training, the pilot will develop the skill to operate the control column with one hand. The upshot of it is that thrust corrections can be made more quickly, all the while not causing disturbances on the control column due to changing of hands.

On the final approach with the runway in sight, there are a few strategies to time the flare maneuver and land safely. One is the *aiming method*: by choosing a virtual aiming point, the pilot has a goal to fly to. At first this point should be the special aiming point markings on the runway, and when just above the runway (according to a specific height on the altimeter), this should change to a point a few thousand feet in the distance (thus initiating the flare). Another strategy is the *sinkrate method*: by looking outside (through the side windows as well) in the distance, a trained pilot is said to be able to observe a change in perceived sinkrate when arriving just above the runway. At this point, the pilot should initiate the flare.

Nomenclature

AFM Aircraft flight manual

AGL (Altitude) above ground level

AR Authorization Required

ATC Air traffic control

CFIT Controlled flight into terrain

CTAF Common traffic advisory frequency

DME Distance measuring equipment

FAA Federal Aviation Administration

GNSS Global navigation satellite system

GPS Global positioning system

HSI Horizontal situation indicator

IAP Instrument approach procedure

ICA Index of cognitive activity

ICAO International Civil Aviation Organization

IenM Ministry of Infrastructure and the Environment

IFR Instrument flight rules

ILS Instrument landing system

IMC Instrument meteorological conditions

IRS Inertial reference systems

LOC Localizer

MAP Missed approach point

NAA National aviation authority

NAVAID Navigation aid, system providing navigation data

NDB Non-directional beacon

nm Nautical mile: non-SI unit equal to 1852 meters

NPAs Non-precision approaches

PAPI Precision approach path indicator

PBN Performance-based navigation

POH Pilot's operating handbook

RNAV Area navigation

RNP Required navigation performance

SAAAR Special Aircraft and Aircrew Authorization Required

SOPs Standard operating procedures

STARs Standard Terminal Arrival Routes

TEPR Task-evoked pupillary response

TPP Terminal Procedures Publication

VFR Visual flight rules

VHF Very-high frequency

VMC Visual meteorological conditions

VOR Omni-directional radio

WAAS Wide area augmentation system

Bibliography

[1] *2013 Safety Report*. ICAO, 2013.

[2] Jans Aasman, Gijsbertus Mulder, and Lambertus J.M. Mulder. Operator effort and the measurement of heart-rate variability. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 29(2):161–170, 1987.

[3] ACT-R Research Group. Department of Psychology, Carnegie Mellon University. URL `http://act-r.psy.cmu.edu/about/`.

[4] Federal Aviation Administration. *Instrument Procedures Handbook*. U.S. Department of Transportation, 2007.

[5] Federal Aviation Administration. *Instrument Flying Handbook*. U.S. Department of Transportation, 2012.

[6] H. Ayaz, P.A. Shewokis, S. Bunce, K. Izzetoglu, B. Willems, and B. Onaral. Optical brain monitoring for operator training and mental workload assessment. *Neuroimage*, 59(1):36–47, 2012.

[7] R. Barhydt and C.A. Adams. *Human factors considerations for performance-based navigation*. National Aeronautics and Space Administration, Langley Research Center, 2006.

[8] H. Bhana. Trust but verify. *AeroSafety world*, 5, 2010.

[9] Kenneth R Boff and Janet E Lincoln. Pupil diameter as an indicator of workload. In *Engineering Data Compendium. Human Perception and Performance*, volume 2, chapter 7.728. Harry G. Armstrong Aerospace Medical Research Laboratory, 1988.

[10] Jeffrey B Brookings, Glenn F Wilson, and Carolyne R Swain. Psychophysiological responses to changes in workload during simulated air traffic control. *Biological psychology*, 42(3): 361–377, 1996.

[11] A.S. Carlin, A.L. Alexander, and N. Schurr. *Modeling Pilot State in Next Generation Aircraft Alert Systems*. TRID, the TRIS and ITRD Database, 2011.

[12] S.M. Casner. Perceived vs. measured effects of advanced cockpit systems on pilot workload and error: Are pilots' beliefs misaligned with reality? *Applied Ergonomics*, 40(3):448–456, 2009.

[13] J.M. Colombi, M.E. Miller, M. Schneider, M.J. McGrogan, C.D.S. Long, and J. Plaga. Predictive mental workload modeling for semiautonomous system design: Implications for systems of systems. *Systems Engineering*, 2012.

[14] Jacques Drappier. Air carrier training insights: Maintaining manual flying skills. In *World Aviation Training Conference and Tradeshow*, Rosen Shingle Creek Resort, Orlando, Florida, USA, 2012.

[15] Jorg Onno Entzinger. *Analysis of Visual Cues for Human Pilot Control Analysis of Visual Cues for Human Pilot Control in the Final Approach and Landing*. PhD thesis, University of Tokyo, 2010.

[16] Flight Safety Foundation. Briefing Note 7.1 Stabilized Approach. *FSF ALAR Tool Kit*, 2009.

[17] D.C. Foyle and B.L. Hooey. *Human Performance Modeling in Aviation*. CRC Press/Taylor & Francis Group, 2008. ISBN 9780805859645.

[18] V.J. Gawron. *Human Performance, Workload, and Situational Awareness Measures Handbook, Second Edition*. Taylor & Francis, 2008. ISBN 9781420064506.

[19] Martin Gestwa. Using fuzzy control for modeling the control behaviour of a human pilot. In L Grigorie, editor, *Fuzzy Controllers, Theory and Applications*. InTech, 2011.

[20] G.H. Gil and D.B. Kaber. An accessible cognitive modeling tool for evaluation of pilot–automation interaction. *The International Journal of Aviation Psychology*, 22(4):319–342, 2012.

[21] K.A. Gluck and R.W. Pew. *Modeling Human Behavior With Integrated Cognitive Architectures: Comparison, Evaluation, and Validation*. Lawrence Erlbaum, 2005. ISBN 9780805850475.

[22] B.F. Gore. Workload as a performance shaping factor for human performance models. In *20th Behavior Representation in Modeling and Simulation (BRIMS) Conference*, Sundance, Utah, 2011.

[23] William R. Gray and James E. Brown. *Time Safety Margin: a generalized method for dive safety planning*. The Society of Experimental Test Pilots, 2010.

[24] Markus Guhe, Wayne D Gray, Michael J Schoelles, Wenhui Liao, Zhiwei Zhu, and Qiang Ji. Non-intrusive measurement of workload in real-time. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 49, pages 1157–1161. SAGE Publications, 2005.

[25] Robert W. Holt, Ronald Chong, Wolfgang Schoppek, Jeffrey T. Hansberger, and Deborah A. Boehm-Davis. Modeling crew interaction. In *Workshop on ACT-R Models of Human-System Interaction*, 2002.

[26] T.T. Ivancevic and B. Jovanovic. Human operator modeling and lie-derivative based control. *arXiv preprint arXiv:0907.1206*, 2009.

[27] Jeff Klingner, Rakshit Kumar, and Pat Hanrahan. Measuring the task-evoked pupillary response with a remote eye tracker. In *Proceedings of ETRA*, pages 69–72, 2008.

[28] Arthur F Kramer. Physiological metrics of mental workload: A review of recent progress. *Multiple-task performance*, pages 279–328, 1991.

[29] Unmesh Kurup and Christian Lebiere. What can cognitive architectures do for robotics? *Biologically Inspired Cognitive Architectures*, 2(0):88 – 99, 2012. ISSN 2212-683X. doi: 10.1016/j.bica.2012.07.004.

[30] Pat Langley, John E. Laird, and Seth Rogers. Cognitive architectures: Research issues and challenges. *Cognitive Systems Research*, 10(2):141 – 160, 2009. ISSN 1389-0417. doi: 10.1016/j.cogsys.2006.07.004.

[31] Ken Leiden and Brad Best. A cross-model comparison of human performance modeling tools applied to aviation safety. *Micro Analysis & Design, Inc. Boulder, CO*, 80301:2005, 2005.

[32] Sandro Leuchter and Leon Urbas. Extending ACT-R for modeling dynamics and timing for operating human-machine systems. *MMI-Interaktiv*, 2004.

[33] T. Liadal. ACT-R: modeling the brain. In *Seminar im Bereich Künstliche Intelligenz*, February 2007.

[34] Sandra P Marshall. *Method and apparatus for eye tracking and monitoring pupil dilation to evaluate cognitive activity*. Google Patents, July 18 2000. US Patent 6,090,051.

[35] Sandra P Marshall. The index of cognitive activity: Measuring cognitive workload. In *Human factors and power plants, 2002. proceedings of the 2002 IEEE 7th conference on*, pages 7–5. IEEE, 2002.

[36] Sandra P Marshall. Identifying cognitive state from eye metrics. *Aviation, space, and environmental medicine*, 78(Supplement 1):B165–B175, 2007.

[37] Matlab Wavelet Toolbox Documentation. The Mathworks, Inc. URL `http://www.mathworks.nl/help/wavelet/`.

[38] Oskar Palinko and Andrew L Kun. Exploring the influence of light and cognitive load on pupil diameter in driving simulator studies. *Driving Assessment*, 2011.

[39] Oskar Palinko, Andrew L Kun, Alexander Shyrokov, and Peter Heeman. Estimating cognitive load using remote eye tracking in a driving simulator. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*, pages 141–144. ACM, 2010.

[40] David Peebles. A cognitive architecture-based model of graph comprehension. In *11th International Conference on Cognitive Modeling*, pages 37–42, April 2012. URL `http://eprints.hud.ac.uk/14247/`.

[41] Elisa Pérez, Luis Miguel Nunes, Ángela Conchillo, and Miguel Ángel Recarte. Mental workload and visual impairment: Differences between pupil, blink, and subjective rating. *The Spanish Journal of Psychology*, 11(2):374–385, 2008.

[42] Robi Polikar. The Wavelet Tutorial. URL `http://users.rowan.edu/~polikar/`.

[43] M. Pomplun and S. Sunkara. Pupil dilation as an indicator of cognitive workload in human-computer interaction. In *Proceedings of the International Conference on HCI*, 2003.

[44] D.D. Salvucci, E.R. Boer, and A. Liu. Toward an integrated model of driver behavior in cognitive architecture. *Transportation Research Record: Journal of the Transportation Research Board*, 1779(-1):9–16, 2001.

[45] Airbus SAS. Getting to grips with RNP AR. *Flight Operations Support & Services*, February 2009.

[46] Maximilian Schwalm. *Pupillometrie als Methode zur Erfassung mentaler Beanspruchungen im automotiven Kontext*. PhD thesis, Universität des Saarlandes, 2009.

[47] Walter Shawlee. RNP: The world of required navigation performance. *Avionics News*, January 2008.

[48] Karen Stevens, Kingsley Fletcher, Alistair Walsh, Vic Demczuk, and Gary Ewing. Comparison of pupil diameter and the index of cognitive activity as measures of task demand. In *40th Australasian Experimental Psychology Conference*, April 2013.

[49] Jan Sietse Stilma and Theodorus Bernardus Voorn. *Oogheelkunde*. Bohn Stafleu van Loghum, isbn: 9789031350803 edition, 2008.

[50] C.B.G. Symank. Real-time workload monitoring: Improving cognitive process models. In *RTO Human Factors and Medicine Panel (HFM) Symposium*. NATO Science and Technology Organization, 2010.

[51] Karl F Van Orden, Wendy Limbert, Scott Makeig, and Tzyy-Ping Jung. Eye activity correlates of workload during a visuospatial memory task. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 43(1):111–121, 2001.

[52] JA Veltman and AWK Gaillard. Physiological indices of workload in a simulated flight task. *Biological psychology*, 42(3):323–342, 1996.

[53] Kim J Vicente, D Craig Thornton, and Neville Moray. Spectral analysis of sinus arrhythmia: A measure of mental effort. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 29(2):171–182, 1987.

[54] Joseph Tao-yi Wang. Pupil dilation and eye tracking. *A Handbook of Process Tracing Methods for Decision Research: A Critical Review and User's Guide, M. Schulte-Mecklenbeck, A. Kühberger and R. Ranyard, eds., Psychology Press, New York*, pages 185–204, 2011.

[55] Glenn F Wilson. An analysis of mental workload in pilots during flight using multiple psychophysiological measures. *The International Journal of Aviation Psychology*, 12(1):3–18, 2002.

[56] Deniz Yilmaz, Michael Jump, Lu Linghai, and Michael Jones. State-of-the-art pilot model for RPC prediction report. *ARISTOTEL, Aircraft and Rotorcraft Pilot Couplings – Tools and Techniques for Alleviation and Detection*, 2011.

[57] G.O.A. Zapata, R. Kawakami, H. Galvao, and T. Yoneyama. Extracting fuzzy control rules from experimental human operator data. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 29(3):398–406, 1999.

[58] Jeffrey C Zemla, Volkan Ustun, Michael D Byrne, Alex Kirlik, Kenyon Riddle, and Amy L Alexander. An ACT-R Model of commercial jetliner taxiing. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 55, pages 831–835. SAGE Publications, 2011.