PILOT TASK DEMAND LOAD DURING RNAV APPROACHES

Monique Heiligers

Design: Legatron Electronic Publishing, Rotterdam, The Netherlands

Cover design: Monique Heiligers

Printed by: Ipskamp Drukkers, Enschede, The Netherlands

ISBN: 978-90-9025943-7

PILOT TASK DEMAND LOAD DURING RNAV APPROACHES

Proefschrift

ter verkrijging van de graad van doctor

aan de Technische Universiteit Delft,

op gezag van de Rector Magnificus prof.ir. K.C.A.M. Luyben,

voorzitter van het College voor Promoties,

in het openbaar te verdedigen op maandag 10 januari 2011 om 12.30 uur

door

Maria Monica VAN EMST – HEILIGERS

ingenieur luchtvaart en ruimtevaart

geboren te Geldrop

Dit proefschrift is goedgekeurd door de promotoren:

Prof. dr.ir. Th. van Holten
Prof. dr.ir. M. Mulder

Samenstelling promotiecommissie:

Rector Magnificus, voorzitter
Prof.dr.ir. Th. van Holten, Emeritus hoogleraar Technische Universiteit Delft, promotor
Prof.dr.ir. M. Mulder, Technische Universiteit Delft, promotor
Prof.dr.ir. M.J.L. van Tooren, Technische Universiteit Delft
Prof.dr.ir. A. Verbraeck, Technische Universiteit Delft
Prof.dr.ir. P.A. Wieringa, Technische Universiteit Delft
Prof.dr. S.W.A. Dekker, Lund University, Sweden
Ir. F.J. Abbink, Voormalig hoogleraar Technische Universiteit Delft, Abbink Aviation Consultancy
Prof.dr. R. Curran, Technische Universiteit Delft, reservelid

This research is supported by the Dutch Technology Foundation STW, which is the applied science division of NWO, and the Technology Programme of the Ministry of Economic Affairs (project number DLR.7071).

Contents

PART II. SCIENTIFIC BACKGROUND

Summary

Introduction

The question that started this research was: "Why is approach A more difficult to fly for a pilot than approach B?". An approach is the final part of the flight. It is a published three-dimensional trajectory that aircraft follow when flying towards an airport. In order to find out why approach A is more difficult to fly than approach B, the factors that complicate an approach for the pilot need to be identified. Once these factors are identified, approaches can be designed based on these factors, thereby keeping the difficulty of flying the approach low.

Knowledge about which factors complicate an approach is important because of two reasons. First, a large percentage of accidents occur during the approach and landing phase. It is hypothesized that, when an approach becomes less difficult to fly, the chance decreases that some of the causal factors (contributing to these accidents) occur. Second, air traffic movements are forecasted to increase considerably over the coming decades, while, at the same time, there is the ambition to reduce aircraft noise as well as aircraft emissions. Part of the solution to bring together these different objectives is to design new approaches. When the factors that complicate an approach are better understood, they can be taken into account during the design of these new approaches.

Based on the above, the three goals of this research were defined as follows:

Goal 1: the factors relevant for approach design, which influence the difficulty experienced by pilots while flying an approach, need to be identified.
Goal 2: a method needs to be developed with which it is possible to predict how difficult it will be for a pilot to fly a particular approach.
Goal 3: this method needs to be captured within a computer simulation that can be used in the early stages of approach design.

Basic principles and assumptions

Since there is a worldwide change towards Area Navigation (RNAV) operations, only RNAV approaches are considered within this research. An RNAV approach is described by a series of waypoints that are defined as latitude and longitude positions in space. At each of these waypoints a required altitude and airspeed can be defined, referred to as altitude and airspeed constraints. When flying an approach the pilot aims to follow the defined 3-D trajectory and aims to meet all airspeed constraints at the waypoints. The approach starts at the Initial Approach Fix (IAF), see Figure S.1, and is (for this research) divided in three parts. The Intermediate Fix (IF) is the first waypoint on runway heading, and is the waypoint where the Localizer is captured. The last characteristic waypoint is the Final Approach Fix (FAF).
The difficulty experienced by the pilot while flying an approach is expressed in terms of the Task Demand Load (TDL). TDL expresses how difficult a task is. It should not be confused with mental load, which is the workload as experienced by the pilot performing the task. In the latter, factors such as fatigue, a pilot's skill and training could have a large effect. In other words, whereas for a given approach (and given circumstances) the TDL is the same, the mental load may vary between pilots, in fact it might even vary between different occasions for the same pilot.

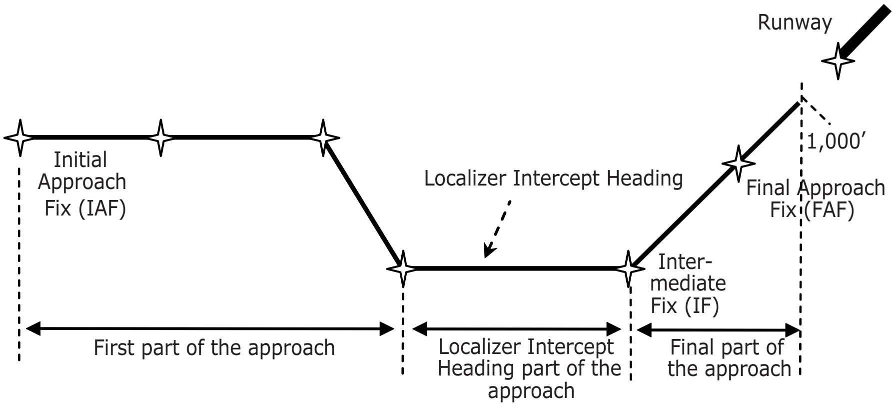


Figure S.1 | Part of flight considered (top view).

The approach is assumed to be flown according to Standard Operating Procedures and pilots are assumed to aim to achieve a stabilized approach at 1,000 feet. The criteria for a stabilized approach are defined in terms of required flap setting, required airspeed, etc. When these criteria are not met at 1,000 feet, the pilot should initiate a go-around.

Within this research, pilot TDL is assessed as a function of: 1. the approach trajectory and its altitude and velocity constraints, 2. the wind speed and wind direction, 3. the aircraft type (Boeing 747 or Cessna Citation) and 4. the aircraft mass.

Results

Since the scanning of the instruments and the continuous control actions were assumed to contribute to pilot TDL, a theoretical scanning and control model for the pilot were developed. These were validated by two flight simulator experiments with eye-tracking equipment. The conclusion, however, was that the scanning and the manual control do not contribute significantly to pilot TDL. It was found that other tasks such as selecting flaps and gear, performing checklists etcetera were more important contributors to pilot TDL. Additionally, pilots seemed to be much more concerned with 'higher level' or, so to say, larger time-scale properties of the approach trajectory, such as for example the amount of track-miles available between important waypoints in the approach, rather than the smaller time-scale properties such as additional control actions necessary after flap deployment.

Therefore, another flight simulator experiment was performed for a Boeing 747 which was used to obtain a first indication of these other factors that contribute to pilot TDL. A rather extensive list of contributing factors was obtained from this experiment. A second flight simulator experiment was performed (also for the Boeing 747), which reduced this list of factors. Based on the results of these experiments, the factors that seem to influence pilot TDL were identified as:

- ❖ For the first part of the approach (see Figure S.1): the fact whether or not the constraints can be met at the waypoints.
- ❖ For the Localizer intercept part of the approach (see Figure S.1): the time available to perform all actions. And to a lesser extent: the Localizer intercept speed, the Localizer intercept angle, and whether the constraints at the waypoints can be met.
- ❖ For the final part of the approach (see Figure S.1): the fact whether or not a stabilized approach can be achieved at 1,000 feet, the distance between IF and FAF and the airspeed on final.

Based on these factors that were found to contribute to pilot TDL, a method to predict pilot TDL was developed, which consists of seven guidelines for the design of approaches. When all guidelines are followed, pilot TDL is predicted to be acceptable for an approach. The guidelines for the contributors to pilot TDL for a Boeing 747 are that:

- aircraft should be able to meet the altitude and airspeed constraints throughout the approach;
- there should be sufficient time for pilots to perform all actions on Localizer intercept heading;
- it should be possible to achieve a stabilized approach;
- the distance between the IF and FAF should be sufficient;
- the aircraft vertical speed should be well below the sink rate warning;
- the Localizer intercept speed should not be too high, and that
- the Localizer intercept angle should not be too large.

Subsequently, a different aircraft type (the Cessna Citation) was considered to check whether the factors and guidelines that were found for the Boeing 747 were also valid for another aircraft. Based on flight simulator tests and real flight tests, it was concluded that this indeed was the case. There was also no difference in factors contributing to pilot TDL between the simulator sessions and the real flights.

Additionally, a highly-detailed computer simulation program for the Boeing 747 and Citation was developed, in order to be able to predict whether the guidelines for pilot TDL are adhered to for a given approach. The computer simulation consisted of an aircraft model, a wind model and a pilot model. The trajectory with its airspeed and altitude constraints formed the input. All pilot models in the computer simulation were kept as simple as possible, whereas the pilot's environment (aircraft and wind) was modeled as detailed as possible. This is done because it is not the intention to model the pilot's actions as accurately as possible, but the aim is to only get an indication of how "hard" the pilot has to work as a consequence of the demands the environment places on the pilot. The pilot models were based on the data recorded during the flight simulator tests. The pilot's actions were modeled using trigger events and reaction times with respect to these trigger events. The non-linear computer simulation for the B747 and Cessna Citation, based on a Monte Carlo simulation technique, could:

- predict the percentage of flights that would meet the altitude and velocity constraints at the waypoints and that would result in a stabilized approach at 1,000 feet as a function of wind direction, wind speed, type of aircraft and aircraft mass;
- provide insight into the busy parts of an approach from the pilot's perspective; and
- predict the aircraft's motion in the longitudinal (vertical) plane.

Additionally, and most importantly, the computer simulation, combined with the regulations in the Procedures for Air Navigation Services – Aircraft Operations (PANS-OPS), provides sufficient information to assess whether the guidelines for the contributors to pilot TDL are adhered to.

Finally, a much simpler computer simulation based on a point mass model representation of the aircraft was developed. This point mass model simulation could generate exactly the same results as the highly-detailed computer simulation but required significantly shorter simulation times. This point mass model simulation can be used by approach designers as an additional tool during the design of approaches.

Recommendations

For further research regarding pilot TDL during approaches it is advised to concentrate on the computer simulation based on the point mass model. It is recommended to incorporate other aircraft types in the point mass model computer simulation in order to arrive at a tool that can be used to analyze an approach for all the aircraft types that will actually fly the approach once it becomes operational. However, when adding an aircraft that requires a very different kind of SOPs, such as an Airbus aircraft, it is advised to perform additional flight simulator tests because of two reasons: 1. to check whether the adopted simulation philosophy (using trigger events and reaction times) also produces reliable results for these kinds of aircraft, and 2. to check whether the same factors influence pilot TDL for these types of aircraft.

The same philosophy that was adopted during this research regarding approaches: using a detailed model of the environment of the pilot and a simple model for the pilot him/herself, modeling the pilots' actions according to trigger events and reaction times, etc. can also be adopted for other flight phases, most notably for departures.

Abbreviations

AAL	Above Airport Level
AAS	Amsterdam Airport Schiphol
ALAR	Approach-and-landing Accident Reduction
ATC	Air Traffic Control
ATM	Air Traffic Management
BADA	Base of Aircraft Data
CAA-NL	Civil Aviation Authority Netherlands
CDA	Continuous Descent Approach
CNS	Communications-Navigation-Surveillance
DME	Distance Measuring Equipment
ECAC	European Civil Aviation Conference
FAF	Final Approach Fix
FMS	Flight Management System
FSF	Flight Safety Foundation
GNSS	Global Navigation Satellite System
GRACE	Generic Research Aircraft Cockpit Environment
IAF	Initial Approach Fix
IAS	Indicated Airspeed
ICAO	International Civil Aviation Organization
IF	Intermediate Fix
ILS	Instrument Landing System
KLM	KLM Royal Dutch Airlines
LNAV	Lateral Navigation
LVNL	Air Traffic Control the Netherlands
MCC	Multiple Crew Coordination
MIDAS	Man-Machine Integrated Design and Analysis System
ML	Mental Load
NDB	Non-Directional Beacon
NASA	National Aeronautics and Space Administration
NLR	National Aerospace Laboratory
PANS-OPS	Procedures for Air Navigation Services Aircraft Operations
PROCRU	Procedure-oriented crew model
RNAV	Area Navigation
RNP	Required Navigation Performance
SIMONA	Simulation Motion and Navigation
SOP	Standard Operating Procedures
SRS	SIMONA Research Simulator
STAR	Standard Arrival Route
TDL	Task Demand Load
TLX	Task Load indeX
VNAV	Vertical Navigation
VHF	Very High Frequency
VOR	VHF Omnidirectional Range
VREF	Reference Speed

Outline of this thesis

The contents of this thesis are organized into two parts. Part I is written for the general reader with a background in aerospace engineering or aviation. Part II treats the subject more in depth, in a scientific manner, and mainly consists of the journal articles that were written throughout this research.

Part I contains a description of the entire research. It is written in the style of an "executive summary", guiding the reader through the process of this Ph.D. research. It clearly indicates the relevance of the research, the assumptions that were made, the scope of the research, the steps that were followed, the experiments that were performed and the results that were obtained. It does so without going into too much detail, and without providing the scientific background for these results. It is intended to give the reader an overview of the research, and will suffice for people who are mainly interested in the results but who require more information than provided in the summary. In this part reference is made to the journal articles that can be found in Part II.

Part II is added for the reader who is interested in the scientific background. Although Part II can be read independently, it is advised to read Part I first in order to obtain a global overview of this Ph.D. research. As stated above, part II comprises the five journal articles that were written on the subject of this thesis, as well as the discussion of the results. The conclusions of part II, obviously, are the same as the conclusions of part I. The literature study is included in the first journal article.

PART I

FOR THE GENERAL READER

GOALS, BASIC PRINCIPLES AND BACKGROUNDS

1 | Introduction

The Flight Safety Foundation Approach-and-landing Accident Reduction (FSF ALAR) Task Force (FSF ALAR Task Force, 1999) has performed a comprehensive research with respect to Approach and Landing Accident Reduction. Approach-and landing phase accidents account for a significant proportion of air transport accidents. Approximately 45 percent of the world jet-fleet accidents until 2009 occurred in these flight phases and accounted for 38 percent of all fatalities (Boeing Commercial Airplane Group, 2010). Table 1.1 states the most frequently identified causal factors in 76 approach-and-landing accidents and serious incidents worldwide in 1984 through 1997. It clearly shows that the causal factors with the highest percentage, and thus the most frequent causal factors, are all crew related factors. It is hypothesized that the chance of these crew related factors occurring increases when an approach becomes more complicated for a pilot to fly.

There are many reasons why an approach might become more difficult to fly. One can think of, for example, emergency situations such as engine failure or having to deal with an uncontrollable aircraft. These are extreme situations that cannot be taken into account during the design of approaches. There are also more common situations such as, for instance, a demanding clearance from Air Traffic Control (ATC), a rushed approach or strong tailwind conditions, that might increase the difficulty (and that are also underlying factors for the results shown in Table 1.1). These type of factors *can* be taken into account when designing an approach. In order to be able to design approaches such that the difficulty of flying the approach (under nominal conditions) is kept low, it is necessary to identify the full set of factors that complicate an approach for a pilot. By designing approaches such that the difficulty of flying the approach is low, it is hypothesized that the chance that some of the causal factors (given in Table 1.1) occur is reduced.

Table 1.1 | Most frequently identified causal factors in 76 approach-and-landing occurrences (FSF ALAR Task Force, 1999). It should be noted that the factors are not mutually exclusive, e.g., "press-on-itis" (continuing toward the destination in spite of a lack of readiness of the airplane or crew) also may have involved being "being high/fast on the approach".

Causal Factor	Yes Percent	No Percent	Unknown Percent
Poor professional judgment/airmanship	73.7	19.7	6.6
Omission of action/inappropriate action	72.4	22.4	5.3
Failure in CRM (cross-check/coordinate)	63.2	25.0	11.8
Lack of positional awareness in air	51.3	42.1	6.6
Lack of awareness of circumstances in flight	47.4	40.8	11.8
Flight-handling difficulties	44.7	34.2	21.1
Slow/delayed crew action	44.7	43.4	11.8
"Press-on-itis"	42.1	42.1	15.8
Deliberate nonadherence to procedures	39.5	48.7	11.8
Slow and/or low on approach	35.5	55.3	9.2
Incorrect or inadequate ATC instruction/advice/service	32.9	60.5	6.6
Fast and/or high on approach	30.3	60.5	9.2
Postimpact fire (as a causal factor of the fatalities)	26.3	71.1	2.6
Aircraft becomes uncontrollable	25.0	69.7	5.3
Lack of qualification/training/experience	22.4	60.5	17.1
Disorientation or visual illusion	21.1	64.5	14.5
Interaction with automation	19.7	65.8	14.5

Another reason to identify the factors that complicate an approach is the fact that a growth in the air transport industry is expected, with forecasts indicating that air traffic movements in the European Civil Aviation Conference (ECAC) Area will increase by over 30% between 2008 and 2015 (EUROCONTROL, 2008). Other forecasts indicate that overall world passenger traffic is expected to increase by 4.7% per annum until 2028 and that, in the same period of time, the number of frequencies offered on passenger routes will more than double (Airbus, 2009). These extra movements are likely to create extra congestion and delays, and mean that there is an ever-growing pressure to upgrade the capacity of the overall system. On the other hand there is the drive to reduce aircraft noise as well as aircraft emissions, as set forth in the goals of "European Aeronautics: A Vision for 2020" (Argüelles et al., 2001). In order to unify these different objectives, new solutions will need to be created that will increase the capacity and at the same time reduce aircraft noise and aircraft emissions. Part of the solution is the design of new approaches, which might include, for instance, the use of higher approach altitudes or so-called continuous descent approaches. Before these new approaches can be introduced it is necessary, to understand the impact these new approaches have on the difficulty as experienced by the pilot when flying the approach. When the factors are known that increase the difficulty of flying an approach, then these can be taken into account during the design of these new approaches.

The goal of this research therefore is to identify the factors in an approach that influence the difficulty as experienced by the pilot and to find a way to predict this difficulty.

2 | Goals of this research resulting from current practice in approach design

To gain insight into the process of approach design and to establish how the difficulty of flying an approach is currently assessed two meetings were held with Air Traffic Control the Netherlands (LVNL). LVNL design approaches for Amsterdam Airport Schiphol (AAS). The discussion below summarizes the findings of these meetings.

2.1 | Current practice in approach design

Firstly, a new procedure is designed taking into account the rules and regulations set forth in the Procedures for Air Navigation Services Aircraft Operations (PANS-OPS) by ICAO (ICAO, 2006). In addition to this, basic rules of thumb about the climb and descent gradients that can be achieved by various types of aircraft are used. In environments that are non-critical for obstacles, flyability related design parameters, e.g. the minimum distance between waypoints and the mimimum turn radius for different phases of flight, may also be based on experience with previous designs.

Subsequently the approach is tested by LVNL with respect to each of the following three areas:
1. Conflicts arising from Air Traffic Control
2. Noise Abatement
3. Flyability of the approach

Regarding conflicts arising from Air Traffic Control one can think of, for example, the crossing of climbing and descending aircraft. The newly designed approach is analysed by air traffic controllers and based on expert judgement the approach is pronounced to be acceptable. Experience has shown that this works well.

The checks that are performed regarding noise abatement will not be discussed here since these are not directly relevant for the subject of this research.
Of main importance for this research are the tests that are performed to demonstrate the flyability of the approach, or in other words: that demonstrate that an aircraft and pilot can actually fly a new approach.

First, a desktop computer simulation is performed to check whether the Flight Management System (FMS) can fly the newly designed approach. To this end desktop software emulating the General Electric FMS of the Boeing 737 is used. Experience has shown that this FMS normally provides a good lateral track keeping, but when waypoints are placed too closely together the predicted path may deviate from the desired route. This happens, for instance, in strong wind conditions. This desktop computer simulation is thus used to check the geometry of the procedure in the lateral (horizontal) plane, no accurate prediction of the flyability in the longitudinal (vertical) plane is available.

Secondly, the approach is tested in a full flight simulator for one or more relevant types. The approach is flown with and without the FMS operable under varying weight and wind conditions to validate the desktop runs.

2.2 | Observations regarding current practice in approach design

Two remarks can be made regarding the process of testing and implementing a new approach as described above. One remark concerns the fact that, before the newly designed approach is tested in the flight simulator, the assessment of the flyability of the approach concentrates on the fact whether the FMS can fly the approach. Whether the pilot can fly the approach is tested during the flight simulator test. This check is performed at the end of the design cycle. If problems are detected at this stage it will be costly and time-consuming to modify the approach since a large part of the design process then needs to be repeated.

The other remark concerns the fact that, before the flight simulator tests are performed, no accurate information is available about the motion of the aircraft in the longitudinal plane. For example: the approach designers have no indication of how 'easy' it is for different aircraft types to meet the velocity and altitude constraints at the waypoints. It is desirable to have this information available since if the constraints are too loose or conservative (i.e., it is easy for the aircraft to meet the constraints) this might, in practice, result in aircraft reaching the required altitude and velocity well before reaching the waypoint. This in its turn results in the situation that they have to level off and have to apply power to the engines in order to maintain the required altitude and velocity, thereby causing noise nuisance on the ground. On the other hand, waypoint constraints that are chosen too tight (such that it is difficult for aircraft types to meet the constraints) might in reality not be met by aircraft, especially when tailwind is present.

2.3 | Goals of this research

With the two remarks given above in mind, the goals of this research can now be more accurately defined. The purpose is threefold:

- Goal 1: the factors relevant for approach design, which influence the difficulty experienced by pilots while flying an approach, need to be identified.
- Goal 2: a method needs to be developed with which it is possible to predict how difficult it will be for a pilot to fly a certain approach.
- Goal 3: this method needs to be captured within a computer simulation that can be used in the early stages of approach design. With this tool, the designer should be able to rapidly evaluate a potential approach from the perspective of the demands it imposes on the pilot's limited resources. Additionally, this computer simulation should be able to give a realistic prediction of the motion of the aircraft in the longitudinal plane.

3 | Types of approaches considered: RNAV approaches

Many different types of approaches exist and within the limited amount of time available for this research not all types of approaches can be considered. Therefore a choice needs to be made. Within this research only Area Navigation (RNAV) approaches will be considered. RNAV approaches provide pilots with lateral and longitudinal guidance based on a series of waypoints. These waypoints are published latitude and longitude positions in space with no associated ground navigational aid. They are pre-programmed into a global positioning satellite (GPS) receiver or flight management system (FMS), which display the aircraft's position relative to these waypoints during the approach (Godley, 2006).

The choice to focus on RNAV approaches is based on the developments for future approaches. In this chapter both the future developments at Amsterdam Airport Schiphol are presented as well as the future developments in approach design for the European Civil Aviation Conference Airspace. At the end of this chapter the reason to focus on RNAV approaches will be clear.

3.1 | Future developments for Amsterdam Airport Schiphol

Around 2011 aircraft that only use conventional navigation aids will no longer be allowed at Amsterdam Airport Schiphol. As of 2011-2012 all aircraft should be capable of achieving Advanced Required Navigational Performance 1 (Advanced RNP1). RNP1 refers to a required navigation performance accuracy within 1 NM (1.85 km) of the desired flight path (ICAO 1999) that is expected to be achieved at least 95 percent of the time by the population of aircraft operating within the airspace (ICAO 1999). For future operations at AAS the carriage of RNAV equipment will be required. This clearly indicates a shift towards RNAV approaches for Amsterdam Airport Schiphol.

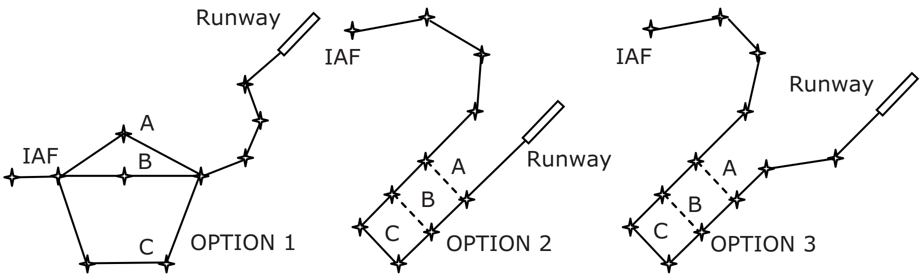


Figure 3.1 | Three options for future approaches at Amsterdam Airport Schiphol

In more detail, it can be stated that at the moment three different options for future approaches are considered for Amsterdam Airport Schiphol, all based on RNAV capability, see figure 3.1. Each of the options incorporates some sort of flexibility in order to provide the air traffic controller with the possibility to guide aircraft along different routes (A, B or C) thereby achieving correct separation between the aircraft. Each option is briefly described below (for a more detailed description see section 8.2).

The first option uses flexible routes at the beginning of the approach and a fixed route at the end. Correct separation of aircraft is achieved by guiding aircraft along different routes with different lengths.

The second option has a fixed route at the beginning of the approach. Flexibility and required separation are obtained relatively close to the runway by using a so-called trombone model (already in use at Frankfurt, Germany): where, by using RNAV capability, a large number of waypoints are defined on the downwind leg, and the air traffic controller can decide at which waypoint the aircraft should initiate the turn (to follow route A, B or C).

The third option bears resemblance with the second option, however, in option 3 a so-called displaced centerline is used. This option can be useful when the 'trombone' in option 2 (shortcuts A, B and C) would be situated over a crowded area which would not be desirable with respect to noise abatement. Option 3 provides the possibility to locate the shortcuts A, B and C over a more favorable area since the endpoints of A, B and C do not have to be located on the extended centerline.

3.2 | Future developments for the European Civil Aviation Conference airspace

The future developments for the European Civil Aviation Conference (ECAC) airspace are described in the Navigation Strategy for ECAC (EUROCONTROL, 1999). This Strategy aims to provide a harmonised and integrated common framework which will allow a cost-effective, customer oriented evolution of the European Air Navigation Systems during the period 2000-2015. Two of the main strategic streams described in the Navigation Strategy are aimed at (EUROCONTROL, 1999):

- ❖ achieving a total RNAV environment with defined RNP values for all operations ECAC wide;
- ❖ implementing 4D RNAV operations, to support the transition to a full gate to gate management of flight by 2015;

These objectives clearly indicate a change towards RNAV operations for the ECAC Airspace, and even 4D RNAV (navigation in 3D and time) for the en-route flight segment. The gradual transition to RNAV is the cornerstone of the ECAC Navigation Strategy from a Navigation Application perspective (EUROCONTROL, 2008). This agrees with Resolution A3-23 adopted by the ICAO 36th General Assembly (ICAO, 2007) which, amongst others, urges states to implement RNAV and RNP operations (where required) for en route and terminal areas, and with Implementation Package 1 identified by the Single European Sky ATM Research Programme, which comprises operational improvements requiring P-RNAV (EUROCONTROL, 2008, SESAR Consortium, 2008).

Since there is a change at Amsterdam Airport Schiphol towards RNAV approaches, and a shift towards RNAV operations within the ECAC airspace, the choice to only consider RNAV approaches within this research is a logical one.

4 | Basic principles and assumptions for this research

In this chapter, an overview is given of the basic principles of this research. The assumptions and the choices that have been made as to what is incorporated within this research are clarified, and the factors that are considered to be beyond the scope of this research are stated.

In short, the basic idea behind this research is the following: when the properties of the approach trajectory, such as location of waypoints, altitude constraints and velocity constraints are known it should be possible to predict the difficulty experienced by the pilot while flying this approach for different aircraft types and different wind- and turbulence conditions. A more elaborate explanation is given below.

4.1 | Difficulty in terms of Task Demand Load

The difficulty of flying an approach can be expressed in many different ways. Therefore, to formulate more accurately, this research aims to develop a method to quantify the Task Demand Load (TDL) of conducting an airport approach. Task demand load is defined as the mental workload *imposed* by the system to be controlled or supervised (Stassen, Johannsen & Moray, 1990), see also Figure 4.1. The task demand load is not to be mistaken for the mental workload *experienced* by the human operator, which is referred to as Mental Load (ML). Many of the well-known methods to measure workload, like the NASA Task Load indeX, measure ML, not TDL.

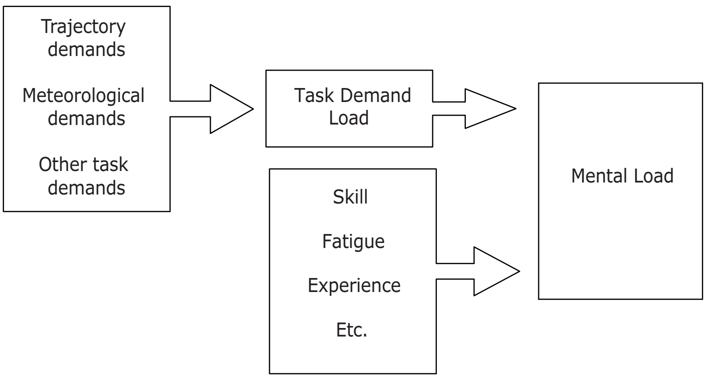


Figure 4.1 | Difference between Task Demand Load and Workload, adapted from (Hilburn and Jorna, 2001).

Within this research several experiments are performed during which pilots are asked to comment on approaches regarding the amount of effort these approaches require, or their effect on the difficulty as experienced by the pilot. When pilots give their opinion on these matters, they obviously base their opinion on the mental workload they experienced. This results in the situation that in order to obtain information about the Task Demand Load, pilots are asked about the mental workload they experienced during the experiments, unfortunately there is no other way. However, by choosing pilots with different levels of experience etc., by testing the approaches in random order and by converting the pilots' ratings to relative ratings, it is assumed that through the comments of the pilots a good indication of the Task Demand Load can be obtained.

4.2 | Approaches considered and automation used

Obviously, pilot TDL depends directly on the type of approach that is considered. As explained in the previous chapter, this research focuses on RNAV Approaches. Although it is appreciated that non-precision approaches such as NDB approaches are, in general, more difficult for a pilot to fly than RNAV approaches (Godley, 2006), a deliberate choice is made to focus on RNAV approaches for reasons given in the previous chapter. The last part of the RNAV approach is assumed to be flown using the Instrument Landing System (ILS). Therefore to formulate more accurately: this research will consider RNAV transitions ending with ILS approaches.

The part of the flight that is considered within this research starts at the Initial Approach Fix (IAF) and comprises the entire approach (Initial Approach, Intermediate Approach and Final Approach) until 1,000 feet above airport elevation, see Figure 4.2. Based on interviews with pilots it was decided to assume two different levels of automation during the approach: until Localizer Intercept Heading the approach is assumed to be flown using the FMS, Autopilots (Lateral Navigation (LNAV) and Vertical Navigation (VNAV) modes) and Autothrottle. At Localizer Intercept Heading (but before Localizer capture) the pilot switches to Flight Director (FD) mode and disconnects the Autothrottle, the remainder of the approach is thus flown using the FMS and FD, which implies manual control by the pilot. Although it is appreciated that pilots might also switch to the FD and disconnect the Autothrottle at a later point in time, or fly the approach completely on autopilots and autothrottle until landing, the situation assumed for this research is considered to be the most demanding in terms of pilot TDL of all common practice

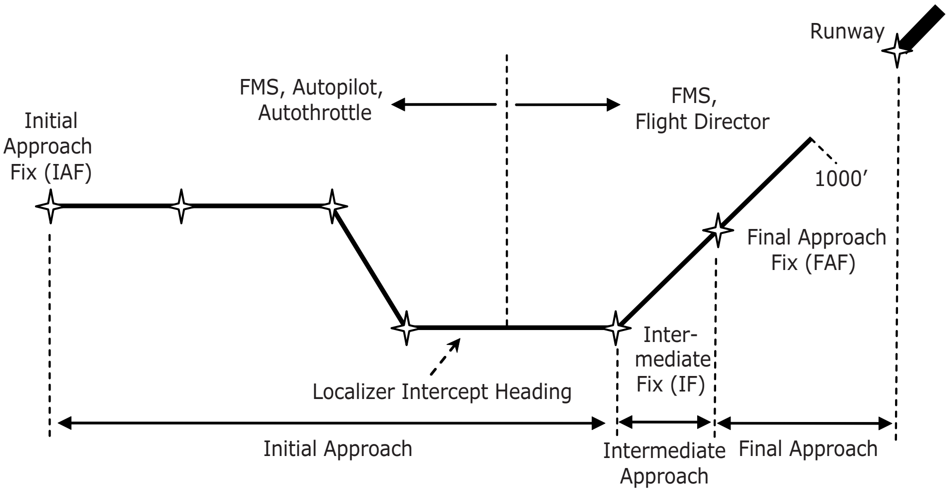


Figure 4.2 | Part of flight considered (top view) and automation used

The Intermediate Fix (IF) in Figure 4.2 is the first waypoint on runway heading, this is the point in the approach where the aircraft captures the Localizer. The Final Approach Fix (FAF) is the waypoint where the aircraft captures the glideslope in case the intermediate approach, and possibly (part of) the initial approach, is flown horizontally. It should be noted, however, that within this research trajectories are also considered that are continuously descending throughout the initial and intermediate approach. These trajectories are more favorable because of the reduction in noise and fuel that can be achieved. For these trajectories the glideslope is captured before the FAF is reached, sometimes even before the localizer is captured.

4.3 | Boundary conditions: Stabilized approach and Standard Operating Procedures

Pilot TDL also depends on the boundary conditions that are set, e.g., the accuracy with which the approach needs to be flown. The boundary conditions chosen for this research are that the approach should be performed according to Standard Operating Procedures and that pilots should aim to achieve a stabilized approach at 1,000 feet above airport elevation. This decision is based on the conclusions of the ALAR Task Force (FSF ALAR Task Force, 1999). The FSF ALAR Task Force concluded that: "Establishing and adhering to adequate standard operating procedures (SOPs) and flight-crew decision making processes improve approach-and-landing safety" (FSF ALAR Task Force, 1999). Additionally, they concluded that "Unstabilized and rushed approaches contribute to approach-and-landing accidents" (FSF ALAR Task Force, 1999). Therefore, in order to improve safety, an approach should be designed such that it can be flown according to SOPs and such that (while flying according to SOPs) a stabilized approach can be achieved. These are therefore the boundary conditions for this research.

To determine whether a stabilized approach is achieved at 1,000 feet, the following nine criteria are recommended by (FSF ALAR Task Force, 1999) and used within this research:
1. The aircraft is on the correct flight path;
2. Only small changes in heading/pitch are required to maintain the correct flight path;
3. The aircraft speed is not more than VREF + 20 knots Indicated Airspeed (IAS) and not less than VREF[1];
4. The aircraft is in the correct landing configuration;
5. Sink rate is no greater than 1,000 feet per minute; if an approach requires a sink rate greater than 1,000 feet per minute, a special briefing should be conducted;
6. Power setting is appropriate for the aircraft configuration and is not below the minimum power for approach as defined by the aircraft operating manual;
7. All briefings and checklists have been conducted;
8. Specific types of approaches are stabilized if they also fulfill the following: instrument landing system (ILS) approaches must be flown within one dot[2] of the glide slope and localizer; a Category II or Category III ILS approach must be flown within the expanded localizer band; during a circling approach, wings should be level on final when the aircraft reaches 300 feet above airport elevation; and
9. Unique approach procedures or abnormal conditions requiring a deviation from the above elements of a stabilized approach require a special briefing.

As an example for SOPs, the SOPs for the Delayed Flap Approach for the B747 are given in figure 4.3. The SOPs prescribe the following: Flaps 1 and flaps 5 should be selected before reaching (Localizer) Intercept Heading. On Localizer Intercept Heading, the pilot should select flaps 10, Arm the Approach, and switch to the Heading Select Mode. At Glide Slope Intercept, the pilot should ask for Gear down, flaps 20 and should arm the speedbrake. Finally, at 1,200 feet the pilot should select flaps LAND. An approach should be designed

1 The Reference Speed (VREF) is defined as 1.3 times the stall speed

2 One dot deviation on the glide slope equals 0.7^{o} beam error, one dot deviation on the Localizer equals 2.5^{o} beam error.

such that the pilots can perform these actions at the prescribed locations, and that they have sufficient time to do so.

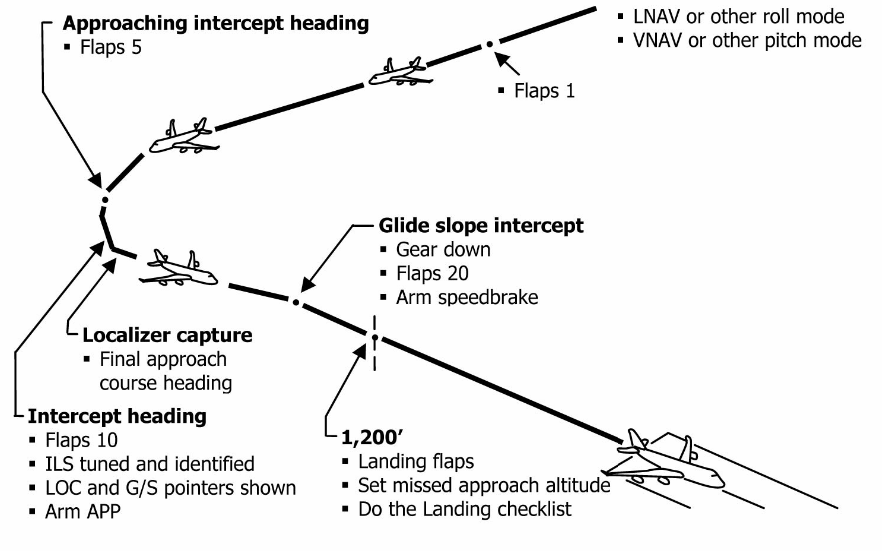


Figure 4.3 | SOPs for B747 Delayed Flap Approach.

4.4 | Factors of the air transport system included

Except for the type of approach and the boundary conditions that are set, many other factors and the interactions between those factors have an influence on pilot TDL during an approach. Figure 4.4 shows the factors that influence the safety of an approach (see journal article 1 for a complete explanation). To determine pilot TDL, this research will only take into account those factors that are labeled "direct influence on airport approach", most importantly the characteristics of the trajectory, the type of aircraft and the meteorological conditions. All factors are briefly explained below.

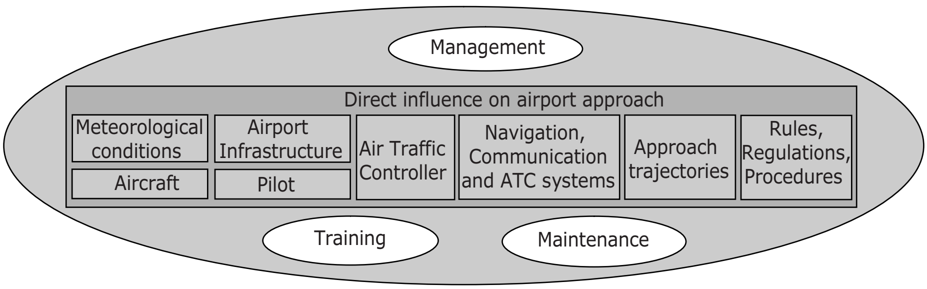


Figure 4.4 | Direct and indirect factors that influence the safety of airport approaches.

Meteorological conditions – The only effects on pilot TDL considered in this research are wind direction, windspeed and turbulence level.

Aircraft – Ultimately it should be possible to analyse approaches with respect to pilot TDL for all types of aircraft. But for this research only the Boeing 747-100 and the Cessna Citation are considered. The mass can be varied for both aircraft.

Airport infrastructure – This is not a variable quantity within this research, it is assumed that the runway is sufficiently long and that ILS equipment is available. Other airport infrastructure related factors, such as taxiways, are not important for this research since the analysis of the approach stops at 1,000’ altitude.

Air traffic controller – Since RNAV approaches are considered only, the influence of the air traffic controller on pilot TDL will be limited. The air traffic controller is assumed to speak audible English, and there is no other traffic. For the first part of this research it is assumed that once cleared for an approach the aircraft will indeed be allowed to actually fly that approach without interference of ATC (except, of course, for the necessary clearances). In the second part of this research ATC is assumed to interfere, and is assumed to re-clear the aircraft for another approach or to vector the aircraft in order to achieve the required separation between aircraft, according to the three different options in figure 3.1. The pilots’ preference for one of these options with respect to pilot TDL is also investigated.

Navigation, communication and ATC systems – These factors are not variable within this research and their effect on pilot TDL is not assessed. Navigation, communication and ATC systems are assumed to work flawlessly, supporting the pilot in his/her task to fly the RNAV approach.

Approach trajectory – As explained before, RNAV approaches are considered, starting at the IAF and ending at 1,000’ altitude. The effect of the approach trajectory on pilot TDL is one of the main focus points. Variables of the approach trajectory important for this research are the location of the waypoints, as well as the altitude and velocity constraints at these waypoints. Together these variables completely define how much energy the aircraft needs to dissipate within a certain amount of time (or distance) and how much time pilots have available to perform their actions according to SOPs.

Rules, regulations, procedures – As already explained in the previous section the boundary conditions are that an approach should be flown according to SOPs and that it should be possible to achieve a stabilized approach at 1,000’. The criteria for a stabilized approach are the same for all aircraft types, the SOPs, however, may differ for each aircraft type.

Summarizing, it can be stated that the approach trajectory (with airspeed and altitude constraints) is the main variable, or, so to say, the input within this research. This approach trajectory can be analyzed with respect to pilot TDL for different aircraft types, different aircraft masses and different wind and turbulence conditions. These factors can thus all be varied within this research.

By choosing a different aircraft type, the SOPs change accordingly. The navigation, communication and ATC systems might also vary with aircraft type, but for the two aircraft considered in this research this is not the case. By choosing a different approach trajectory, although the standard communication with ATC will still occur at the same characteristic points in the approach, the additional communication resulting from the three options to include flexibility (Figure 3.1) can differ. These factors (SOPs, navigation systems and communication with ATC) thus vary due to a variation in other factors.

Then there are the factors that are actually kept constant: the airport infrastructure in terms of ILS equipment and runway length, the standard communication with ATC, other meteorological conditions than wind and turbulence conditions and the fact that there is no other traffic.

By making the above decisions, some factors in Figure 4.4 labelled "direct influence on airport approach" that might actually influence pilot TDL are not considered in this research. These factors are: the presence of other traffic, difficult meteorological conditions such as mist, air traffic controllers that are badly audible and minor flaws in supporting systems.

4.5 | Non-nominal conditions and emergencies

Non-nominal conditions and emergencies such as engine failure are not considered in this research. The goal is to determine pilot TDL for published RNAV approaches under nominal conditions. When any emergencies such as engine failure occur, the crew will most likely not be required to follow the RNAV approach anyway, but will be vectored to the runway in the most convenient way.

Additionally, the assumption for less severe non-nominal situations is that when flying under nominal conditions, the RNAV approach should provide enough 'margin' with respect to pilot TDL, such that the pilot has enough spare capacity and time to deal with non-nominal conditions. This implies that the TDL that will be predicted by this research for a certain approach is valid for nominal conditions and should be well below the absolute maximum TDL a pilot can cope with in order to guarantee this margin.

4.6 | Computer simulation

As stated in goal #3, the method to predict pilot TDL needs to be captured within a computer simulation. This computer simulation will simulate a pilot flying an RNAV approach, under the conditions mentioned in the previous sections. The basic principles and assumptions made for this computer simulation are described in this section.

4.6.1 | Input

The only necessary information to give an indication of pilot TDL should be a definition of the approach trajectory in terms of a list of waypoints, defined by their lat-lon coordinates, and the altitude and speed constraints at these waypoints. Based on this information it should be possible to give an indication of pilot TDL when flying this approach.

4.6.2 | Level of detail of computer simulation models

It is the goal to incorporate very detailed models of the *environment* of the pilot in the computer simulation, and to add to this a rather simple model of the pilot (the reason for this choice will be explained extensively in the next chapter). Therefore, the aircraft with its kinematic and dynamic constraints, the 3-D properties of the trajectory, the velocity profile, turbulence, wind, etcetera, in other words: the factors that have a direct influence on an approach as given in Figure 1, are modeled as detailed and accurate as possible. Whereas the pilot model is kept as simple as possible.

4.6.3 | Pilot model

The relatively simple pilot model will consist of three sub-models: 1. a manual control model to simulate the pilot's control actions for elevator and ailerons, 2. a visual scanning model to replicate the pilot's sampling of the flight instruments and 3. a model for performing actions such as selecting flaps and gear according to the Standard Operating Procedures (SOPs) for a specific type of aircraft. It is noted that subjective factors such as fatigue, memory capacity, etc., are not incorporated in the pilot model.

4.6.4 | Output

The computer simulation will give a prediction for those factors that will be shown to contribute to pilot TDL. Additional to this, or actually as will become clear later on: in support of this, the computer program will predict whether the constraints at the waypoints can be met and whether a stabilized approach can be achieved at 1,000', and if so, under which wind conditions this is still possible. Linked to this, the computer simulation will also give a realistic prediction of the motion of the aircraft in the longitudinal plane. This together would provide an answer to goal #3 as defined in chapter 2.

Next to aiding the design of approaches, the computer simulation will provide approach designers with an indication of the 'busy' parts of an approach thereby improving mutual understanding between ATC specialists and flight crews of each other's operational environment. According to the conclusions of the ALAR Task Force (FSF ALAR Task Force, 1999) this will also improve safety.

The basic principles of the method and computer simulation to predict pilot TDL thus are: to use a detailed model for the environment of the pilot (such as the aircraft, trajectory, etc.) and to add to this a fairly simple model for the pilot. Pilot TDL will be established for nominal conditions only, while flying according to SOPs and aiming to achieve a stabilized approach at 1,000'. RNAV approaches are considered which are assumed to be flown with a high level of automation.

5 | Why use a simple pilot model combined with a highly detailed model for the aircraft?

In order to develop a method with which it is possible to predict pilot TDL during a certain RNAV approach, some sort of model will be needed which is able to predict the pilot's actions. Based on the predictions of such a model it should be possible to derive an indication of pilot TDL. In the previous chapter it was stated that a detailed model of the pilot's environment would be used combined with a relatively simple pilot model. The reasons for this choice are clarified in this chapter, after having explained the existing pilot models.

5.1 | Existing pilot models

Pilot models, which aim to analyse and/or predict the pilot's tasks during flight already exist. Some of these models specifically focus on the pilot, some models intend to comprise (parts of) the total air transport system and as a result the pilot model is then only a smaller part of a larger model. Pilot models have been used for conceptual design purposes, safety studies and risk assessment methodologies. A brief overview of two of these pilot models is given below (for a more extensive overview see the introduction in Journal Article 1).

The first pilot model is a model developed in the late 1970s which is a model for analyzing flight crew procedures in approach to landing: the procedure-oriented crew model, PROCRU for short (Baron et al., 1980). Based on concepts of optimal control and estimation theory, it provided results in terms of, among others, each crew member's estimate of the aircraft state, their attention allocation, and their control actions, all as a function of time. PROCRU was never validated experimentally, however. And because much experience was required in dealing with the many control- and estimation-theoretical aspects of PROCRU in order to use it confidently, it was rarely ever used (Baron, 1990).

The second pilot model, MIDAS, the Man-Machine Integrated Design and Analysis System, was developed by the US Army, NASA, and Sterling Software (Corker, 1999, Gore and Corker, 2000, Smith and Tyler, 1997). MIDAS is a simulation system that includes human performance models that can be used to evaluate candidate crew procedures, controls, and displays. It is used in the early stages of conceptual (cockpit) design, and allows designers to use computational representations of the crew station and operator, instead of conducting hardware simulations and man-in-the-loop studies. MIDAS allows an analysis of the crew station layout for assessments of visibility, legibility, anthropometric aspects, and analyses of cockpit topology and configuration. It also provides facilities to run human operators models, cockpit equipment and mission procedures in an integrated fashion, resulting in activity traces, task-load timelines, information requirements, and mission performance measurements (Gore and Corker, 2000).

Common to all these methods and the human operator models incorporated in them, however, is that they have become quite complex, and require considerable experience to apply them properly. The models often contain many "parameters" that need to be "tuned", and because most of them are not validated experimentally, it is sometimes difficult to initialize the model, as typical values for these parameters are unknown. Furthermore, in attempts to increase their applicability and versatility, the models often include components that account for particular aspects of human behavior, sub-models that have often been developed and validated, however, in a different context. As a consequence, many of these methods and models often do not allow useful predictions to be made about situations that are beyond the conditions for which the models have been tuned. Finally, some of the methods are composed of proprietary parts of software, sometimes developed for military applications, and therefore (parts of them) are simply not available for others, or service in using the methods and models is not available.

5.2 | Pilot model used for this research

For these reasons, a new pilot model will be developed for this research. The intention is to keep the pilot model as simple as possible: the aim is not to replicate the exact actions of the pilot as they are performed in real flight, but only to obtain an "on average" indication of "how hard" pilots have to work. This is a deliberate deviation from previous human performance modeling approaches, such as PROCRU, where one attempts to construct a detailed model of the human operators involved, including all their limitations, both physically and mentally. Rather, a simpler model is developed and used to understand how the environmental constraints (for instance, the aircraft kinematic constraints, the approach trajectory, ….) affect the difficulty a pilot experiences during an approach. In this respect, the approach follows the principles of cognitive work analysis (Vicente, 1999), where emphasis is shifted from detailed investigations of the human operator limitations (like memory capacity, time delay, etc.) to analyzing and describing how the operator's *environment* affects human behavior.

The reasoning behind this shift in emphasis is that it are the constraints in the *environment* that 'shape' the behavior of the human working in that environment. When one is investigating this behavior, and putting this behavior in a mathematical model, which is the classical human performance modeling approach as mentioned above, one is in fact modeling the 'consequences' of the environmental constraints on the ever-adapting human operator. The environmental factors considered in this research are all factors labeled as 'direct influence on aircraft approach' in Figure 4.4.

These are the reasons to use a simple pilot model and a detailed model of the environment within this research as already indicated in Section 4.8.

5.3 | Why use a pilot model at all?

A valid question would be why a pilot model is used at all within this research. Since the focus is on pilot TDL and on how the environment affects the pilot's behavior, would it not be possible to completely eliminate the model for the pilot from the computer simulation models? The answer is no. In order to simulate an approach, the pilots' actions need to be modeled, since without those actions, the approach cannot be flown, and thus not simulated. The question is in how much detail these pilot actions need to be modeled. As explained before, the goal within this research is to keep these pilot models as simple as possible. The following chapters will show that, as this research progressed, the pilot models became more and more simplified.

COMPUTER MODELS AND EXPERIMENTS

Outline of computer models and experiments

As explained in chapter 2 the goal of this research is three-fold:

- ❖ Goal 1: the factors relevant for approach design, which influence the difficulty experienced by pilots while flying an approach, need to be identified.
- ❖ Goal 2: a method needs to be developed with which it is possible to predict how difficult it will be for a pilot to fly a certain approach.
- ❖ Goal 3: this method needs to be captured within a computer simulation that can be used in the early stages of approach design. With this tool, the designer should be able to rapidly evaluate a potential approach from the perspective of the demands it imposes on the pilot's limited resources. Additionally, this computer simulation should also be able to give a realistic prediction of the motion of the aircraft in the longitudinal plane.

The following chapters in this dissertation aim to describe how the answers to these three goals were found, based on the assumptions and basic principles as explained in chapter 4.

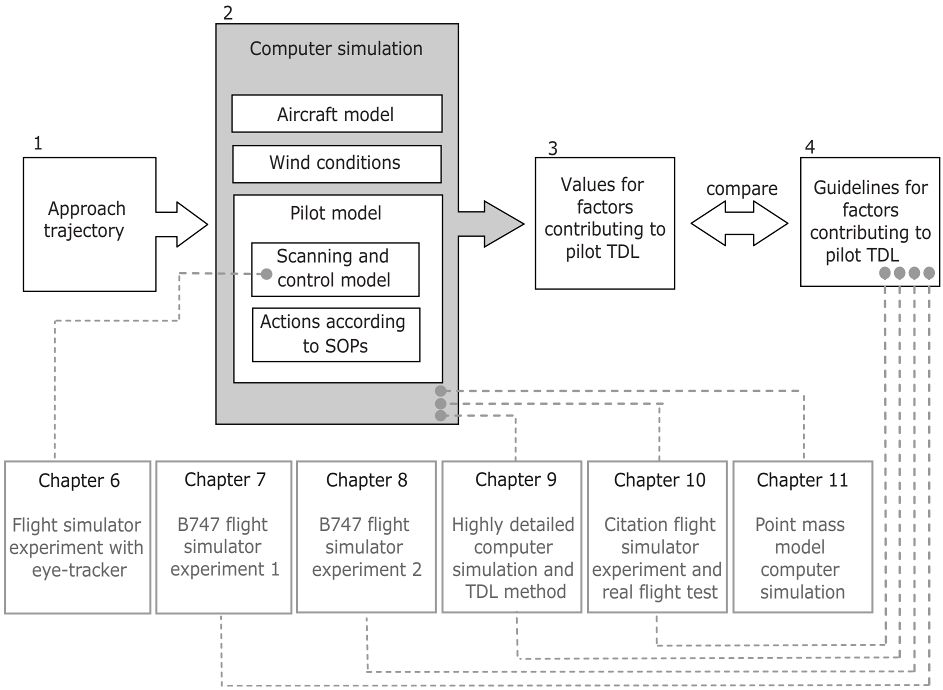


Figure 1 | Systematic overview of research

Figure 1 shows a schematic overview of the research. On the one hand, the factors that contribute to pilot TDL need to be identified, and guidelines (for example maximum values) for these factors need to be determined (block 4). On the other hand, a computer simulation (block 2) needs to be developed which can predict, for a given approach, the values for the factors that contribute to pilot TDL (block 3). For example: say, the time available in part X of the approach is found to be a factor that contributes to pilot TDL. One of the guidelines will then read that in order to keep pilot TDL at an acceptable level there should be sufficient time available in part X to perform all required actions. The computer simulation then predicts whether all actions in part X can indeed be performed within the time available. If yes, then pilot TDL for this factor is acceptable, if no, then pilot TDL for this factor is too high. The method to predict pilot TDL for an approach will be shown to consist of seven such guidelines, if all guidelines are adhered to, pilot TDL is considered to be acceptable.

The input for the computer simulation, as explained in chapter 4, is the approach trajectory (block 1 in Figure 1). The output consists of the values for factors contributing to pilot TDL. The computer simulation itself consists of an aircraft model, the wind conditions and a pilot model. The pilot model in its turn consists of a scanning and control model, as well as a model which simulates the pilots' actions according to the SOPs, see Figure 1.

The following chapters aim to describe the components given in Figure 1. The structure is as follows: Chapter 6 starts with a description of theoretical scanning and control models for the pilot which are then validated by two flight simulator experiments with eye-track equipment. The conclusion of this chapter, however, is that the scanning of the instruments and the manual (continuous) control of the aircraft do not contribute significantly to pilot TDL. Therefore, the scanning model is removed from the computer simulation, and the manual control model is kept as simple as possible.

Chapter 7 describes a flight simulator experiment which was used to obtain a first indication of the factors that contribute to pilot TDL for a B747. A rather extensive list of factors is obtained from this experiment. Chapter 8 presents a second flight simulator experiment for the B747, which reduced this list of factors to the factors that were actually shown to contribute to pilot TDL. Goal 1 is then achieved for the B747.

Chapter 9 presents the method to predict pilot TDL (goal 2). The method consists of seven guidelines for the design of approaches, based on the factors that were found to contribute to pilot TDL in chapters 7 and 8. When all guidelines are adhered to, pilot TDL will be acceptable for an approach. Chapter 9 also gives the description of the highly detailed computer simulation for the B747, and presents it output (goal 3).

Chapter 10 then considers a different aircraft type (the Cessna Citation) to check whether the factors and guidelines that were found for the B747 are also valid for another aircraft. It will be shown that this indeed is the case.

Chapter 11 explores the possibility to replace the highly detailed computer simulation by a much simpler computer simulation based on a point mass model representation of the aircraft. It will be shown that this simpler computer simulation can indeed provide the same output, but in only a fraction of the computation time. This simpler computer simulation can actually be used during the design of approaches (the final step towards goal 3).

Finally, the conclusions are presented in Chapter 12, and the recommendations in Chapter 13.

6 | A simple pilot model consisting of a scanning model and a manual control model

As a first step towards developing a simple pilot model that can be used in the computer simulation to predict pilot TDL, a pilot model is considered that consists of a scanning model and a manual control model. The scanning model represents the pilot's scanning (or sampling) of the flight instruments and/or outside world, whereas the manual control model replicates the pilot's elevator and aileron control actions. It was hypothesized that the rate at which the flight instruments need to be scanned and the amount of manual control actions that are necessary in order to perform the task adequately could serve as (part of) an indication of pilot TDL.

6.1 | Literature

It is, as explained before, the intention to keep the computer models as simple as possible and instead of using proposals for highly detailed, complex scanning models from literature (Senders (1964) or Carbonell, 1966 and Carbonell, Ward & Senders, 1968), the choice was made to model an "average" scanning sequence that might not replicate reality for the full 100%, but can give a fair indication of the amount of scanning the pilot has to perform. The same approach is chosen for the pilot control model: it was decided to get a first indication of how hard the pilot has to work, rather than to mimick the pilot's actual behavior as closely as possible. And, therefore, instead of using the crossover model (McRuer and Jex, 1967) or Optimal Control model (Baron, Kleinman and Levison, 1970 and Kleinman, Baron and Levison, 1970) a simple control model was chosen based on a simplified version of an altitude-controller in which the gains depend on the rate at which the instruments are scanned by the pilot. See sections 2.1 and 2.3 of Journal Article 1 for a more extensive overview of literature.

6.2 | A scanning model and manual control model for horizontal flight

As a first test case, to test whether such a simple computer model can yield predictions with respect to pilot TDL, a horizontal flight through turbulent air was considered. The computer model is used to predict how fast the pilot needs to scan the instruments and how much corrective control actions are required in order to stay within 50ft deviation from the required altitude, this is predicted as a function of turbulence intensity.

The average scanning sequence that was used in the computer simulation for horizontal flight is given in Figure 6.1 (based on [KLM Luchtvaartschool, 1990]). The predictions of the computer simulation indicate that, in order to remain within a predetermined 50ft deviation from the required altitude, the time pilots can take for one scanning cycle needs to decrease with increasing turbulence level (see Figure 6.2). Or, in other words: pilots have to increase their scanning rate with increasing turbulence level in order to stay within 50ft altitude deviation. However, at a certain turbulence level (let's call this the 'transition turbulence level') they are predicted to reach their maximum scanning rate, and as a result for higher turbulence levels the time for one scanning cycle is predicted to remain constant. This is caused by the fact that pilots need a minimum amount of time to complete one scanning cycle.

As a consequence, the task error, which is the deviation from the desired altitude, will remain within 50ft as long as the pilot can increase his/her scanning rate, and will increase for turbulence levels higher than the transition turbulence level, see Figure 6.3. The same is true for the control workload (expressed as the root mean square of the power spectral density of the control column deflection (Padfield (1996)), see Figure 6.4. These predictions were validated by a small experiment with four pilots in a Frasca 121 fixed-base flight simulator (a single engine light aircraft with fixed pitch propeller, cruise speed 120 knots). The experiment showed that the trends predicted in Figures 6.2 to 6.4 could indeed be observed during the simulated flights (see section 4.2 of Journal article 1).

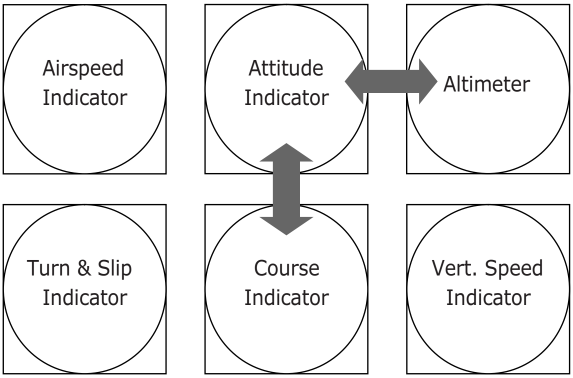


Figure 6.1 | Average scanning cycle during horizontal flight. One scanning cycle exists of attitude indicator – course indicator – attitude indicator – altimeter

A relation is thus found between the pilot's scanning rate, the control activity and the task error. A measure for pilot TDL could be derived by using the inverse of the spare capacity of the time to perform one average scanning cycle in the region to the left of the transition turbulence level, and the task error in the region to the right of the transition turbulence level. For a more elaborate explanation of this part of the research the reader is referred to Journal Article 1.

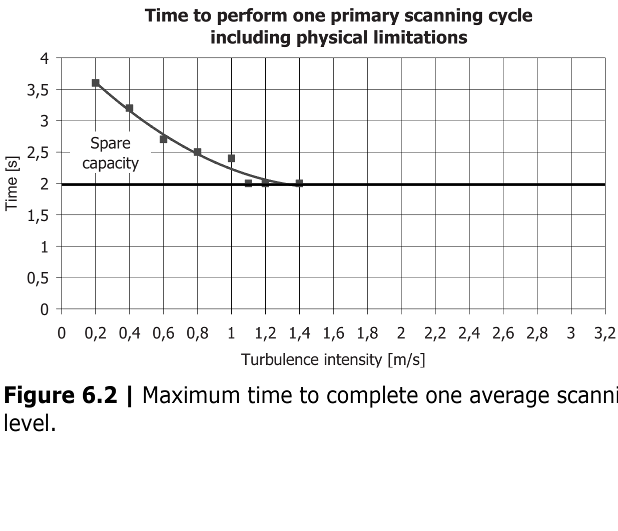


Figure 6.2 | Maximum time to complete one average scanning cycle as a function of turbulence level.

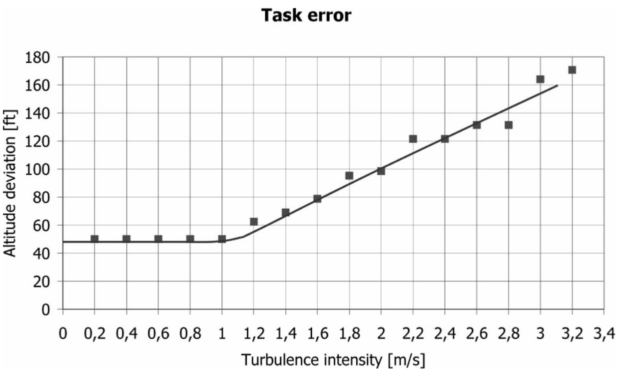


Figure 6.3 | Task error, expressed in altitude deviation, as a function of turbulence level.

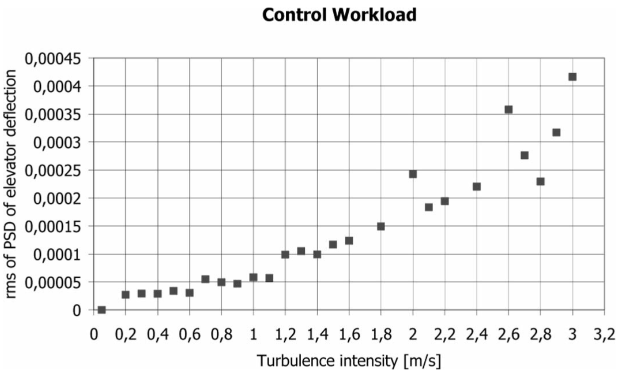


Figure 6.4 | Control workload as a function of turbulence level.

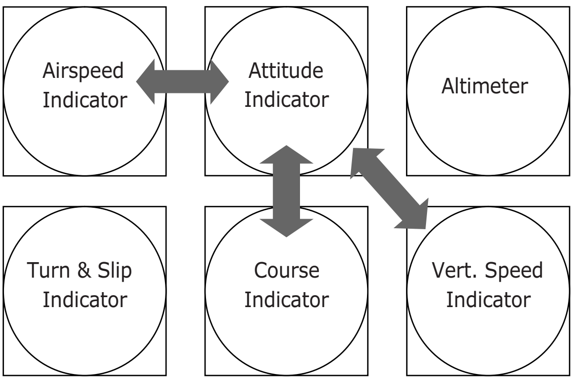


Figure 6.5 | First average scan cycle during approach phase (for descent): Airspeed Indicator → Attitude Indicator → Course Indicator → Attitude Indicator → Vertical Speed Indicator → Attitude Indicator → Airspeed Indicator → Attitude Indicator → etc.

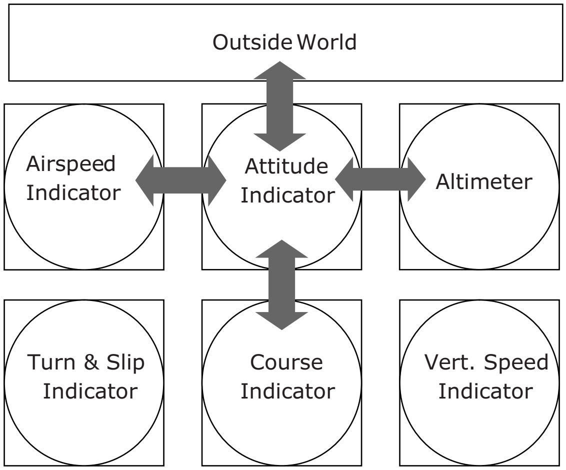


Figure 6.6 | Second average scan cycle during approach phase: Attitude Indicator → Outside World → Attitude Indicator → Airspeed Indicator → Attitude Indicator → Outside World → Attitude Indicator → Altimeter → Attitude Indicator v Outside World → Attitude Indicator → Course Indicator → Attitude Indicator → Outside World → etc.

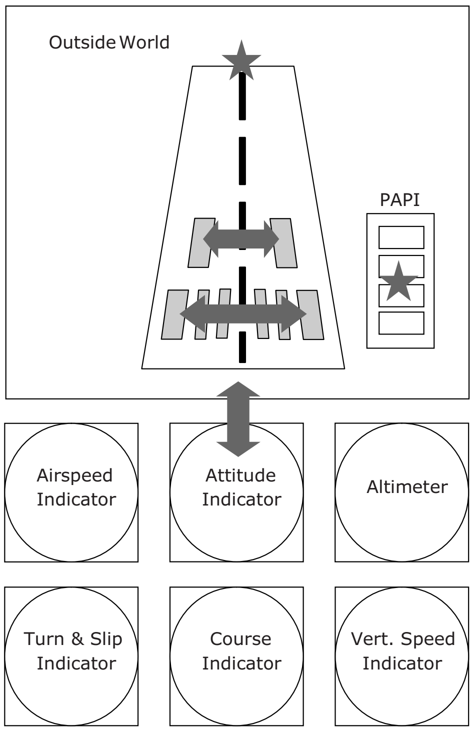


Figure 6.7 | Third average scan cycle during approach phase: Horizon/end of runway → Landing blocks → Horizon/end of runway → Threshold → Horizon/end of runway → PAPI → Horizon/end of runway → Course Indicator (Localizer indication) → Horizon/end of runway → Landing blocks → etc.

6.3 | Scanning models and control model for approach phase

Encouraged by the flight simulator test results, the next step was to expand the scanning and control models such that they could also represent pilot behavior during the approach phase of the flight. To this end, a method to predict pilot control activity during approach was developed (Blok, 2005). Flight simulator tests to validate the predictions of this model were performed in the six-degree-of-freedom TU Delft SIMONA Research simulator (SRS), using a Cessna Citation aircraft model, with five pilots participating in the experiment. The pilots were instructed to fly a manual approach, while they were vectored by ATC towards the Localizer. The final part of the approach was flown using the ILS. Additionally, average scanning cycles for the approach phase were derived from these flight simulator tests (Nijsten, 2005). These average scanning cycles were envisaged to be used in the computer simulation to predict pilot TDL. Note that the approaches considered here are not RNAV approaches as is the case in the remainder of this research.

It was found that during the approach (starting at the IAF) pilots subsequently used three different scanning cycles: the first scanning cycle only incorporated the flight instruments, in the second scanning cycle, which is started after the call 'runway in sight', the outside view was added to the scanning cycle, and during the third scanning cycle almost all attention went to the outside world with a sporadic scan on the course indicator (Nijsten, 2005). The three scanning cycles are depicted in Figures 6.5 to 6.7, and together with the values for the minimum dwell times for all instruments (i.e., the time a pilot looks at an instrument) given in Table 1, these scanning cycles can be used in the computer simulation.

Table 6.1 | Minimum dwell times in seconds for each instrument during the approach phase for three different scanning cycles.

	Minimum dwell time during scanning phase		
Area of Interest	**First**	**Second**	**Third**
Attitude Indicator	1.2	1.2	/
Airspeed Indicator, Altimeter, Vertical Speed Indicator	0.6	0.4	/
Course Indicator	0.7	0.4	0.4*
Outside World	/	1.0	4.0

* Not calculated due to lack of data, but this value can be assumed in a model.

With respect to the control workload during the approach phase (Blok, 2005) it was found that the control activity increased with a factor 1.2 to 1.9 during the 10 seconds following flap deployment and gear selection, and that the control activity for the ailerons increases 1-2 times faster with turbulence level than the control activity for the elevator. It is important to note that the entire flight was flown manually.

6.4 | Conclusions

Although these results are interesting and can be used for a simulation of the approach phase with the computer simulation, it became clear from the flight simulator experiment and conversations with pilots that it is not the scanning of the instruments nor the manual (continuous) control of the aircraft that constitute the largest contribution to pilot TDL. This is especially true when considering modern approaches that now are largely flown automatically using the FMS with autopilots and autothrottle, as is the purpose within this research. Pilots seemed to be much more preoccupied with more ‘discrete’ tasks such as selecting flaps and gear down, meeting the constraints at the waypoints, performing checklists etc. than with the ‘continuous’ scanning and controlling tasks. Additionally, pilots seemed to be much more concerned with ‘higher level’ or, so to say, larger time-scale properties of the approach trajectory, such as for example the amount of trackmiles available between important waypoints in the approach, or the amount of time available to perform the required actions, instead of with small time-scale properties such as additional control actions necessary after flap deployment, or additional control actions due to turbulence.

In this respect these observations correspond with Rasmussen’s skill- rule- and knowledge-based performance model (Rasmussen, 1983), which states that rule based actions (such as selecting flaps, gear down, etc.) require more attention from a human operator than skill based actions (such as scanning the flight instruments and controlling the aircraft with elevator and aileron). Additionally, a high level of automation (as assumed and applied in this research) results in less skill based tasks to be performed and moves the operator towards higher levels of control.

It was therefore decided to no longer pursue the development of scanning models nor to quantify the contribution of scanning or manual control to pilot TDL. Accordingly, the decision was made to no longer include a scanning model in the computer simulation. Instead, the focus was fully shifted towards analyzing and modeling the ‘discrete’ actions (such as selecting flaps and gear, meeting the constraints at the waypoints, etc.) and the larger time scale properties of the approach trajectory (such as the amount of trackmiles available between important waypoints in the approach) and their effect on pilot TDL.

7 | A flight simulator experiment to obtain a first indication of factors influencing pilot TDL

Having concluded that the scanning of the instruments and the pilot manual control of the aircraft do not contribute significantly to pilot TDL, the focus from then on shifted towards the influence on pilot TDL of factors related to the 'discrete' actions (such as selecting flaps and gear) and the larger time scale properties of the approach trajectory.

7.1 | Independent variables

Based on literature (ICAO, 2006, Godley, 2006, Vormer, 2005) and knowledge gained from conversations with pilots as well as by examining the SOPs, the following set of factors is composed as an initial subset of elements that might influence pilot TDL during an approach:

- The number of heading changes in an approach;
- Incorporating many altitude steps in an approach compared to a Continuous Descent Approach (CDA);
- The value of the energy rate demand (when the energy rate demand becomes larger than 1, the velocity and altitude constraints at the waypoints can no longer be met);
- Applying a horizontal approach instead of a CDA;
- The distance available on Localizer Intercept Heading;
- The Localizer intercept speed;
- The aircraft mass;
- The line-up distance (distance between IF and runway); and
- The heading change when turning towards Localizer Intercept Heading.

These factors were tested as independent variables[3] during a B747 flight simulator experiment in the six-degree-of-freedom SIMONA research simulator. Additional to testing these independent variables, many other factors, such as the Final Approach Fix (FAF, see figure 4.2) altitude or the Localizer intercept angle, were changed as well during the experiment to see whether pilots would comment on these aspects. The goal of the experiment was to find some definitive answers for the independent variables: whether they affected pilot TDL or not, and simultaneously to obtain an as complete as possible list with factors that might influence pilot TDL by changing many other factors as well during the experiment and analyzing the pilots comments regarding these factors.

7.2 | Experiment set-up

Twenty different approaches were designed, using a preliminary version of the computer simulation that will be presented in chapter 9. Nine professional B747 pilots participated in the experiment. The pilots were asked to adhere very strictly to SOPs (see Figure 4.3), even if they could foresee that by adhering to SOPs they would not meet certain constraints at waypoints or would end up unstabilized at 1,000'. Additionally, they were asked to perform their tasks according to the principles of Multiple Crew Coordination (MCC) and to fly passenger comfort.

---

3 An independent variable is another name for a predictor variable. It is the variable that is manipulated by the experimenter and so its value does not depend on any other variables (Field, 2005).

7.3 | Results

The results of this flight simulator experiment are given in Table 7.1: a list consisting of 23 factors was obtained, of which 9 were tested as independent variables. For some of these factors it could be established that these factors indeed affected pilot TDL, these factors are indicated by a black dot in Table 7.1. For some factors it could be established that they did not influence pilot TDL, these factors are denoted by an 'x' in Table 7.1. Then there were factors that did influence pilot TDL, but these factors were already covered by another factor, for instance, the time available between the FAF and 1,000' (factor 13) is totally determined by the FAF altitude (factor 12) when ignoring the relatively small effect of the airspeed. These factors are indicated by an '-'. Finally, the fourth group of factors, denoted by an 'o' in Table 7.1, are factors that might possibly have an influence on pilot TDL. These factors were not tested as independent variables, but were commented on by the pilots in the questionnaires, and there were no further data to either support or oppose their effect on pilot TDL.

Table 7.1 shows that some of the factors were also linked to another factor when they were tested. For example: when testing the effect of the distance available on Localizer intercept heading (factor 8) this was tested as an independent variable between two different approaches, but for these same two approaches there was also a difference in the groundspeed during Localizer capture. The effect on TDL could thus be caused by either the difference in distance on Localizer intercept heading, or by the difference in groundspeed.

The notion is that the list of factors in Table 7.1 in itself is fairly complete (given the assumptions in Chapter 4). In the questionnaires and during the experiment pilots did not mention any other factors that are relevant for this research.

A fairly complete list of factors influencing pilot TDL during RNAV approaches has thus been obtained from this flight simulator experiment. From this experiment it could for some factors be established that these indeed *did* or *did not* influence pilot TDL. For some factors it could only be concluded that they *might* influence pilot TDL. An additional flight simulator test needs to be performed to be able to arrive at definitive conclusions regarding the influence on pilot TDL of this second group of factors.

For a more elaborate description of this flight simulator experiment and the results the reader is referred to Journal Article 3.

Table 7.1 | Overview of factors that were found to affect pilot TDL during the flight simulator experiment.

#	Factor	Linked factor	Effect on TDL
1	Number of heading changes	Number of waypoints	x
2	CDA compared to horizontal		x
3	Heading change towards LOC Intercept Heading		x
4	Energy rate demand too high	Localizer Groundspeed	•
5	LOC intercept speed (IAS)/LOC Groundspeed	Energy rate demand IF-FAF	•
6	Mass		•
7	More altitude steps compared to CDA		x
8	Distance available on LOC Intercept Heading	Localizer Groundspeed	•
9	Time available for actions on LOC intercept leg	Localizer Groundspeed	-(1)
10	Line-up distance	IF-FAF distance	•
11	Time available for actions on final	IF-FAF distance	-(2)
12	FAF altitude		•
13	Time available between FAF and 1,000’		-(3)
14	LOC intercept angle		o
15	Tailwind		- (4)
16	Vertical speed on final		- (5)
17	Turbulence		o
18	Trackmiles		o
19	More altitude steps compared to horizontal		o
20	Airspeed on final		o
21	Time available during first part of approach		o
22	Increase in time spent manoevring		- (6)
23	Stabilized at 1,000’		o

• = proven effect on TDL, x = proven that no effect on TDL, - = effect on TDL already covered by another factor, o = factor that could have an influence on pilot TDL.

(1) directly related to distance available on LOC intercept heading

(2) directly related to line-up distance

(3) directly related to FAF altitude.

(4) already incorporated in energy rate demand

(5) directly related to airspeed on final

(6) directly related to number of heading changes, more altitude steps compared to CDA and more altitude steps compared to horizontal

8 | A second simulator experiment to obtain the full set of factors influencing pilot TDL

The purpose of the flight simulator experiment presented in this chapter is to get a more comprehensive overview of the factors that influence pilot TDL or, in other words, to 'fill in' the gaps in Table 7.1 as much as possible.

8.1 | Independent variables

The factors that were linked to another factor and proved to have an effect on pilot TDL are tested as separate independent variables in this flight simulator experiment, making sure that these are not linked to any other factor again. The following independent variables thus arise:

1. Energy rate demand too high in the first part of the approach;
2. Energy rate demand too high in the Localizer intercept heading part of the approach;
3. Energy rate demand too high in the final part of the approach;
4. Localizer intercept speed;
5. Time available for actions on Localizer intercept heading;
6. Line-up distance; and
7. Distance between Intermediate Fix (IF) (the first waypoint on runway heading, see Figure 1) and FAF.

Some of the definitions in the list above require explanation. In the previous flight simulator experiment it was found that, when asked to give an opinion about the TDL during the approach, pilots tended to split the approach in multiple parts, giving a separate description for pilot TDL for each of the parts. The same division is adopted for this flight simulator experiment resulting in three parts (see Figure 8.1): 1. the first part of the approach, 2. the Localizer intercept part of the approach, and 3. the final part of the approach. It is hypothesized that the effect of the energy rate demand will be different for the three approach parts, hence a split up of the factor 'energy rate demand too high' for the three approach parts.

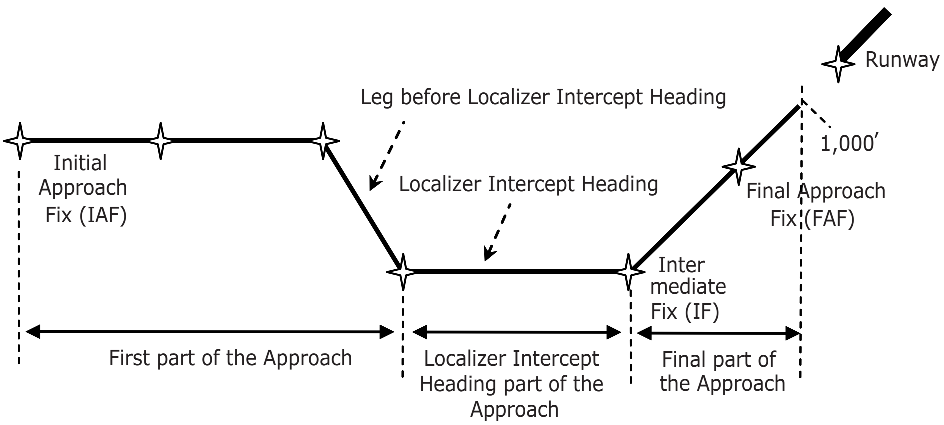


Figure 8.1 | Division of approach in three parts.

Now, returning to the list of independent variables for this flight simulator experiment: the factors that proved not to have an effect on pilot TDL in Table 7.1, or the factors that are already represented by another factor are obviously not tested again as independent measures in this flight simulator experiment. This leaves factors 14, 17 to 21, and 23 in Table 7.1 as candidates for independent variables. Due to the limited amount of approaches that can be tested during the current flight simulator experiment, a choice was made to also include the following factors as independent measures:

8. Localizer intercept angle; and
9. Airspeed on final.

This implies that, amongst others, the following factors in Table 7.1 are not tested as independent measures: amount of trackmiles, more altitude steps compared to horizontal, and time available during first part of approach. These three factors all relate to the first part of the approach, which is hypothesized to have the least influence on pilot TDL. Additionally, the factors CDA compared to horizontal, and more altitude steps compared to CDA did not influence pilot TDL, therefore the chance that more altitude steps compared to horizontal would influence pilot TDL is relatively small. Other factors from Table 1 not included as independent variables in this experiment are: aircraft mass, since the effect of aircraft mass is incorporated in the energy rate demand, stabilized at 1,000′, since this is a *result* of other factors and will therefore vary as a result of these other factors, and turbulence, since this only has an effect in the sense that it can become more difficult to read the flight instruments (which cannot be taken into account during the design of approaches). For a more elaborate explanation of the choice for these independent variables the reader is referred to Journal Article 4.

8.2 | Three options to include flexibility

Next to the factors in Table 7.1, an additional factor was introduced for this flight simulator experiment: the inclusion of a certain kind of flexibility in an approach in order to give ATC the possibility to correctly sequence the aircraft. Three possible options to include flexibility were considered, as already explained in Section 3.1, they are depicted again in Figure 8.2. For each of these options in Figure 8.2, the aircraft was always initially cleared for the longest route, route C in the examples, this was also the active route in the FMS. During the flight, the pilots could be instructed to follow a different route. For option 1, the pilots were re-cleared for a different, shorter, route before they reached the IAF. This other route (route A or B in Figure 3) was in all cases a published route and already pre-programmed in the FMS. The Pilot Monitoring (PM) had to select this other route in the FMS and make it the active route. The new route would be flown in VNAV and LNAV mode.

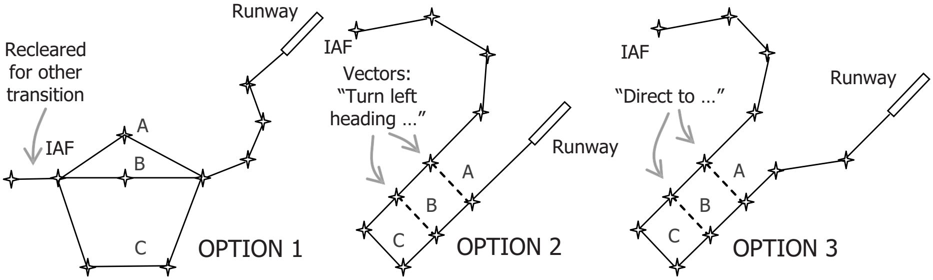


Figure 8.2 | Three options to include flexibility in approaches

For option 2, the crew was vectored from the downwind leg towards the final leg (via route A or B in Figure 3), or was allowed to fly according to the original route C. On the approach charts there was a note warning to 'expect vectors on final'. When the crew was told, for instance, to 'turn left heading 150, further constraints as published', they had to select the Heading Select mode, thereby deactivating the LNAV mode, and had to switch off the VNAV mode. As a result they could no longer use the route information from the FMS, since the active route in the FMS would still be the initial route C.

In option 3, the crew was told to fly "direct to" a specified waypoint. In all cases, this waypoint was already part of the active route in the FMS, the PM had to select the specified waypoint in the FMS, thereby changing the route in the FMS, and the new route would be flown in VNAV and LNAV mode.

The preference of pilots with respect to TDL regarding these three options is also determined during the current flight simulator experiment. Note that the flexibility is only applied in the approach part of the flight, for flexibility options during the arrival phase of the flight see, for instance, the sixth framework programs OPTIMAL (Ferro (2005), Verhoeven and de Gelder (2009)) or ERAT (de Jonge and Törner, 2009).

8.3 | Experiment set-up

The flight simulator experiment was performed in the six-degree-of-freedom Generic Research Aircraft Cockpit Environment (GRACE) flight simulator at the National Aerospace Laboratory NLR. Nine B747 pilots participated in the experiment. As with the previous experiment, the pilots were asked to adhere very strictly to SOPs, even if they could foresee that by adhering to SOPs they would not meet certain constraints at waypoints or would end up unstabilized at 1,000'. Additionally, they were asked to perform their tasks according to the principles of Multiple Crew Coordination (MCC) and to fly passenger comfort.

8.4 | Results

The majority of the pilots indicated that they, in general, preferred either Option 2 or Option 3 to include flexibility in the approach, whereas all pilots preferred Option 2 or 3 when asked to only consider the adjustments needed in the FMS.

The following could be concluded from the simulator experiment with respect to the factors that have an influence on pilot TDL, note that these conclusions are based on a small data set:

For the first part of the approach:

- The major contributor to pilot TDL seems to be the energy rate demand, or, the fact whether or not constraints can be met at the waypoints. This is only true when the effect of not meeting the constraints continues into the Localizer part or final part of the approach. If the consequences of the energy rate demand remain within the first part of the approach this does not influence pilot TDL.
- The number of waypoints, number of heading changes and the altitude profile (horizontal approach, CDA, stepped approach) do not seem to influence pilot TDL. This is due to the fact that this part of the approach is flown in LNAV and VNAV modes with autopilot and autothrottle.
- The time available to perform all actions during this part of the approach is important when flexibility (see Figure 8.2) is introduced, pilots should then have sufficient time to make all necessary adjustments.

For the Localizer intercept part of the approach:

- The time available to perform all actions (which is directly related to the distance available on Localizer intercept heading) seems to be the most important factor for pilot TDL.
- Next to this, pilot TDL is also influenced by the Localizer intercept speed, the Localizer intercept angle, and whether the constraints at the waypoints can be met (the energy rate demand).

For the final part of the approach:

- The most important factors influencing pilot TDL seem to be whether or not a stabilized approach can be achieved at 1,000’, the distance between IF and FAF and the airspeed on final. Whether an approach is stabilized can, for a B747, be determined from: 1. the energy rate demand during the final part of the approach, 2. the value of the vertical speed (should be below the sink rate warning) which in itself is a function of airspeed on final and glideslope angle, and 3. the FAF altitude and distance between IF and FAF (resulting in the line-up distance) since these two factors together determine whether there is enough time available to perform all actions required for a stabilized approach. All these factors thus influence pilot TDL during the final part of the approach.

A more complete list of factors that have been shown to influence pilot TDL has thus been obtained for the B747.

For a more detailed discussion of this flight simulator experiment the reader is referred to Journal Article 4.

4 It should be noted that the approaches during the experiment were flown without other traffic present. Pilots were asked to comment on the approaches and rate the approaches as if they were flown in a real life situation (that is, with other traffic present). They commented that also in real life situations, not meeting the constraints in the first part would not increase pilot TDL, because it would be apparent to ATC that they would not be able to meet the constraints.

9 | Method to predict pilot TDL and highly detailed computer simulation

Having identified the factors that influence pilot TDL during an approach, the two remaining goals of this research are to develop a method to predict pilot TDL and to capture this method within a computer simulation. Since the development of the method and the computer simulation are closely linked, these two subjects are covered together in this chapter. First the method to predict pilot TDL is introduced. Subsequently a highly detailed computer simulation, using Monte Carlo simulation, is presented.

9.1 | A method to predict pilot TDL consisting of seven guidelines

The method to predict pilot TDL based on the results of the flight simulator experiments consists of some guidelines for the design of approaches. When these guidelines are followed, pilot TDL during the approach will be acceptable. The method will thus not give an indication of pilot TDL on a numerical scale possibly even predicting a continuously changing numerical value for pilot TDL during the entire approach. Next to the fact that it will be extremely difficult (if not impossible) to rate the different factors that increase pilot TDL with respect to each other (which would be necessary to achieve such a continuous scale), it is also not necessary to unite the different factors in one rating. Knowing the factors that have a major effect on pilot TDL, and the ones that have a minor effect, it can be stated that an approach should at least meet the guidelines for the major effects (regardless of their relative influence on pilot TDL), and if possible, preferably, also meet the guidelines for the minor effects, a more detailed (numerical) scale is not needed.

The guidelines for the design of approaches with respect to pilot TDL are an accumulation of what has been presented in the preceding chapters. Starting point for the guidelines is that pilots should fly the approach according to SOPs and that they should aim to achieve a stabilized approach at 1,000′. The guidelines for the contributors to pilot TDL for the B747 then are that:

- aircraft should be able to meet the altitude and airspeed constraints throughout the approach, especially during the final part of the approach, and during the first part of the approach if this has consequences for the subsequent parts of the approach;
- there should be sufficient time to perform all actions on Localizer intercept heading;
- it should be possible to achieve a stabilized approach. Whether a stabilized approach can be achieved depends on 1. the energy rate demand during the final part of the approach, 2. the value of the vertical speed (should be below the sink rate warning) which in itself is a function of airspeed on final and glideslope angle, and 3. the FAF altitude and distance between IF and FAF (resulting in the line-up distance);
- the distance between IF and FAF should be sufficient;
- the vertical speed should be below the sink rate warning;
- the Localizer intercept speed should not be too high, and that
- the Localizer intercept angle should not be too large.

These guidelines should be adhered to for all aircraft that will use the approach, the most common prevailing windconditions at the airport, and for the majority of pilots. How these guidelines can be quantified, and how a prediction can be obtained whether these guidelines are followed for an approach is explained in the following sections which describe the computer simulation.

It should be noted that the seven guidelines are based on a small sample of approaches and pilots. Therefore, the decision that some factors from Table 7.1 do not influence pilot TDL was also based on a small data set, and might not be conclusive for all possible circumstances. In that respect, when also wanting to incorporate these factors in the design of the approach, an additional guideline can be specified that: the number of waypoints and heading changes should not be too high, that there should be sufficient time or trackmiles during the first part of the approach, and that the altitude profile preferably is a Continous Descent Approach.

9.2 | The comprehensive Monte Carlo computer simulation

The development of the computer simulation was an iterative process. Starting with a first guess for all the variables needed for the simulation, a first preliminary simulation was developed with which the approaches for the first flight simulator experiment were designed (see chapters 3 and 5 of Journal article 2). Based on the results of the first experiment the computer simulation was improved (see chapter 8 of Journal article 3) and used to design the approaches for the second experiment, after which it was adjusted and updated again (see chapter 6 of Journal article 4). This chapter only presents the final and most complete version of the computer simulation.

The highly detailed computer simulation is based on a Monte Carlo simulation. This means that it evaluates the guidelines for pilot TDL by simulating the approach many times, and using a different windcondition and a different way of performing the pilots' actions for each run. After an x number of runs, the percentage of flights can be established for which the guidelines as described above were indeed met.

9.2.1 | Elements of the highly detailed computer simulation

Computer simulation input – The input of the Monte Carlo computer simulation consists of a list of waypoints, defined by their lat-lon coordinates, and the altitude and speed constraints at these waypoints.

Aircraft (B747-100), Autopilot and Flight Director models – Although the Monte Carlo simulation should eventually work for any aircraft type, for now only a B747 aircraft model is used in the simulation. The non-linear aircraft model is based on the Boeing 747-100 documentation by Hanke and Nordwall (1970,1971) and is modeled as detailed as possible. Autopilot, Autothrottle and Flight Director (FD) models are also derived from Hanke and Nordwall (1970). The hierarchy in meeting the constraints at the waypoints is as follows: the Autopilot and FD modes will always aim to meet the altitude constraints at the waypoints, second to this, the Autothrottle controls the airspeed. This results in the situation that the altitude constraint at the next waypoint will always be met, while the speed constraint might not be met (i.e., the airspeed might be higher than required).

Pilot model and Standard Operating Procedures for the B747 – To these highly detailed, non-linear models a relatively simple pilot manual control model for the flight director task is added, consisting of only a time delay (equal to 0.3 seconds) and pure gain. All other pilot actions such as selecting flaps are modeled according to the SOPs. As explained in chapter 6 the scanning of the flight instruments in no longer a part of the pilot model.

Since the aircraft model in the computer simulation is a Boeing 747-100, the SOPs that are modeled in the computer simulation are based on the SOPs for a Boeing 747, more specifically: the SOPs for the Delayed Flap Approach with flaps LAND equal to flaps 25 (see figure 4.3).

Table 9.1 | Trigger events for pilot actions in Monte Carlo simulation.

Pilot action	Trigger Event	Action modeled as function of
Approach Checklist	Transition level	Altitude
Flaps 1	UP mark(1)	IAS
Flaps 5	1 mark(2)	IAS
Flaps 10	End of turn to Localizer Intercept Heading	time
Heading Select	End of turn to Localizer Intercept Heading	time
ARM Approach	End of turn to Localizer Intercept Heading	time
Autopilot Off	End of turn to Localizer Intercept Heading	time
Autothrottle Off	Autopilot Off	time
Gear Down	Reaching FAF	time
Flaps 20	Gear Down	time
Landing Checklist part 1	Latest of Gear Down or Flaps 20	time
Flaps 25 (Flaps LAND)	Reaching 1,200′	time
Landing Checklist part 2	Flaps 25 (Flaps LAND)	time

(1) UP mark: the Indicated Airspeed (IAS) below which flaps 1 should be selected

(2) 1 mark: the IAS below which flaps 5 should be selected

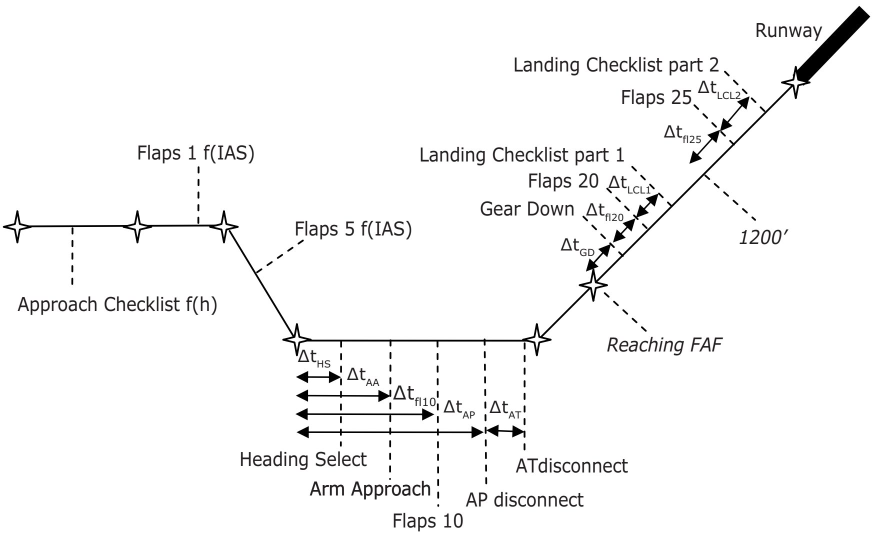


Figure 9.1 | Visualization of trigger events and reaction times (Δt) for pilot actions in the Monte Carlo computer simulation.

Each of these pilot actions prescribed by the SOPs is modeled using a 'trigger' event (e.g., reaching 1,200 feet) and a reaction time representing the time between reaching the trigger event and actually performing the action (e.g., 2 seconds after reaching 1,200 feet, flaps 25 are selected). Some pilot actions are modeled as a function of IAS or altitude instead of as a function of time. These trigger events and reaction times are modeled based on data obtained during the two flight simulator experiments described in Chapters 7 and 8. The trigger events as used in the computer simulation are given in table 9.1, see also Figure 9.1. The reaction time distributions can be found in Appendix I. For a more detailed explanation of the reaction times, see section 3.3 in journal article 2, section 5.3 in journal article 3, and sections 4.3 and 6.4 in journal article 4.

Turbulence and Wind models – Turbulence is modeled according to the Dryden spectra [15], the wind speed is modeled using a Weibull distribution (λ = 12.5 kts, k = 2.0). The wind direction is varied around the runway heading by applying a normal distribution with μ = runway heading and σ =30 deg. During one Monte Carlo simulation run the turbulence intensity, wind direction and wind speed are constant throughout the entire approach, between different Monte Carlo runs these are varied.

9.2.2 | Output of the computer simulation

The computer simulation can now be used, amongst others, to predict under what wind conditions the constraints at the waypoints can be met and a stabilized approach can be achieved. Both factors were shown to influence pilot TDL.

To determine whether the constraints at the waypoints were met the following was assumed: the constraints at a waypoint were considered to be met when the actual Indicated Airspeed (IAS) at that waypoint was less than the required IAS plus 10 knots, and the actual altitude at that waypoint was less than the required altitude plus 100 feet. A lower boundary for these constraints is not necessary since the Monte Carlo simulation always regulates the airspeed and altitude towards the constraints at the next waypoint, when the required airspeed and altitude are attained, the Monte Carlo simulation maintains the required airspeed and altitude until the waypoint is reached. Therefore, in the Monte Carlo simulation, the altitude and airspeed will never be too low at a waypoint.

To determine whether a Monte Carlo simulation run of the approach resulted in a stabilized approach at 1,000 ft above airport elevation the criteria from section 4.4 were quantified as follows:
- ❖ Heading change and pitch change are within 5 deg/s;
- ❖ The IAS is not more than VREF + 20 knots;
- ❖ Flaps 25 are selected, landing gear is down;
- ❖ Sink rate is not larger than 1,000 feet per minute;
- ❖ Localizer and glide slope are within one dot; and
- ❖ All checklists are completed.

As an example of the computer simulation output, see the results in Figure 9.3 for the approach defined in Figure 9.2 and Table 9.2. Figure 9.3 clearly shows that the possibility of achieving a stabilized approach depends on the wind direction (a strong headwind on final results in a stabilized approach). Other reasons for ending stabilized or unstabilized are the moment in time at which flaps 20 and/or gear down are selected and the completion of the checklist. From this plot it can be concluded that this approach is a very bad design, and should not be implemented in reality. A similar plot as in Figure 9.3 can be generated for meeting the constraints at each waypoint. The results of the computer simulation also provide insight into the locations in the approach where pilots are performing many actions, or where they are performing checklists, thereby providing approach designers with an indication of the 'busy' parts of an approach, see Figure 9.4 for an example.

The predictions of the computer simulation have been checked against the data from the flight simulator experiments and agree very well. For a more elaborate discussion of the Monte Carlo simulation see chapters 3 and 5 of Journal article 2, chapter 8 of Journal article 3 and chapter 6 of Journal article 4.

By predicting the aircraft's altitude and velocity profiles, the percentage of flights that meet the constraints at the waypoints, and the percentage of flights that achieves a stabilized approach, the computer simulation provides a prediction of the aircraft's motion in the longitudinal plane.

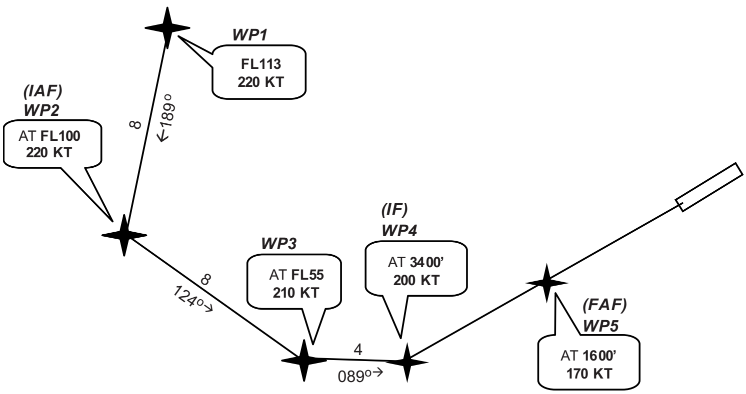


Figure 9.2 | Approach chart.

Table 9.2 | Waypoints with altitude and airspeed constraints.

Waypoint		**Altitude [ft]**	**IAS [knots]**
WP1	×	11,300	220
WP2	IAF	10,000	220
WP3	×	5,500	210
WP4	IF	3,400	200
WP5	FAF	1,600	170
RWY	×	0	VREF + 5

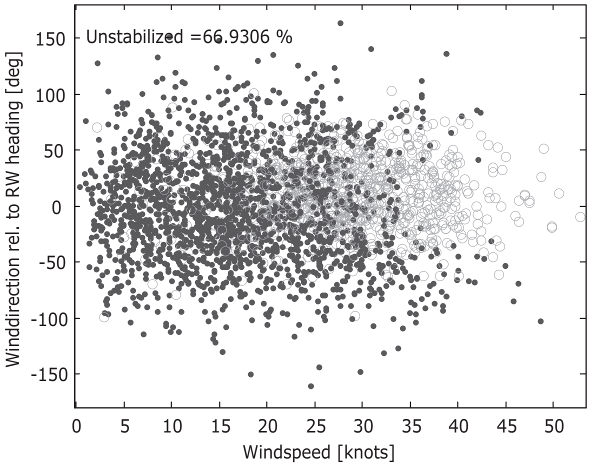


Figure 9.3 | Example of the results of the computer simulation (3,000 runs) with respect to the possibility of achieving a stabilized (grey circle) or unstabilized (black dot) approach at 1,000′ as a function of wind speed and wind direction.

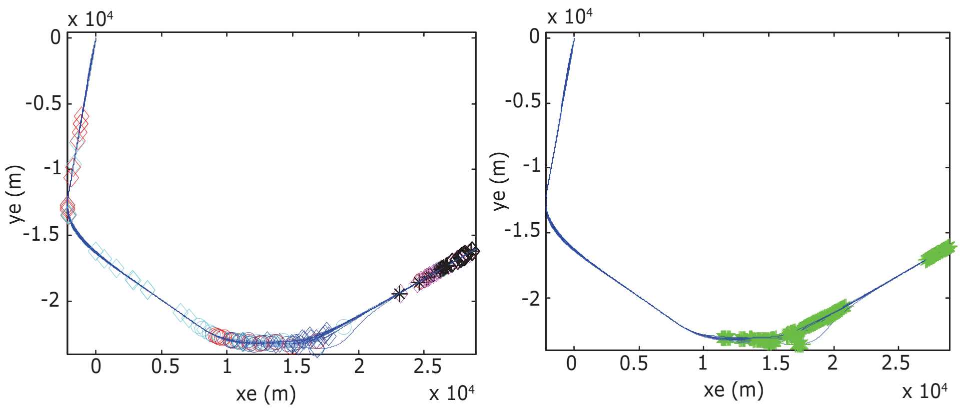


Figure 9.4 | Example of the results of the computer simulation, providing insight into the locations in the approach where pilots are performing actions such as selecting flaps, gear, etc. (left) or are performing checklists (right).

9.2.3 | Does the computer simulation predict the information for all guidelines?

The question now is, whether the output of this computer simulation provides sufficient information in order to assess whether the approach meets all guidelines for the contributors to pilot TDL as given earlier in this chapter. The simulation obviously predicts whether a stabilized approach can be achieved (guideline 3), and whether the constraints at the waypoints can be met (guideline 1). It also provides insight whether there is sufficient time on Localizer intercept heading since the moments in time at which all actions are performed is predicted (guideline 2), and it can easily predict the sink rate (guideline 5).

However, although the simulation can predict or calculate the numerical values for the (actual) Localizer intercept speed (guideline 6), the Localizer intercept angle (guideline 7), and the distance between IF and FAF (guideline 4), it does not give a qualitative indication of whether these numerical values are sufficiently high or low. Fortunately, the minimum or maximum values for these factors are very accurately prescribed in the PANS-OPS (ICAO, 2006). The PANS-OPS prescribe a minimum straight distance between IF and FAF of 2nm with an additional turning distance (which depends on the airspeed and intercept angle), and recommend an interception angle at the Localizer not exceeding 30 degrees. Actually, the PANS-OPS and the predictions of the computer simulation complement each other very nicely regarding factors contributing to pilot workload, since what is not prescribed in the PANS-OPS is predicted by the computer simulation and vice versa.

The conclusion thus is that the predictions of the computer simulation combined with the regulations in the PANS-OPS together provide sufficient information to assess whether the approach adheres to all the guidelines for pilot TDL given earlier in this chapter. Additionally, the computer simulation gives a realistic prediction of the aircraft's motion in the longitudinal plane. Therefore, for the B747 aircraft, the goals 2 and 3 as defined in chapter 2 are met.

10 | Expanding the method to incorporate another type of aircraft: the Cessna Citation

To check whether the factors that were found to influence pilot TDL for the B747 also apply to other aircraft types, the same research steps followed for the B747 were repeated for a Cessna Citation. The highly detailed Monte Carlo computer simulation was adjusted to incorporate the aerodynamics, autopilots, flight director and SOPs of the Citation, and was used to design 10 different approaches. These 10 approaches were flown in the SIMONA flight simulator by 6 Citation pilots. To enable these tests, the autopilots, flight director and FSM logic in the SIMONA flight simulator were adjusted to suit the Cessna Citation.

Additional to flying the 10 approaches in the flight simulator, the same 6 pilots flew the same 10 approaches during real flight in the Cessna Citation laboratory aircraft, which is jointly owned by the National Aerospace Laboratory (NLR) and Delft University of Technology. The goal was to check whether the same factors influence pilot TDL during real flight as during the flight simulator tests.

In this chapter first the relevant differences between the B747 and the Cessna Citation are discussed, followed by the results for factors influencing pilot TDL obtained from the tests in the flight simulator and tests with the laboratory aircraft. To conclude it presents the differences between the flight simulator tests for the Cessna Citation and the real flight tests for the Cessna Citation.

10.1 | Differences between the B747 and the Cessna Citation

Except for the obvious differences between the B747 and Cessna Citation (size, number of passengers, etc.) there are some differences that are important when comparing the results for factors that influence pilot TDL between the flight simulator tests for both aircraft types.

First, it should be mentioned that an approach which results in high energy rate demands for the B747, meaning that the B747 cannot meet the constraints at the waypoints, might be relatively easy or difficult to fly for the Citation due to different kinetic properties. Therefore, for the tests with the Citation *new* approaches were designed that were different from the approaches used in the B747 tests but were based on the same principles. For example: an approach was designed such that the Citation would not be able to meet the constraints at the waypoints.

Second, the Citation does not have an autothrottle or a VNAV mode. This most probably will result in a higher pilot TDL in the absolute sense. It should be noted, however, that this situation is the same for all approaches flown with the Citation, and for these experiments the focus is on the relative difference in pilot TDL *between* these approaches.

Third, the SOPs for the Citation differ from the SOPs for the B747, see Figure 10.1. For example: for the Cessna Citation there are slightly less actions to be performed on Localizer intercept heading, due to different procedures for flap selection, and due to the fact that there is no autothrottle to switch off. Another example for a difference in SOPs is that for the Citation flaps sometimes need to be selected between the IF and FAF, whereas there were no actions required in this part of the approach for the B747. This might result in a different outcome for the factors that influence pilot TDL. When designing the approaches for the Citation the factor 'sufficient time on Localizer intercept heading' was therefore more generally used in the sense that there should be sufficient time to perform the actions, regardless of the exact location in the approach.

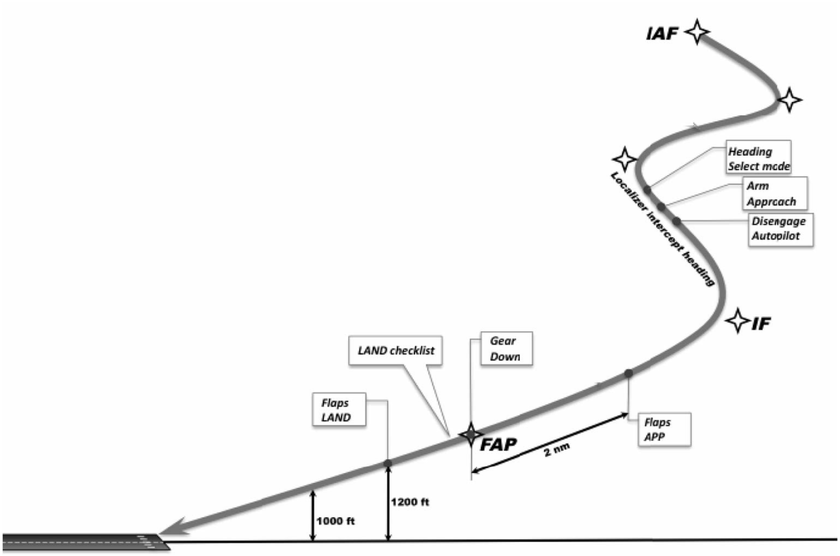


Figure 10.1 | SOPs for the Cessna Citation

10.2 | Pilot TDL results for SIMONA flight simulator tests and laboratory aircraft flight tests

Analysis of the results of the flight simulator tests and the real flight tests with the laboratory aircraft showed that both test series resulted in the same list of factors that influence pilot TDL. In other words: there was no difference between the real flight tests and flight simulator tests regarding the factors that influence pilot TDL. The factors that seem to influence pilot TDL, again based on a small data set, are mentioned below.

For the first part of the approach (all minor effects on pilot TDL)
- The only contributor to pilot TDL seems to be the energy rate demand. Note that the flexibility of approaches (as in Figure 8.2) was not part of the Citation experiments, therefore nothing can be concluded about the influence of this flexibility on pilot TDL.

For the Localizer Intercept part of the approach (all minor effects on pilot TDL)
- The time (or distance) available on Localizer intercept heading to perform all required actions;
- The Localizer intercept angle; and
- The Localizer intercept speed.

For the final part of the approach (all major effects on pilot TDL):
- The line-up distance;
- The airspeed at the FAF;
- The FAF altitude, since this altitude defines the amount of time available to perform all required actions between FAF and 1,000’;
- The energy rate demand;

- The distance between IF and FAF; and
- The fact whether a stabilized approach can be achieved at 1,000'.

It should be noted that the factors that are mentioned for the final part of the approach, had a larger influence on pilot TDL than the factors listed for the first part and Localizer intercept part of the approach.

Comparison of these results with the results found for the B747 in Chapter 8 shows that all factors mentioned above were also found to influence pilot TDL for the B747. There are some minor differences: for example the fact that the time on Localizer intercept heading had a major effect on pilot TDL for the B747 and only a minor effect for the Citation. This might be due to the fact that for the Citation less actions had to be performed on Localizer intercept heading.

For an extensive description of the flight simulator tests and real flight test the reader is referred to Journal Article 5 and van Bennekom and van Tuinen, 2010.

10.3 | Differences between the SIMONA flight simulator tests and laboratory aircraft flight tests

Although the results for the factors influencing pilot TDL were the same for both the flight simulator tests for the Citation and the real flight tests with the Citation, two significant differences between the 'real life' situation and the situation in the flight simulator could be observed.

The first difference concerns the communication with ATC. During the flight simulator tests all communication was standard and the same for all approaches (in order not to add yet another variable to the test). As a result pilots would know, after having flown a couple of approaches, what ATC was going to say. Consequently, after a while, they would continue performing checklists even if ATC was giving them instructions. During the real flights this was not the case: once the pilot monitoring received a call from ATC all attention was diverted to ATC contact and all other activities (such as performing checklists, selecting flaps, corresponding to calls from the pilot flying) stopped.

The second difference regards the time needed to perform the checklists. In the flight simulator not all dials and knobs were built in, and although pilots were asked to 'act the part' and pretend to check all dials and knobs during checklists, they were much quicker in performing the checklists than in reality.

It can thus be concluded that the same factors that were shown to influence pilot TDL for the B747 were also found to influence pilot TDL for the Cessna Citation. Additionally, the flight simulator tests and the real flight tests with the laboratory aircraft both resulted in the same list of factors that influence pilot TDL, there was thus, in this respect, no difference between the 'real life' situation and the flight simulator.

11 | A simple computer simulation to predict pilot TDL based on a point mass model

The computer simulation that has been used to give an indication of pilot TDL, and which was presented in the chapter 9, is based on highly detailed aircraft models. Additionally, it is based on a Monte Carlo simulation. These two factors together result in a computer simulation that does generate reliable results regarding the percentage of flights that can achieve a stabilized approach and factors that influence pilot TDL, but that takes a long time to produce these results (in the order of several hours per approach that is analyzed). This is not very practical when the computer simulation is intended to be used as a tool during the design of approaches.

The reason to choose for these highly detailed models was that it was hypothesized at the start of the research that short time scale factors such as manually controlling the aircraft after disturbances (e.g., turbulence gusts, deployment of flaps) and scanning the flight instruments would influence pilot TDL. In order to model these effects on pilot TDL highly detailed models would indeed be necessary. However, in chapter 6, it was concluded that these short time scale factors do actually not influence pilot TDL, and that all factors that *do* are, so to say, long time scale factors. The list of factors influencing pilot TDL as found during this research might therefore not require highly detailed (non-linear) models in order to predict them accurately.

For these reasons it is investigated whether a much simpler model based on a point mass model, with a considerably shorter calculation time, can generate results as reliable as the highly detailed computer simulation. If this is true, then this point mass model can be used as a tool during the design of approaches.

It was found that a point mass model can indeed generate the same results as the highly detailed computer simulation. This is true as long as the point mass model contains:
1. a detailed lift-drag polar for all flap settings and gear up/down setting,
2. a detailed model of the flight idle thrust,
3. an accurate model to simulate the lateral track, specifically the distance of turn anticipation since this influences the amount of trackmiles available between two waypoints, and
4. a model to simulate the pilots' actions according to the trigger events and reaction time distributions found in this research.

The fact that a point mass model can be used already significantly reduces the run time of the computer simulation. An additional gain in calculation time can be obtained by using a different simulation technique. Instead of using a Monte Carlo simulation and simulating all possible pilot reaction times (resulting in many runs) and afterwards checking the results to identify how many pilots met the constraints and achieved a stabilized approach, a different kind of simulation is applied. Now, the approach designer has to specify upfront the percentage of flights that should be able to meet the constraints and achieve a stabilized approach, for instance 95% of all flights. The simulation then predicts whether 95% of all flights can indeed achieve this or not. This requires only *one* computer simulation run with only the pilot actions with the slowest (worst) reaction times for 95% of the pilots. In order to determine for which wind conditions the constraints can still be met and a stabilized approach can be achieved the simulation automatically performs some additional runs. For a more detailed explanation of the point mass model simulation see section 6.4 in Journal Article 4.

Thus, by using a computer simulation based on a point mass model and using a different simulation technique, a simulation tool is obtained that takes only a couple of seconds to analyze an approach and can provide the same predictions as the highly detailed Monte Carlo computer simulation.

The ideas outlined above have been captured in a computer program that enables the user to analyze an approach. This computer program was developed in cooperation with the company To70. After the waypoints and corresponding constraints are entered into the simulation the approach is plotted on a map. On this map the location of the waypoints can be altered interactively (if desired) by clicking and dragging the waypoints. The program then calculates in a couple of seconds whether the constraints can be met and a stabilized approach can be achieved for the required percentage of flights. Additionally, it predicts for which wind conditions this can still be achieved, and calculates the amount of thrust that is applied on each leg. It also visualizes at which locations in the approach the pilots are performing which actions. At this moment the computer simulation encompasses a model for the B747-100 and the Cessna Citation. Other aircraft models, however, can be added relatively easily. Currently, the reaction time distributions used in the point mass model computer simulation are the normal curve approximations of the reaction times obtained from the flight simulator tests and real flight tests. An example output of this program is enclosed in Appendix II .

CONCLUSIONS AND RECOMMENDATIONS

12 | Conclusions

The goals of this research were defined as follows:

- ❖ First, the factors relevant for approach design, which influence pilot Task Demand Load (TDL) while flying an approach, need to be identified.
- ❖ Second, a method needs to be developed with which it is possible to predict pilot TDL during approach.
- ❖ Third, this method needs to be captured within a computer simulation that can be used in the early stages of approach design. With this tool, the designer should be able to rapidly evaluate a potential approach from the perspective of the demands it imposes on the pilot's limited resources. Additionally, this computer simulation should also be able to give a realistic prediction of the motion of the aircraft in the longitudinal plane.

The conclusions of this research are grouped per goal.

12.1 | Factors that influence pilot TDL

Contrary to what was hypothesized at the very beginning of this research, continuous skill-based actions such as manually controlling the aircraft or scanning the flight instruments do not influence pilot TDL. This is true for aircraft with a high level of automation and for pilots who are used to flying these aircraft types.

The list of factors that were found to influence pilot TDL can, per approach part, be given as:

For the first part of the approach:

- ❖ The major contributor to pilot TDL seems to be the energy rate demand, or, the fact whether or not constraints can be met at the waypoints. This is only true when the effect of not meeting the constraints continues into the Localizer part or final part of the approach. If the consequences of the energy rate demand remain within the first part of the approach this does not influence pilot TDL.

For the Localizer intercept part of the approach:

- ❖ The time available to perform all actions (which is directly related to the distance available on Localizer intercept heading) seems to be the most important factor for pilot TDL.
- ❖ Next to this, pilot TDL is also influenced by the Localizer intercept speed, the Localizer intercept angle, and whether the constraints at the waypoints can be met (the energy rate demand).

For the final part of the approach:

- ❖ The most important factors influencing pilot TDL seem to be whether or not a stabilized approach can be achieved at 1,000', the distance between IF and FAF and the airspeed on final. Whether an approach is stabilized can, for a B747 and a Cessna Citation, be determined from: 1. the energy rate demand during the final part of the approach, 2. the value of the vertical speed (should be below the sink rate warning) which in itself is a function of airspeed on final and glideslope angle, and 3. the FAF altitude and distance between IF and FAF (resulting in the line-up distance) since these two factors together determine whether there is enough time available to perform all actions required for a stabilized approach. All these factors thus influence pilot TDL during the final part of the approach.

This list of factors applies to the B747-100 as well as to the Cessna Citation, it is again noted that these results are based on small data sets. It is hypothesized that the same factors will also determine pilot TDL for other aircraft types. In this respect it should be noted that the factors 'FAF altitude' and 'time available on Localizer intercept heading' are factors that originate from the fact that both for the B747 and the Citation the Standard Operating Procedures require pilots to perform a number of actions on Localizer intercept heading and between the FAF and 1,000ft, and that they should have sufficient time to do so. If, for another aircraft type, these actions are required to be performed in another part of the approach, then care should be taken that sufficient time is available in that particular part of the approach. In that case, the factors 'time available on Localizer intercept heading' and 'FAF altitude' might not influence pilot TDL for that particular aircraft type. All other factors in the list above are assumed to be valid for all aircraft types.

Test results showed that the list of factors influencing pilot TDL was the same for tests performed in flight simulators and tests performed during real flight. It might be interesting to note that differences that *did* occur were related to communication with Air Traffic Control (which was taken much more seriously during the real flight tests) and the amount of time needed to perform checklists (which took longer during the real flight tests).

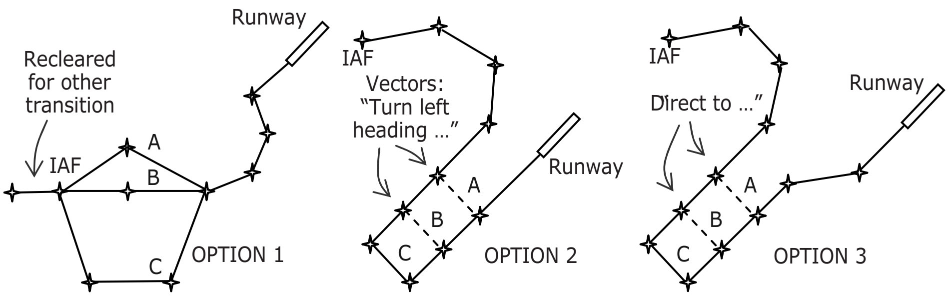


Figure 12.1 | Three options to include flexibility in approaches.

Three possible options to include flexibility were considered, in order to give Air Traffic Control (ATC) the possibility to correctly sequence the aircraft, they are depicted again in figure 12.1. The majority of the pilots indicated that they, in general, preferred either Option 2 or Option 3 to include flexibility in the approach, whereas all pilots preferred Option 2 or 3 when asked to only consider the adjustments needed in the FMS.

When such a flexibility is introduced in the approach, an additional factor influencing pilot TDL can be identified: pilots should have sufficient time to make all necessary adjustments.

12.2 | A method to predict pilot TDL

The method to predict pilot TDL following from the results of the flight simulator experiments consists of some guidelines for the design of approaches. When these guidelines are adhered to, pilot TDL during the approach will be acceptable. The method does not give an indication of pilot TDL on a numerical scale. Next to the fact that it will be extremely difficult (if not impossible) to rate the different factors that increase pilot TDL with respect to each other (which would be necessary to achieve such a continuous scale), it is also not necessary (for the specific goals in this research) to unite the different factors in one rating.

Starting point for the guidelines is that pilots should fly the approach according to SOPs and that they should aim to achieve a stabilized approach at 1,000'. The guidelines for the contributors to pilot TDL then are that:

- aircraft should be able to meet the altitude and airspeed constraints throughout the approach, especially during the final part of the approach, and during the first part of the approach if this has consequences for the subsequent parts of the approach;
- there should be sufficient time to perform all actions on Localizer intercept heading;
- it should be possible to achieve a stabilized approach. Whether a stabilized approach can be achieved depends on 1. the energy rate demand during the final part of the approach, 2. the value of the vertical speed (should be below the sink rate warning) which in itself is a function of airspeed on final and glideslope angle, and 3. the FAF altitude and distance between IF and FAF (resulting in the line-up distance);
- the distance between IF and FAF should be sufficient, and that
- the vertical speed should be below the sink rate warning.
- the Localizer intercept speed should not be too high, and that
- the Localizer intercept angle should not be too large.

It should be noted that the seven guidelines are based on a small sample of approaches and pilots. Therefore, the decision that some factors do not influence pilot TDL was also based on a small data set, and might not be conclusive for all possible circumstances. In that respect, when also wanting to incorporate these factors in the design of the approach, an additional guideline can be specified that: the number of waypoints and heading changes should not be too high, that there should be sufficient time or trackmiles during the first part of the approach, and that the altitude profile preferably is a Continous Descent Approach.

12.3 | A computer simulation to predict pilot TDL

A highly detailed non-linear computer simulation based on a Monte Carlo simulation technique was designed and built. This computer simulation assumes flight according to SOPs where the pilots' actions are modeled by using trigger events and distributions for reaction times relative to these trigger events. This computer simulation can:

- predict the percentage of flights that will result in a stabilized approach at 1,000' as a function of wind direction, wind speed and timing of pilot actions;
- predict the percentage of flights that will meet the altitude and velocity constraints at the waypoints as a function of winddirection, windspeed and timing of pilot actions;
- provide insight into the busy parts of an approach from the pilot's perspective;
- predict the aircraft's motion in the longitudinal plane

Additionally, and most importantly, the computer simulation, combined with the regulations in the PANS-OP, provides sufficient information to assess whether the guidelines for the contributors to pilot TDL are adhered to. The predictions of the simulation provide a good, average indication of pilot TDL.

Next to the highly detailed computer simulation, a much simpler computer simulation based on a point mass model of the aircraft was developed. This point mass model simulation can generate exactly the same results but requires significantly shorter simulation times, and can be used by approach designers as an additional tool during the design of approaches.

Both computer simulations were developed and validated for the B747 and Cessna Citation aircraft. They are, however, set-up in a modular way which makes it easy to incorporate other aircraft types in the simulation. For other aircraft types the SOPs might (and probably will) deviate from the SOPs for the B747 and Citation, the list of trigger events and reaction times obtained for the B747 and Citation might, however, provide a good starting point to also model pilot actions according to different SOPs.

13 | Recommendations

13.1 | Further research on approaches

For further research regarding pilot TDL during approaches it is advised to concentrate on the computer simulation based on the point mass model. When compared to the highly detailed Monte Carlo simulation it can provide the same predictions in less time, is much easier to use during approach design and is easier to adapt for other SOPs or other aircraft.

It is recommended to incorporate other aircraft types in the point mass model computer simulation in order to arrive at a tool that can be used to analyze an approach for all the aircraft types that will actually fly the approach once it becomes operational. For aircraft with comparable SOPs to the B747 and Cessna Citation, the trigger events and reaction times as found during this research can be used in the computer simulation. However, when adding an aircraft that requires a very different kind of SOPs, such as an Airbus aircraft, it is advised to perform additional flight simulator tests to check whether the adopted simulation philosophy (using trigger events and reaction times) also produces reliable results for these kinds of aircraft.

It is always desirable to obtain an as large as possible set of experimental data on which the computer simulation models can be based, and against which they can be validated. Therefore, more flight simulator tests and real flight tests are always advisable. However, for the B747 and the Cessna Citation a large dataset for the pilots' reaction times has already been obtained, and it is questionable whether a more extensive set of reaction times will produce significantly better (more realistic) results. It is very well probable that a more accurate set of data, which is very expensive to obtain, will not improve the predictions of the computer simulation since there are other factors, such as small deviations between the actual aircraft thrust and the modeled aircraft thrust, that will completely annihilate possible improvements due to more accurate data for pilot actions. Therefore it is not recommended to perform additional flight simulator tests for the B747 or the Cessna Citation for the purpose of obtaining reaction times. Additional tests should be performed, however, to obtain reaction times for aircraft with very different SOPs, as indicated above.

From the viewpoint of obtaining more subjective pilot data about which factors influence pilot TDL, it *is* advised to perform more flight simulator tests for the B747 and Cessna Citation. It is especially useful to explore the influence of the factors that were assumed to be constant in this research (see section 4.4) on the pilots' reaction times and on the factors that influence pilot TDL. It should be investigated whether the presence of other traffic, difficult meteorological conditions such as mist or air traffic controllers that are badly audible have an influence. Additionally, also for this purpose, flight simulator tests for aircraft with different SOPs should be performed, in order to discover whether different factors have an effect on pilot TDL for aircraft with different SOPs.

The point mass model computer simulation should be further developed into a user-friendly software package, incorporating help functions, etc., in order for it to be used by approach designers who were not involved in this research and therefore do not have any background information. Other aircraft should be added to the simulation and the reaction time distributions should be based on the actual reaction times obtained from the flight simulator tests and real flight tests instead of on the normal curve approximations.

13.2 | Extending this research to other flight phases

The same philosophy that has been adopted during this research regarding approaches: using a detailed model of the environment of the pilot and a simple model for the pilot him/herself, modeling the pilots' actions according to trigger events and reaction times, etc. can also be adopted for other flight phases, most notably for departures. In order to extend the research towards departures, the point mass model computer simulation can be used as a starting point, and should be augmented with SOPs for departures, all other necessary sub models (such as a thrust model, lift-drag polar, etc.) are already part of the simulation model. Additional flight simulator tests will have to be performed for departures, but the factors influencing pilot TDL as found for approaches can serve as a starting point, and together with the knowledge that the continuous scanning of the instruments and continuous control actions do not contribute significantly to pilot TDL, the amount of simulator tests can be much smaller than was the case for this research.

13.3 | Connection to other research

The guidelines to keep pilot TDL at an acceptable level during approaches, as found during this research, are not numerous in number and are easy to comprehend. Therefore they can relatively easily be applied to other research areas as well. For example: within research regarding optimization of approaches with respect to noise abatement, the guidelines can be incorporated in the optimization procedures as additional constraints. By doing so, the optimization will produce approach trajectories that are favorable both in terms of noise abatement and in terms of pilot TDL. The guidelines can also serve as a starting point (or as boundary conditions) for research projects that aim to design new or innovative procedures for the approach phase of the flight.

On the other hand, it would be very interesting to acquire the data of approaches that were flown in flight simulators for other research projects, and to analyze these approaches with the method and computer simulation developed within this research. Without performing extra flight simulator tests, this would provide the possibility for additional validation of the models.

With respect to research projects that aim to incorporate any of the options to include flexibility in an approach, it is advised to also consider the pros and cons of each of the three options from the viewpoint of pilot TDL, before making a definitive choice for one of the options.

References

Amalberti, R., Wioland, L. (1997) Human error in aviation. In: Soekkha H., editor. *Aviation safety: Human factors, system engineering, flight operations, economics, strategies, management,* VSP Utrecht; The Netherlands, pp. 91-108.

Argüelles, P. et al. (2001), Report of the Group of Personalities, *European Aeronautics: A Vision for 2020, Meeting Society's Needs and Winning Global Leadership*, The European Commission, Belgium.

Baren, G.B. van, Speijker, L.J.P., Bruin, A.C. de (2002), *Wake vortex safety evaluation of single runway approaches under different weather and operational conditions*, Proc. 6th Int. Conf. Probabilistic Safety Assessment and Management (PSAM6), June 23-28, 2002, San Juan, Puerto Rico.

Baron, S., Kleinman, D.L., Levison, W.H. (1970), *An Optimal Control Model of Human Response Part II: Prediction of Human Performance in a Complex Task,* Automatica, Vol.6, pp. 371-383.

Baron, S., Zacharias, G., Muralidharan, R., Lancraft, R. (1980), PROCRU: A model for analyzing flight crew procedures in approach to landing. In *NASA-MIT 16th Annual Manual 1980*, pp 488-520.

Baron, S., Kruser, D.S., Huey, M.B., *Quantitative Modeling of Human Performance in Complex, Dynamic Systems*,National Academy Press, Washington D.C., 1990.

Bennekom, K.T.P. van, Tuinen, T.J. van (2010), *Pilot task demand load considerations and a flight mechanical assessment tool for RNAV approaches with a Cessna Citation,* Master thesis report, Delft University of Technology.

Blok, G. (2005), *Prediction of pilot control workload due to the use of flaps and landing gear, based on a simulated approach through turbulent air,* Master thesis report, Delft University of Technology.

Blom, H.A.P., Bakker, G.J., Blanker, P.J.G., Daams, J., Everdij, M.H.C., Klompstra, M.B. (2001), *Accident risk assessment for advanced ATM*, In: Air Transportation Systems Engineering, edited by G.L. Donohue and A.G. Zellweger, Progress in Astronautics and Aeronautics Vol. 193, Paul Zarchan, Editor-in-Chief, Chapter 29, pp. 463-480, 2001

Blom, H.A.P., Klompstra, M.B., Bakker, G.J. (2001), *Accident risk assessment of Simultaneous Converging Instrument Approaches*, Proc. 4th USA/Europe ATM R&D Seminar FAA/Eurocontrol, December 2001.

Boeing Commercial Airplane Group (2010), *Statistical Summary of Commercial Jet Airplane Accidents, Worldwide Operations, 1959-2009,* Seattle, Washington, United States, July 2010.

Carbonell, J.R. (1966), A Queuing Model of Many-Instrument Visual Sampling, *IEEE Transactions on Human Factors in Electronics*, HFE-7.

Carbonell, J.R., Ward, J.L., & Senders, J.W. (1968). A queuing model of visual sampling: Experimental validation. *IEEE Transactions on Man-Machine Systems*, MMS-9.

Corker, K.M. (1999), Human Performance Simulation in the Analysis of Advanced Air Traffic Management, In *Proceedings of the 1999 Winter Simulation Conference,* Farrington, P.A., Nembhard, H.B., Sturrock, D.T., Evans, G.W. eds.

Everdij, M.H.C., Blom, H.A.P., and Bakker, G.J. (2002), *Accident risk assessment for airborne separation assurance*, Proc. Advanced workshop on air traffic management, Capri, Italy, 23-26 September 2002.

EUROCONTROL (1999), *Navigation Strategy for ECAC,* European Organisation for the Safety of Air Navigation, European Air Traffic Control Harmonisation and Integration Programme, Edition 2.1, 15 March 1999.

EUROCONTROL (2008), *Navigation Application & Navaid Infrastructure Strategy for the ECAC Area up to 2020,* European Organisation for the Safety of Air Navigation, Edition 2.0, 15 May 2008.

Ferro, D. (2005), *OPTIMAL Operational Concept Deliverable D1.2,* WP1.2-EEC-001-V1.0-ED-PU

Flight Safety Foundation Approach-and-landing Accident Reduction (ALAR) Task Force (1999), Killers in Aviation: FSF Task Force Presents Facts about Approach-and-landing and Controlled-flight-into-terrain Accidents*, Flight Safety Digest*, November 1998 - February 1999.

Godley, Stuart T (2006), *Perceived Pilot Workload and Perceived Safety of RNAV (GNSS) Approaches,* Australian Transport Safety Bureau, ATSB Transport Safety Investigation Report, Aviation Safety Research and Analysis Report 20050342.

Gore, B.F., Corker, K.M. (2000), A Systems Engineering Approach to Behavioral Predictions of an Advanced Air Traffic Management Concept, *19th Digital Avionics Systems Conference.* Philadelphia, Pennsylvania.

Hanke, C. Rodney and Nordwall, Donald R. (1970), *The Simulation of a Jumbo Jet Transport Aircraft, Volume II: Modeling Data,* The Boeing Company, Wichita Division Wichita, Kansas, USA, September 1970.

Hanke, C. Rodney (1971), *The Simulation of a Jumbo Jet Transport Aircraft, Volume I: Mathematical Model,* The Boeing Company, Wichita Division Wichita, Kansas, USA, March 1971.

Hilburn, B. and Jorna, P. (2001), Workload and Air Traffic Control. In PA Hancock and PA Desmond (Eds.) *Stress, Workload and Fatigue: Theory, Research and Practice*, Hillsdale, New Jersey, USA: Erlbaum

Hollnagel, E. (1993), *Human Reliability Analysis, Context and Control,* Academic Press, London

ICAO (1999), *Manual on Required Navigation Performance (RNP),* Doc 9613-AN/937, Second Edition, 1999.

ICAO (2006), *Procedures for Air Navigation Services Aircraft Operations, Volume II Constructions of Visual and Instrument Flight Procedures,* Fifth edition 2006.

ICAO (2007), *Resolutions Adopted by the Assembly,* Assembly – 36th Session, Montréal, 18 – 28 September 2007.

Jonge, H. de, Törner, S. (2009), *ARN D4-1 Concept of Operations Arlanda Airport – improvement of efficiency and the environment by advanced departure and arrival procedures-.*

Kleinman, D.L., Baron, S., Levison, W.H. (1970), *An Optimal Control Model of Human Response Part I: Theory and Validation,* Automatica, Vol.6, pp 357-369

KLM Luchtvaartschool (KLM Flight Academy) (1990), *Basic Attitude Flying*, Eelde

McRuer, D.T. and Jex, H.R. (1967), A Review of Quasi-Linear Pilot Models, *IEEE Transactions on Human Factors in Electronics,* Vol. HFE-8, No. 3, September 1967

Nijsten, K., Modelling the pilot's scanning behaviour, Monitoring and processing the available information during the approach phase, Master thesis report, Delft University of Technology, August 2005.

Nuic, A. (2004), *User Manual for The Base of Aircraft Data (BADA) Revision 3.6,* EEC Note No. 10/04, Eurocontrol Experimental Centre, July 2004.

Nuic, Angela, Poinsot, Chantal, Iagaru, Mihai-George, Gallo, Eduardo, Navarro, Fransisco A., Querejeta, Carlos (2005), *Advanced Aircraft Performance Modeling for ATM: Enhancements to the BADA Model,* 24th Digital Avionics System Conference, Washington D.C, October 30 - November 3, 2005.

Rasmussen, J. (1983), Skills, Rules and Knowledge; Signals, Signs, and Symbols, and Other Distinctions in Human Performance Models, *IEEE Transactions on Systems, Man and Cybernetics,* Vol. SMC-13, No. 3, May/June 1983.

Reason, J.T. (1990), *Human Error*, Cambridge University Press, U.K.

Roelen, A.L.C., Cacciabue, P.C., Hayes, P., Molemaker, R.J., Joyce, A. (2002), *A method for setting and optimizing safety targets in the air transport system,* Proc. 6th Int. Conf. Probabilistic Safety Assessment and Management (PSAM6), June 23-28, 2002, San Juan, Puerto Rico.

Roelen, A.L.C., Wever, R., Cooke, R.M., Lopuhaä, R., Hale, A.R., Goossens, L.H.J., Simons, M., Valk, P.J.L., (2003), Causal modeling for integrated safety at airports, Proc. European Safety and Reliability Conference ESREL 2003, June 15-18, Maastricht, The Netherlands. Eds: T. Bedford & P.H.A.J.M. van Gelder, Balkema Publishers.

Senders, J.W. (1964), The human operator as a monitor and controller of multidegree of freedom systems, *IEEE Transactions on Human Factors in Electronics, 5,* 2-6.

SESAR Consortium (2008), *The ATM Deployment Sequence D4,* DLM-0706-001-02-00.

Smith, B.R., Tyler, S.W. (1997), *The Design and Application of MIDAS: A Constructive Simulation for Human-System Analysis,* 2nd Simulation Technology & Training Conference, Canberra, Australia.

Stassen, HG, Johannsen, G, & Moray, N. (1990), Internal Representation, Internal Model, Human Performance Model and Mental Workload. *Automatica, 26(4)*, 811-820.

Verhoeven, R.P.M., Gelder, N. de (2009), *Time-based navigational and asas interval managed CDA procedures,* NLR-TP-2009-477

Vormer, F.J.P. (2005), *Theoretical and Operational Aspects of Optimal Airport Arrival Trajectories,* TUD Technische Universiteit Delft.

Wickens, C.D. and Hollands, J.G. (2000), *Engineering Psychology and Human Performance,* Third Edition, Prentice-Hall International

APPENDICES

Appendix I | Reaction time distributions for pilot actions during RNAV approaches

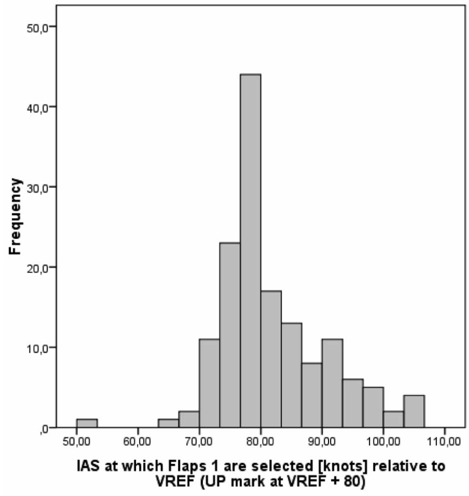
The reaction time distributions given in this appendix are based on the combined data of the two flight simulator experiments for the B747 in the SIMONA and GRACE flight simulators. The sample size *N* differs among the histograms, although all histograms are based on the same experiments. The reasons for these different sample sizes are given below.

Flaps 1 and flaps 5 were not selected during all of the approaches. Some approaches started at a relatively low IAS such that flaps 1, and sometimes flaps 5, were the initial condition, and hence not selected during the approach.

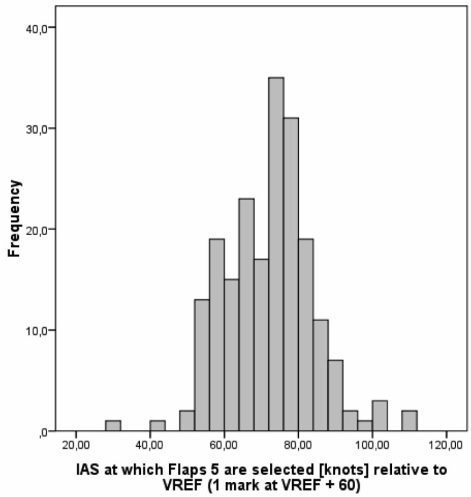
For the actions on Localizer intercept heading some approaches were not taken into account for these histograms. These are the approaches for which pilots were forced (due to the airspeed constraints at waypoints) to select flaps 10 before they reached Localizer intercept heading. The sample size for the action "Autothrottle off" is even smaller, due to the fact that this action was not always logged during the second flight simulator experiment.

With respect to the histograms for the actions "Flaps 20" and "Gear Down" some approaches were disregarded because for these approaches flaps 20 had to be selected before the FAF was reached due to the airspeed constraints.

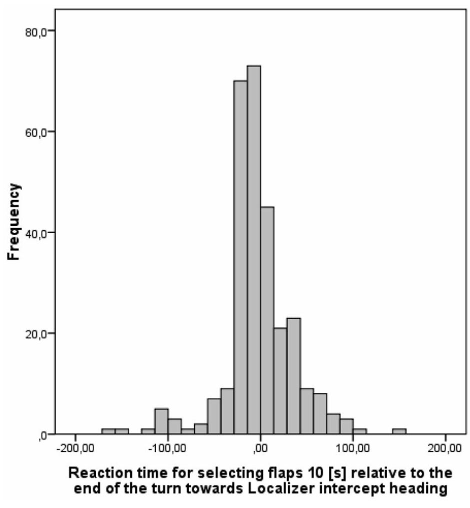
The reaction times and duration of the approach and landing checklists were logged manually during the second experiment, and were not recorded during the first experiment. During the second experiment they were also not always consistently logged, which results in a rather small sample size. The fact that for the approach checklist the sample size for the altitude at which the approach checklist is started is smaller than the sample size for the duration of the approach checklist is caused by the following: some approaches started at 2,000'. Some pilots decided to perform the approach checklist for these approaches anyway, although in reality they would have performed the approach checklist earlier in the flight, when passing transition level. For these situations, the duration of the approach checklist is used in the histograms, but the altitude at which the approach checklist is started (2,000') is not.



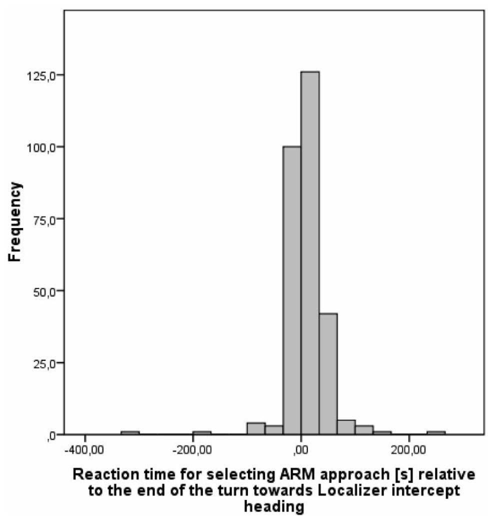
Flaps 1
N = 148
Pilot action is function of IAS



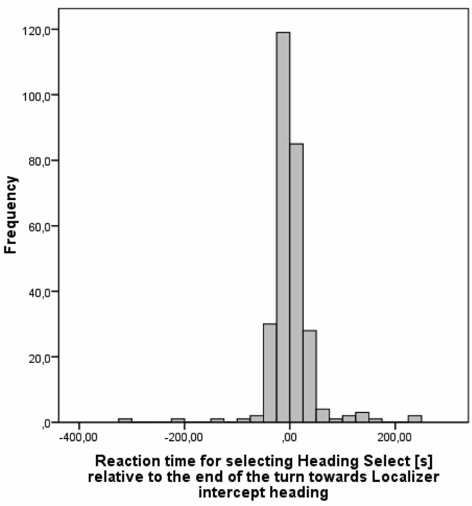
Flaps 5
N = 202
Pilot action is function of IAS



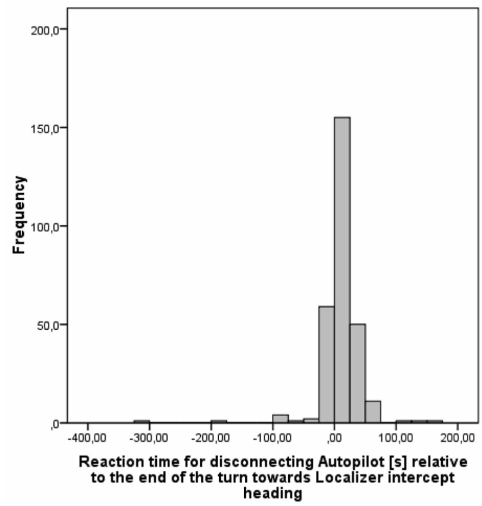
Flaps 10
N = 288
Pilot action is function of time, Δt_{fl10}



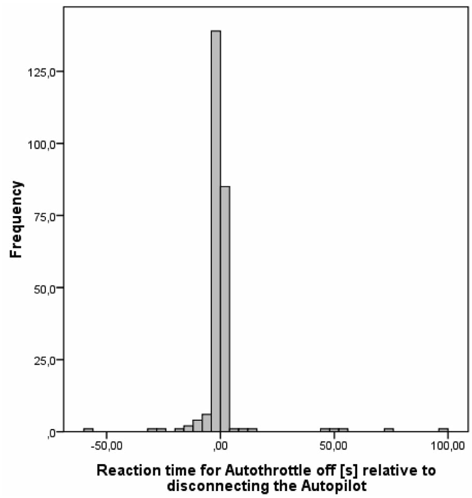
ARM Approach
N = 287
Pilot action is function of time, Δt_{AA}



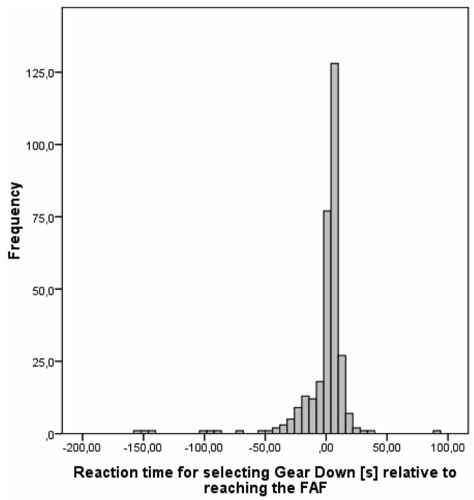
Heading Select
N = 281
Pilot action is function of time, Δt_{HS}



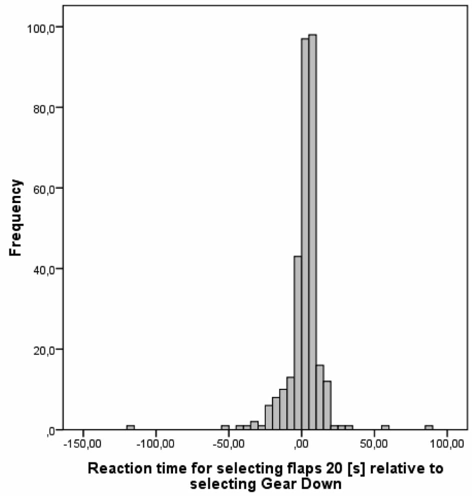
Disconnect Autopilot
N = 287
Pilot action is function of time, Δt_{AP}



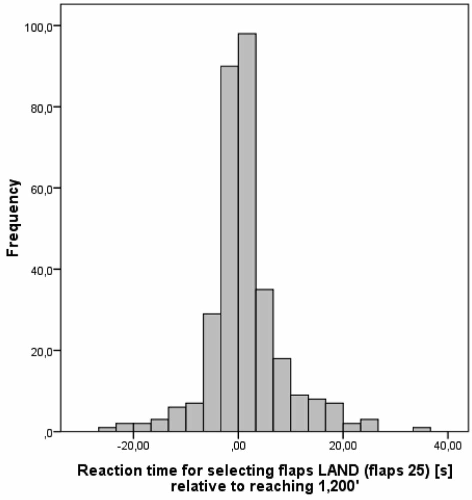
Autothrottle Off
N = 248
Pilot action is function of time, Δt_{AT}



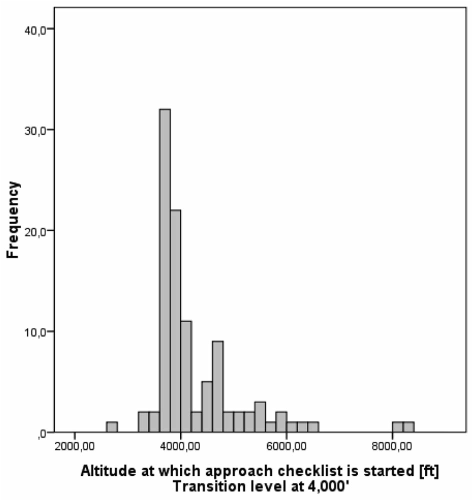
Gear Down
N = 315
Pilot action is function of time, Δt_{GD}



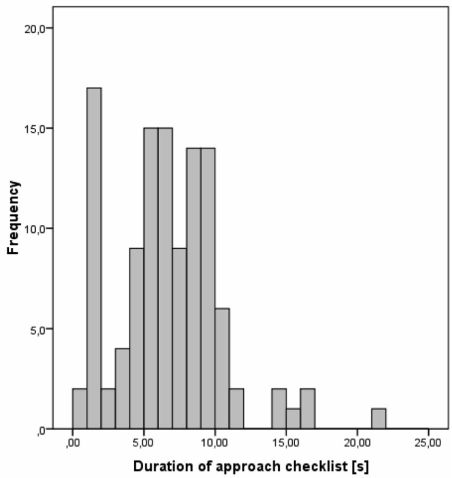
Flaps 20
N = 315
Pilot action is function of time, Δt_{fl20}



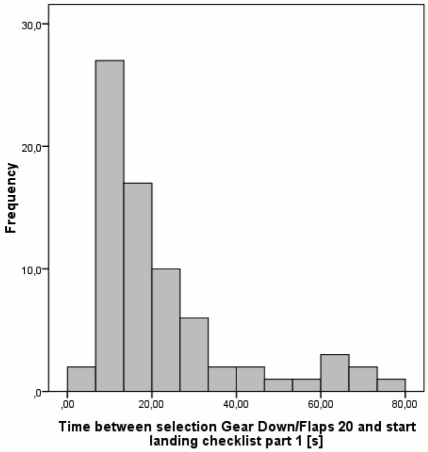
Flaps 25 (Flaps LAND)
N = 321
Pilot action is function of time, Δt_{fl25}



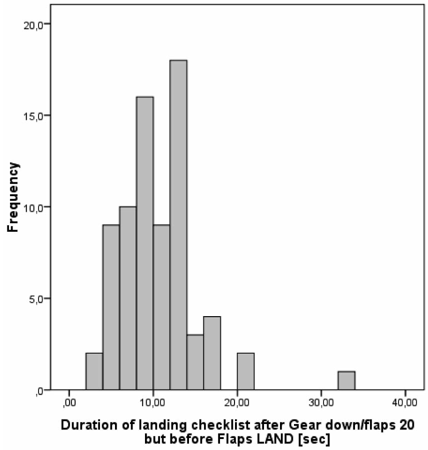
Start Approach Checklist
N = 103
Pilot action is function of altitude



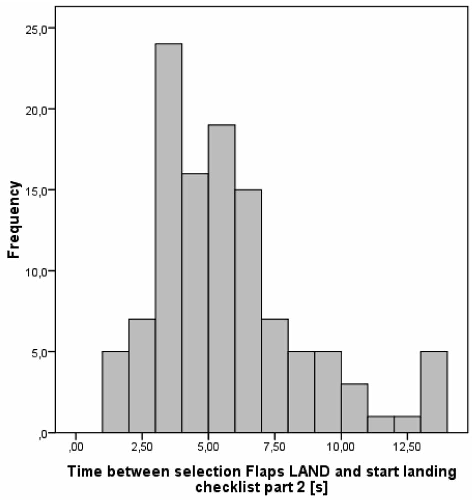
Duration Approach Checklist
N = 115



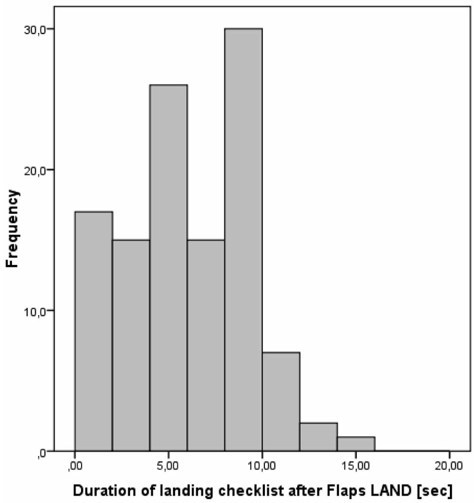
Start Landing Checklist part 1
N = 74
Pilot action is function of time



Duration Landing Checklist part 1
N = 74



Start Landing Checklist part 2
N = 113
Pilot action is function of time



Duration Landing Checklist part 2
N = 113

Appendix II | Example output of the computer simulation to predict pilot task demand load based on a point mass model

This appendix contains the standard report provided for each analyzed approach by the computer simulation based on the point mass model.

Report – APPR103– *12-Jan-2010*

Aircraft Specifications
Aircraft type: **Boeing 747-100**
Aircraft weight: **573397.69 lbs (2550600 N / 260088.8173 kg)**
Speedbrakes: **Off**

Trajectory

The trajectory consists of some waypoints that have a location (consisting of longitude and latitude in degrees), an altitude (ft), an Airspeed (kts) and in some cases a marker which indicates the type of waypoint (IAF = Initial Approach Fix, IF = Intermediate Fix and FAP = Final Approach Point).

Below they are tabulated (see Table A).

Table A | Waypoints and constraints defining the approach.

Waypoint	Lattitude (deg)	Longitude (deg)	Altitude (ft)	Callibrated Air Speed (kts)	Marker
1	52.3287	4.5087	9000	220	IAF
2	52.2907	4.4192	8000	220	-
3	52.2526	4.3298	7000	220	-
4	52.2145	4.2406	6000	220	-
5	52.1602	4.3026	5000	210	-
6	52.1985	4.3915	4000	200	-
7	52.2367	4.4805	3000	200	-
8	52.2084	4.5538	2000	180	IF
9	52.2446	4.6215	1600	170	FAP
10	52.2892	4.7372	0	160	-

The next figure (Figure A) shows the trajectory.

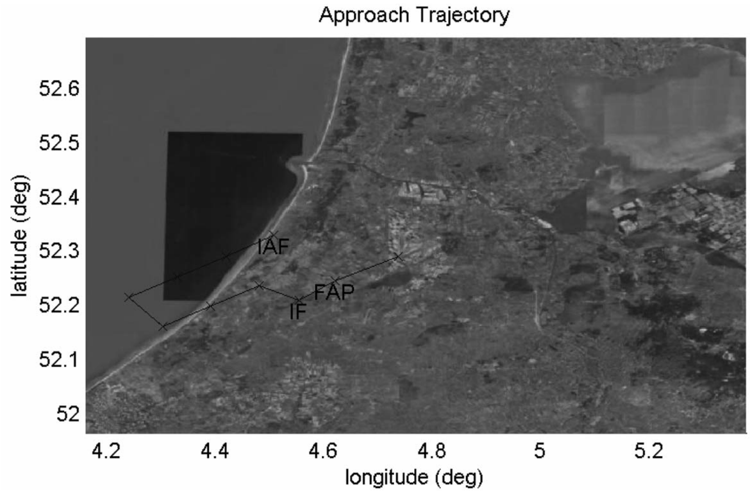


Figure A | Plot of approach trajectory.

Flyability

It is of vital importance to know whether or not every segment of the trajectory can be flown. For each of the segments of the approach trajectory it was checked whether or not the altitude and velocity constraints could be met. If not at every waypoint these constraints are defined, the altitude will be on the glide slope defined by the waypoints around it and the velocity will be whatever velocity it can reach. Optionally this was repeated for different pilot percentages. The pilot percentage determines the time at which the pilot performs his actions, those actions (Flaps 5, Autopilot Off, Gear Down, etc.) are normally distributed around some reference. For example: *'if the pilot percentage is 95, then the action is performed at the moment that (according to research by M. Heiligers, 2008) 95 percent of all pilots would have done so at that time'*.

If the segment was flyable at zero wind conditions, a tailwind was introduced and the segment was analyzed again, to see if with this tailwind it is still flyable. If there was no velocity constraint defined, this was not done. However, in that case, analyzing the next segment that has a velocity constraint requires going back to the last waypoint with a defined velocity before that, to calculate the velocity at the beginning of the segment.

The following table (Table B) shows the maximum allowable tailwinds for all segments of the trajectory versus all pilot percentages (a 'X' indicates that the segment was not flyable and a '-' means that there was no velocity constraint for the waypoint, the maximum tailwind used is 100 kts):

In addition a wind rose can be made. The wind rose shows the maximum wind speeds from all directions, see Figure B. *(Beware of the fact that segments that are not flyable are ignored).*

If the segment was flyable with idle thrust, a bit of thrust was added and the segment analyzed again, to see if with this thrust it is still flyable. If there was no velocity constraint defined, this was not done.

Table B | Maximum allowable tailwinds for all segments of the approach.

Pilots	Segment 1	Segment 2	Segment 3	Segment 4	Segment 5	Segment 6	Segment 7	Segment 8	Segment 9
10%	> 100 kts	> 100 kts	> 100 kts	44 kts	43 kts	89 kts	X	> 100 kts	92
30%	55 kts	50 kts	48 kts	44 kts	43 kts	89 kts	X	> 100 kts	92
50%	55 kts	50 kts	48 kts	X	43 kts	89 kts	X	> 100 kts	99
70%	55 kts	50 kts	48 kts	X	X	89 kts	X	> 100 kts	95
90%	72 kts	65 kts	60 kts	X	X	24 kts	X	> 100 kts	88
95%	72 kts	65 kts	60 kts	X	X	24 kts	X	> 100 kts	85

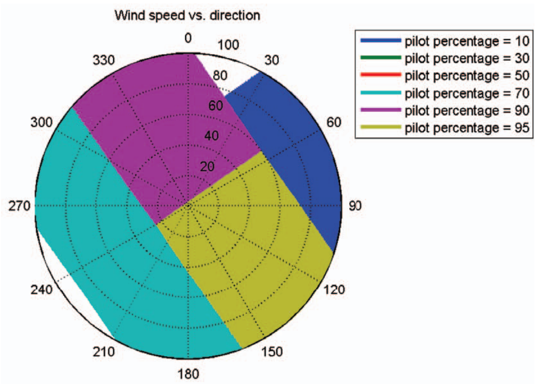


Figure B | Wind rose with maximum allowable wind speeds during the approach.

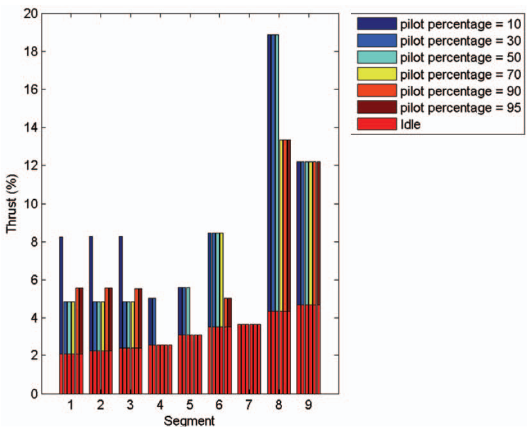


Figure C | maximum allowable thrust as a percentage of the maximum possible thrust for all pilot percentages at every segment of the trajectory.

Figure C shows the maximum allowable thrust as a percentage of the maximum possible thrust for all pilot percentages at every segments of the trajectory. If only the idle part is visible, it means that either the segment was not flyable or there was no velocity defined.

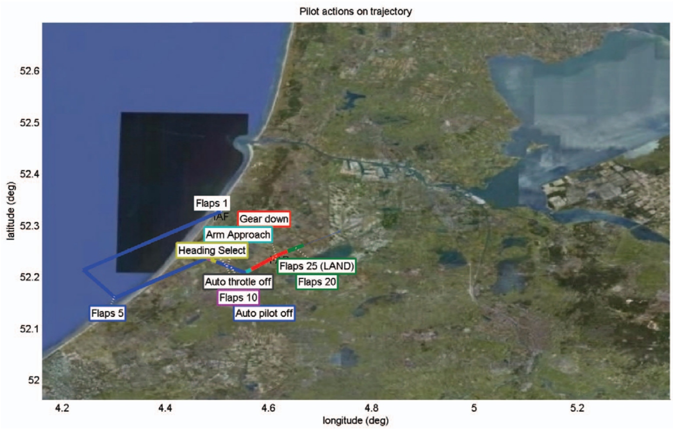
Stabilized Approach

An important factor in determining the flyability of an approach is whether or not the approach is stabilized. In order for the approach to be stabilized it should meet the following requirements at an altitude of 1000 ft:

- Heading and pitch changes are smaller than deg/sec
- CAS is smaller than V_{REF} + 20 and larger than V_{REF} (V_{REF} = 149.2501 kts)
- Flaps are in landing configuration
- Landing gear is down
- Localizer and Glide Slope both are within 1 dot
- Descent rate is smaller than 1000 ft/min

Table C | Percentage of pilots that will achieve a stabilized approach.

Pilots	Stability	Reasons
10%	Stable	-
30%	Stable	-
50%	Stable	-
70%	Stable	-
90%	Stable	-
95%	Unstable	Flaps LAND too late



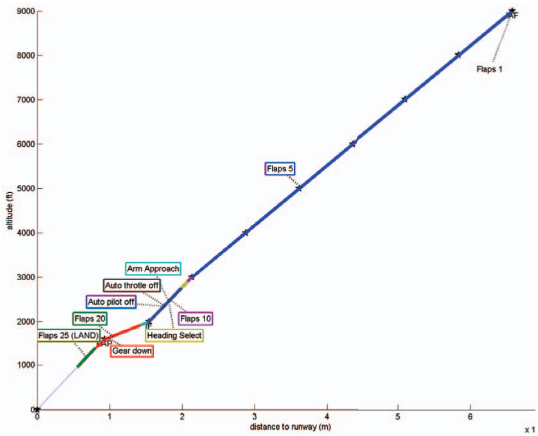


Figure D | Overview of location in approach at which pilot actions are performed.

Pilot Actions

Another convenient thing to know is where and when the pilot will perform the actions specified by the Standard Operating Procedures (SOPs). Figure D will show you at what location on the trajectory each action is performed and the figure after that shows you the same thing on the altitude profile. This is based on a normal distribution around a reference (e.g. a velocity or time of passing a reference point). The colored bars indicate the distribution of positions where the pilot could perform the actions. The labels are placed at the mean to indicate where the 'average' pilot would perform the action.

To show <u>when</u> pilots will perform the SOPs, a timeline iincluded (see Figure E). Here the black areas indicate the range of time different pilots will perform their actions. The earliest time corresponds to 5% of the pilots and the latest to 95% of the pilots. This was done because a normal distribution is used which extends into infinity, so that it is impossible to include all pilots. However, in reality the latest pilot would also perform his action before infinity (i.e. the normal distribution is not valid far away from the mean), so 5-95% is quite a representative range.

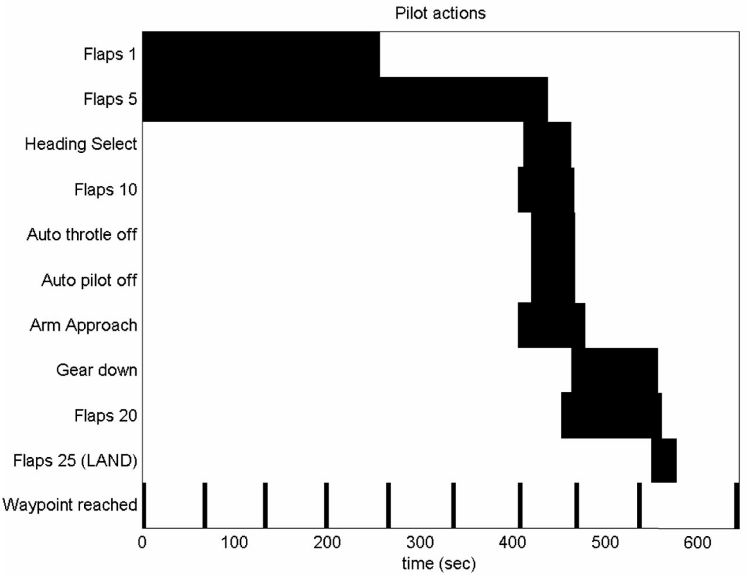


Figure E | Timeline of pilot actions.

PART II

Scientific background

Journal Article 1

Predicting pilot task demand load during final approach

M.M. Heiligers, Th. Van Holten, M. Mulder

The International Journal of
Aviation Psychology, 19(4), 391-416,
October-December 2009

1

Abstract

This research aims at developing a method to assess the safety of airport approaches from the perspective of airline pilots. The main hypothesis is that safety is inversely proportional to the Task Demand Load (TDL) experienced by the pilots conducting it. The Task Demand Load is the mental workload imposed by the system to be controlled, and is not to be mistaken for the mental workload experienced by the operator, referred to as Mental Load. This paper presents the results of preliminary research that focused on how some of the environmental factors that determine the approach, such as the type of aircraft and the meteorological conditions, affect the TDL. A 'paper pilot' model, consisting of a scanning model, mental model and control model, was used to quantify how "hard" an "average" pilot has to work to conduct the specified approach. Off-line computer simulations with this model revealed some clear trends in pilot scanning rate, task error and control activity, as a function of the environmental factors, that will eventually allow for a prediction and quantification of the TDL. Depending on the approach performance requirements on the one hand, and the pilot scanning capacity on the other, increasing turbulence intensity showed a change in trend for all TDL-related parameters, at a particular turbulence intensity. Results from a pilot-in-the-loop evaluation, conducted in a fixed-base flight simulator, supported the main trends of the off-line simulation trials. They also indicated that through adopting different scanning strategies, pilots can leverage their workload; more research is necessary to clarify this result, however.

1 | Introduction

The safety of an airport approach depends on many factors (Figure 1). Factors that affect the execution of an approach directly are the aircraft and its avionics involved, the meteorological conditions, the airport infrastructure, the approach procedures, and last but not least, pilots and air traffic controllers. Factors that affect safety indirectly are maintenance, training and management. Safety has increased tremendously over the past decades, through the use of more advanced technology but also through other, non-technological developments. Nowadays, the majority of incidents and accidents that still occur are attributed to "human error" (NTSB, 2001), and to further enhance safety it is mandatory to better understand the various roles of the human operators involved.

When considering the air transport system as a whole, i.e., everything that is enclosed by the ellipse in Figure 1, several studies have been performed to assess its safety. Roelen et al. (2002) discuss quantitative risk assessment models of air transport that allow a cost-benefit analysis of safety measures. Roelen et al. (2003[a,b]) aim to construct causal safety models that provide objective, quantitative, safety-related information for managerial decision making. Safety studies have also been performed on the human factors aspects of maintenance and management, introducing concepts like "safety cultures" (Hudson, 2001[a,b], Westrum, 1997, Westrum et al., 1999, Reason, 1990).

Several risk assessment methodologies are reported (Blom, Klompstra, Bakker, 2001, Blom et al., 2001, Everdij et al., 2002, van Baren et al., 2002) that aim to analyze the safety of an approach from the perspective of primarily the direct factors of Figure 1. In (Suarez, 2003) the safety of a new approach is assessed through flight tests, a qualitative study of identified hazards, and conversations with pilots and air traffic controllers. Blom et al. (2001) assess safety both qualitatively as well as quantitatively, including human performance models formulated in terms of dynamically-colored Petri-nets that combine the cognitive modes of Hollnagel (1993) with multiple resource theory (Wickens and Hollands, 2000), Reason's classical slips/lapses model (Reason, 1990), and the human capability to recover from errors (Amalberti and Wioland, 1997).

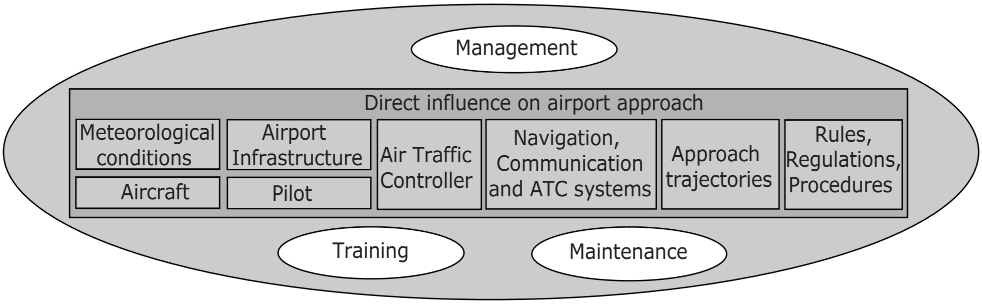


Figure 1 | Direct and indirect factors that influence the safety of airport approaches.

Further narrowing the focus to the *human operators* that have a direct influence on the approach, it can be noted that including elaborate human performance models in safety assessment methods is not new. Already in the late 1970s a model for analyzing flight crew procedures in approach to landing has been developed, the procedure-oriented crew model, PROCRU for short (Baron et al., 1980). Based on concepts of optimal control and estimation theory, it provided results in terms of, among others, each crew member's estimate of the aircraft state, their attention allocation, and their control actions, all as a function of time. PROCRU was never validated experimentally, however. And because much experience was required in dealing with the many control and estimation-theoretical aspects of PROCRU in order to use it confidently, it was rarely ever used.

More recent efforts to model human operators in approach-related tasks have been conducted as well. An extensive overview of methods and tools that have been developed for the air traffic controller can be found in (GAIN working group B, 2003). The US Army, NASA, and Sterling Software developed MIDAS, the Man-Machine Integrated Design and Analysis System (Corker, 1999, Gore and Corker, 2000, Smith and Tyler, 1997). MIDAS is a simulation system that includes human performance models that can be used to evaluate candidate crew procedures, controls, and displays. It is used in the early stages of conceptual design, and allows designers to use computational representations of the crew station and operator, instead of conducting hardware simulations and man-in-the-loop studies. MIDAS allows an analysis of the crew station layout for assessments of visibility, legibility, anthropometric aspects, and analyses of cockpit topology and configuration. It also provides facilities to run human operators models, cockpit equipment and mission procedures in an integrated fashion, resulting in activity traces, task-load timelines, information requirements, and mission performance measurements (Gore and Corker, 2000).

Common to all these safety assessment methods and the human operator models incorporated in them, however, is that they have become quite complex, and require considerable experience to apply them properly. The models often contain many "parameters" that need to be "tuned", and because most of them are not validated experimentally, it is sometimes difficult to initialize the model, as typical values for these parameters are unknown.

Furthermore, in attempts to increase their applicability and versatility, the models often include components that account for particular aspects of human behavior, sub-models that have often been developed and validated, however, in a different context. As a consequence, many of these methods and models often do not allow useful predictions to be made about situations that are beyond the conditions for which the models have been tuned. Finally, some of the methods are composed of proprietary parts of software, sometimes developed for military applications, and therefore (parts of them) are simply not available for others, or service in using the methods and models is not available.

For these reasons, stimulated by the need of local authorities, airlines and air traffic control organizations, and funded by the Dutch government, Delft University of Technology has initiated a project to develop a new method to address the safety of airport approaches from the perspective of pilots. The method should allow designers of airport approach trajectories to analyze how "safe" a new approach trajectory is with respect to existing approaches, like Standard Arrival Routes (STARs). The main goal of the project is therefore to develop a method, captured within a computerized analysis tool, that allows a rapid assessment of the safety of approach trajectories.

The fundamental hypothesis in this project is that the safety of flying a particular approach is inversely proportional to the workload imposed on the pilots conducting it. We believe this hypothesis is a reasonable one. Obviously, workload must remain within certain limits. When pilots are very busy with their primary task of controlling the aircraft along a complicated approach trajectory in adverse weather conditions, the heavy load imposed on them by the task leaves them with little time to do other tasks, not to mention the case where they have to deal with abnormal situations like an engine failure. Heavy loads generally increase the likeliness of human error (Reason, 1990). When the load imposed by the task becomes very low, however, the probability of human error is also likely to increase (Campbell, 1999). It is assumed, however, that during an approach, workload is always sufficiently large to sustain the inversely proportional relationship between safety and workload.

Many definitions and interpretations exist of (mental) workload, accompanied with various methods to quantify it. Therefore, to formulate more accurately, our research aims to develop a method to quantify the Task Demand Load (TDL) of conducting an airport approach. Task demand load is defined as the mental workload *imposed* by the system to be controlled or supervised (Stassen, Johannsen & Moray, 1990). The task demand load is not to be mistaken for the mental workload *experienced* by the human operator, which is referred to as Mental Load (ML). Many of the well-known methods to measure workload, like the NASA Task Load indeX, measure ML, not TDL. A common approach to measure task demand load is modeling, and that is the approach adopted here. But in order to make the model more accessible and easier to use for approach designers, that have often little background in human operator modeling (or none), our aim is to keep the model as simple as possible, and also to validate its parameters experimentally both in flight simulators as in real flight. Moreover, by specifically focusing on assessing TDL during approaches, and limiting the research to this particular aspect of human performance, the pitfall of making a comprehensible, "all-including" human performance model is prevented.

The TDL assessment method is intended to act as an additional tool that facilitates the design of new approach trajectories and procedures. With this tool, the designer should be able to rapidly evaluate a potential approach from the perspective of the demands it imposes on the pilot's limited resources, as a function of time. It will allow the designer to adapt existing or prospective approaches in such a way that the task demand load during the entire approach is kept at an acceptable level, before it is tested by flight crews in flight simulators. The method should be able to incorporate all factors that have a direct influence on an airport approach

(Figure 1), most importantly the characteristics of the trajectory, the type of aircraft and the meteorological conditions. All indirect factors in Figure 1 are not included.

Note that our approach results in an estimate of the "nominal" TDL, and does not include situations like engine-failure. It goes without saying that TDL increases significantly in these cases, and properly modeling events like these is also much more difficult. When such abnormal events occur, however, it is very likely that the approach procedure, and the approach trajectory that corresponds to it, is changed, and the aircraft is directed to the runway as soon as possible.

At the heart of the project lies the development of a simulation program that incorporates the aspects that may affect the task demand load of conducting an approach, including procedures, altitude-profiles, velocity-profiles, but also alternative techniques for conducting noise-abatement maneuvers, such as delayed gear, reduced flaps, steeper approach angles, higher approach altitudes and curved approaches. It should also be possible to enter different types of aircraft, to change the meteorological conditions (turbulence intensity, amount of crosswind), the airport infrastructure and emulate communication with air traffic control. These properties are the descriptors of the environment that form the "input" of the computer program as they constitute the specific characteristics of the approach to be evaluated. The "output" of the program will be a quantitative indication of how the task demand load evolves during (specific parts of) the approach.

The question then becomes how to arrive at an objective prediction of task demand load of a particular airport approach, as a function of the various environmental factors (trajectory, aircraft, weather, …) that together constitute an approach. Our approach is to employ a relatively simple pilot model to simulate manual flight along the proposed approach trajectory, in order to quantify how "hard" an "average" pilot has to work to conduct the approach. Manual flight control is chosen rather than automatic control, as the manual control task can be considered a worst-case scenario. It will indicate the approach designer the TDL that is likely to be the maximum level in nominal situations, and that in normal flight when pilots use their automatic controllers to conduct the approach, TDL is generally (much) lower. As will be discussed in more detail, the pilot model used in this preliminary phase of the project contains a mental model, scanning model, and control model. Our intention has been to keep these components as simple as possible, as we do not aim at replicating the exact control actions and eye movements as they are performed in real flight, but only to obtain an "on average" indication of "how hard" pilots have to work. Thus, we deliberately deviate from previous human performance modeling approaches, such as PROCRU, where one attempts to construct a detailed model of the human operators involved, including all their limitations, both physically and mentally. Rather, we intend to use a simpler model and use it to understand how the environmental constraints affect the task demand load. In this respect, our approach was influenced by the principles of cognitive work analysis (Vicente, 1999), where emphasis is shifted from detailed investigations of the human operator limitations (like memory capacity, time delay, etc.) to analyzing and describing how the environment affects human behavior.

The goal of this paper is to present the preliminary results of our project on how to measure the task demand load of pilots during manual approaches. It is structured as follows. First the pilot model will be discussed in detail, followed by a description of results obtained with a series of off-line simulations. Second, the results of an experimental evaluation of our approach to determine the task demand load, conducted in a fixed-base flight simulator, will be discussed. The paper ends with the conclusions.

2 | A preliminary pilot model to determine task demand load

The aim of this project is to investigate how the various direct factors that define an airport approach affect the TDL. Because we intend to focus particularly on how the environmental constraints affect TDL, our aim is to incorporate a pilot model that remains relatively simple. To verify whether such a model can give useful insights into the task demand load of an approach, a first test case has been considered (Heiligers, 2003). With regard to the direct factors, Figure 1, the meteorological conditions are included (in terms of turbulence level), the aircraft dynamics are modeled and the trajectory that needs to be flown is defined (a horizontal recti-linear flight). This is a limited set-up, and will be of limited use to analyze the whole approach, but it can serve very well as a first reality-check. The pilot model consists of three sub-models, Figure 2, that are included to describe pilot scanning, memory and control behavior. Each of these three sub-models will be introduced below.

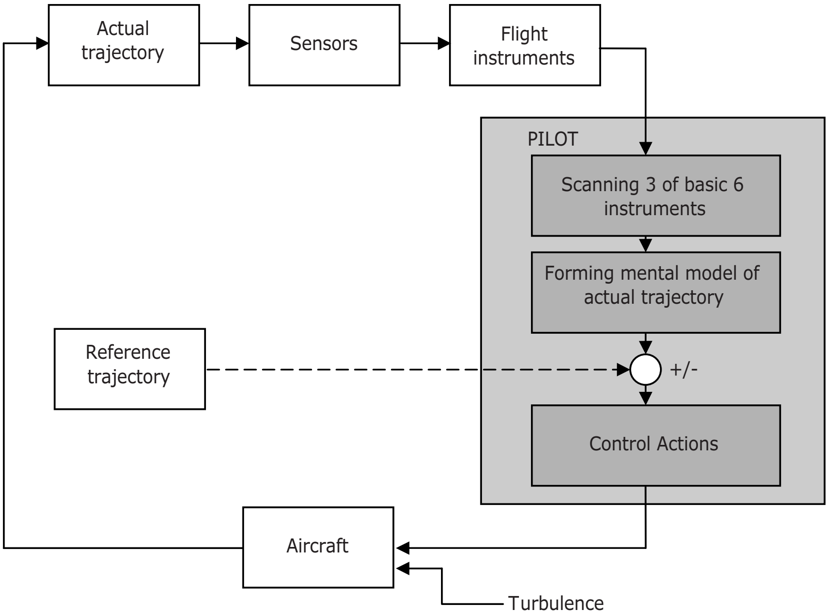


Figure 2 | Model of the pilot/aircraft system for a trajectory-following task in the presence of turbulence.

2.1 | Scanning model

Results from the literature. The visual sampling, or scanning, of the cockpit instruments is a process that consists of saccades, i.e., jerky eye movements where the eye fixation jumps from one point in the visual field to the other. Each fixation can be characterized by, among others, a location (the fixation center) and a dwell time (the fixation duration) (Wickens and Hollands, 2000). To model pilot scanning it is thus necessary to be able to give a reliable prediction of the fixation *sequence* among the instruments, the fixation frequency (how often a specific instrument is fixated on) and the dwell time for each instrument.

Senders (1964) derived such a model, based on the hypothesis that the fixation frequency on a certain instrument only depends on the bandwidth of the signal displayed by that instrument. In the case of an ideal observer the fixation frequency should increase linearly with bandwidth. Shannon's theorem states that the fixation frequency should equal two times the signal cut-off frequency. In Senders' model the dwell time depends on the amount of information to be taken at each observation, and the fixation sequences are predicted by using transition probabilities between instruments. Senders' experiments showed that the predicted transition probabilities were very close to the values obtained experimentally, and that the observed fixation frequencies were monotonically increasing with signal bandwidth, although not entirely according to two times the cut-off frequency.

Carbonell extended Senders' model with the concept of 'relative costs', arriving at a "queuing model" of visual sampling (Carbonell, 1966 and Carbonell, Ward & Senders, 1968). At each sampling interval, it is decided which instrument will be sampled next by calculating the costs of not looking at an instrument for each and every instrument. "Cost" was defined as the product of the probability of a signal exceeding some particular 'threshold', times the 'real cost' of exceeding that threshold. E.g., exceeding an altimeter threshold has a high 'real cost' during landing, but a low 'real cost' during cruise flight, see (Carbonell, 1966). The instrument that has the highest cost will be the next instrument to be sampled. Experiments showed that the model was capable of accurately representing the behavior of pilots visually sampling their instruments during an instrumented flight.

Bellenkes, Wickens & Kramer (1997) compared scanning behavior of novice and experienced pilots during different flight phases. They report that experts universally tend to visit instruments more frequently, while novices tended to dwell for a longer time on each instrument. Experts also differed from novices in terms of a fixation pattern that guarded against unwanted "tunneling". That is, whereas novices tended to focus on the instrument that indicates the main changing variable during a maneuver, experts remained a fixation pattern across more instruments, especially those indicating cross-coupling effects related to the maneuver (e.g., a loss of altitude during a heading change).

Wickens, Goh, Helleberg, Horry & Talleur (2003) present a model for visual scanning, not across separate flight instruments but across three primary areas of interest (AOIs): the instrument panel, the outside world and the cockpit display of traffic information. The proportion of percentage dwell time allocated to each of these three AOIs is related to the bandwidth along the AOI, the degree of relevance (R) of each AOI to the most critical task and the value of the task supported by the AOI. By assigning values to these factors they were able to obtain a good fit of the experimental data.

Scanning model implemented. Previous models developed by Senders and Carbonell aim at modeling the scanning behavior as precisely as possible. This requires a number of variables that needs to be determined for each instrument that is sampled, a number that increases rapidly when more instruments are to be considered. In the present study it is not the intention, however, to model scanning behavior as precisely as possible, but merely to model an "average" scanning sequence that is able to predict the amount of information pilots need to derive out of their instruments in order to conduct their task. Therefore, a fixed scanning sequence is chosen that corresponds with the scanning cycles pilots are taught during their training (KLM Luchtvaartschool, 1990). The scanning cycle for straight and level flight is depicted in Figure 3. The attitude indicator is the main instrument and confirmation about the selected attitude is given by the altimeter and the course indicator. The fixation sequence is: attitude indicator – course indicator – attitude indicator – altimeter – attitude indicator –course indicator, etcetera. Thus, in terms of relevance (Wickens et al., 2003) the attitude indicator is twice as 'relevant' as the altimeter and course indicator. The fixation sequence is fixed, but the scanning frequency can still be changed. Senders' model (1964) predicts that with increasing bandwidth (e.g., increasing turbulence intensity in Figure 2) the fixation frequency must increase in order to stay within the specified accuracy margins.

In our case, however, the instruments are not considered separately but rather in a series of fixations. Hence, Sender's prediction in our context means that the sequence of fixations across instruments will be repeated more frequently when bandwidth increases.

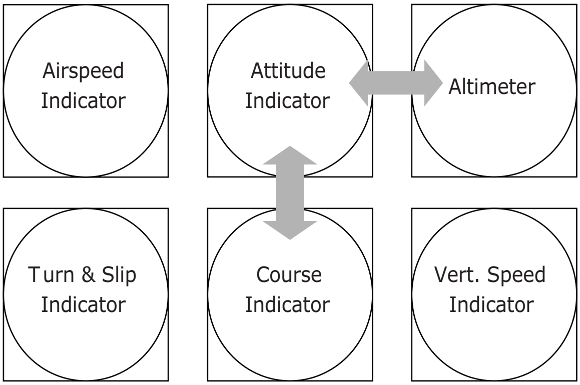


Figure 3 | The primary scanning cycle for horizontal, level flight.

The dwell times on each instrument are chosen as fixed constants. Harris (1980) reported that the dwell times of altimeter, attitude indicator and heading indicator all equal approximately 0.5 seconds. Bellenkes et al. (1997), however, reported a dwell time for the attitude indicator of approximately 0.6-0.75 seconds for experienced pilots. Since dwell times on altimeter and course indicator were not given in Bellenkes et al. (1997), as a starting point, an average dwell time of 0.5 seconds is assumed in the simulation for *all* instruments, test results will show whether this value will need to be adjusted. Note that although the course indicator is part of the scanning cycle, and time is reserved in the simulation for the paper pilot to 'look' at it, it will not be used to perform correcting control actions since the task does not include lateral control. Further, it is assumed that no false readings or misinterpretations of the flight instruments occur.

It can now be determined how fast the pilot has to scan the instruments in order to perform a horizontal flight within certain accuracy margins. When the pilot needs to scan faster this can be seen as an indication of the fact that the pilot needs more information to achieve a certain level of performance. Although the scanning model does not represent exact scanning behavior, the scanning rate can be considered to say something about the amount of information needed and the accompanying task demand load.

2.2 | Mental model

The mental model represents the pilot's memory. Naturally, pilots will not only react on the values that they read from the instruments at a certain moment in time, but rather combine these values with values scanned previously. To model this process, a First Order Hold (FOH) mechanism is applied for each instrument. The pilot remembers the values of the new and previous scans and, based on these two values, computes an expected value by linear extrapolation in time. This extrapolation ends when a new scan is available, at that time a new extrapolation starts again, see Figure 4.

Figure 5 shows how the variables are transformed while they pass through the computer simulation. The aircraft model generates a continuous signal that is displayed on the flight

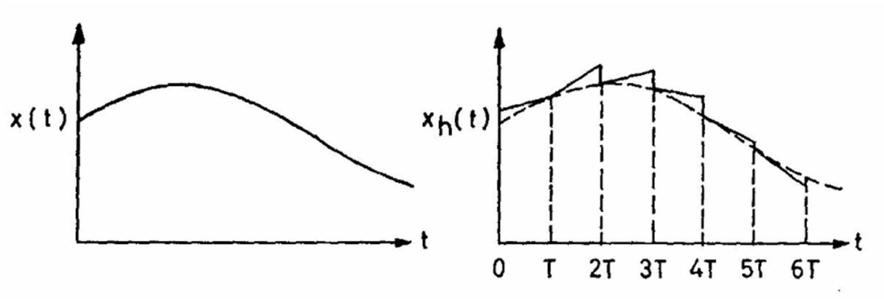


Figure 4 | Example of first order hold system (Verbruggen, 1982).

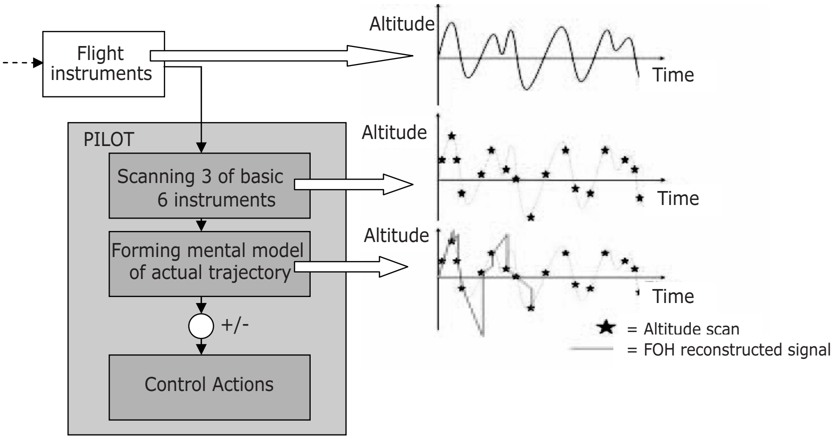


Figure 5 | Variable transformation when passing through the computer simulation model.

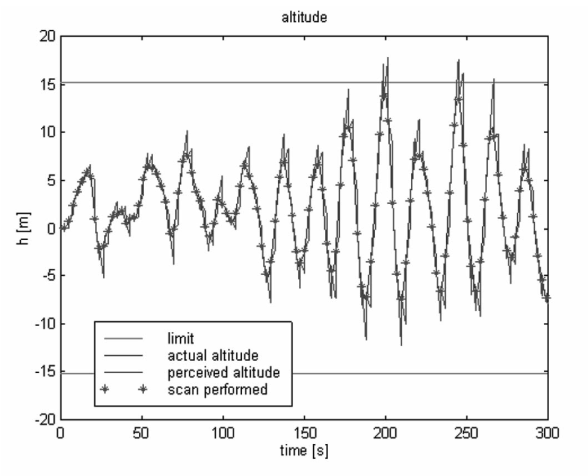


Figure 6 | Example of the accuracy of the scanning model and mental model.

instruments and is sampled at discrete moments in time by the scan model, the mental model then transforms these discrete samples into a new continuous signal which can be used by the control model. The accuracy of the values generated by the mental model depends on the scan frequency and the frequency-content of the signal being perceived. In our simulations, the first order hold approximation performed reasonably well, see Figure 6.

2.3 | Control model

A similar approach is chosen for the pilot control sub-model as for the scanning sub-model. Again it was decided to keep the model simple, to get a first indication of how hard the pilot has to work, rather than mimicking the pilot actual behavior as closely as possible. Instead of using the crossover model (McRuer and Jex, 1967) or Optimal Control model (Baron, Kleinman and Levison, 1970 and Kleinman, Baron and Levison, 1970) in this preliminary survey a basic control model is chosen.

In the simulation, the 'paper' pilot will perform correcting control actions based on the values that are generated by the mental model. Since the pilot scans the attitude indicator and altimeter these values will be altitude (h) and pitch angle (θ). The only control that is taken into account is the aircraft elevator as, in this simulation, only longitudinal motion is considered. Preliminary tests with the simulation models showed that the turbulence did not cause the aircraft airspeed to deviate more than 10 knots (the established accuracy margin explained in the next section), even without pilot controlling throttle. Therefore the control of the aircraft velocity could be discarded in this preliminary study, and the only variable to be controlled by the 'paper pilot' was altitude.

The following equations show in what way elevator deflection (δ_e) was computed from the altitude and pitch angle that the pilot scanned from the flight instruments:

$$\delta_e = K_{\delta h}\left(h_{ref} - h\right) + K_{\delta\theta}\left(\theta_{des} - \theta\right), \tag{1}$$

$$\theta_{des} = K_\theta\left(h_{ref} - h\right), \tag{2}$$

These equations represent a simplified version of an altitude-controller reported in (Etkin & Reid, 1996). The reference altitude (h_{ref}) is defined by the reference trajectory and the desired pitch angle (θ_{des}) depends on the altitude deviation from the reference altitude. The altitude h and the pitch angle q represent the altitude and pitch angle generated by the pilot mental model, perceived and extrapolated from the altimeter and attitude indicator, respectively. The controller gains $K_{\delta h}$, $K_{\delta\theta}$ and K_θ are functions of the scanning interval, i.e., the time between two successive fixations on the same flight instrument. This is based on the Ziegler-Nichols rules which state that for a discrete system the effect of the proportional gains needs to be diminished and the effect of the integration and differentiation gains needs to be amplified when compared to a continuous system (Verbruggen, 1982). The controller gains can be expressed as, with the numerical values valid for this particular simulation set-up:

$$K_{\delta h} = \frac{-9\cdot 10^{-3}}{40\cdot T_h}, \text{ in [rad/m]} \tag{3}$$

$$K_{\delta\theta} = \frac{-\ 0.7}{20\cdot T_\theta}, \text{ in [-]} \tag{4}$$

$$K_\theta = \frac{0.006}{20\cdot T_\theta}, \text{ in [rad/m]} \tag{5}$$

In these equations, T_h and T_θ represent the scanning intervals, i.e., the time between two successive scans, for the altimeter and attitude indicator, respectively. In a qualitative sense, this means that as the scanning interval for an instrument becomes larger, the first order hold model will extrapolate the scanned values over a longer period of time (see Figures 4, 5 and 6) and the original signal will be reconstructed less accurately. The mental model becomes less reliable, especially when rapid deviations occur, and therefore the control model does not react as 'confidently' on the data provided by the mental model as compared to the situation where the instruments are scanned more frequently, i.e., the control gains become smaller.

Similar to the scanning rate, the control activity of the paper pilot can be considered a metric of the task demand load, as it indicates the amount of control actions necessary to perform the task.

Summarizing, since we focus on how the environmental constraints affect task demand load the pilot model is kept relatively simple and only consists of a scanning model, mental model and control actions model. The scanning rate and control activity can be used as metrics for the task demand load. In the next section, it will be examined whether the preliminary pilot model can generate useful results with respect to the task demand load, using off-line simulations.

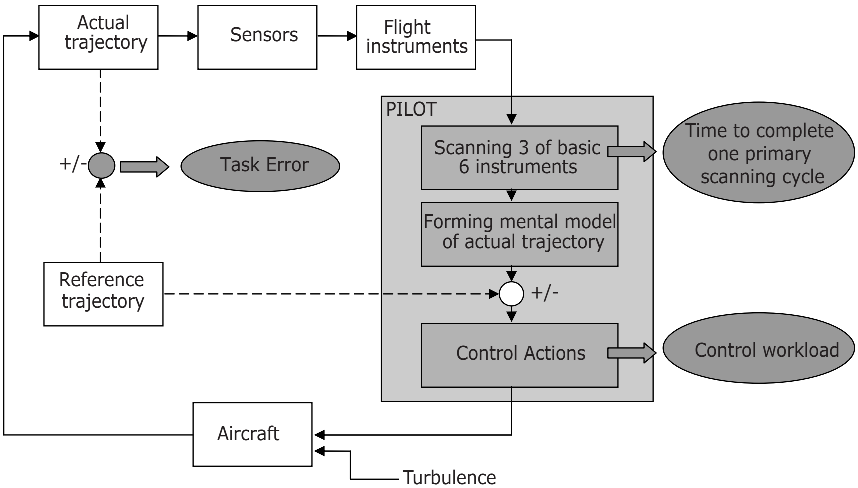


Figure 7 | Schematic representation of computer simulation model and metrics for the Task Demand Load

3 | Off-line simulation results with the preliminary pilot model

The preliminary pilot model described in the previous section is incorporated in a computer simulation program, to check whether the results of the model can be used to predict the task demand load. The factors that have a 'direct' influence on the airport approach (Figure 1) included in the computer simulation are the aircraft and the meteorological conditions in terms of turbulence, illustrated in Figure 7. The computer program simulates a horizontal, recti-linear flight through turbulent air. The aircraft is disturbed by atmospheric turbulence, and the pilot compensates for the effects in a closed loop manual control task. The 'paper pilot' "scans" the flight instruments, and based on the deviations from the reference trajectory performs correcting control actions. In this section, first the simulation components other than the pilot model will be discussed in more detail, followed by a description of the metrics for the task demand load that are derived, and will finally present the results of off-line simulations.

3.1 | Modeling the 'direct' factors that affect an airport approach

Aircraft dynamics. The aircraft was modeled according to the linearized longitudinal equations of motion (Etkin, 1972). Thus, in this preliminary set-up only motion in the vertical plane was considered. The choice for linear equations of motion and only motion in the vertical plane was made in order to keep the computer simulation simple. If this test case proves its usefulness, it will be extended to the six degree-of-freedom non-linear equations of motion. The aircraft used in the computer simulation is a Cessna Ce-500 (Cessna Citation I), a small two-engine jet (airspeed equal to 128 m/s (256 knots), altitude 5,000 m).

Atmospheric turbulence. Turbulence was modeled according to the Dryden spectra (Etkin, 1972). Generally speaking, the strength or "level" of turbulence is determined through two parameters, i.e., the turbulence intensity σ and the turbulence scale length L_g. The quotient of both relates to the variance of the gust velocities encountered by the aircraft. That is, when the turbulence intensity increases, so does the variance of the gust velocities. When the scale length increases, the variance of the gust velocities becomes smaller. Levels of turbulence intensity applied in the computer simulation ranged from 0 m/s to 3.2 m/s, with the turbulence scale length equal to 300 m. According to Houbolt (1964) turbulence intensities can be categorized as follows: clear sky turbulence = 1 m/s, cumulus = 1 – 3 m/s and a thunderstorm = 2 - 5 m/s.
Approach trajectory and accuracy. In this preliminary test-case, the *reference trajectory* (see Figure 7) is a horizontal recti-linear trajectory, demanding the pilot to maintain a level flight and correct for disturbances, i.e., a typical disturbance-rejection task. It is clear that the task demand load depends on, among other things, the accuracy with which the turbulence rejection task needs to be performed. To determine reasonable accuracy margins, necessary to "tune" our pilot model in the off-line simulations, a questionnaire was handed out to eight professional airline pilots, inquiring about what deviations in altitude and velocity they considered acceptable during initial and intermediate approach. Based on this questionnaire, it was concluded that altitude should remain within 50 feet of the reference altitude, and airspeed should stay within 10 knots of the initial airspeed.

3.2 | Metrics for the task demand load

Three metrics for the task demand load are derived from the computer simulation:

Time to complete one primary scanning cycle. For every turbulence intensity it is investigated how fast the paper pilot has to scan the flight instruments in order to perform the turbulence rejection task within the established accuracy margins. This is expressed in terms of 'maximum time to complete one primary scanning cycle', indicating how much time the pilot can allow himself to perform one primary scanning cycle: the higher this amount of time the slower the pilot can scan the instruments. A limit is imposed by the fact that the pilot needs at least 0.5 seconds per flight instrument, and therefore at least 2 seconds to complete a primary scanning cycle. A decrease in time to complete one primary scanning cycle indicates an increase in task demand load.

Task error. The task error is defined as the difference between the actual trajectory and the reference trajectory in terms of altitude and airspeed. An increase in task error indicates an increase in task demand load.

Control workload. According to Padfield (1996) the root mean square of the control activity can be used as a control workload metric, where control activity is given by the power spectral density function of the control column deflection. Control column deflection is not known within the computer simulation, but assuming linearity between elevator deflection and control column (tests on a Cessna Citation I proved that this is a valid assumption for this type of aircraft), the control activity can be derived from the power spectral density of the elevator deflection. An increase in control workload indicates an increase in task demand load.

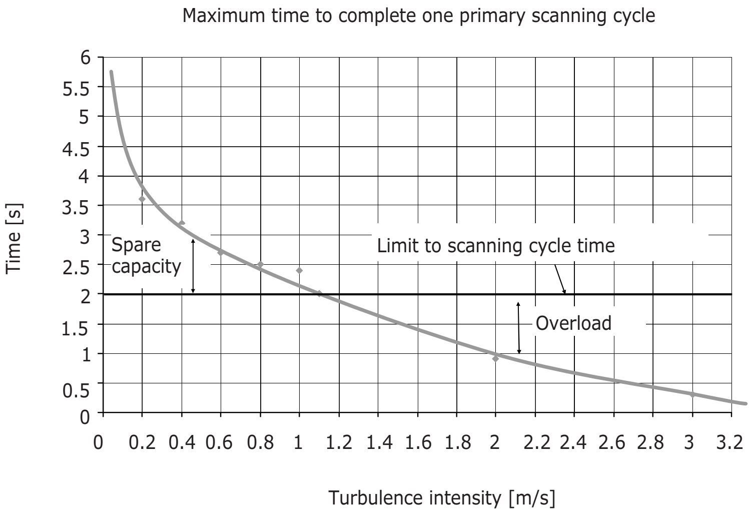


Figure 8 | Maximum time to perform one scanning cycle as a function of turbulence intensity in order to stay within altitude and velocity limits

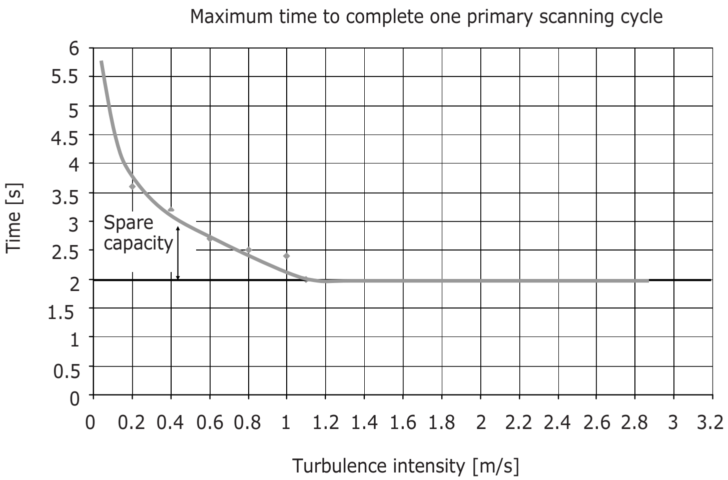


Figure 9 | Maximum time to perform one scanning cycle as a function of turbulence intensity in order to stay within altitude and velocity limits when physical limitations are taken into account.

3.3 | Results of the computer simulation

Time to complete one primary scanning cycle. Figure 8 shows the results for the time to complete one primary scanning cycle: for very small turbulence intensities (smaller than 0.05 m/s), the 'pilot' does not need to scan the flight instruments, for the turbulence will not cause the aircraft (here the Cessna Citation I) to exceed the performance requirements for altitude and airspeed. The higher the turbulence intensity, however, the less time the pilot has to perform a scanning cycle. Taking the lower limit of 2 seconds for a primary scanning cycle into account, Figure 8 transforms into Figure 9: with increasing turbulence the time to complete one primary scanning cycle will decrease to 2 seconds, and from there on will remain constant. In the remainder of this paper, the turbulence intensity where this occurs will be referred to as the 'transition turbulence intensity'.

Task error. The fact that from the transition turbulence intensity the time to complete one primary scanning cycle remains constant will have an influence on the task error. The pilot cannot scan the flight instruments as fast as necessary in order to remain within the accuracy margins, and as a consequence, task error in terms of altitude will exceed 50 feet, see Figure 10. Note that the task error increases linearly because both the aircraft and pilot models in the simulation are linear. The applied turbulence did not cause the aircraft airspeed to deviate more than 10 knots from the initial airspeed, even without the pilot controlling throttle.

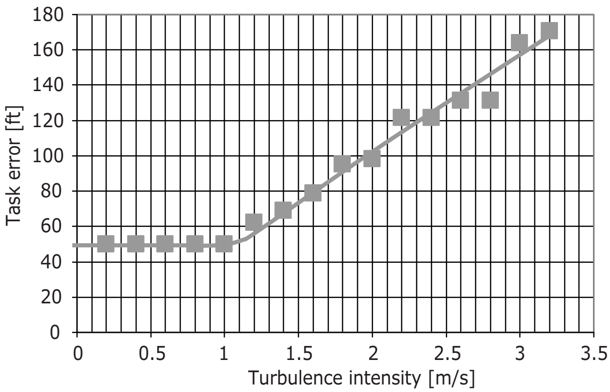


Figure 10 | Task error as a function of turbulence intensity

Control workload. The results for the control workload as a function of turbulence intensity are given in Figure 11: control workload rises slowly as long as the turbulence intensity is lower than the transition turbulence intensity. For turbulence intensities larger than the transition turbulence intensity, the control workload rises much faster. The scattering of the data points in Figure 13 can be explained by the fact that turbulence is a random process. When repeating the experiment the scattering would average to what is, similarly to the task error, a linear increase of control workload with turbulence intensity.

Overall results of computer simulation. The computer simulation predicts a relation between the time the pilot has to perform one scanning cycle, the task error and the control workload. With increasing turbulence intensity the time to perform one scanning cycle becomes smaller and will finally stabilize at a minimum value which results in a bend or knee in the graph. At the same turbulence intensity at which this bend occurs, a bend also occurs in the graph for task error. In the region to the left of this bend a constant task error is obtained equal to the given accuracy margin, in the region to the right of this bend the task error will increase linearly.

Control workload as a function of turbulence intensity is relatively constant (or increases slightly) before the bend occurs. Beyond the bend, control workload shows a linear relation with turbulence intensity. In each graph, the bend occurs at the same turbulence intensity: the transition turbulence intensity. Up to this intensity, part of the turbulence increase can be taken care of by increasing the scanning rate. Above the transition turbulence intensity, an increase in turbulence intensity results in an increase in control workload.

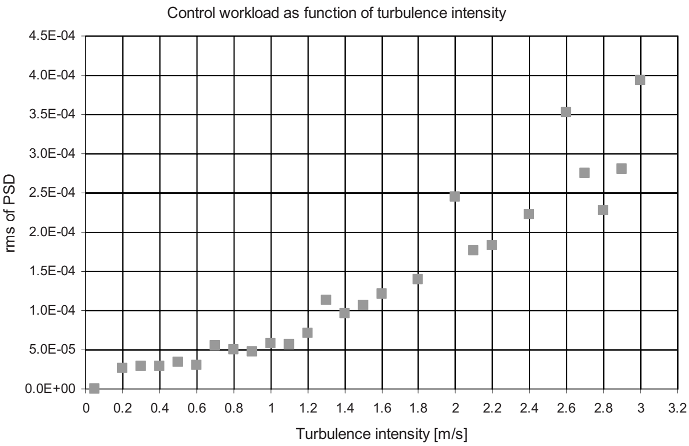


Figure 11 | Indication of control workload as a function of turbulence intensity

The transition turbulence intensity depends on many factors. For example, if the imposed accuracy margins are increased, i.e., there is more room for trajectory deviations, the transition intensity moves to the right as the pilot will still be able to achieve adequate performance for higher turbulence intensities. Another factor that influences the location of the bend is the minimum time the 'paper pilot' needs to perform one primary scanning cycle. When the lower limit in the scanning figure (Figures 8 and 9) moves downward, i.e., when the pilot can scan faster and less time is needed to perform a scanning cycle, the bend in all graphs will move to the right.

As mentioned before, the time to complete one primary scanning cycle, task error and control workload are metrics for the task demand load. All metrics indicate that task demand load increases with increasing turbulence intensity. A clear limit for the task demand load for which the task error remains within limits can be derived from the location of the transition turbulence intensity.

Since the main hypothesis is that safety diminishes when the task demand load increases, the graphs also provide a (relative) measure for the safety of a certain flight. The simulation shows that safety diminishes with higher levels of turbulence. In fact, when the imposed accuracy margins for altitude are substituted by the altitude margins necessary for safe separation, to the right of the bends in the graphs the separation criteria are no longer met and the situation can be labeled "unsafe".

4 | Flight simulator validation experiment

The flight simulator experiment, performed to validate the results of the off-line computer simulation, will be described in this section. First the pilot's task and the flight simulator will be described followed by an explanation of the independent variables, dependent measures and experiment hypotheses. Thereafter, the results of the flight simulator evaluation will be described. The results are divided into three parts and will be presented in this order: first a check whether the pilots adhere to the primary scanning cycle, second the tests results per pilot regarding task error, scanning behaviour and control workload, and third the general test results.

4.1 | Method

Subjects and instructions. Four pilots with Instrument Flight Rules (IFR) experience (ranging from 400 hours to 10,000 hrs) participated. They were instructed to fly level at 2,000 ft under IFR conditions, keep the altitude variations within +/- 50 ft, and maintain a heading of their choice within 10 deg. At all times, altitude control was to be regarded more important than heading control. The airspeed was set at 120 knots. Before the tests started, pilots familiarized themselves with the simulator, and were required to adjust the throttle setting and trim the aircraft in order to achieve the desired airspeed equal to 120 knots. After that, pilots were no longer allowed to change the throttle and could only use elevator-, aileron- and rudder control. Pilots were instructed to scan the flight instruments as they would normally do during real flight. This implies that their scanning behavior does not necessarily correspond to the primary scanning cycle in Figure 3.

Apparatus and setup. A Frasca 121 fixed-base flight simulator (a single engine light aircraft with fixed pitch propeller, cruise speed 120 knots) was used to perform the flight simulator tests. During the tests, aircraft and pilot performance data were logged. Furthermore, pilot eye movements were tracked by a remote eye-tracking device (SensoMotoric Instruments), see Figure 12.

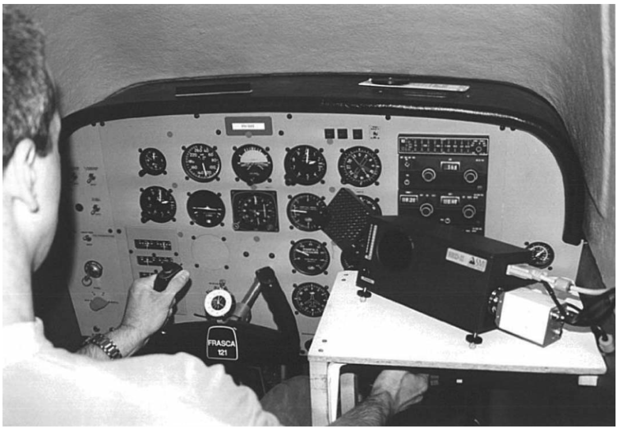


Figure 12 | Frasca 121 flight simulator and eye tracking device.

Independent variables. During the tests, six different levels of turbulence were applied, in longitudinal, lateral and vertical direction. Unfortunately, it was not possible to measure the exact intensity of the turbulence, but a qualitative sense of the magnitude of the turbulence was obtained through pilot opinion. Level 1 was agreed upon as 'normal weather'; level 2: 'thermal weather'; level 3: 'strong thermal weather'; level 4: 'thunderstorms'; level 5: 'thunderstorms, more severe than level 4'; level 6: 'thunderstorms, more severe than level 5'. During the tests each turbulence level was applied once to each pilot for a period of 5 minutes of which 3.5 minutes could be used for analysis. This yields a limited set of data and no strong conclusions can be drawn from these test results.

Dependent measures. The three dependent measures corresponded to the three metrics for Task Demand Load of Figure 7: (i) control workload, (ii) the time a pilot takes to perform one primary scanning cycle, and (iii) the task error. The control workload was calculated by taking the root mean square (RMS) of the elevator control activity. The time to perform a primary scanning cycle was determined through calculating the average dwell time for each instrument that is part of the primary scanning cycle and adding these dwell times.

Task error was expressed in various metrics that concern the altitude deviation from the reference of 2,000 ft. As long as altitude remained within 50 ft of the reference, task performance was considered adequate. When the altitude deviation exceeded 50 ft, its highest peak value was registered for analysis, see Figure 13. Additionally, the amount of time (t) the pilot spent outside the altitude limit of 50 ft was registered. The pilot's task error for each turbulence level was then determined by taking into account: (1) the *number* of registered peaks; (2) the *value* of registered peaks and (3) the time spent outside the altitude limits.

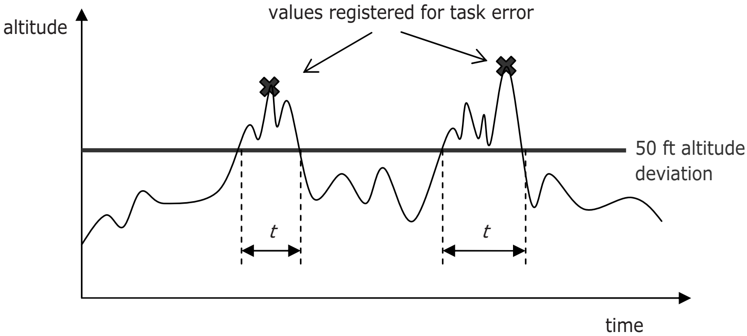


Figure 13 | Altitude deviation task error metrics.

Experiment hypotheses. Our main hypothesis was that the same trends in the dependent measures would occur as they were found in the computer simulation, summarized in Figures 9, 10 and 11. In summary, it is expected that for turbulence levels lower than the transition turbulence level the time to complete one scanning cycle will decrease, the task error will remain adequate and the control workload will remain relatively constant. For turbulence levels higher than the transition turbulence level it is expected that the time to complete one scanning cycle will remain constant at its minimum value, the task error will increase linearly and the control workload will increase linearly with increasing turbulence.

Since the aircraft used during the flight simulator tests differs from the aircraft modeled in the off-line simulations, and operates at a different airspeed, and because the turbulence levels applied during the flight simulator tests could not be quantified, the focus of our investigation will be on evaluating the trends in both sets of data.

4.2 | Results of the flight simulator evaluation

Primary scanning cycle. To check whether the pilots adhered to the primary scanning cycle as hypothesized in the preliminary pilot model, and recognized the three instruments in this cycle as the most important ones, the amount of time the pilots looked at them was examined. Figure 14 shows that about 85 to 98 percent of the total dwell time was spent on the instruments of the primary scanning cycle; this number increases with increasing turbulence intensity.

Of all fixations on altimeter, course indicator and attitude indicator together, in theory, 25% of these fixations should be on the altimeter, 25% on the course indicator and 50% on the attitude indicator, see Figure 3. Simulator results show that pilots 1, 2 and 4 indeed show a distribution of fixations over the instruments, which approaches this theoretical distribution (Figure 15). Pilot 3 however, spent a disproportionally high percentage of fixations on the attitude indicator, increasing to almost 80% for the higher turbulence levels.

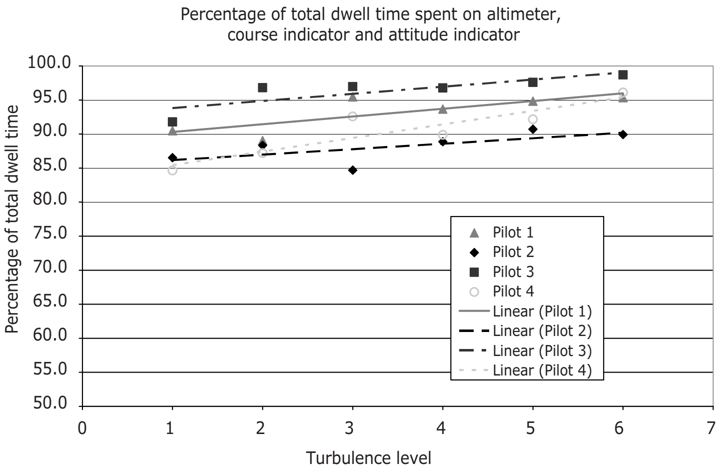


Figure 14 | Percentage of total dwell time spent on the instruments of the primary scanning cycle.

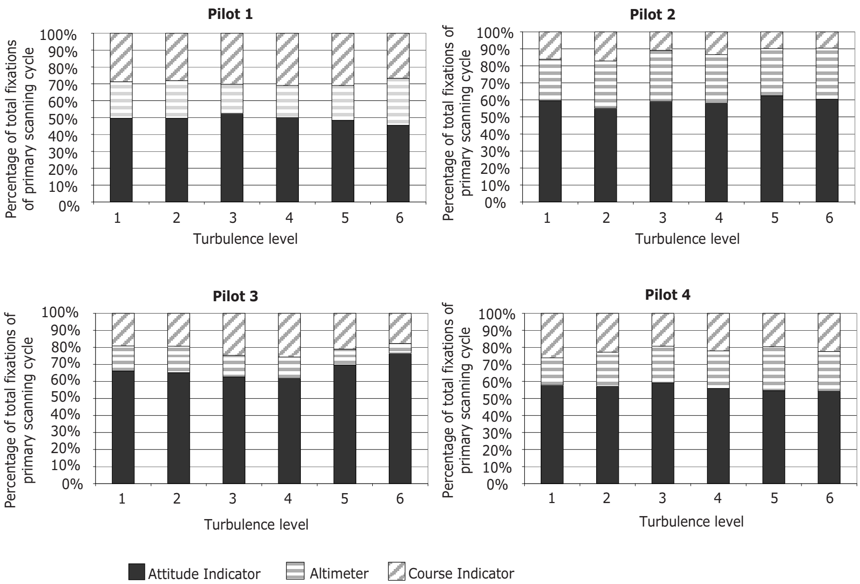


Figure 15 | Percentage of total number of fixations on each instrument, given for every turbulence level and every pilot.

In this respect it is interesting to note that pilot 3 was the least-experienced pilot. The results therefore agree with the findings of Bellenkes et al. (1997), who reported that whereas expert pilots maintain a certain fixation pattern, novice pilots tend to focus on one instrument. It can be concluded that for three of the four pilots in this experiment, the scanning cycle for straight and level flight as implemented in the preliminary pilot model, approximates ‘realistic’ scanning behavior adequately.

Test results per pilot. The task error for pilot 1 is given in Figure 16a. Each dot in the graph represents one registered peak value for the altitude deviation as explained previously. The triangles represent the average of the registered peak values, and the line is the linear regression for these average values. It can be seen that the altitude limit of 50 ft was exceeded for every turbulence level for large percentages of the time, and that there are many peaks for each turbulence level. As hypothesized, task error increases with increasing turbulence level. When compared to the other turbulence levels the task error for turbulence level 2 was disproportionately high, which might be due to the fact that this level was the first level applied to pilot 1 during the flight simulator tests; a learning effect could still be present during the first turbulence level. This however, is not certain.

The fixation results show that the time pilot 1 takes to perform one primary scanning cycle remains approximately constant or decreases slightly, see Figure 17. Figure 18 shows that control workload for pilot 1 increased with higher turbulence levels, with again a relatively high value for turbulence level 2.

These results indicate that the behavior of pilot 1 corresponds to the region to the right of the bend (or transition turbulence level) in Figures 9, 10 and 11: the altitude deviation and control workload increase with increasing turbulence level while the time to perform one primary scanning cycle remains constant.

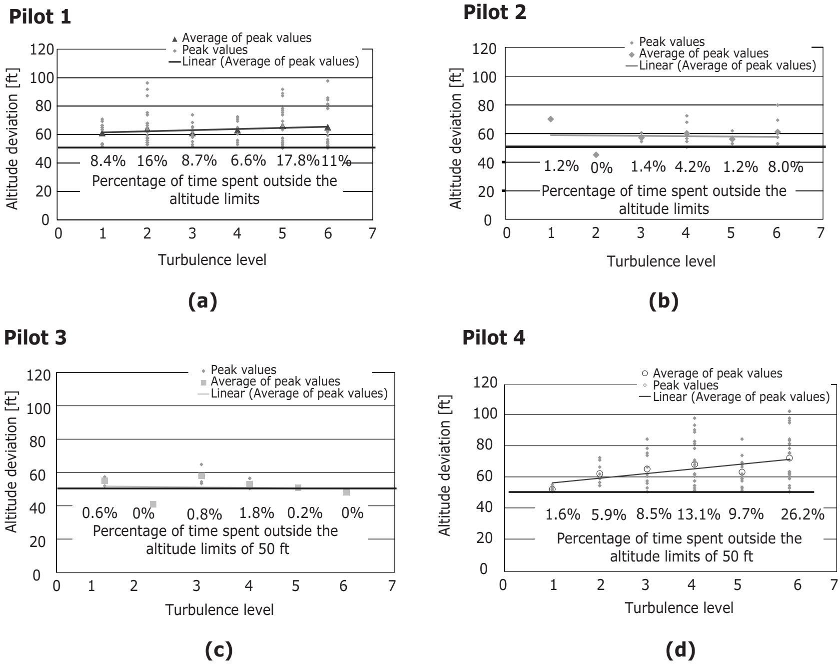


Figure 16 | Task error, in terms of peak values in altitude deviation, as a function of turbulence level.

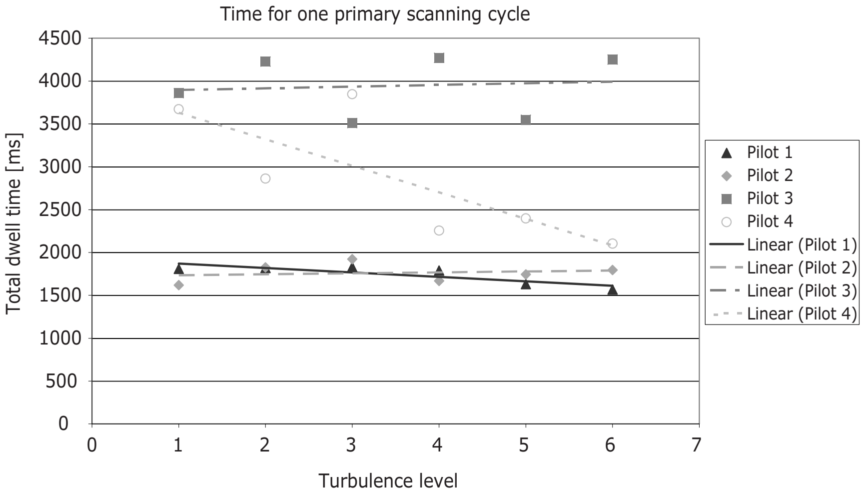


Figure 17 | The time for one primary scanning cycle, as a function of turbulence level.

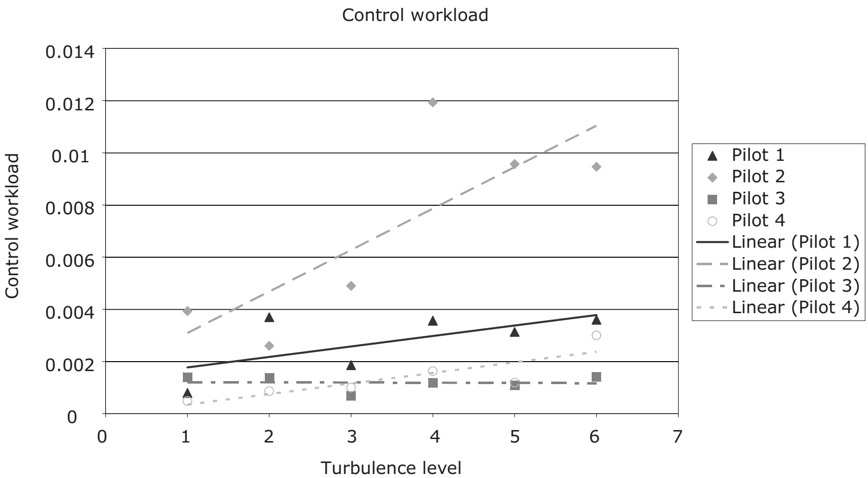


Figure 18 | Control workload, as a function of turbulence level.

The task error for pilot 2, Figure 16b, shows that this pilot also exceeded the altitude limits for each turbulence level, except for turbulence level 2. The percentage of time that was spent outside the limits, however, is larger for turbulence levels 4 and 6 than for the lower levels, and the number of registered peak values increases for the higher turbulence levels. In short, the task error again increased for higher turbulence levels. The time pilot 2 took to perform one primary scanning cycle, Figure 17, remained approximately constant, while the control workload increased considerably with increasing turbulence level, Figure 18. Again, the results indicate that the pilot operates in the region to the right of the transition turbulence intensity.

When neglecting the very small percentages of time pilot 3 spends outside the altitude limits, see Figure 16c, it can be stated that pilot 3 managed to keep the aircraft within the set altitude limits for all turbulence levels. The time pilot 3 took to perform one primary scanning

cycle was significantly larger than for pilots 1 and 2, Figure 17, and maintained a value of approximately 4000 ms, independent of the level of turbulence. Control workload for pilot 3, Figure 18, also remains at the same value, independent of the turbulence level.
Summarizing, pilot 3 remained within the altitude limits, with a constant time for one primary scanning cycle, and a constant control workload. These metrics do not show that the pilot has to work 'harder' for higher turbulence intensities. However, when looking closer at the pilot's scanning behavior, it was found that, with increasing turbulence level this pilot focused more and more on the attitude indicator, increasing to a maximum of 93% of the total dwell time and 80% of the total number of fixations with only a sporadic (short dwell time) scan on the Course Indicator, while all other instruments were virtually neglected. The primary scanning cycle was reduced to only one instrument which is scanned at an infinite scanning rate (almost continuously) and it is no longer correct to speak of an actual scanning *cycle*.

If the behavior of pilot 3 had to be compared to the graphs in Figures 9, 10 and 11, his behavior would be located to the left of the bend, as he remained the aircraft within the altitude limits and changed his scanning behavior, while control workload remains constant.

Figure 16d shows that for pilot 4 there are many registered peaks for each turbulence level and that a considerable percentage of time was spent outside the altitude limits. Task error increased for higher levels of turbulence. The time it took pilot 4 to perform one primary scanning cycle decreased with increasing turbulence level, Figure 17. Control workload, Figure 18, increased slightly for the turbulence levels 1-5 and jumps to a higher value for turbulence level 6.

By decreasing the time to perform one scanning cycle, this pilot managed to keep control workload at a low level until it finally starts to increase at turbulence level 6. This could indicate that the behavior of the pilot is located to the left of the bend in Figures 9, 10 and 11. Task error, however, does not correspond to this region. The question remains whether the altitude deviation would have remained within the 50 ft limit if the pilot would have decreased the time for one primary scanning cycle sooner (for lower turbulence levels). It is impossible to obtain an answer to this question with the present data, however.

General test results. Although the amount of data was limited, an attempt can be made to derive some general test results. Figure 17 shows that the lower limit on the amount of time needed to perform one primary scanning cycle of 2 seconds is not too far off from reality when looking at the results for pilots 1, 2 and 4. Dwell times on the separate instruments of the primary scanning cycle, however, differed per pilot and were not necessarily equal to the 500 ms assumed for the theoretical pilot model.

Note that the differences in absolute values for the control workload found during the flight simulator tests (Figure 18) and those found for the computer simulation (Figure 11) is due to the fact that different aircraft were used.

With respect to the hypothesis it can be said that the experiment showed a clear relation between pilot scanning behavior and control workload. When pilots adapt their scanning cycle, like pilots 3 and 4, control workload can be kept at a lower level. When pilots do not change their scanning behavior, pilots 1 and 2, control workload increases for higher levels of turbulence. The behavior of three of the pilots seems to be located to the right of the transition turbulence level that was found in Figures 9, 10 and 11. The behavior of one pilot seems to be to the left of the transition turbulence level. There are not sufficient results to clearly indicate the location of the bend in each graph for each pilot. Although the results indicate a correspondence between control workload, scanning behavior and task error, more tests are mandatory to further investigate this relation.

5 | Concluding remarks

A preliminary method to predict pilot task demand load when conducting an airport approach has been presented. The method is intended to employ a fairly-detailed model of the environment (e.g., type of aircraft, level of turbulence, crosswind, approach geometry that needs to be flown, etc.) in combination with a rather elementary pilot model that itself consists of a scanning model, a mental model and a control model. To check whether this combination can generate useful predictions, a computer simulation has been developed of a straight and level flight through turbulent air.

The computer simulation predicts a relation between the time the pilot has to perform one scanning cycle, the task error and the control workload. With increasing turbulence intensity the time to perform one scanning cycle decreases and will finally stabilize at a minimum value which results in a bend or knee in the graph, at the so-called 'transition' turbulence intensity. At this transition, a bend also occurs in the graph for task error: to the left of the bend a constant task error is obtained that is equal to the defined accuracy margin, to the right of the bend task error will increase. Control workload as a function of turbulence intensity increases slightly before the transition turbulence intensity, and increases linearly beyond this transition with increasing turbulence intensity.

A preliminary flight simulator evaluation has been performed to investigate whether the trends predicted by the computer simulation also occur in reality. The evaluation revealed that there is indeed a relation between pilot scanning behavior, task performance and control workload. When pilots change their scanning behavior control workload can be kept at a lower level. When pilots do not change their scanning behavior control workload appears to rise. The behaviour of three of the four pilots that participated fits fairly well to the model predictions for the region to the right of the transition turbulence intensity. The behavior of one pilot can be described with our model acting for lower levels of turbulence intensity, i.e., to the left of the bend. There are, however, not sufficient test results to prove our hypotheses, and more tests are mandatory to explore the relationship between turbulence intensity level, pilot scanning and control behaviour, and task demand load in more detail.[1]

Acknowledgments

The authors wish to thank Dr. Th. Boersema and C.W. Jorens for sharing their expertise on scanning behavior and the use of the eye track equipment. Thanks also go to the pilots participating in the simulator tests and Mr. van Spijker of IST Flight Training for adapting the Frasca 121 to fit the requirements of our experiments.

1 For this Ph.D. thesis a description of additional tests is added in Appendix A to Journal Article 1.

References

1. Amalberti R, Wioland L. (1997) Human error in aviation. In: Soekkha H., editor. *Aviation safety: Human factors, system engineering, flight operations, economics, strategies, management,* VSP Utrecht; The Netherlands, pp. 91-108.
2. Baren, G.B. van, Speijker, L.J.P., Bruin, A.C. de, *Wake vortex safety evaluation of single runway approaches under different weather and operational conditions*, Proc. 6th Int. Conf. Probabilistic Safety Assessment and Management (PSAM6), June 23-28, 2002, San Juan, Puerto Rico.
3. Baron, S., Kleinman, D.L., Levison, W.H. (1970), *An Optimal Control Model of Human Response Part II: Prediction of Human Performance in a Complex Task,* Automatica, Vol.6, pp. 371-383.
4. Baron, S., Zacharias, G., Muralidharan, R., Lancraft, R. (1980), PROCRU: A model for analyzing flight crew procedures in approach to landing. In *NASA-MIT 16th Annual Manual 1980*, pp 488-520.
5. Baron, S., Kruser, D.S., Huey, M.B. (1990), *Quantitative Modeling of Human Performance in Complex, Dynamic Systems,* National Academy Press
6. Bellenkes, A.H., Wickens, C.D., Kramer, A.F. (1997), Visual scanning and Pilot Expertise: The Role of Attentional Flexibility and Mental Model Development*, Aviation, Space and Environmental Medicine,* Vol 68(7), pp 569-579
7. Blom, H.A.P., Bakker, G.J., Blanker, P.J.G., Daams, J., Everdij, M.H.C., Klompstra, M.B., *Accident risk assessment for advanced ATM*, In: Air Transportation Systems Engineering, edited by G.L. Donohue and A.G. Zellweger, Progress in Astronautics and Aeronautics Vol. 193, Paul Zarchan, Editor-in-Chief, Chapter 29, pp. 463-480, 2001
8. Blom, H.A.P., Klompstra, M.B., Bakker, G.J., *Accident risk assessment of Simultaneous Converging Instrument Approaches*, Proc. 4th USA/Europe ATM R&D Seminar FAA/ Eurocontrol, December 2001.
9. Campbell, R.D. and Bagshaw, M. (1999), *Human Performance and Limitations in Aviation,* 2nd edition, Blackwell Science Ltd.
10. Carbonell, J.R. (1966), A Queuing Model of Many-Instrument Visual Sampling, *IEEE Transactions on Human Factors in Electronics*, HFE-7.
11. Carbonell, J.R., Ward, J.L., & Senders, J.W. (1968). A queuing model of visual sampling: Experimental validation. *IEEE Transactions on Man-Machine Systems*, MMS-9.
12. Carolan, T., Scott-Nash, S., Corker, K., Kellmeyer, D. (2000), *An application of human performance modeling to the evaluation of advanced user interface features,* Human Factors and Ergonomics Society.
13. Corker, K.M. (1999), Human Performance Simulation in the Analysis of Advanced Air Traffic Management, In *Proceedings of the 1999 Winter Simulation Conference,* Farrington, P.A., Nembhard, H.B., Sturrock, D.T., Evans, G.W. eds.
14. Corker, K.M. (1999), Human Performance Simulation in the Analysis of Advanced Air Traffic Management, In *Proceedings of the 1999 Winter Simulation Conference,* In: Farrington, P.A., Nembhard, H.B., Sturrock, D.T., Evans, G.W. Eds.
15. Dahn, D., Laughery, R., Belyavin, A. (1997), The integrated performance modeling environment a tool for simulating human-system performance, In *Proceedings of the 41st Human Factors and Ergonomics Society Conference,* Albuquerque, NM.
16. Etkin, B. and Reid, L.D. (1996), *Dynamics of Flight, Stability and Control,* John Wiley & Sons

17. Etkin, B. (1972), *Dynamics of Atmospheric Flight,* John Wiley & Sons.
18. Everdij, M.H.C. , Blom, H.A.P., and Bakker, G.J., *Accident risk assessment for airborne separation assurance*, Proc. Advanced workshop on air traffic management, Capri, Italy, 23-26 September 2002.
19. GAIN Working Group B, Analytical Methods and Tools, *Guide to Methods & Tools for safety analysis in air traffic management,* First edition-June 2003.
20. Gore, B.F., Corker, K.M. (2000) A Systems Engineering Approach to Behavioral Predictions of an Advanced Air Traffic Management Concept, *19th Digital Avionics Systems Conference.* Philadelphia, Pennsylvania.
21. Harris, R.L.,& Christhilf, D.M. (1980). What do pilots see in displays? In G. Corrick, E. Hazeltine & R. Durst (eds), *Proceedings of the 24th annual meeting of the Human Factors Society.* Santa Monica, CA: Human Factors Society.
22. Houbolt, J.C., Steiner, R. & Pratt, K.G. (1964). *Dynamic Response of Airplanes to Atmospheric Turbulence Including Flight Data on Input and Response.* NASA TR-R-199.
23. Hollnagel, E. (1993), *Human Reliability Analysis, Context and Control,* Academic Press, London
24. Hudson, P.T.W. & Parker, D. (2001) *Profiling safety culture: The OGP study.* Report for the Oil and Gas Producers Association OGP, London.
25. Hudson, P.T.W. (2001a) Safety Management and Safety Culture: The Long, Hard and Winding Road. In W. Pearse, C. Gallagher & L. Bluff (Eds.) *Occupational Health and Safety Management Systems.* Melbourne: Crowncontent.
26. Hudson, P.T.W. (2001b), *Aviation Safety Culture,* Safeskies 2001.
27. Kleinman, D.L., Baron, S., Levison, W.H. (1970), *An Optimal Control Model of Human Response Part I: Theory and Validation,* Automatica, Vol.6, pp 357-369
28. KLM Luchtvaartschool (KLM Flight Academy) (1990), *Basic Attitude Flying*, Eelde
29. McRuer, D.T. and Jex, H.R. (1967), A Review of Quasi-Linear Pilot Models, *IEEE Transactions on Human Factors in Electronics,* Vol. HFE-8, No. 3, September 1967
30. Moray, N. (1986), Monitoring behavior and supervisory control. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of Perception and Performance*, Vol II, pp 40-1 – 40-51, New York: Wiley & Sons.
31. Mulder, J.A. & Vaart, J.C. van der (1998), *Aircraft responses to atmospheric turbulence*, Lecture Notes D-4, Faculty of Aerospace Engineering, Delft University of Technology, Delft, The Netherlands.
32. National Transportation Safety Board, *Annual Review of Aircraft Accident Data NTSB/ARC-06/01, US Air Carrier Operations, Calendar Year 2001,* Washington, D.C., USA
33. National Transportation Safety Board, *Annual Review of Aircraft Accident Data, US General Aviation, Calendar Year 1997*
34. Padfield, G.D. (1996), *Helicopter Flight Dynamics, The Theory and Application of Flying Qualities and Simulation Modelling*, Blackwell Science Ltd
35. Pharmer, J.A., Freeman, J.T., Scott-Nash, S., Santoro, T.P., Kieras, D. (2001), *Complementary Methods of modeling team performance,* Panel discussion during the Human Factors and Ergonomics Society, 45th Annual Meeting.
36. Reason, J.T. (1990), *Human Error*, Cambridge University Press, U.K.
37. Roelen, A.L.C., Cacciabue, P.C., Hayes, P., Molemaker, R.J. , Joyce, A., *A method for setting and optimizing safety targets in the air transport system,* Proc. 6th Int. Conf. Probabilistic Safety Assessment and Management (PSAM6), June 23-28, 2002, San Juan, Puerto Rico.

38. Roelen, A.L.C., Wever, R., Cooke, R.M., Lopuhaä, R., Hale, A.R., Goossens, L.H.J. (2003a). Aviation causal model using Bayesian Belief Nets to quantify management influence, Proc. European Safety and Reliability Conference ESREL 2003, June 15-18, Maastricht, The Netherlands, Eds: T. Bedford & P.H.A.J.M. van Gelder, Balkema Publishers.
39. Roelen, A.L.C., Wever, R., Cooke, R.M., Lopuhaä, R., Hale, A.R., Goossens, L.H.J., Simons, M., Valk, P.J.L., (2003b), Causal modeling for integrated safety at airports, Proc. European Safety and Reliability Conference ESREL 2003, June 15-18, Maastricht, The Netherlands. Eds: T. Bedford & P.H.A.J.M. van Gelder, Balkema Publishers.
40. Sheridan, T.B., & Stassen, H.G. (1979). Definitions, Models and Measures of Human Workload. In: Moray, N. (Ed.) *Mental Workload: Its Theory and Measurements*, pp 219-234. Plenum Press, New York.
41. Smith, B.R., Tyler, S.W. (1997), *The Design and Application of MIDAS: A Constructive Simulation for Human-System Analysis,* 2nd Simulation Technology & Training Conference, Canberra, Australia.
42. Stassen, H.G., Johannsen, G., & Moray, N. (1990). Internal Representation, Internal Model, Human Performance Model and Mental Workload. *Automatica, 26*(4), 811-820.
43. Suarez, N. (2003), *Advanced Airport Approach Procedures Including Validation and Elaboration*, APPROVE Final Report, EC DG-TREN TEN-T
44. Verbruggen, H.B. (1982), *Digitale regelsystemen*, Delftse Universitaire Pers, Delft.
45. Vicente, K.J. (1999), *Cognitive work analysis: Towards safe, productive, and healthy computer-based work*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
46. Westrum, R. & Adamski, A.J. (1999) Organizational Factors Associated with Safety and Mission Success in Aviation Environments. In D.J.Garland, J.A. Wise & V.D.Hopkin (Eds.) *Handbook of Aviation Human Factors.* Lawrence Erlbaum, Mahwah, NJ. pp 67-104
47. Westrum, R. (1993) Cultures with Requisite Imagination. In J.Wise, P.Stager & J.Hopkin (Eds.) *Verification and Validation in Complex Man-Machine Systems.* Springer, New York pp 401-416.
48. Westrum, R. (1997) Social Factors in Safety-Critical Systems. In F. Redmill & J.A. Ragan (Eds.) *Human Factors in Safety-Critical Systems.* Butterworth/Heineman, New York. pp 233-256.
49. Wickens, C.D. and Hollands, J.G. (2000), *Engineering Psychology and Human Performance,* Third Edition, Prentice-Hall International
50. Wickens, C.D., Goh, J., Helleberg, J., Horry, W.J. and Talleur, D.A. (2003), Attentional Models of Multitask Pilot Performance Using Advanced Display Technology, *Human Factors*, 45(3), pp. 360-380.

Appendix A to Journal Article 1
Change in focus with respect to factors that influence pilot task demand load and consequences for the remainder of this research

As stated in the conclusions of the first journal article, more tests are mandatory to explore in more detail the relationship between turbulence intensity level, pilot scanning and control behaviour, and task demand load. For this purpose, a flight simulator experiment was performed. This section will briefly describe the set-up of the experiment, the results of the experiment with respect to scanning cycles and control workload, and, most importantly, the overall conclusion resulting from this experiment, which is that the scanning of the flight instruments and manually controlling the aircraft are not the major contributors to pilot task demand load during approaches. Based on this conclusion the focus of the research has shifted considerably, the consequences of this change in focus are also presented.

This section thereby provides a bridge between the first journal article (based on the 'old' focus) and the second and subsequent journal articles (based on the 'new' focus).

Experiment Set-up and hypothesis

The experiment was performed in Delft University of Technology's six-degree-of-freedom SIMONA Research Simulator, see Figure 1. The aircraft implemented in the simulator was the Cessna Citation I. The relevant flight instruments that were available were the airspeed indicator, attitude indicator, altimeter, vertical speed indicator and course indicator. Localizer and Glide slope deviation indications were available on the course indicator and attitude indicator respectively. No flight director guidance was available.

The pilots' eye movements were recorded by the Remote Eye-Tracking Device of the SMI iView System, the same system used in the experiment described in the first journal article. The pilots' continuous control actions are logged, such that these data are available for further analysis.

Five pilots participated in the experiment (flight hours ranging from 1300 to 7200 hours), three of these pilots were current on the Cessna Citation, one pilot was a Boeing 737 pilot and one pilot an MD-11 pilot. Each of the pilots flew 12 approaches. During the experiment the pilots were vectored towards the Localizer by Air Traffic Control (ATC), the final part of the approach was flown using the Instrument Landing System (ILS). Before the experiment started, the pilots were instructed to fly the approach manually, and were asked to fly 'best performance', which implied that they did not have to take into account passenger comfort. Their task was to keep the velocity and the path of the aircraft as close as possible to the velocity and path as instructed by ATC.

For the experiment two different approach trajectories with corresponding velocity profiles were used (each pilot flew each of these trajectories six times). For both approach trajectories the velocity and altitude requirements could be met. Each approach consisted of the initial, intermediate and final approach segments.

Figure 1 | The SIMONA Research Simulator

The independent variable was the turbulence intensity (modeled according to the Dryden spectra), and was varied in six discrete steps, ranging from 0.0 m/s to 1.5 m/s. During four or five of the total of 12 approaches per pilot an additional interruption was added to the approach. This could be: a late landing clearance, a short approach (the vector route is cut short such that the pilot has less time available to prepare for landing), or a last minute change of runway.

The dependent measures were: 1. the pilot's control workload, 2. the pilot's dwell time per flight instrument.

The main hypothesis was that for turbulence levels lower than the transition turbulence level the dwell times for all instruments would decrease, the task error would remain adequate and the control workload would remain relatively constant. For turbulence levels higher than the transition turbulence level it was expected that the dwell times for all flight instruments would remain constant at their minimum values, the task error would increase and the control workload would increase with increasing turbulence. Additionally it was hypothesized that the control workload would increase momentarily after flap and gear deployment, and it was hypothesized that pilots would use different scanning cycles during the approach.

Experiment Results

When analyzing the flight simulator data and data from the eye-tracking device, no clear relation could be derived between the turbulence intensity and the dwell times on the flight instruments, the number of fixations per flight instrument or the control workload for elevator or aileron. The trends predicted in the first journal article are thus not found during this experiment. This might be due to the fact that, when compared to the experiment in the first journal article, pilots now had more tasks to perform, such as planning the approach and landing, selecting flaps and gear, communication with ATC, and, additionally, the three interruptions (late landing clearance, short approach or last minute change of runway). All these additional tasks might have greatly influenced the pilots' scanning behaviour. Another factor that might have influenced the results is the fact that some pilots discovered that over longer periods of time the effect of the turbulence would cancel out (there is no patchy turbulence incorporated in the Dryden spectra), this also influenced their control and scanning behaviour.

With respect to the control workload during the approach phase after flap and gear deployment, it was found that the control activity increased with a factor 1.2 to 1.9 during the 10 seconds following flap deployment and gear selection, and that the control activity for the ailerons increased 1-2 times faster with turbulence level than the control activity for the elevator (Blok, 2005).

It was found that during the approach pilots subsequently used three different scanning cycles: the first scanning cycle only incorporated the flight instruments, in the second scanning cycle, which is started after the call 'runway in sight', the outside view was added to the scanning cycle, and during the third scanning cycle almost all attention went to the outside world with a sporadic scan on the course indicator (Nijsten, 2005). The three scanning cycles are depicted in Figures 2 to 4. The values for the minimum dwell times for all instruments are given in Table 1.

Overall conclusion resulting from the experiment

Although these results are interesting and can be used for a simulation of the approach phase within a computer simulation, it became clear from the flight simulator experiment and conversations with pilots that it is not the scanning of the instruments nor the manual (continuous) control of the aircraft that constitute the largest contribution to pilot TDL. This is especially true when considering modern approaches that are now largely flown automatically

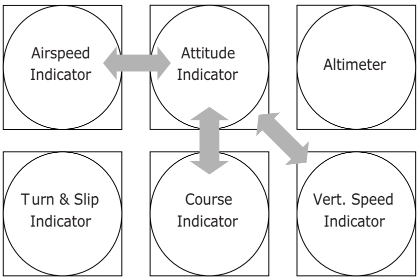


Figure 2 | First average scan cycle during approach phase (for descent): Airspeed Indicator → Attitude Indicator → Course Indicator → Attitude Indicator → Vertical Speed Indicator → Attitude Indicator → Airspeed Indicator → Attitude Indicator → etc.

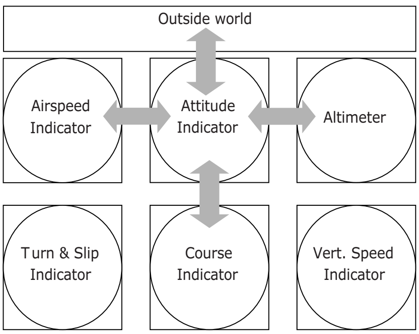


Figure 3 | Second average scan cycle during approach phase: Attitude Indicator → Outside World → Attitude Indicator → Airspeed Indicator → Attitude Indicator → Outside World → Attitude Indicator → Altimeter → Attitude Indicator → Outside World → Attitude Indicator → Course Indicator → Attitude Indicator → Outside World → etc.

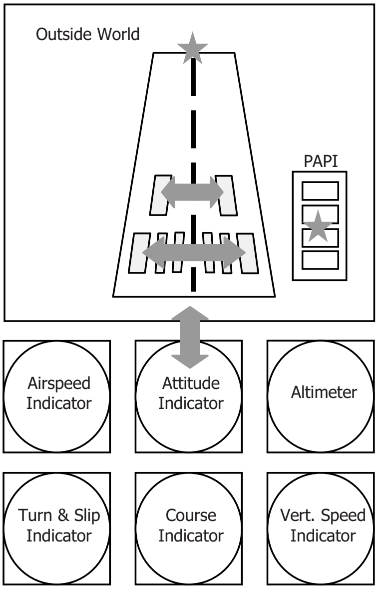


Figure 4 | Third average scan cycle during approach phase: Horizon/end of runway → Landing blocks → Horizon/end of runway → Threshold → Horizon/end of runway → PAPI → Horizon/end of runway → Course Indicator (Localizer indication) → Horizon/end of runway → Landing blocks → etc.

using the FMS with autopilots and autothrottle. Pilots seemed to be much more preoccupied with more 'discrete' tasks such as selecting flaps and gear down, meeting the constraints at the waypoints, performing checklists etc. than with the 'continuous' scanning and controlling tasks. Additionally, pilots seemed to be much more concerned with 'higher level' or, so to say, larger time-scale properties of the approach trajectory, such as for example the amount of trackmiles available between important waypoints in the approach, or the amount of time available to perform the required actions, instead of with small time-scale properties such as additional control actions necessary after flap deployment, or additional control actions due to turbulence.

Table 1 | Minimum dwell times in seconds for each instrument during the approach phase for three different scanning cycles.

Area of Interest	**Minimum dwell time during scanning phase**		
	First	**Second**	**Third**
Attitude Indicator	1.2	1.2	/
Airspeed Indicator, Altimeter, Vertical Speed Indicator	0.6	0.4	/
Course Indicator	0.7	0.4	0.4*
Outside World	/	1.0	4.0

* Not calculated due to lack of data, but this value can be assumed in a model.

In this respect these observations correspond with Rasmussen's skill- rule- and knowledge-based performance model (Rasmussen, 1983), which states that rule based actions (such as selecting flaps, gear down, etc.) require more attention from a human operator than skill based actions (such as scanning the flight instruments and controlling the aircraft with elevator and aileron). Additionally, a high level of automation results in less skill based tasks to be performed and moves the operator towards higher levels of control.

Consequences for this research

It was therefore decided to no longer pursue the development of scanning models nor to quantify the contribution of scanning or manual control to pilot TDL. Instead, the focus was fully shifted towards analyzing and modeling the 'discrete' actions (such as selecting flaps and gear, meeting the constraints at the waypoints, etc.) and the larger time scale properties of the approach trajectory (such as the amount of trackmiles available between important waypoints in the approach) and their effect on pilot TDL.

Due to this shift in focus, the set-up of the computer simulation to predict pilot TDL as presented in the first journal article and repeated here in Figure 5, also had to be changed. The decision was made to no longer incorporate a scanning model, or mental model in the offline computer simulation. The new set-up of the computer simulation is elaborately explained in the second journal article (which can be found directly following this section).

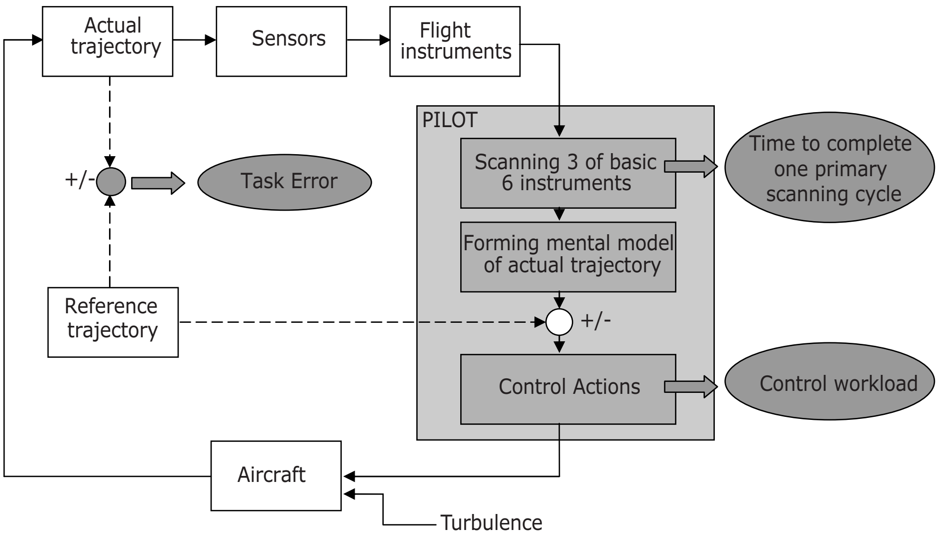


Figure 5 | Schematic representation of computer simulation model and metrics for the Task Demand Load

Based on the conversations with the pilots it was also decided to no longer consider manual control of the aircraft during the entire approach phase. Instead, in order to make the topic of this research more compatible with the reality of how approaches are flown, a higher level of automation is assumed: in the first part of the approach it is assumed that autopilots and autothrottle are used, in the final part the use of the flight director is assumed. This decision implied that, in order to be able to simulate these types of approaches, models for the autopilot, autothrottle and flight director had to be developed which could be used in an offline computer simulation as well as in the SIMONA Research Simulator.

References

1. Blok, G. (2005), *Prediction of pilot control workload due to the use of flaps and landing gear, based on a simulated approach through turbulent air,* Master thesis report, Delft University of Technology.
2. Nijsten, K., Modeling the pilot's scanning behaviour, Monitoring and processing the available information during the approach phase, Master thesis report, Delft University of Technology, August 2005.
3. Rasmussen, J. (1983), Skills, Rules and Knowledge; Signals, Signs, and Symbols, and Other Distinctions in Human Performance Models, *IEEE Transactions on Systems, Man and Cybernetics,* Vol. SMC-13, No. 3, May/June 1983.

Journal Article 2

Flight mechanical evaluation of approaches

M.M. Heiligers, Th. van Holten, M. Mulder

Accepted by the Journal of Aircraft

Abstract

The task demand load experienced by a pilot while flying an approach is, amongst others, influenced by the following two factors: first, the fact whether it is possible to meet the altitude and airspeed constraints defined at the waypoints, and second, the fact whether it is possible to achieve a stabilized approach at 1,000ft. In order to be able to analyze an approach with respect to these two factors a flight mechanical assessment tool based on a Monte Carlo simulation is presented in this paper. The Monte Carlo simulation predicts, given the aircraft, the Standard Operating Procedures, the wind conditions and the approach trajectory, the percentage of flights that will not meet the constraints at the waypoints and will not achieve a stabilized approach. The Monte Carlo simulation can be used by approach designers as it will provide them insight into the flight mechanical feasibility of an approach as well as give an indication of the task demand load that will be experienced by pilots. To demonstrate the results of the Monte Carlo simulation a fictitious approach is considered in a case study.

1 | Introduction

APPROACH-and-landing phase accidents account for a significant proportion of air transport accidents. The Flight Safety Foundation Approach-and-Landing Accident Reduction (FSF ALAR) Task Force [1] has performed comprehensive research with respect to Approach and Landing Accident Reduction. Approximately 59 percent of the world jet-fleet accidents to date occurred in these flight phases and accounted for 29 percent of all fatalities [2]. The most frequent causal factors are related to crew performance, emphasizing the need to better understand which factors complicate an approach for a flight crew and which factors in an approach increase the chance of accidents.

In addition, a considerable growth in the air transport industry is expected, with forecasts indicating that air traffic movements around the world will increase significantly [3]. This growth is likely to cause extra congestion and delays, which in turn calls for upgrading the overall system capacity. Part of the capacity increase could come from the design of new approach trajectories, possibly including curved approaches, the use of higher approach altitudes, continuous descent approaches et cetera. Before these new approaches can be introduced it is mandatory to investigate whether these approaches are more complicated for a pilot to fly.

The overall goal of the research presented in this paper and its companion [4], is to develop a method that can predict how complicated an approach will be for a pilot. That is, our research should be able to provide an accurate prediction of what we call the pilot Task Demand Load (TDL), for any (existing or to-be-designed) approach. Note that task demand load is defined as the mental workload *imposed* by the system to be controlled or supervised [5], and is not to be mistaken for the mental workload as *experienced* by the human operator, which is referred to as Mental Load (ML). Whereas the latter measure depends on, for instance, an individual operator's training, motivation, and mental capacities, the task demand load attempts to capture the external demand, the effort needed to reach the goals defined.

Overall, our hypothesis is that the task demand during an approach is mainly affected by the external constraints put on the pilot/aircraft system. In the accompanying paper [4] it is demonstrated experimentally that pilot mental workload (as reported subjectively by professional airline pilots) is influenced by: the localizer intercept speed, the localizer intercept angle, the line-up distance (distance between runway threshold and intermediate fix), the altitude of the final approach fix, the fact whether the altitude and velocity constraints at the waypoints can be met and the fact whether a stabilized approach at 1,000' can be achieved. Clearly then, these external constraints form (some of) the systematic elements that (together) make an approach more difficult to perform, or not. In other words, these are the factors that determine the external demand, and therefore the pilot task demand load.

The accompanying paper thus shows that pilot mental load increases significantly when the aircraft is unable to meet the altitude and velocity constraints at the waypoints, and also when a stabilized approach at 1,000' cannot be achieved. This paper presents a flight mechanical assessment tool that allows the approach designer to study in advance the effects of these two factors. It predicts the percentage of flights where aircraft are unable to, first, meet the constraints at the waypoints and, second, will not achieve a stabilized approach. The tool allows the approach designer to freely vary aircraft characteristics (e.g., weight), the wind conditions, the approach trajectory, the constraints at the waypoints and the Standard Operating Procedures (SOPs). Apart from the fact that this information is useful for the flight mechanical design of the approach, it will also give the approach designer an indication of pilot TDL.

The paper starts with an explanation of the basic principles and assumptions for the method to predict pilot TDL. Next, the Monte Carlo computer simulation that forms the flight mechanical assessment tool is explained in detail. Since the Monte Carlo computer simulation is quite elaborate, and requires some time to run, a point mass model for initial estimates (regarding the possibility to meet the constraints at the waypoints and to achieve a stabilized approach) is subsequently presented. Then, in order to demonstrate the output and predictions of the flight mechanical assessment tool a case study is presented, followed by the conclusions and recommendations.

2 | Basic principles of the method to predict pilot TDL

This section explains the basic principles of the method to predict pilot TDL: the assumptions and the choices that have been made as to what is incorporated within this research, and also what is currently considered to be beyond the scope of this research. As explained before, the flight mechanical assessment tool is linked to the research on pilot TDL and is therefore based on the same assumptions.

2.1 | Factors of the air transport system included

Many different factors and the interactions between those factors have an influence on the execution of an approach, see Figure 1. The research on pilot TDL obviously concentrates on the "pilot" box in Figure 1. To determine pilot TDL, this research will only take into account the factors that have a direct influence on an approach (see Figure 1), most importantly the characteristics of the trajectory, the type of aircraft and the meteorological conditions.

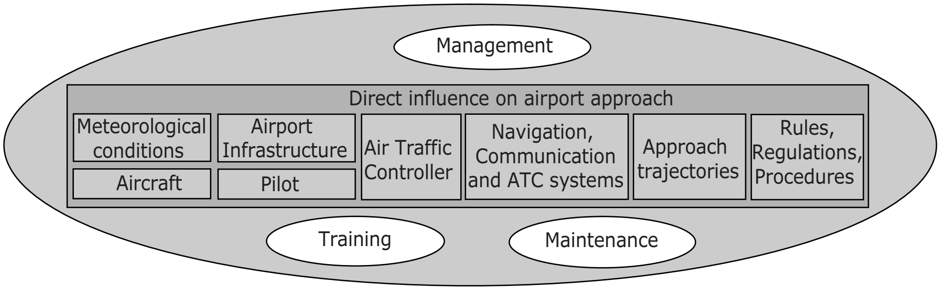


Figure 1 | Direct and indirect factors that influence the safety of airport approaches.

2.2 | Approaches considered and automation used

Obviously, pilot TDL directly depends on the type of approach that is considered. This research focuses on Area Navigation (RNAV) Approaches. Although it is appreciated that non-precision approaches such as NDB approaches are, in general, more difficult for a pilot to fly than RNAV approaches [6], a deliberate choice is made to focus on RNAV approaches since these are expected to become more and more frequently used in the future. The last part of the RNAV approach is assumed to be flown using the Instrument Landing System (ILS).

The part of the flight that is considered in the method to predict pilot TDL and therefore also in the flight mechanical assessment tool, starts at the Initial Approach Fix (IAF) and comprises the entire approach (Initial Approach, Intermediate Approach and Final Approach) until 1,000 feet above airport elevation, see Figure 2. Based on interviews with pilots it was decided to use two different levels of automation during the approach: until Localizer Intercept Heading the approach is flown using the FMS, Autopilots and Autothrottle. At Localizer Intercept Heading (but before Localizer capture) the pilot switches to Flight Director (FD) mode and disconnects the Autothrottle, the remainder of the approach is thus flown using the FMS and FD, which implies manual control by the pilot.

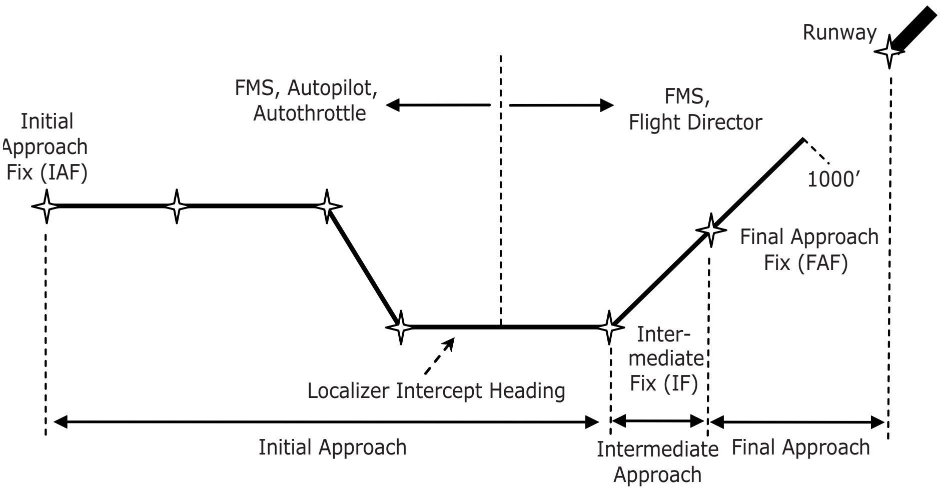


Figure 2 | Part of flight considered (top view) and automation used

2.3 | Non-nominal conditions and emergencies

Non-nominal conditions and emergencies such as engine failure are not considered in this research. The goal is to determine pilot TDL for published RNAV approaches under nominal conditions. When any emergencies such as engine failure occur, the crew will most likely not be required to follow the RNAV approach anyway, but will be vectored to the runway in the most convenient way.

Additionally, the assumption for less severe non-nominal situations is that when flying under nominal conditions, the RNAV approach should provide enough 'margin' with respect to pilot TDL, such that the pilot has enough spare capacity and time to deal with non-nominal conditions. This implies that the TDL that is predicted by this research for nominal conditions should be well below the absolute maximum TDL a pilot can cope with in order to guarantee this margin. Therefore the simulation of non-nominal conditions is not part of the flight mechanical assessment tool.

2.4 | Boundary conditions: Stabilized approach and Standard Operating Procedures

The TDL experienced by the pilot also depends on the boundary conditions that are set, e.g. the accuracy with which the approach needs to be flown. The boundary conditions chosen for this research and for the flight mechanical assessment tool are that the approach should be performed according to Standard Operating Procedures and that pilots should aim to achieve a stabilized approach at 1,000 feet above airport elevation. This decision is based on the conclusions of the ALAR Task Force [1] which stated that: "Establishing and adhering to adequate standard operating procedures (SOPs) and flight-crew decision making processes improve approach-and-landing safety and that "Unstabilized and rushed approaches contribute to approach-and-landing accidents" [1].

To determine whether a stabilized approach is achieved at 1,000 feet, the following nine criteria are used [1]:

1. The aircraft is on the correct flight path;
2. Only small changes in heading/pitch are required to maintain the correct flight path;
3. The aircraft speed is not more than VREF + 20 knots Indicated Airspeed (IAS) and not less than VREF[1];
4. The aircraft is in the correct landing configuration;
5. Sink rate is no greater than 1,000 feet per minute; if an approach requires a sink rate greater than 1,000 feet per minute, a special briefing should be conducted;
6. Power setting is appropriate for the aircraft configuration and is not below the minimum power for approach as defined by the aircraft operating manual;
7. All briefings and checklists have been conducted;
8. Specific types of approaches are stabilized if they also fulfill the following: instrument landing system (ILS) approaches must be flown within one dot[2] of the glide slope and localizer; a Category II or Category III ILS approach must be flown within the expanded localizer band; during a circling approach, wings should be level on final when the aircraft reaches 300 feet above airport elevation; and
9. Unique approach procedures or abnormal conditions requiring a deviation from the above elements of a stabilized approach require a special briefing.

1 The Reference Speed (VREF) is defined as 1.3 times the stall speed

2 One dot deviation on the glide slope equals 0.7º beam error, one dot deviation on the Localizer equals 2.5º beam error.

2.5 | Level of detail of computer simulation models

The approach we have chosen to predict pilot TDL [7] deviates from the philosophy behind human operator models such as the Procedure-Oriented Crew model (PROCRU) [8] or the Man-Machine Integrated Design and Analysis System (MIDAS) [9-10]. Our approach to predict pilot TDL is based on the principles of cognitive work analysis [13]. The main characteristic of cognitive work analysis is that it shifts the emphasis from investigating the constraints of the human operator (like memory capacity, time delay, etc.) to analyzing and describing the operator's environment (like the trajectory, the aircraft dynamics, the wind conditions, etc.). The reason for this choice is that the constraints in the environment actually 'shape' the behavior of the human working in that environment. By choosing the approach to focus on the operator's environment instead of the human operator, the model will be more accessible and easier to use for approach designers, who often have little background in human operator modeling (or none).

Therefore, the goal is to incorporate detailed models of the *environment* of the pilot in the flight mechanical assessment tool, and to add to this a rather simple model for the pilot. Consequently, the aircraft with its kinematic and dynamic constraints, the 3-D properties of the trajectory, the velocity profile, turbulence, wind, etcetera, in other words: the factors that have a direct influence on an approach as given in Figure 1, are modeled as detailed and accurate as possible. The pilot model consists of a continuous manual control model (which in effect only contains a pure gain plus time delay) and a model for performing discrete actions such as selecting flaps and gear according to the Standard Operating Procedures (SOPs).

3 | Monte Carlo computer simulation

The flight mechanical assessment tool consists of a Monte Carlo computer simulation based on the basic principles and assumptions as explained in the previous section. When a (newly designed) approach is entered into the computer simulation, the simulation predicts whether the constraints at the waypoints will be met, and whether a stabilized approach can be achieved at 1,000 feet. It also predicts under what circumstances (e.g., wind conditions) this can be achieved. This section will describe the aircraft model, pilot model, SOPs, wind model and turbulence model that are used within the Monte Carlo computer simulation.

3.1 | Computer simulation input

The input of the Monte Carlo computer simulation exists of a list of waypoints, defined by their lat-lon coordinates, and the altitude and speed constraints at these waypoints. At the moment, the only constraints that can be handled by the simulation are 'AT' constraints, meaning that at a certain waypoint the aircraft should be *at* a certain altitude and *at* a certain speed ('at or above' constraints are not yet possible). Additionally, the user has to define which waypoint in the list is the Final Approach Fix (FAF).

3.2 | Aircraft (B747-200), Autopilot and Flight Director models

Although the Monte Carlo simulation should eventually work for any aircraft type, for now a B747 aircraft model is used in the simulation. The non-linear aircraft model is based on the Boeing 747-200 documentation by Rodney and Hanke [14] and is modeled as detailed as possible. Autopilot, Autothrottle and Flight Director (FD) models are also derived from [14]. Autopilot modes included are: Lateral Navigation (LNAV), Altitude hold, Altitude select, Glideslope, Vertical Navigation (VNAV), Heading Select and Localizer modes. The LNAV mode is based on the VOR modes described in [14], the VNAV mode is based on the vertical speed mode described in [14] where the selected (calculated) vertical speed depends on the constraints at the waypoints and the wind conditions.

The hierarchy in meeting the constraints at the waypoints is as follows: the Autopilot and FD modes will always aim to meet the altitude constraints at the waypoints, second to this, the Autothrottle controls the airspeed. This results in the situation that the altitude constraint at the next waypoint will always be met, while the speed constraint might not be met (airspeed might be higher than required).

3.3 | Pilot model and Standard Operating Procedures for the B747

The pilot manual control model for the flight director task consists of a continuous time delay of 0.3 seconds and a pure gain. All other pilot actions such as selecting flaps are modeled according to the SOPs.

Since the aircraft model in the computer simulation is a Boeing 747-200, the SOPs that are modeled in the computer simulation are based on the SOPs for a Boeing 747, more specifically: the SOPs for the Delayed Flap Approach with flaps LAND equal to flaps 25. The items of the SOPs that are of importance for the computer simulation are the following (see Figure 3): Flaps 1 and flaps 5 should be selected before reaching (Localizer) Intercept Heading. On Localizer Intercept Heading, the pilot should select flaps 10, Arm the Approach, and switch to the Heading Select Mode. At Glide Slope Intercept, the pilot should ask for Gear down, flaps 20 and should arm the speedbrake. Finally, at 1,200 feet the pilot is required to select flaps LAND (equal to flaps 25).

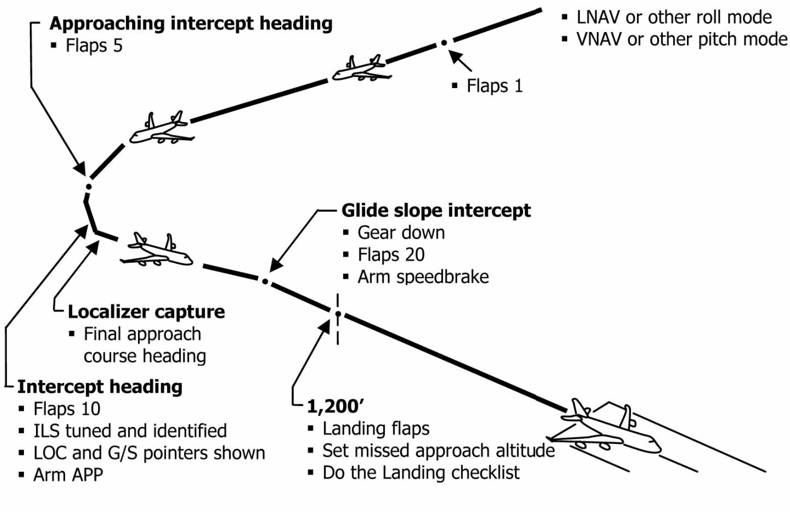


Figure 3 | SOPs for B747 Delayed Flap Approach

Each of these pilot actions prescribed by the SOPs is modeled using a 'trigger' event (e.g., reaching 1,200 feet) and a reaction time representing the time between reaching the trigger event and actually performing the action (e.g., 2 seconds after reaching 1,200 feet, flaps 25 are selected). These reaction times are modeled based on normal distributions. The trigger events and corresponding reaction times as used in the computer simulation are given in

Table 1, see also Figure 4. Since no information could be found regarding the reaction times, the values in Table 1 are a first guess. The trigger events for all actions are explained below.

Table 1 | Trigger events and reaction time distributions for pilot actions in Monte Carlo simulation.

Pilot action	Trigger Event	Mean	Standard deviation
Flaps 1	Reaching First waypoint of the leg on which aircraft is 20nm from field	T/3(1)	T/3(1)
Flaps 5	Reaching first waypoint of the leg before Localizer Intercept Heading	2×T/3(2)	T/3(2)
Autothrottle / Autopilot Off	Turn to Localizer Intercept Heading	T/2(3)	T/2(3)
Flaps 10	Autothrottle / Autopilot Off	2 sec	0.5 sec
ARM Approach	Selecting Flaps 10	4 sec	1 sec
Gear Down	Glideslope Intercept / Reaching FAF	2 sec	0.5 sec
Flaps 20	Gear Down	2 sec	0.5 sec
Flaps 25	Reaching 1,200’	2 sec	0.5 sec

(1) T = amount of time spent on leg on which the aircraft is 20nm from field
(2) T = amount of time spent on leg before Localizer Intercept Heading
(3) T = amount of time spent on Localizer Intercept Heading

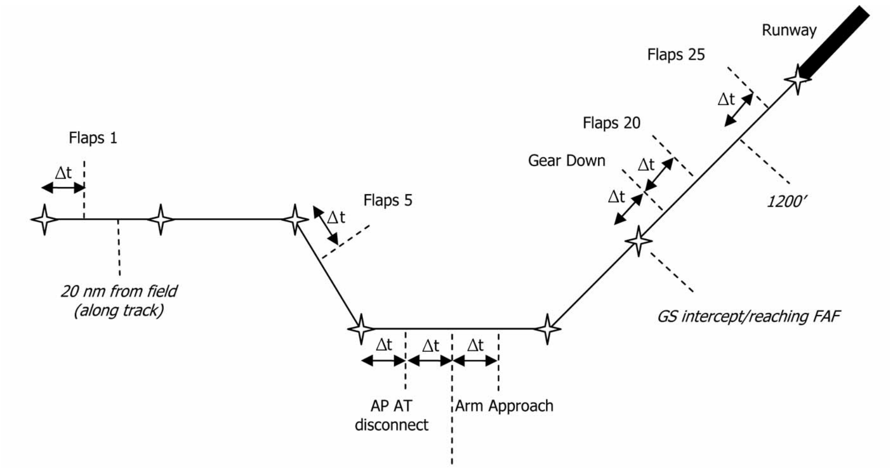


Figure 4 | Visualization of trigger events and reaction times (Δt) for pilot actions in the Monte Carlo computer simulation.

The SOPs do not give a specified location in the approach where flaps 1 should be selected. Based on conversations with pilots it was concluded that pilots would select flaps 1 when they are approximately 20 nm (along track) from the field. Therefore it was decided to use the moment of reaching the starting waypoint of the leg on which the aircraft is 20nm from the field (along track) as a trigger event. Based on the same conversations, the trigger event for selecting flaps 5 was chosen to be the moment of reaching the first waypoint of the leg *before* Localizer Intercept Heading (see also Figure 4).

The first pilot action on Localizer Intercept Heading modeled in the Monte Carlo simulation is not prescribed by the SOPs, but follows from the chosen level of automation that is used during the approach (see Figure 2), this action is Autothrottle and Autopilot disconnect (for now assumed to take place simultaneously). The trigger event for this first action on Localizer Intercept Heading is starting the turn to Localizer Intercept Heading. The trigger event for the second action on Localizer Intercept Heading (selecting flaps 10) is the first action on Localizer Intercept Heading (Autothrottle and Autopilot disconnect), and so on, see also Figure 4 and Table 1. The actions on Localizer Intercept Heading are thus modeled sequentially, always in the same order, although the SOPs do not prescribe a fixed order. Officially, pilots should also switch from LNAV mode to Heading Select mode on Localizer Intercept Heading, this switch however does not have any impact on the Monte Carlo simulation, and is therefore, for now, not modeled. In reality, however, the effect of not switching to Heading Select mode would cause the pilot considerable difficulty in easily and properly establishing a correct Localizer Intercept Heading.

Just as for the actions on Localizer Intercept Heading, the actions Gear Down and flaps 20 are also always modeled in the same order while the SOPs do not prescribe this. The trigger event for Gear Down is Glide Slope Intercept, however, when the approach is designed as a Continuous Descent Approach (implying that the Glide Slope could be captured many miles out), the trigger event is reaching the FAF. The trigger event for selecting flaps 20 is the selection of Gear Down.

The trigger event for selecting flaps 25 (flaps LAND) is reaching 1,200 feet altitude.

It is important to note that, due to simulation technical reasons, (for now) in the computer simulation only positive reaction times could be modeled. If any of the distributions for the reaction times in Table 1 yielded a negative number, a new value was created using the same distribution until the value for the reaction times was a positive number. The reaction times are thus based on normal distributions, but are not necessarily normally distributed. Another important remark concerns the selection of flaps: if the airspeed constraints at the waypoints required an airspeed lower than the instantaneous flap speed mark, the next flap setting is selected in the Monte Carlo simulation irrespective of SOPs. For example if the UP mark (speed below which flaps 1 should be selected) is at 220 knots, and the aircraft decelerates to 210 knots because that is the speed constraint at the next waypoint, then flaps 1 are selected when the airspeed becomes less than 220 knots, even if the aircraft is still 50 nm from the runway.

3.4 | Turbulence and Wind models

Turbulence is modeled according to the Dryden spectra [15], the longitudinal scale (L_g) is fixed in the simulation at 300 m whereas the turbulence intensity (σ) is varied according to a Weibull distribution, with λ = 2 m/s, k = 2. The wind speed is modeled using a Weibull distribution (λ = 12.5 kts, k = 2.0). The wind direction is varied around the runway heading by applying a normal distribution with μ = runway heading and σ = 30 deg. During one Monte Carlo simulation run the turbulence intensity, wind direction and wind speed are constant throughout the entire approach, between different Monte Carlo runs these are varied.

3.5 | Outputs of the computer simulation

The output of the Monte Carlo simulation is a prediction whether the constraints at the waypoints can be met, and whether a stabilized approach can be achieved at 1,000 feet.

The constraints at a waypoint were considered to be met when the actual Indicated Airspeed (IAS) at that waypoint was less than the required IAS plus 10 knots, and the actual altitude at that waypoint was less than the required altitude plus 100 feet. A lower boundary for these constraints is not necessary since the Monte Carlo simulation always regulates the airspeed and altitude towards the constraints at the next waypoint, when the required airspeed and altitude are attained, the Monte Carlo simulation maintains the required airspeed and altitude until the waypoint is reached. Therefore, in the Monte Carlo simulation, the altitude and airspeed will never be too low at a waypoint.

To determine whether a Monte Carlo simulation run of the approach resulted in a stabilized approach, the criteria from paragraph 2.5 were quantified as follows:

- Heading change and pitch change are within 5 deg/s;
- The IAS is not more than V_{REF} + 20 knots;
- Flaps 25 are selected, landing gear is down;
- Sink rate is not larger than 1,000 feet per minute; and
- Localizer and glide slope are within one dot;

These criteria are evaluated at 1,000 ft above airport elevation. An average value is calculated for each of the criteria: for the interval starting 5 seconds before reaching 1,000 ft and ending at 1,000 ft. A larger time interval is taken to calculate the average sink rate, since this criterion does not directly refer to the 1,000 ft point in itself, but refers to the approach. The time interval chosen to calculate the sink rate starts 1 minute before reaching 1,000 ft, and ends at 1,000 ft. The completion of all checklists and the power setting are, at the moment, not included as criteria for a stabilized approach.

4 | Point mass model for initial estimates

For rapid evaluation of an approach, a point mass model (PMM) has been developed. This point mass model is less accurate but much faster than the non-linear aircraft model used in the Monte Carlo simulation, and is used to obtain an initial estimate whether constraints at the waypoints can be met at all, and whether a stabilized approach can be achieved. It does so by using the energy rate demand which is defined as the ratio of the energy rate as commanded by the trajectory, $\dot{E}_{cmd}$, and the minimum energy rate that can be achieved by the aircraft $\dot{E}_{ac}$. The PMM calculates the energy rate demand for each leg of the approach for a predefined aircraft weight, and predefined wind speed and wind direction. The PMM assumes SOPs, implying that flap and gear settings during the approach are performed according to SOPs.

4.1 | Expression for the energy rate demand

The general formula for the energy rate demand is derived below, and the simplifications that result in a formula for the energy rate demand applicable to the PMM are explained subsequently. Additionally, some comments are given on the influence of the aircraft weight on the energy rate demand, since this influence might differ from general perception. The limitations of the predictions of the PMM due to the simplifications that are made are summarized at the end of this section.

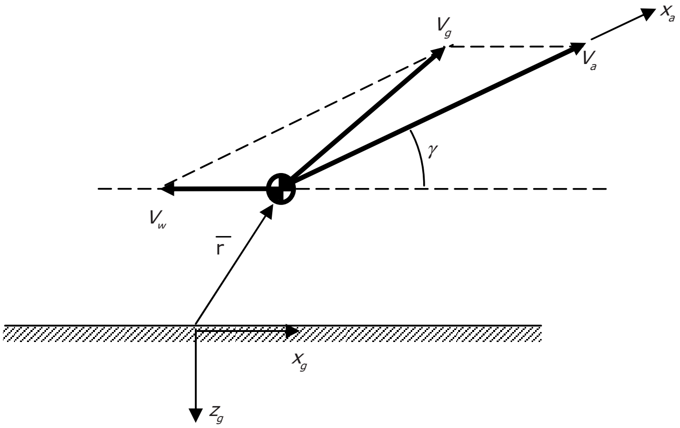


Figure 5 | Definition of air path reference frame and geodetic reference frame.

Starting with the equation of motion in the direction of flight for an aircraft in a horizontal wind field and assuming small angles it is possible to write (see also Figure 5):

$$\frac{W}{g}\frac{d}{dt}\left(V_a - V_w\right) = T - D - W\gamma \tag{1}$$

$$\frac{dx_g}{dt} = V_a - V_w \tag{2}$$

$$\frac{dH}{dt} = V_a\gamma \tag{3}$$

Here, W is the aircraft weight, g the gravitational acceleration, V_a the true airspeed in the air path system, V_w the horizontal wind speed in the geodetic reference frame, V_g the aircraft's speed in the geodetic reference frame, γ the flight path angle, T the aircraft thrust and D the aerodynamic drag (see Figure 1). For the purpose of the PMM it is necessary to express the energy rate demand in terms of the geodetic reference frame; the kinematic equations (Equations (2) and (3)) are therefore used:

$$\frac{W}{g}\frac{dV_a}{dx_g}\left(V_a - V_w\right) - \frac{W}{g}\frac{dV_w}{dt} = T - D - \frac{W}{V_a}\left(V_a - V_w\right)\frac{dH}{dx_g} \tag{4}$$

Assuming a uniform wind field ($V_w = \overline{V}_w$ = *constant,* $\frac{dV_w}{dt} = 0$) and re-ordering terms yields*:*

$$\frac{W}{g}\frac{dV_a}{dx_g}\left(V_a - V_w\right) + \frac{W}{V_a}\left(V_a - V_w\right)\frac{dH}{dx_g} = T - D \tag{5}$$

The energy rate demand can now be expressed in the geodetic reference frame as:

$$\hat{E} = \frac{\frac{(V_a - V_w)}{g}\frac{dV_a}{dx_g} + \frac{(V_a - V_w)}{V_a}\frac{dH}{dx_g}}{\frac{T_0}{W} - \frac{C_D}{C_L}} \tag{6}$$

In this equation the numerator represents the energy rate as commanded by the trajectory (which will be a negative number in the case of approaches). In the denominator, the value for the flight idle thrust T_0 is used, such that the denominator represents the maximum energy rate (decrease) that can be achieved by the aircraft. Note that in the case of approaches this will also be a negative number.

Within the PMM an *average* value for airspeed ($\bar{V}_a$) and altitude ($\bar{H}$) is used for each leg of the approach, instead of the instantaneous value, see Figure 6 for an example. This logically results in an average value for the flight idle thrust ($\bar{T}_0$), aerodynamic drag coefficient ($\bar{C}_D$) and aerodynamic lift coefficient ($\bar{C}_L$) for each leg. With this assumption, the expression for the energy rate demand for each leg of the approach (using Equation (6)) simplifies to:

$$\hat{E} = \frac{\frac{\bar{V}_a - V_w}{g}\Delta V_a + \left(1 - \frac{V_w}{\bar{V}_a}\right)\Delta H}{\left(\frac{\bar{T}_0}{W} - \frac{\bar{C}_D}{\bar{C}_L}\right)\Delta x_g} \tag{7}$$

This is the expression for the energy rate demand used in the PMM. The PMM also assumes that flaps and gear are selected at the earliest moment possible according to SOPs. For example: flaps 10 should be selected on Localizer Intercept Heading, therefore the PMM assumes that flaps 10 are selected at the first waypoint on Localizer Intercept Heading, see also Figure 6. Additionally, the PMM does not consider the flap speed marks, flaps are always selected according to the SOPs.

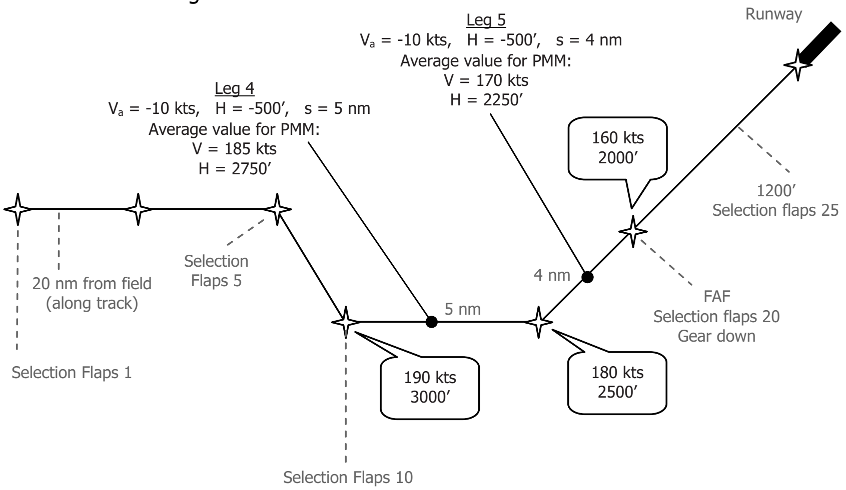


Figure 6 | Example of average values used in the PMM for each leg, and SOPs as assumed in the PMM.

4.2 | Influence of aircraft weight

As stated before, the PMM calculates the energy rate demand for each leg of an approach as a function of, among others, aircraft weight. To understand the predictions of the PMM, it is necessary to understand the influence of aircraft weight on the energy demand ratio. In this light, a common remark heard from pilots is that it becomes more difficult to dissipate energy (and thus to meet the constraints at waypoints) with increasing aircraft weight. To verify this remark, equation (6) is used.

It can be seen that the rate at which the aircraft can dissipate energy (the denominator in Equation (6) and (7)) depends on the ratios of flight idle thrust and weight, and lift coefficient (C_L) and drag coefficient (C_D). An increase in weight decreases the flight idle thrust over weight ratio, and in that respect increases the energy rate demand. Note that the idle thrust is a function of atmospheric pressure, Mach number, ambient temperature and altitude. However, the aircraft weight also influences the C_D/C_L ratio. When considering a given approach with defined airspeed and altitude constraints, and assuming that the approach is flown according to SOPs, this implies that at each point of the approach the airspeed, altitude, flap setting and gear setting are known and with this the values of C_L and C_D for a given aircraft and aircraft weight. When the aircraft weight is increased, the effect on the energy rate demand depends on whether the C_D/C_L ratio decreases or increases (becomes more or less optimal) as a result of this increase in weight for a particular segment of the approach.

For some altitude and airspeed constraints it is the case that a higher aircraft weight brings about an increase in C_D/C_L ratio larger than the decrease in idle thrust over weight ratio, and thus diminishes the energy rate demand. For other altitude and airspeed constraints both ratios will decrease with higher aircraft weight, and as a result the energy rate demand will increase. This will be illustrated in the case study presented in the next section. It is thus not true that a higher aircraft weight by definition increases the energy rate demand, it depends on the conditions. This is important to keep in mind when analyzing the results of the PMM and Monte Carlo simulation.

4.3 | Reliability of PMM predictions

The assumptions made for the PMM do influence the reliability of the energy rate demands that are calculated by the PMM. The fact that, during most approaches, flaps and gear will be selected at a later point in time than assumed in the PMM will influence the energy rate demand. Additionally, the assumption of average airspeeds and altitudes has a large effect on the value of the idle thrust, and with that on the value of the energy rate demand. Comparison of the energy rate demands calculated by the PMM, and the energy rate demands that occurred in the Monte Carlo simulation indicated that in some (extreme) cases an energy rate demand as high as 1.2 (as predicted by the PMM) would for the same approach result in energy rate demands smaller than 1 in the Monte Carlo simulation. In other words: the PMM as it is at the moment can give a global indication, which serves well as a first estimate, but is certainly not accurate. That is also its intended use: when a newly designed approach is analyzed, it will first be analyzed using the PMM for a global indication. Some modifications can then already be made based on these global predictions. After this a more accurate analysis can be performed using the Monte Carlo computer simulation.

5 | Case Study

To demonstrate the predictions of the PMM and the Monte Carlo simulation, the fictitious approach as illustrated in Figure 7 (the details of which are summarized in Table 2) is considered. Note that the information as given in Table 2 is the only input required for the Monte Carlo simulation. It can be seen that this is a very short continuous descent approach, with a very low FAF at 1350′. The only function of the FAF in this case is that pilots are required to select flaps 20 and gear down at the FAF, therefore this approach with low FAF can also be regarded as a modified procedure for late selection of flaps 20 and gear down. Table 2 shows that the Monte Carlo simulation and PMM will use an indicated airspeed equal to VREF plus 5 knots at the threshold. Moreover, the Monte Carlo and PMM simulation will also aim for VREF + 5 knots at 1,000′, which is well within the margins for a stabilized approach.

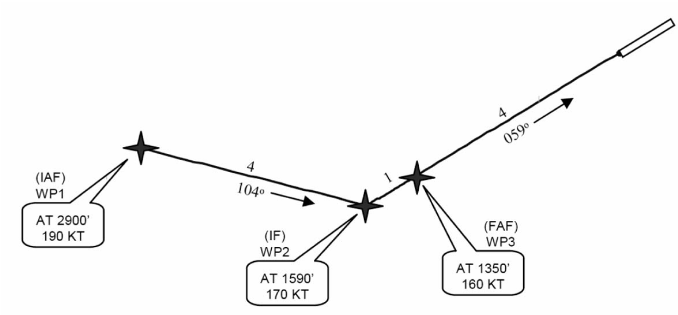


Figure 7 | Approach chart

Table 2 | Waypoints with altitude and airspeed constraints

Waypoint		Altitude [ft]	IAS [knots]
WP1	IAF	2,900	190
WP2	IF	1,590	170
WP3	FAF	1,350	160
RWY	×	0	VREF + 5

5.1 | Point Mass Model predictions

To obtain a first indication, this approach is analyzed using the PMM, the results for two different aircraft weights *for zero wind conditions* are given in Table 3. The PMM assumes that flaps and gear are selected according to SOPs, it does not take into account whether the required speeds are too low for the flap settings prescribed by the SOPs. Table 3 shows that the energy demand ratios (denoted 'E-ratios' in the table) are calculated for five subsequent parts of the approach (note that an E-ratio smaller than 1 indicates that the constraints at the waypoint can be met), these parts are:

1. when flying from WP1 to WP2, energy demand ratios 1.20 and 1.14
2. when flying from WP2 to WP3, energy demand ratios 1.11 and 1.01
3. when flying from WP3 to 1,200′, energy demand ratios 1.27 and 1.20
4. when flying from 1,200′ to 1,000′, energy demand ratios 1.15 and 0.61
5. when flying from 1,000′ to the runway, energy demand ratios 0.61 and 0.61

It can be seen that for all parts of the approach the energy demand ratio for the highest aircraft weight (295,000 kg) is *lower* than the energy demand ratio for the lowest aircraft weight (238,000 kg). This implies that it would be easier to meet the constraints at the waypoints and to achieve a stabilized approach (only considering the airspeed) with a higher aircraft weight. For the first three parts of the approach (WP1 to 1,200′) this is entirely due to the different location on the C_L/C_D curve due to the aircraft weight, see Figure 8. It can be seen that for each of these three approach parts the location of the value of C_L/C_D for the higher aircraft weight is located further away from the maximum (or, optimal) C_L/C_D value when compared to the location of the value of C_L/C_D for the lower aircraft weight. This implies that the C_L/C_D value for the higher aircraft weight is smaller than for the lower aircraft weight, which results in a decrease of the energy demand ratio.

Table 3 | Predictions of the PMM for zero wind conditions and two different aircraft weights.

				238,000 kg			295,000 kg		
Waypoint		**Flaps**	**Gear**	C_L	C_D	**Eratio**	C_L	C_D	**Eratio**
WP1	IAF	10	UP						
WP2	IF	10	UP	0.95	0.071	1.20	1.04	0.079	1.14
WP3	FAF	20	DOWN	1.13	0.087	1.11	1.24	0.10	1.01
1,200′		25	DOWN	1.20	0.12	1.27	1.32	0.14	1.20
1,000′		25	DOWN	1.30	0.14	1.15	1.32	0.14	0.61
RWY		25	DOWN	1.40	0.15	0.61	1.32	0.14	0.61

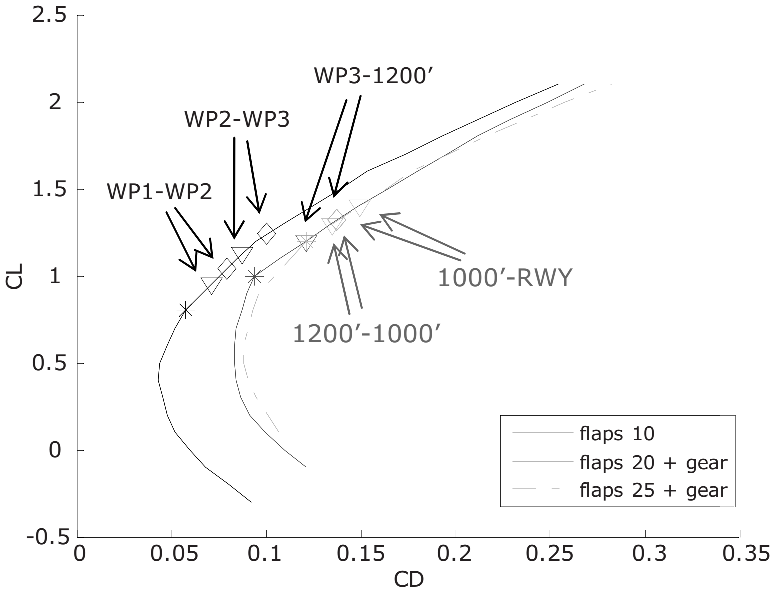


Figure 8 | Lift drag polar for different flap and gear settings. The asterisks represent the maximum C_L/C_D value for the lift drag polars. The triangles represent C_L/C_D values for the lowest aircraft weight (238,000 kg), the diamonds represent C_L/C_D values for the highest aircraft weight (295,000 kg).

Now, for the second to last part of the approach (1,200′-1,000′), another factor contributes to the smaller energy demand ratio for the larger aircraft weight than for the lower aircraft weight. The value for VREF for the higher aircraft weight is 161 knots, whereas the value for the lower aircraft weight is 143 knots. Meaning that when flying with the higher aircraft weight, it is not necessary to decelerate anymore after reaching WP3 (which has an airspeed constraint equal to 160 knots), whereas, when flying with the lower aircraft weight, a deceleration of 12 knots is required to reach the final approach speed equal to VREF + 5 knots.

Judging from the results of the PMM it can thus be expected that (for zero wind conditions) it will probably not be possible to achieve a stabilized approach and meet the constraints at the waypoints for the lower aircraft weight, since all values for the energy rate demand are (much) larger than 1. For the higher aircraft weight it might be possible to meet the constraints and to achieve a stabilized approach (only considering the airspeed), since the values for the energy rate demand are closer to 1. Note that when the PMM would given an exact prediction of the energy rate demand, a value larger than 1 would mean that the constraints could not be met. However, due to the assumptions in the PMM a value slightly larger than 1 could in reality still result in meeting the constraints.

5.2 | Monte Carlo simulation results

The same approach is analyzed using the Monte Carlo simulation. As explained before, the variables within the Monte Carlo simulation are the reaction times for all pilot actions, aircraft weight, the wind direction, the wind speed and the turbulence intensity. Compared to the PMM, the Monte Carlo simulation uses a more realistic simulation of the pilot's actions which also vary in time for each run, and incorporates the effects of varying wind and turbulence.

The trajectory and location of the pilot actions resulting from the Monte Carlo simulation are given for five runs in Figure 9, for both the low aircraft weight (left) and the high aircraft weight (right). It can be seen that the Monte Carlo simulation stops at 1,000', the remainder of the trajectory, until the runway threshold, is indicated in light grey. Since this approach starts at Localizer Intercept Heading, the Monte Carlo simulation is initialized to start with flaps 5, unless the required airspeed is lower than can be achieved with flaps 5.

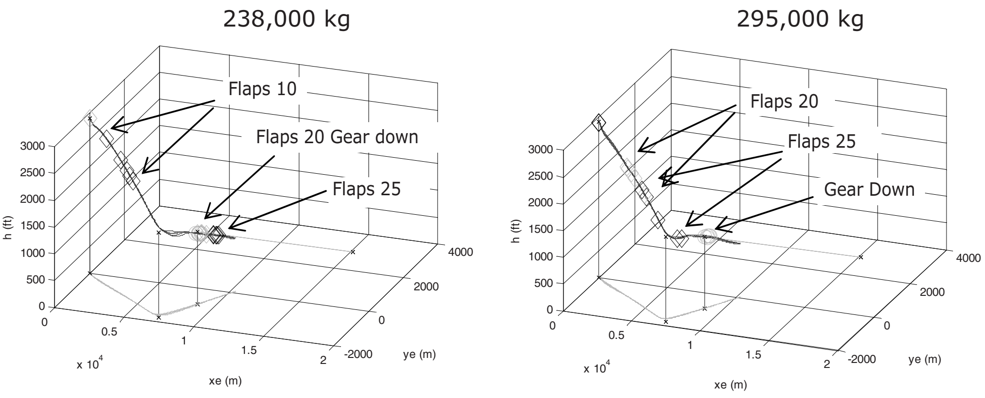


Figure 9 | Trajectory and pilot actions according to Monte Carlo simulation for five runs.

For the low aircraft weight, Figure 9 shows that flaps 10 are indeed selected on Localizer Intercept Heading, the actions flaps 20 and gear down are performed around the FAF, and flaps 25 are selected at 1,200', all according to the SOPs. For the high aircraft weight, however, Figure 9 indicates that the pilot actions will deviate from the SOPs. The airspeed constraints at the waypoints are, for the higher aircraft weight, too low to adhere to SOPs. Due to the airspeed constraints, the Monte Carlo will start with flaps 10, additionally, flaps 20 and flaps 25 need to be selected (at 181 and 171 knots respectively) due to the flap speed marks, which results in the selection of flaps well before the SOPs require them to be selected. The Monte Carlo simulation thus immediately provides the insight that it is impossible to fly this approach according to SOPs with an aircraft weight as high as 295,000 kg.

Whether a stabilized approach was achieved at 1,000’ is depicted in Figure 10 for 1,000 runs with the Monte Carlo simulation. It should be noted that in Figure 10, *all* unstabilized approaches are due to the fact that the IAS was higher than VREF + 20 knots, i.e., for all approaches the flap setting, gear setting, Localizer deviation, etcetera were all according to the criteria given in paragraph 4.5 when assuming trigger events and reaction times as defined in Table 1. It can be seen that for the lower aircraft weight a stabilized approach can be achieved for wind speeds larger than 10 knots coming from (approximately) runway heading or from a direction slightly larger than runway heading (from the ‘right hand side’ of the runway). This makes sense because, in this case, the aircraft would have a considerable headwind both on Localizer Intercept Heading as well as on final (flying from WP2 to WP3), resulting in more time to meet the constraints and to achieve a stabilized approach.

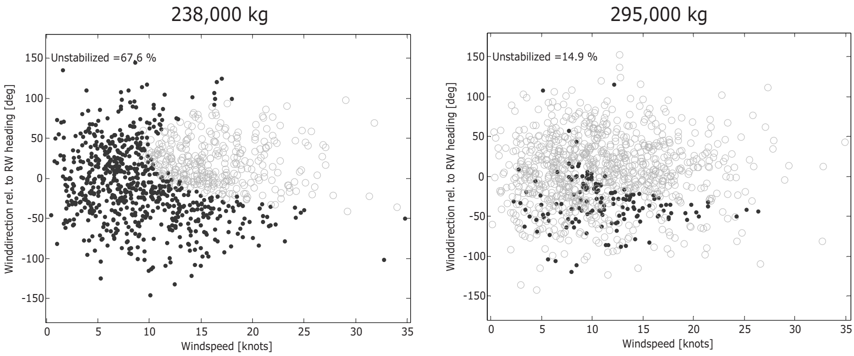


Figure 10 | Results of the Monte Carlo simulation (N = 1,000), for different wind speeds and different wind directions (relative to runway heading). In this figure and the following, a grey circle indicates a stabilized approach was achieved at 1,000’, a black dot indicates that the approach was not stabilized at 1,000’.

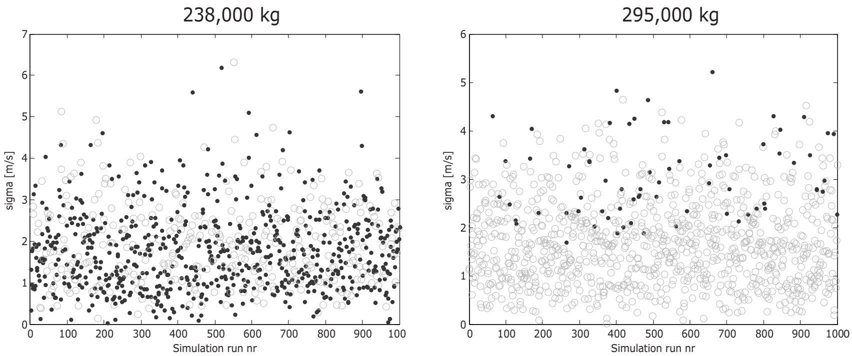


Figure 11 | Turbulence intensity factors (s) as used for each run of the Monte Carlo simulation.

For the higher aircraft weight, Figure 10 shows that a stabilized approach can be achieved for all wind speeds and all wind directions. However, in some cases an unstabilized approach occurs. This is due to the turbulence: when an unfavorable wind direction is combined with a large turbulence gust in the ‘wrong’ direction, this will result in an unstabilized approach.

Figure 11 indeed shows that all unstabilized approaches occur for runs with a very high turbulence intensity factor (σ). When running the Monte Carlo simulation again, and using a turbulence intensity factor equal to zero (no turbulence) for all runs, this yields the results given in Figure 12. Figure 12 indeed demonstrates that all approaches are now stabilized for the higher aircraft weight.

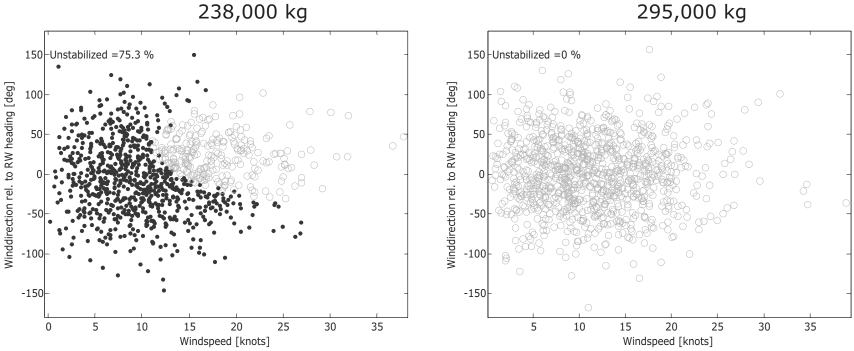


Figure 12 | Results of the Monte Carlo simulation (*N* = 1,000), for different wind speeds and different wind directions (relative to runway heading), with turbulence intensity factor equal to zero in all runs.

The fact that, for this fictitious approach, more approaches are stabilized for the higher aircraft weight than for the lower aircraft weight is due to the fact that the reference speed VREF is higher, the fact that flaps 20 and 25 are selected much earlier in the approach (adding more drag to the aircraft for a larger part of the approach), and might, additionally, be due to the smaller energy rate demand as predicted by the PMM. For a stabilized approach the airspeed at 1,000ft needs to be lower than VREF plus 20 knots. This implies that the airspeed should be lower than 181 knots for the high aircraft weight, and lower than 163 knots for the low aircraft weight. The aircraft with the higher weight thus needs to decelerate less in the last part of the approach than the aircraft with the lower weight. This effect is entirely due to the different numerical values that result from the airspeed criterion for a stabilized approach, it has, in itself, nothing to do with the ability of the aircraft to dissipate energy.

Table 4 | Percentage of approaches during which the constraints at the waypoints WP2 were met according to the Monte Carlo simulation (*N* = 1,000), results given for simulations with turbulence, and without turbulence (σ = 0).

	238,000 kg	**295,000 kg**	**238,000 kg σ = 0 m/s**	**295,000 kg σ = 0 m/s**
WP2	11.4 %	90.9 %	10.7 %	93.5 %

The Monte Carlo simulation also calculates whether the constraints at the waypoints are met. As an example, Table 4 gives an overview of the percentage of flights during which the constraints were met at waypoint WP2, for flights with turbulence, and flights without turbulence (σ = 0 m/s). It can be seen that for the higher aircraft weight a higher percentage of flights meets the constraints at the waypoints. The fact that the constraints are not met is always due to the fact that the airspeed is too high. The VNAV mode will direct the aircraft towards the required altitude, and second to this the autothrottle controls the airspeed.

When the path is too steep (i.e. the energy rate demand is too high), the autothrottle can not regulate the airspeed towards the desired value even when applying idle thrust. In reality, when this occurs, a 'drag required' message appears (see also the accompanying paper [4]) and pilots will use speedbrakes in an attempt to meet the constraints. However, within this flight mechanical assessment tool that is to be used during the design of approaches, speedbrakes are not incorporated since the premise is that approaches should be designed such that they can be flown without the use of speedbrakes.

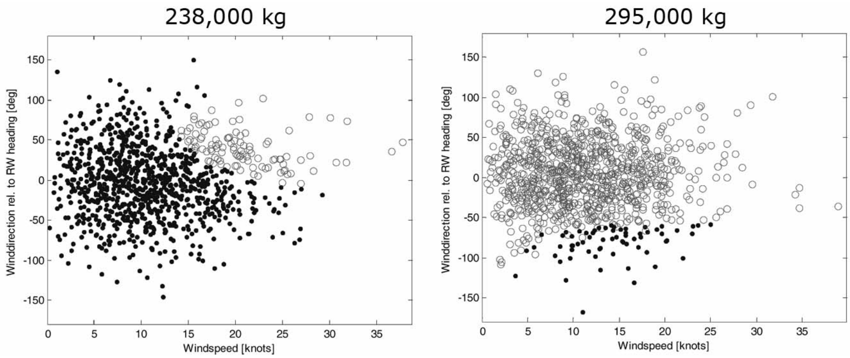


Figure 13 | Results of the Monte Carlo simulation (N = 1,000) for WP2, for different wind speeds and different wind directions (relative to runway heading). A grey circle indicates the altitude and airspeed constraints were met at WP2, a black dot indicates that the constraints were not met at WP2.

The results for waypoint WP2 are visualized as a function of wind speed and wind direction in Figure 13. It shows that for the lower aircraft weight, the constraints at WP2 are not met, unless there is a considerable headwind when flying from WP1 to WP2. For the higher aircraft weight, the constraints at WP2 are met, unless the aircraft experiences a tailwind while flying from WP1 to WP2.

Concluding, it can be stated that for this particular approach with these airspeed and altitude constraints, the aircraft with the higher weight performs better than the aircraft with the lower weight. The fact that the higher aircraft weight shows better results, however, is not necessarily true for other approaches as well. It depends on the combination of three factors, which can combine into better or worse performance. The three factors are: 1. the value of the energy rate demand for the particular aircraft weight, 2. the value of the reference speed VREF for the particular aircraft weight (which determines how much the aircraft needs to decelerate in the last part of the approach), and 3. the fact whether or not the approach can be flown according to SOPs for that aircraft weight. (With increasing aircraft weight, the chance increases that flaps need to be selected before the SOPs require them to be selected. This in itself is not desirable since the approach can in this case not be flown according to SOPs, however, it will result in better deceleration characteristics due to the additional drag caused by the flaps).

5.3 | Regression

A stepwise linear regression analysis was performed for this particular approach on the results of the Monte Carlo simulation in order to determine which variables that are modeled in the Monte Carlo simulation have an influence on whether the constraints at the waypoints can be met and whether or not a stabilized approach was achieved at 1,000'.

Table 5 | Results of a stepwise linear regression for the dependent measure 'IAS at 1,000ft' (probability of F to enter <=.05, probability of F to remove >=.01). The predictors are listed vertically.

	238,000 kg σ = 0	295,000 kg σ = 0	Both masses σ = 0
Flaps 10 reaction time	×	×	×
Approach ARM reaction time	×	×	×
APFD switch reaction time	×	**.138 (3)**	**.623 (5)**
Flaps 20 reaction time	**.951 (5)**	×	×
Gear down reaction time	.949 (3)	×	.622 (4)
Flaps 25 reaction time	.951 (4)	×	×
Headwind on final	.948 (2)	.135 (2)	.621 (3)
Headwind on LOC int. HDG	.915 (1)	.049 (1)	.616 (2)
σ	n.a.	n.a.	n.a.
Aircraft mass	n.a.	n.a.	.479 (1)

The first regression presented in this section concerns the stabilized (or unstabilized) approaches at 1,000'. Since the only reason for the unstabilized approaches in the Monte Carlo simulation was the fact that the IAS was higher than VREF + 20 at 1,000 ft, a prediction of whether the approach is stabilized can be directly derived from a prediction of the IAS at 1,000'. Therefore, the 'IAS at 1,000ft' is chosen as dependent measure for the regression analysis. A regression is performed for the two aircraft weights separately as well as for the results of the two aircraft weights combined, see Table 5. The predictors for the regression are the inputs for the Monte Carlo simulation, and are listed vertically in the first column of Table 5. The entries show the value of R^2, the number in parenthesis indicates the step at which the predictor was added to the regression. The maximum difference between R^2 and adjusted R^2 values in this table is 0.2%. No multicollinearity can be assumed: for all predictors the Variance Inflation Factor (VIF) is smaller than 1.8 (and hence the tolerance larger than .55).

It can be seen that the R^2 value for the lower aircraft weight (.951) is substantially larger than the R^2 value for the higher aircraft weight (.138). This can be explained by the fact that for the higher aircraft weight more approaches are stabilized, and therefore a large percentage of the approaches will achieve an IAS equal to VREF + 5 knots at 1,000', independent of the value of the predictor variables. As a result there will be less correlation between the IAS at 1,000' (largely constant between many runs) and the predictors. This effect can also be seen in the results of the regression for both aircraft weights.

Therefore, in the authors' opinion, the regression model for the lower aircraft weight gives the best indication of predictors that influence the IAS at 1,000', and thus the predictors that influence the chance of achieving a stabilized approach (first column in Table 5). This regression model is given in Table 6. It can be seen that the wind direction (represented by the predictors 'headwind on final' and 'headwind on Localizer Intercept Heading') has the largest influence on the IAS at 1,000', which agrees with the discussion of the Monte Carlo simulation results shown in Figure 12. Additionally, the reaction time in pilot actions for the actions flaps 20, flaps 25 and gear down also influence the IAS at 1,000' but to a much lesser extent. However, for these pilot reaction times one would expect positive standardized β values, since a larger reaction time represents a delay in adding more drag to the aircraft and is thus expected to result in a higher IAS at 1,000′. This is not the case for the standardized β value for the flaps 20 reaction time, therefore it is better to use the regression model given in step 4 (Table 6), or even use the regression model given in step 2 since with only two predictors an R^2 value of .948 is already achieved. For this particular approach, pilot actions that occur earlier in the approach such as selecting flaps 10, do not appear to influence the IAS at 1,000′.

Table 6 | The stepwise linear regression model for the dependent measure 'IAS at 1,000ft', aircraft mass 238,000kg, without turbulence.

	B	***SE B***	***β***	***R²***	**Adjusted *R²***
Step 1					
Constant [knots]	167.53	.043			
Headwind on LOC int. HDG [knots]	-4.78	.005	-.957*	.915	.915
Step 2	×	×	×	×	×
Constant [knots]	168.22	.044			
Headwind on LOC int. HDG [knots]	-.402	.005	-.803*		
Headwind on final [knots]	-.136	.005	-.237*	.948	.947
Step 3					
Constant [knots]	167.798	.083			
Headwind on LOC int. HDG [knots]	-.401	.005	-.802 *		
Headwind on final [knots]	-.137	.005	-.237 *		
Gear down reaction time [sec]	.210	.035	.042 *	.949	.949
Step 4					
Constant [knots]	137.43	.109			
Headwind on LOC int. HDG [knots]	-.401	.005	-.802 *		
Headwind on final [knots]	-.136	.005	-.236 *		
Gear down reaction time [sec]	.210	.035	.042 *		
Flaps 25 reaction time [sec]	.182	.035	.037 *	.951	.950
Step 5					
Constant [knots]	167.61	.131			
Headwind on LOC int. HDG [knots]	-.402	.005	-.803 *		
Headwind on final [knots]	-.136	.005	-.236 *		
Gear down reaction time [sec]	.208	.035	.042 *		
Flaps 25 reaction time [sec]	.181	.035	.037 *		
Flaps 20 reaction time [sec]	.088	.035	-.018 **	.951	.951

* $p < .001$, ** $p = .012$

The second linear stepwise regression is performed for meeting the constraints at WP2. As explained before, the altitude constraints are always met, so the reason for not meeting the constraints is an airspeed that is too high. Therefore the dependent measure is chosen to be the 'IAS at WP2'. Table 7 shows that the R^2 value is again considerably lower for the higher aircraft weight than for the lower aircraft weight, this is because of the same reasons as given above (for the higher aircraft weight a large percentage of the flights did meet the constraints, and therefore a large percentage of the flights arrived at and remained at 170 knots). Therefore, the regression model for the lower aircraft weight is again considered, it is given in Table 8. It can be seen that the amount of headwind on Localizer Intercept Heading is by far the most important predictor. This agrees with the results shown in Figure 13 and the discussion of the results of the Monte Carlo simulation.

Table 7 | Results of a stepwise linear regression for the dependent measure 'IAS at WP2'. The predictors are listed vertically. The maximum difference between R^2 and adjusted R^2 values is 1.5%. The VIF is smaller than 1.1.

	238,000 kg σ = 0	**295,000 kg σ = 0**	**Both masses σ = 0**
Flaps 10 reaction time	×	×	×
Approach ARM reaction time	.946 (2)	×	×
APFD switch reaction time	×	.426 (2)	.882 (3)
Headwind on LOC int. HDG	.946 (1)	.421 (1)	.881 (2)
σ	n.a.	n.a.	n.a.
Aircraft mass	n.a.	n.a.	.673 (1)

Table 8 | The stepwise linear regression model for the dependent measure 'IAS at WP2', aircraft mass 238,000kg, without turbulence.

	B	***SE B***	***β***	***R^2***	**Adjusted *R^2***
Step 1					
Constant [knots]	186.66	.032			
Headwind on LOC int. HDG [knots]	-.451	.003	-.973 *	.946	.946
Step 2	×	×	×	×	×
Constant [knots]	186.46	.100			
Headwind on LOC int. HDG [knots]	-.451	.003	-.973*	.946	.946
Approach Arm delay	.049	.024	.015**	.946	.946

* $p < .001$, ** $p = .04$

5.4 | Comparison of results

The results of the Monte Carlo simulation seem credible and can be explained when considering the influence of wind speed, wind direction and turbulence on meeting the constraints at waypoints and achieving a stabilized approach at 1,000'. This is true for both aircraft weights.

The results of the regression analysis correspond very well with the explanation of the results of the Monte Carlo simulation for the lower aircraft weight. The regression analysis indicated that the wind speed and wind direction had the largest influence on meeting the constraints at waypoints and achieving a stabilized approach at 1,000’ whereas the pilot reaction times for selecting flaps and gear according to SOPs did not have a large influence or no influence at all. This can, however, be due to the fact that the reaction times for all pilot actions in the Monte Carlo simulation (see Table 1) were chosen too small.

For the higher aircraft weight, the regression analysis did not yield good results in terms of R^2, and did not correspond with the explanation of the results of the Monte Carlo simulation. This is because for the higher aircraft weight most of the time the constraints could be met and a stabilized approach was achieved, resulting in a constant value for the dependent measure (IAS) for a very large percentage of all approaches, which logically influenced the regression analysis.
The predictions of the PMM approximately agreed with the results of the Monte Carlo simulation for the lower aircraft weight, since for this weight the approach could indeed be performed according to SOPs. If the airspeeds required by the approach were too low (i.e., when due to the required airspeed flaps needed to be selected earlier in the approach than prescribed by SOPs) the predictions of the PMM were based on an unrealistic flap schedule (strictly adhering to the SOPs in Figures 3 and 4) and did therefore not correspond to the results of the Monte Carlo simulation. This was the case for the higher aircraft weight.

6 | Conclusions and Recommendations

A flight mechanical assessment tool based on a Monte Carlo simulation was presented which predicts, given the aircraft, the SOPs, the wind conditions and the approach trajectory, the percentage of flights that will not meet the constraints at the waypoints and will not achieve a stabilized approach. These two factors affect, amongst others, pilot TDL during approach [4]. The flight mechanical assessment tool is intended to be used during the design of approaches.

The predictions of the Monte Carlo simulation seem plausible. The predictions were analyzed by visually inspecting the plots with results and by regression analyses. For the approach considered in the case study, the dominating factors influencing the fact whether the constraints at waypoints could be met and whether a stabilized approach could be achieved were the wind speed, wind direction and aircraft weight. The moments in time at which the pilots performed all their actions (such as selecting gear or flaps) appeared to have only a very small influence or no influence at all. This could, however, be due to the assumptions that were made regarding these moments in time where the actions are performed; it can be that these were chosen too optimistically.

The predictions of the Monte Carlo simulation, as well as the modeling of the pilots’ actions within the Monte Carlo simulation have to be validated by flight simulator tests. This is done in the accompanying paper [4]. The paper will show that reliable predictions by the Monte Carlo simulation can indeed be obtained by modeling the pilots’ actions according to SOPs and applying a distribution in time for all these actions. This conclusion is based on the comparison of the predictions of the Monte Carlo simulation to the data from the flight simulator tests. However, the accompanying paper [4] will also demonstrate that better results can be obtained when some of the trigger events as assumed in this paper are replaced by different trigger events. For example: in the current paper the selection of flaps 1 was based on the trigger event “reaching 20nm from field”, whereas the results of the flight simulator data showed that reaching a certain Indicated Airspeed was a better trigger event. Additionally, the flight simulator test results in the accompanying paper [4] will demonstrate that the distribution of the reaction times for most of the pilots’ actions were more widely spread than assumed in the current paper.

The PMM presented in this paper will need to be refined, for, as it is, its predictions are not accurate enough to be used confidently during the design of approaches. The goal is to use the Monte Carlo simulation and flight simulator tests to identify the factors that have an influence on whether the constraints can be met and a stabilized approach can be achieved, and to incorporate these factors in detail in the PMM. This refined PMM is expected to yield more reliable results.

Acknowledgements

This research was financially supported by the Dutch Technology Foundation STW (DLR7071).

Literature

1. Flight Safety Foundation Approach-and-landing Accident Reduction (ALAR) Task Force. Killers in Aviation: FSF Task Force Presents Facts about Approach-and-landing and Controlled-flight-into-terrain Accidents, *Flight Safety Digest*, November 1998 - February 1999
2. Boeing Commercial Airplane Group. *Statistical Summary of Commercial Jet Airplane Accidents, Worldwide Operations, 1959-2005*, Seattle, Washington, United States, May 2006.
3. EUROCONTROL. *Navigation Strategy for ECAC*, European Organisation for the Safety of Air Navigation, European Air Traffic Control Harmonisation and Integration Programme, Edition 2.1, 15 March 1999.
4. Heiligers, M.M., Holten, Th. Van, Mulder, M., Factors that influence pilot task demand load during RNAV approaches, recommended for publication in Journal of Aircraft (Journal Article 3 in this dissertation).
5. Stassen, HG, Johannsen, G, & Moray, N. Internal Representation, Internal Model, Human Performance Model and Mental Workload. Automatica, 26(4), 811-820, 1990.
6. Godley, Stuart T. Perceived Pilot Workload and Perceived Safety of RNAV (GNSS) Approaches, Australian Transport Safety Bureau, ATSB Transport Safety Investigation Report, Aviation Safety Research and Analysis Report 20050342, 2006.
7. Heiligers, M.M., Holten, Th. Van, Mulder, M, Predicting Pilot Task Demand Load During Final Approach, *The International Journal of Aviation Psychology,* 19(4), 391-416, October-December 2009.
8. Baron, S., Zacharias, G., Muralidharan, R., Lancraft, R. (1980), PROCRU: A model for analyzing flight crew procedures in approach to landing. In NASA-MIT 16th Annual Manual 1980, pp 488-520.
9. Corker, K.M. (1999), Human Performance Simulation in the Analysis of Advanced Air Traffic Management, In Proceedings of the 1999 Winter Simulation Conference, Farrington, P.A., Nembhard, H.B., Sturrock, D.T., Evans, G.W. eds.
10. Gore, B.F., Corker, K.M. (2000) A Systems Engineering Approach to Behavioral Predictions of an Advanced Air Traffic Management Concept, 19th Digital Avionics Systems Conference, Entering the Second Generation of Powered Flight, Philadelphia, Pennsylvania.
11. Smith, B.R., Tyler, S.W. (1997), The Design and Application of MIDAS: A Constructive Simulation for Human-System Analysis, 2nd Simulation Technology & Training Conference, Canberra, Australia
12. Baron, S., Kruser, D.S., Huey, M.B., Quantitative Modeling of Human Performance in Complex, Dynamic Systems,National Academy Press, Washington D.C., 1990.

13. Vicente, K.J., Cognitive Work Analysis, Towards Safe, Productive, and Healthy Computer-Based Word, Lawrence Erlbaum Associates, Publishers, Mahwah, New Jersey, 1999
14. Hanke, CR and Nordwall, DR. The simulation of a jumbo jet transport aircraft Volume II: Modeling Data, The Boeing Company, Wichita Division Wichita, Kansas, September 1970.
15. Etkin, B. Dynamics of Atmospheric Flight, John Wiley & Sons, 1972.

Journal Article 3

Predicting pilot task demand load during RNAV approaches: flight simulator tests

M.M. Heiligers, Th. van Holten, M. Mulder

Accepted by the Journal of Aircraft

Abstract

The goal of this research is to develop a method which can predict pilot task demand load (TDL) during approaches. This paper presents the results of a flight simulator experiment that aimed to give a first indication of the factors related to the approach trajectory that influence pilot TDL. Analysis of the flight simulator data showed that the following factors influence pilot TDL: the energy rate demand during the last parts of the approach, the Localizer Intercept Speed, the distance available on Localizer intercept heading, the line-up distance, the required altitude at the Final Approach Fix, the Localizer intercept angle, and the fact whether the approach is stabilized at 1,000 ft. The results of the flight simulator tests are also used to validate the predictions of a Monte Carlo simulation. The Monte Carlo simulation predicts whether it is possible to meet the constraints at the waypoints and whether it is possible to achieve a stabilized approach, which are two of the factors that were shown to influence pilot TDL. A significant correlation was found between these Monte Carlo predictions and the flight data from the flight simulator experiment.

1 | Introduction

3

This research aims to develop a method which predicts the task demand load (TDL) as experienced by the pilot while flying an approach. TDL is defined as the mental workload *imposed* by the system to be controlled or supervised on the human operator [1]. First, this will yield insight in which aspects of an approach actually influence pilot TDL. And second, during the design of approaches this method can be used to rapidly evaluate a potential approach and to 'optimize' an approach with respect to pilot TDL.

The method should predict pilot TDL when flying a published approach according to Standard Operating Procedures (SOPs) when using Autopilots, Autothrottle and the Flight Management System (FMS) and while aiming to achieve a stabilized approach (the criteria for a stabilized approach will be explained in section 4.2). The goal is to link pilot TDL first of all to the properties of the approach trajectory (for example, the number of waypoints, the Localizer intercept speed, etc.), next to this, the effects of other factors such as wind conditions and aircraft weight are considered. For a definition of a stabilized approach, as well as the basic principles of the method, the assumptions and the choices that have been made as to what is, and what is not incorporated in the scope of the research see [2] and [3].

As explained in [2] and [3] the approach we have chosen to predict pilot TDL deliberately deviates from the idea behind models such as the Procedure-Oriented Crew model (PROCRU) [4, 5] or the Man-Machine Integrated Design and Analysis System (MIDAS) [6-8] that use human operator models. The approach we have chosen is based on the principles of cognitive work analysis [9]. The main characteristic of cognitive work analysis is that it shifts the emphasis from investigating the constraints of the human operator (like memory capacity, time delay, etc.) to analyzing and describing the operator's environment (like the trajectory, the aircraft dynamics, the wind conditions, etc.). The reason for this choice is that the constraints in the environment actually 'shape' the behavior of the human working in that environment. By choosing the approach to focus on the operator's environment instead of the human operator, the model will be more accessible and easier to use for approach designers, who often have little background in human operator modeling (or none)

The research is split up into five steps: the first step is to determine whether, for a specified approach, it is possible to meet the constraints at the waypoints and to achieve a stabilized approach, since it will be shown in this paper that, among others, pilot TDL during approach is influenced by these two factors. This first step was the topic of the accompanying paper [2] which, to this end, described a Monte Carlo simulation based on an aircraft model for the B747-100 and an initial pilot model. This pilot model consisted of a simple control model for the flight director task, and a model for the pilots' actions such as selecting flaps and gear which were modeled using trigger events (for example reaching 1,200ft) and distributions for reaction times (for instance 2 seconds after reaching 1,200ft flaps 25 are selected), see Figure 1.

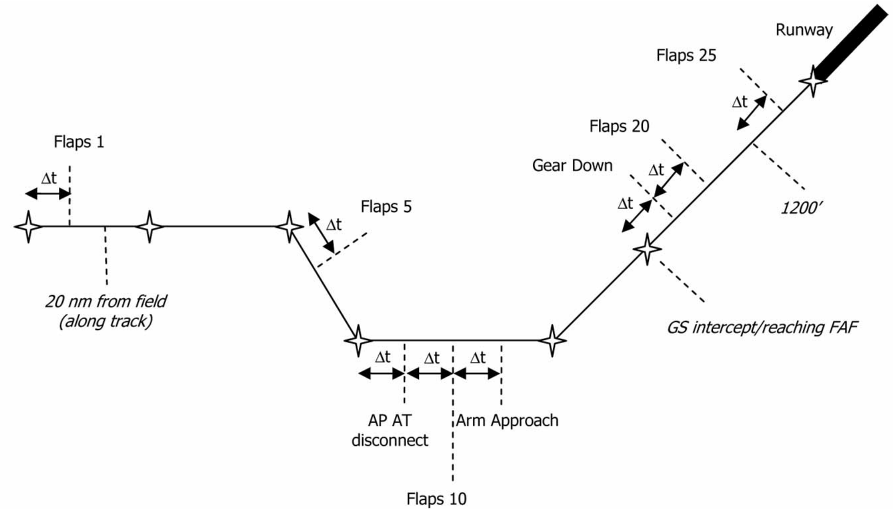


Figure 1 | Visualization of trigger events and reaction times (Δt) for pilot actions in the Monte Carlo computer simulation.

Based on assumed reaction time distributions, the Monte Carlo simulation [2] predicted for a given approach the percentage of flights that can meet the altitude and airspeed constraints at the waypoints and the percentage of flights that can achieve a stabilized approach as a function of wind direction, wind speed and aircraft mass.

The second step is to validate the predictions of the B747-100 Monte Carlo simulation for a given approach, and to validate the assumptions that were made for the Monte Carlo simulation with respect to the modeling of all pilot actions. This is done by a first set of flight simulator experiments for the B747-100. These flight simulator experiments will also give a first indication of the other factors of an approach that influence pilot TDL. This second step is the topic of this paper.

In the third step a second flight simulator experiment for the B747-100 will be performed to obtain a more detailed understanding of the factors that influence pilot TDL during approach. To check the general applicability of the factors found to influence pilot TDL, a Monte Carlo simulation, a flight simulator experiment and real flight tests are performed for a different aircraft (a Cessna Citation) in the fourth step. The fifth and final step is to gather all relevant data, and all factors that influence pilot TDL, and to incorporate these in an easy-to-use tool, that can be used during the design of approaches.

This paper starts with a review of literature (section 2), based on this review the factors assumed to influence pilot TDL that are considered in this paper are presented in section 3. Section 4 presents a human-in-the-loop flight simulator experiment, the results of which (both relating to the validation of the Monte Carlo simulation as well as relating to the factors assumed to influence pilot TDL) are given in section 5. A discussion of the results can be found in section 6. In section 7, a preliminary regression model is proposed which gives an indication of pilot TDL based on the numerical values of the factors that proved to influence pilot TDL. Section 8 then provides a comparison between the human-in-the-loop experiment and the predictions of the Monte Carlo computer simulation. Finally, The conclusions are presented in section 9.

2 | Results of literature survey

A review of literature revealed that there are three documents (or researches) that provide background for this research with respect to predicting pilot TDL during RNAV approaches. The first document describes international standards for RNAV approaches, with guidelines for an optimum RNAV approach. The two other researches consider pilot TDL during (RNAV) approaches. Each is described briefly below.

The international standards for an RNAV approach using basic Global Navigation Satellite System (GNSS) receivers are defined in ICAO document 8168 Volume 2 (PANS-OPS) [10]. These standards are given per approach segment:

- Initial approach segment – When used, the central initial approach segment has no maximum length [10]. The optimum length is 5.0 NM [10]. The initial approach is the segment from the Initial Approach Fix (IAF) until the Intermediate Fix (IF), the IF is the first waypoint on runway heading, see also Figure 2.
- Intermediate segment – The length is variable but will not be less than 2.0 nm allowing the aircraft to be stabilized prior to overflying the Final Approach Fix (FAF) [10]. The intermediate approach is the segment from IF until the FAF. If the approach is flown horizontally, the FAF is the location in the approach where the glideslope is captured.
- Final approach segment – The optimum length is 5.0 nm but it should normally not exceed 10 nm [10]. The final approach is the segment from the FAF until the Runway Threshold.
- The profile descent path should have an angle no greater than 3.7 degrees, with an optimum descent angle of 3 degrees [10].

It is not stated whether these optimum lengths and optimum descent angle are also 'optimal' with respect to pilot TDL.

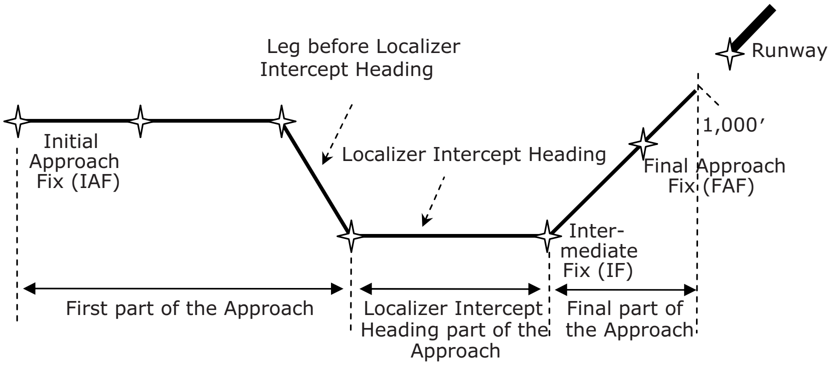


Figure 2 | Definitions relating to the approach trajectory used in this research.

The Australian Transport Safety Board has performed an extensive survey of pilots to gain an understanding of pilot perceptions of RNAV approaches [11]. This survey was held among pilots with a Civil Aviation Safety Authority (Australian) pilot's license, not all respondents were flying VNAV equipped aircraft. The findings of the Australian survey for the category of airline pilots are reported in Table 1. It should be noted that only factors that relate to the nominal trajectory are mentioned, factors concerning, for example, the naming convention of waypoints, improvement of approach charts, etcetera although reported in [11] are not mentioned here. Table 1 only reports the most frequently mentioned answers.

Table 1 | Findings of the Australian survey of pilots to gain an understanding of pilot perceptions of RNAV approaches [11]

Aspects of an RNAV approach that contribute to mental workload, physical workload or time pressure
Programming Flight Management Computer (FMC), setting up approach
Varying/irregular segment lengths/Many (close) steps
Descent and Position monitoring/situational awareness
Aircraft configuration late increases/Aircraft configuration early reduces workload
GPS/Flight Management System (FMS) manipulation
Late decision/clearance to fly RNAV approach
Briefing
Early preparation reduces workload/time pressure
Aspect of an RNAV approach that can be improved
FAF (and steps after FAF) removed from design (10 nm last segment)
Reduction in number of waypoints / steps or Removal of short steps
Standard (PANS-OPS) distances between all waypoints / Standard Missed Approach Point (MAPt) position
Reduced GPS/FMS inputs
3 degree slope only
Runway alignment on all approaches
Overlayed approaches / waypoints matching ground based aids
Most difficult circumstances to conduct an RNAV approach
Weather conditions poor
Turbulent conditions
Night
Instrument Meteorological Conditions (IMC)
Terrain is significant
Speed too fast (rushed or tailwind)
Short sectors (limited preparation time)
Multiple (short) limiting steps / complex approach design
Approach not runway aligned
Short notice from Air Traffic Control (ATC) or limited preparation time
Traffic

Table 2 | Metrics evaluated for estimating pilot TDL in [12], an asterisk indicates that this factor influences pilot TDL according to [12], metrics without an asterisk do not influence pilot TDL according to [12].

Nr	Effect on TDL	Metric
1		Number of vertical path changes
2		Maneuver time for vertical path changes
3	*	Cumulative size of vertical path changes
4	*	Maximum vertical acceleration for vertical path changes
5		Number of speed changes
6	*	Maneuver time for speed changes
7		Maximum size of change for speed changes
8	*	Maximum acceleration for speed changes
9		Number of combined vertical path and speed changes
10		Maneuver time for path-speed changes
11	*	Cumulative size of path changes for path-speed changes
12		Maximum size of speed change for path-speed changes
13	*	Maximum acceleration for path-speed changes
14	*	Maximum vertical acceleration for path-speed changes
15		Number of track changes
16		Lateral maneuver time
17		Cumulative size of track changes
18		Maximum rate of turn
19	*	Maximum energy rate demand
20		Total number of maneuvers
21	*	Longitudinal maneuver time
22		Total maneuver time

Vormer [12] studied pilot TDL as a function of different approach trajectories and different 4-D Guidance Displays. Twenty-two metrics, all related to the approach trajectory, were evaluated for estimating TDL during flight simulator tests. The metrics are given in Table 1. During the flight simulator tests pilots were required to fly manually, using 4-D guidance displays and were instructed to minimize the deviation from the reference flight path and reference speed. Based on TLX workload ratings and control activity, 9 metrics of the total of 22 were found to influence pilot TDL (see Table 2). The energy rate demand in Table 2 is the ratio between the energy rate commanded by the trajectory and the maximum energy rate that can be achieved by the aircraft. If the value for the energy rate demand becomes larger than 1 this implies, in the case of approaches, that the decrease in energy required by the trajectory can not be met by the energy decrease of the aircraft, and as a consequence the altitude and/or airspeed constraints at the next waypoint can not be met.

3 | TDL factors considered in this paper

Within the present research we specifically focus on pilot TDL during RNAV approaches, during which the part of the approach until Localizer Intercept Heading is flown using the autothrottle and autopilot in Vertical Navigation (VNAV) and Lateral Navigation (LNAV) modes, and at Localizer Intercept Heading the pilot switches to Flight Director and disconnects the autothrottle (see Section 2 of [2]). The last part of the approach is flown using the Instrument Landing System (ILS). In this respect, the results of the studies described above [11][12] can not be directly applied to this research due to the different types of automation used: the study of Vormer [12] considered manual control throughout the flight, and the survey described in [11] included aircraft that were not VNAV equipped. These different levels of automation might result in different factors influencing pilot TDL. However, the two studies considered do provide a good basis to start from. Therefore we have chosen to concentrate on the following nine independent variables that might influence pilot TDL for the experiment described in this paper:

- The number of heading changes. Although Vormer [12] found that lateral maneuvers did not appear to increase workload, the number of heading changes might contribute to the complexity of approaches which was found to be one of the most difficult circumstances for RNAV approaches in [11].
- Incorporating many altitude steps in an approach compared to a Continuous Descent Approach (CDA). Many altitude steps will increase the cumulative size of vertical path changes, the longitudinal maneuver time and the maximum vertical acceleration for vertical path changes, which were all found to increase pilot TDL [12].
- The maximum energy rate demand. By applying a strong tailwind the maximum energy rate demand is increased, this would increase pilot TDL [11] and is mentioned as one of the most difficult circumstances to conduct an RNAV approach [12].
- Applying a horizontal approach instead of a CDA. According to [12] this will increase pilot TDL by increasing the cumulative size of vertical path changes, the longitudinal maneuver time and the maximum vertical acceleration for vertical path changes. On the other hand, pilots might be more accustomed to flying a horizontal approach and intercepting the glideslope from below, which might decrease the pilot TDL compared to flying a CDA.

These four factors were based on the findings in literature [11, 12]. Next to these factors, other factors that might contribute to pilot TDL during approaches are considered as well in this paper. These other factors are selected based on conversations with pilots and by examining the SOPs (see [2], paragraph 4.3). The additional factors used as independent variables are:

- The distance available on Localizer Intercept Heading, since according to the SOPs many actions need to be performed on Localizer Intercept Heading decreasing the distance (and thus time) available is assumed to affect pilot TDL;
- The Localizer intercept speed, this is equal to the airspeed at the IF;
- The aircraft mass;
- The line-up distance, this is the distance between IF and runway, and
- The heading change when turning towards Localizer Intercept Heading, in Figure 2 this is the heading change when turning from the "Leg before Localizer Intercept Heading" to "Localizer Intercept Heading".

To test the effect of the aforementioned factors on pilot TDL during approach, twenty different approaches have been designed. These were flown in a flight simulator by nine pilots. All approaches were also simulated by the Monte Carlo simulation [2].

4 | Human-in-the-loop experiment

A human-in-the-loop experiment was performed which involved pilots flying different approaches under varying conditions in a six-degree-of-freedom flight simulator.

4.1 | Experiment goal

The experiment was designed to test the influence of the nine independent variables (as explained in the previous section) on pilot TDL during an approach. Next to testing the influence of these independent variables, additional goals of the experiment were:

1. to validate the Monte Carlo simulation (presented in the accompanying paper [2]) with respect to the simulation of the pilot's actions;
2. to validate the Monte Carlo simulation with respect to the prediction whether or not the constraints at the waypoints are met;
3. to validate the Monte Carlo simulation with respect to the prediction whether or not the approach is stabilized at 1,000';
4. to get an indication of which other factors besides the independent variables have an effect on pilot TDL; and
5. to test whether pilots are aware of the existence of an 'optimal' RNAV approach according to the 'standard' or 'optimal' PANS-OPS distances as described in Section 2. Since designing approaches according to the optimal approach criteria in [10] is one of the most common improvements for RNAV approaches mentioned in [11].

4.2 | Method

4.2.1 | Apparatus and B747 model

The experiment was performed in Delft University of Technology's six-degree-of-freedom SIMONA Research simulator (SRS) with out-the-window view, see Figure 3. The Boeing 747 aerodynamic models as well as the autopilots, flight director, autothrottle and yaw damper are based on the B747 model with JT9D-3 engines given in [13], and are identical to the models used in the Monte Carlo simulation [2]. Autopilot modes available during the experiment were: VNAV, LNAV, Heading Hold, Heading Select, Altitude Hold and Vertical Speed. In Flight Director operation additional modes available were Glide slope mode and Localizer mode. In VNAV mode a 'drag required' message appeared on the Navigation Display when altitude and speed constraints at the next waypoint were calculated not to be met (energy rate demand too high) with the current aircraft configuration. The approaches were flown in Visual Flight Rules conditions. During the experiment there was no other traffic, and no emergencies (for instance, engine fire) occurred.

All approaches were pre-programmed in the Flight Management System (FMS)/Control Display Unit (CDU). The appropriate approach was loaded in the FMS before the start of the approach, and during the experiment pilots could switch between the 'Progress' and 'Legs' pages, but could not use the CDU interactively, or modify the approach.

There were some discrepancies between the SRS and a B747 that are of importance for the experiment. First, the cockpit lay-out in the SRS differed from reality (see Figure 3). Second, the altitude setting on the Mode Control Panel (MCP), did not overrule the VNAV mode. If the required altitude at the next waypoint was 5,000', and the altitude set on the MCP was 6,000', the aircraft would continue to descend to 5,000' instead of leveling off at 6,000'. Third, the autothrottle did not consider the Flap Speed Marks. For example: the aircraft was flying 230 knots with flaps UP and the 'UP mark' (the Flap Speed Mark for flaps UP) was at 220 knots. The required airspeed at the next waypoint was 200 knots. In reality the autothrottle would not decelerate to 200 knots, but keep the airspeed at the Flap Speed Mark (in this case at 220 knots) until flaps 1 would be selected. This feature was not incorporated in the SRS, and as a result the aircraft would decelerate to 200 knots, regardless of flap setting or Flap Speed Marks. Fourth, the power levers did not move during autothrottle operation, and as a result, the power lever position at autothrottle disconnect was not necessarily the correct setting

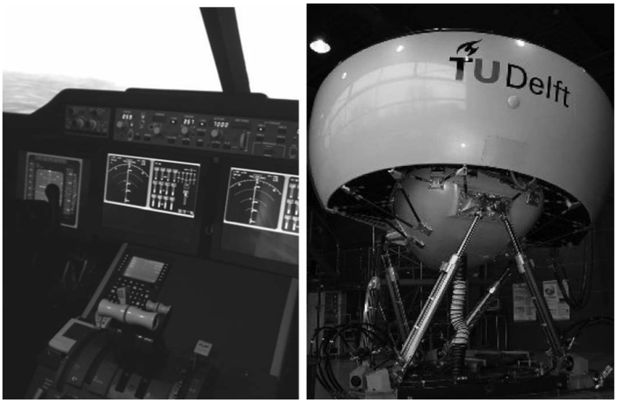


Figure 3 | The SIMONA Research Simulator

but needed adjustment by the pilot. These differences between the SRS and reality were the same for all pilots and all approaches.

4.2.2 | Subjects and instructions

Nine B747 pilots participated in the experiment as Pilot Flying (PF), total flight hours ranging from 360 to 18,500 hours (M = 11,793 hours, s = 6,708 hours). Their flight hours on the B747 ranged from 200 to 8,000 hours (M = 4,228 hours, s = 2,800 hours). Two students of the Faculty of Aerospace Engineering were instructed and trained to act as Pilot Monitoring (PM) during the experiment. Each crew consisted of one B747 pilot, and one student. The task of the crew was to fly 18 approaches, starting at the Initial Approach Fix (IAF) and ending at 800’ above airport level (AAL). Some crews flew two additional approaches.

Two weeks before the experiment the pilots (PF) received a briefing by mail. On the day of the experiment they were briefed as well. The pilots were asked to adhere very strictly to SOPs as explained in Section 4 of the accompanying paper[1], even if they could foresee that by adhering to SOPs they would not meet certain constraints at waypoints or would end up unstabilized at 1,000’. The pilots were briefed that the only situation in which they were allowed to deviate from the flap settings according to SOPs, was when the airspeed required by the approach was lower than the flap setting according to SOPs could accommodate, in that case they were allowed to select the next flap setting. Additionally, they were asked to perform their tasks according to the principles of Multiple Crew Coordination (MCC) and to fly passenger comfort. They were briefed about the discrepancies between the SRS and the B747 (as explained in the previous paragraph), and were informed that there would be no emergencies during the flight. They were told that they could fly the approach as published on the approach and landing charts, implying that ATC would not interfere.

4.2.3 | Procedure

Before starting the experiment the pilots could familiarize themselves with the SRS and their task during three to five (depending on the pilot) practice approaches. After that the

1 Pilots were briefed that they were free to select flaps 1 and flaps 5 at a location in the approach they considered appropriate. They were told that they should select flaps 10, Arm the approach, switch to Heading select mode and disconnect the autothrottle and autopilot at Localizer Intercept Heading. Pilots were additionally briefed that they should select flaps 20 and gear down at glideslope intercept, or in the case of a Continous Descent Approach that they should do so at the FAF, and that they should select flaps LAND (flaps 25) at 1200’ altitude.

experiment started. Before every approach pilots could take as much time as they thought necessary to study the approach and landing charts, to brief the approach and to prepare the SRS for the next approach. The simulation was started when pilots indicated that they were ready.

After every approach the PF was asked to fill in a feedback form. Each feedback form consisted of three parts: the first part required a rating of the approach on the Rating Scale Mental Effort (RSME)[14], the second part was an open format question asking pilots to indicate which factors made this specific approach difficult or easy to fly, and the third part contained two questions asking the pilot's opinion on whether or not the approach was stabilized at 1,000', and whether the pilot would have adhered to SOPs during real flight.

As stated, in the first part of the feedback form the pilots were asked to rate each approach on an RSME scale. They were free to give one RSME rating for the entire approach or to divide the approach in multiple parts, giving each part a separate RSME rating. The RSME is constructed according to the 'magnitude estimation' method [15] and the Dutch version of the scale (which was also used for this research) was used and validated in [16, 17]. It is used here because of its simplicity and ease of use when compared to, for example, a NASA TLX rating procedure [18].

At the end of the day, after all approaches were flown, the pilot PF filled in an end-of-day questionnaire. The first part of the end of day questionnaire regarded the realism of the SRS and the realism of the experiment as a whole. The second part contained general questions about factors that might possibly influence pilot TDL during approach.

Table 3 | Independent variables and hypothesized TDL effect.

Approach pair	Independent variable	Linked factor	Approaches	TDL effect
A	1. Number of heading changes	·	1 & 2	=
B	2. CDA compared to horizontal	·	3 & 4	+
C	3. Heading change towards Localizer Intercept Heading	·	9 & 10	=
D	4. Energy rate demand too high	Localizer groundspeed	17 & 18	+
E	5. Localizer intercept speed (IAS)/ Localizer ground speed	Energy rate demand IF-FAF	12, 13 & 14	+
F	6. Mass	·	20 & 21	+
G	7. More altitude steps compared to CDA	·	2 & 16	=
H	8. Distance available on Localizer Intercept Heading	Localizer ground-speed	3 & 9	+
I	9. Line-up distance (distance between IF and RWY)	IF-FAF distance and Energy rate demand IF-FAF	3 & 6	+

4.2.4 | Independent variables and approaches

The approaches (cases) were designed in pairs to test the independent variables, see Table 3. As explained before, the selection of the independent variables was partly based on [11, 12]. All approaches were designed for Amsterdam Airport Schiphol Runway 06. Within an approach pair the independent variable was the only changing variable, except when the independent variable was inevitably linked to another factor. Between pairs many variables were changed to explore whether pilots would comment on these variables in the feedback forms, for the same reason approaches 5, 7, 8 and 11 were added to the experiment as additional approaches.. The independent variables and corresponding approach pairs are given in Table 3, and are explained in Appendix A[2]. Variables that were changed between approach pairs were: the required altitude at the FAF (from now on referred to as FAF altitude), the IAS at the FAF, the distance between IAF and IF, the distance between IF and FAF, the possibility to be stabilized at 1,000′ and the energy rate demand. As a final note it is mentioned that approach 6 was designed according to guidelines of the 'optimal' RNAV approach as described in [10].

4.2.5 | Dependent measures

For every pilot the RSME ratings for all approaches given on the feedback forms were transformed into z-scores. The reason is the following: one pilot might rate all approaches during the experiment high on the RSME scale, whereas another pilot might rate all approaches low on the RSME scale, the absolute values are thus far apart, and difficult to compare. However, we are only interested to find out whether both pilots rated approach 'x' lower than approach 'y', irrespective of the absolute values. Therefore, the RSME scores of one pilot, are converted to z-scores for this one pilot using all RSME scores given by this pilot. For each approach pair the RSME z-scores are compared to establish whether there was a significant difference between the RSME z-scores for both approaches.

In addition, as explained before, pilots were free to give one rating for the entire approach, or to divide the approach up in parts and give multiple RSME ratings, one for each self-assigned part. If pilot A gave one RSME rating for the entire approach (RSME_A), and pilot B decided to divide the same approach in two parts (B1 and B2), resulting in two RSME ratings (RSME_B1 and RSME_B2, one for each part), then RSME_A was assumed to be valid both for part 1 and 2 and was counted as two separate (although similar) ratings, and treated as such for the calculation of the z-scores per pilot. It is noted that this is an assumption, and it cannot be checked whether pilot A would indeed have given the same RSME rating to both separate parts.

Further subjective data were obtained from the pilots' answers to the open format question why the approach was difficult or easy to fly, and the pilots' answers to the questions whether the approach was stabilized, and whether they would adhere to SOPs during real flight. Additionally, subjective data regarding the influence of specified factors on the difficulty of flying an approach were gained from the end of day questionnaire. All subjective data are compared, analyzed for inconsistencies and, whenever possible, compared to the objective flight data. These subjective data are used to determine which factors have an effect on pilot TDL.

Two factors that will be shown to influence pilot TDL are whether the approach is stabilized at 1,000ft and whether the constraints at the waypoints are met. In order to establish whether the approach was stabilized at 1,000ft the following criteria were used (based on [19]):

- Heading change and pitch change are within 5 deg/s;
- The IAS is not more than V_{REF} + 20 knots;
- Flaps 25 are selected, landing gear is down;
- Sink rate is not larger than 1,000 feet per minute; and
- Localizer and glide slope are within one dot;

2 This is Appendix A to Journal Article 3

An average value is calculated for each of the criteria: for the time slot starting 5 seconds before reaching 1,000 ft and ending at 1,000 ft. A larger time slot is used to calculate the average sink rate: it starts 1 minute before reaching 1,000 ft, and ends at 1,000 ft.

The constraints at the waypoints are considered to be met when the airspeed is within 10 knots of the required airspeed and the altitude is within 100ft of the required altitude.

In order to validate the Monte Carlo simulation (presented in the accompanying paper [2]) with respect to the simulation of the pilot's actions, the following pilot actions were logged during the experiment: selection of flaps, Autothrottle off, Autopilot off, ARM approach, Heading Select and Gear Down, as well as the use of the speedbrakes.

4.3 | Hypotheses

Regarding the influence of the independent variables the following factors were hypothesized to *not* influence pilot TDL (see also last column of Table 3 and the explanation in Section 3):

- ❖ The number of heading changes (independent variable 1)
- ❖ The heading change towards Localizer Intercept Heading (independent variable 3)
- ❖ Whether the approach is a stepped approach (approach with several altitude steps) or a CDA (independent variable 7)

Factors that were hypothesized to increase pilot TDL are:

- ❖ A CDA compared to a horizontal approach (independent variable 2)
- ❖ An energy rate demand that is too high (independent variable 4)
- ❖ A higher Localizer intercept speed (independent variable 5)
- ❖ A lower aircraft mass combined with the speed restrictions for this particular approach pair (pair F) (independent variable 6)
- ❖ A decrease of the distance available on Localizer Intercept Heading (independent variable 8)
- ❖ A shorter line-up distance (independent variable 9)

3

5 | Results

The experiment results consist of subjective data (collected using feedback forms after each approach and an end of day questionnaire) and of objective flight data, the results are presented below.

5.1 | Feedback forms

The feedback forms were filled in after each approach. The number of pilots that flew each approach is given in Table 4, and a total of 172 feedback forms were available for analysis (approach 15 could not be simulated in the SRS, approach 11 was flown by only 5 pilots due to time considerations, and approach 5 and 21 were flown by all pilots but only yielded usable results for 7 and 8 pilots respectively). As explained before, each feedback form consists of three parts (a rating of the approach on the RSME scale, an open format question asking pilots to indicate which factors made this specific approach difficult or easy to fly, and two questions asking the pilot's opinion on whether or not the approach was stabilized at 1,000', and whether the pilot would have adhered to SOPs during real flight) the results for each of these parts are discussed below.

Table 4 | Approaches and number of pilots that flew each approach (total N = 172)

Approach #	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**
Nr. of pilots	9	9	9	9	6	9	9	9	9	9	5	9	9	9	0	9	9	9	9	9	8

5.1.1 | RSME scores

As explained earlier, pilots were free to give one RSME rating for the entire approach or to divide the approach in multiple parts, giving each part a separate RSME rating. It is interesting to note that, when pilots decided to divide the approach in multiple parts, they almost always split the approach at the IF and/or at the first waypoint on Localizer Intercept Heading. A distinction can therefore be made between three approach parts (see also Figure 2): 1. the 'first part of the approach' which is the segment from the IAF until the first waypoint on Localizer Intercept Heading, 2. the 'Localizer Intercept Heading part' which is the segment from the first waypoint on Localizer Intercept Heading until the IF (this part includes Localizer capture), and 3. the 'final part of the approach' which is the segment from the IF until 1,000 feet.

The results of the average RSME z-scores for all approaches flown by 9 pilots during the experiment (but excluding approaches 5, 11, 15 and 21 since these were flown by less than 9 pilots) are given in Figure 4. Approach 21 was flown by only eight pilots, therefore, in order to be able to compare approach 20 and 21 (approach pair F) all RSME z-scores were also calculated for only eight pilots and all approaches except approach 5, 11 and 15, see Figure 4. Given the small sample size of average RSME z-scores for each approach it is noted that the following statistical analysis of these scores should be considered merely an indication of possible effects.

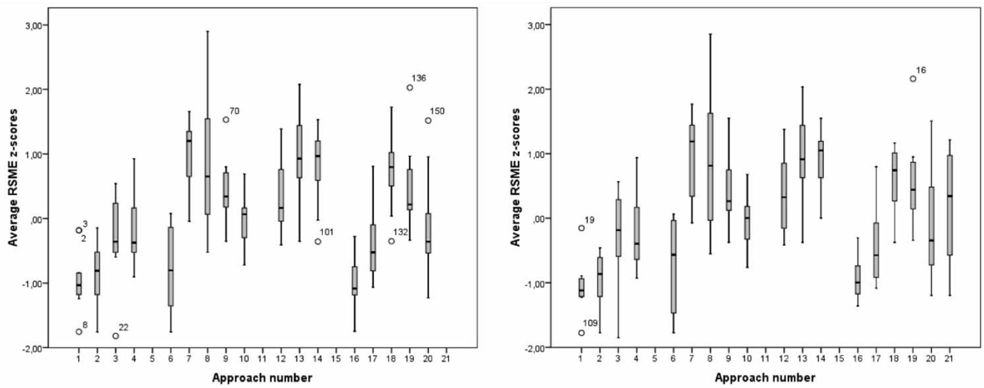


Figure 4 | Pilots' average RSME z-scores for all approaches (*N* = 9) except approaches 5, 11, 15 and 21 (left) and for all approaches (*N* = 8) except approaches 5, 11 and 15 (right).

The RSME z-scores for all approaches (see Figure 4) were tested for normality using Kolmogorov-Smirnov tests. Only the RSME z-score for approach 4, $D(9) = 0.29$, $p<0.05$ was significantly non-normal. Levene's test showed that the variances could be considered equal, $F(16,136)=1.18$, *ns*. The RSME z-scores for all independent variables were compared using t-tests for the normally distributed RSME z-scores, and Wilcoxon signed rank tests for the non-normally distributed RSME z-scores. The RSME z-scores for approach pair E were analyzed using a one-way repeated-measures ANOVA. The results are shown in Table 5 .

The pilots' average RSME z-scores indicate that the following independent variables have a significant effect on pilot TDL[3]:

❖ A *smaller* heading change towards Localizer Intercept Heading, a surprising result which will be further elaborated upon in paragraph 6.2 (independent variable 3).

3 For this Ph.D. thesis a more elaborate explanation of the RSME z-scores is added in Appendix B to Journal Article 3.

❖ An energy rate demand that is too high (independent variable 4)
❖ A smaller distance available on Localizer Intercept Heading (independent variable 8)
❖ A shorter line-up distance (independent variable 9)

The other independent variables did not cause a significant difference in the pilots' RSME z-scores and thus in pilot TDL.

Table 5 | Independent variables and results of comparison of RSME z-scores.

Approach pair	Independent variable	Statistic	Effect size	Significant effect?
A	1. Number of heading changes	$t(8) = -0.851$	$r = .29$	no
B	2. CDA compared to horizontal	T=15	$r = .10$	no
C	3. Heading change towards Localizer Intercept Heading	$t(8) = 2.57$*	$r = .67$	yes
D	4. Energy rate demand too high	$t(8) = -5.03$***	$r = .87$	yes
E	5. Localizer intercept speed (IAS)/Localizer ground speed	$F(2, 16) = 2.06$	$\omega^2 = .08$	no
F	6. Mass	$t(7) = -1.14$	$r = .4$	no
G	7. More altitude steps compared to CDA	$t(8) = -0.58$	$r = .2$	no
H	8. Distance available on Localizer Intercept Heading	$t(8) = -2.28$*	$r = .63$	yes
I	9. Line-up distance (distance between IF and RWY)	$t(8) = 1.89$*	$r = .55$	yes

*p <.05, *** p <.001

3

Additionally, it can be noted that the RSME scores for approach 6, which was designed according to the optimum design criteria for an RNAV approach in the PANS-OPS, are among the lower ratings on the RSME scale. Approach 6 is thus one of the approaches with lower TDL compared to the other approaches, but can not be regarded the approach with the lowest TDL.

5.1.2 | Factors influencing the difficulty of an approach

After each approach the pilots were asked to explain in writing why, in their opinion, the approach was difficult or easy to fly. The most frequently mentioned answers for all pilots and all approaches are summarized in Figure 5. It is important to note that this was an open format question, and none of the factors in Figure 5 were preprinted on the feedback form. Pilots could mention more than one factor for each approach, in total there were 238 comments relating to 24 factors. A factor was included in Figure 5 when it was mentioned by at least three different pilots, consequently, Figure 5 shows a total of 208 comments relating to 11 factors.

It is interesting to note that only few comments, 13 out of 208 total (6%) in Figure 5, are related to the first part of the approach (the factor 'Time available during first part of approach'). The factors 'Wind' and 'Energy rate demand' apply to the entire approach and comprise 71 comments (34%) of the total amount of 208. The majority of the factors (8 out of 11), and the majority of the comments (124 out of 208 which is 60%) relate uniquely to the last part of the approach and regard the phase from Localizer Intercept Heading to 1,000'.

It is interesting to note that only few comments, 13 out of 208 total (6%) in Figure 5, are related to the first part of the approach (the factor 'Time available during first part of approach'). The factors 'Wind' and 'Energy rate demand' apply to the entire approach and comprise 71 comments (34%) of the total amount of 208. The majority of the factors (8 out of 11), and the majority of the comments (124 out of 208 which is 60%) relate uniquely to the last part of the approach and regard the phase from Localizer Intercept Heading to 1,000'.

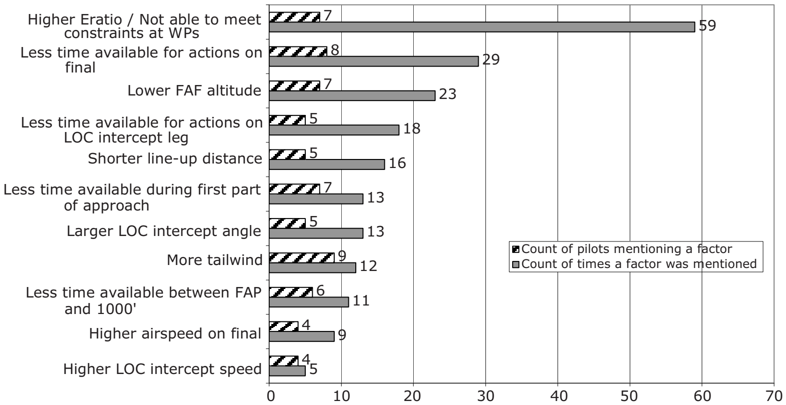


Figure 5 | Factors increasing pilot TDL during approach as mentioned by at least 3 pilots on the feedback forms.

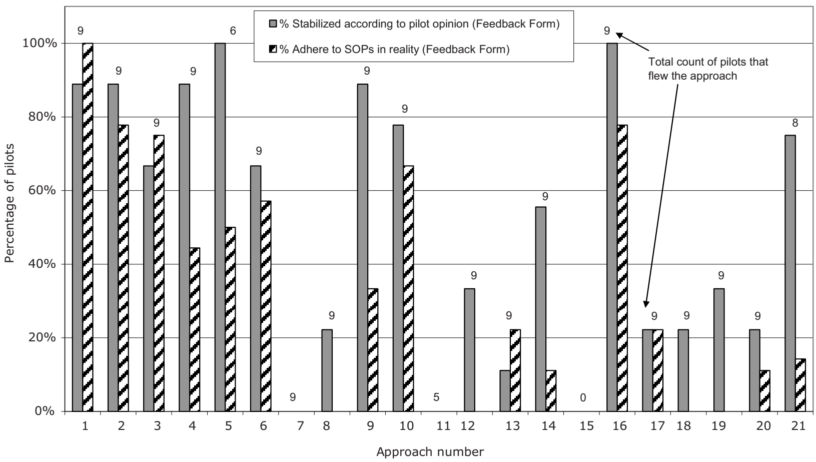


Figure 6 | Percentage of pilots indicating (on feedback forms) that the approach was stabilized at 1,000', and percentage of pilots indicating that they would fly the approach according to SOPs in reality.

5.1.3 |Stabilized at 1,000' and SOPs

To complete the feedback form the pilots were asked whether the approach (in their opinion) was stabilized at 1,000', response options "yes" or "no", and whether they would have flown this approach according to SOPs in reality (again, response options "yes" or "no"). Figure 6 shows the results. There was a positive relationship between the pilot indicating the approach was stabilized, and the same pilot indicating that he would adhere to SOPs in reality when flying that approach, τ = .45 (Kendall's tau), p(one-tailed) < .01.

5.2 | End of day questionnaire

5.2.1 | Reality of SRS and reality of experiment

Pilots were asked their opinion regarding the reality of the SRS by answering closed format questions with categorical response options for certain aspects of the simulator. For example: "In your opinion, how realistic was the autothrottle in SIMONA?", the corresponding response options were: "Very unrealistic", "Unrealistic", "Average", "Realistic", "Very realistic". Figure 7 shows that especially the Flight Director and the Autothrottle were regarded to be unrealistic by some of the pilots. When asked, two pilots (out of nine) answered that pilot TDL increased due to the fact that some of the aspects in Figure 7 were unrealistic.

All pilots indicated that ATC provided them with sufficient information regarding the approach, and that the communication between ATC and their flight was "Average" to "Very realistic". Nevertheless, two pilots (out of nine) reported that the TDL was influenced by the fact that the contact with ATC differed from reality.

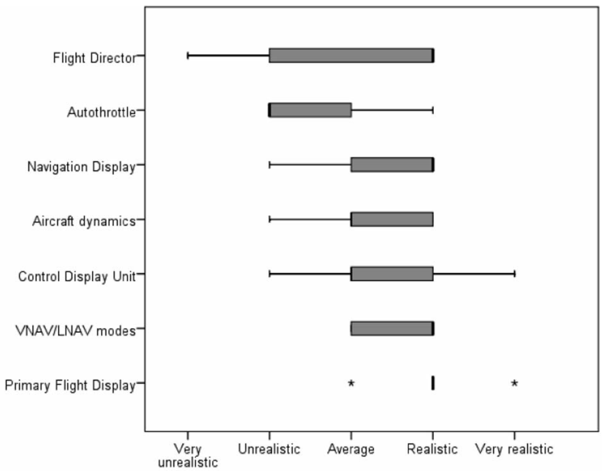


Figure 7 | Pilot answers (N=9) regarding the reality of the SRS.

The pilots answered that the communication between PF and PM was "Average" to "Realistic", and that the way the PM performed his tasks was "Average" to "Very realistic". However, one pilot indicated that pilot TDL increased due to the fact that the PM was not a 'real' pilot.

All pilots answered that they had sufficient time to prepare for the approach. Of all nine pilots, one pilot reported that the approach and landing charts provided "Insufficient" information, while eight pilots reported that the charts provided "Sufficient" or "More than sufficient" information. According to two pilots the TDL during the approach was influenced by these factors.

To conclude this part of the end of day questionnaire, the pilots were asked two more questions. The first question asked whether the fact that the cockpit layout of the SRS did not resemble the actual layout of a B747 influenced the difficulty of flying the approaches. One pilot replied "yes", all other pilots replied "no". The second question was whether the fact that the altitude and speed settings on the MCP did not overrule the VNAV mode had an influence on the difficulty of flying the approaches, three pilots replied "yes" six pilots replied "no".

It should be noted that although some of the aspects mentioned above deviated from reality and therefore might have had an influence on the difficulty of flying the approach, all these aspects were the same for all pilots and remained constant during all approaches.

5.2.2 | Factors influencing the difficulty of an approach

In the second part of the end of day questionnaire pilots were asked to indicate the factors that influence pilot TDL while flying an approach by answering closed format questions. An example of such a question is: "Considering an RNAV approach: when the Localizer intercept angle becomes larger, the approach becomes:" provided response options: "A lot easier", "Easier", "No influence", "More difficult", "A lot more difficult". The pilots' responses are summarized in Figure 8.

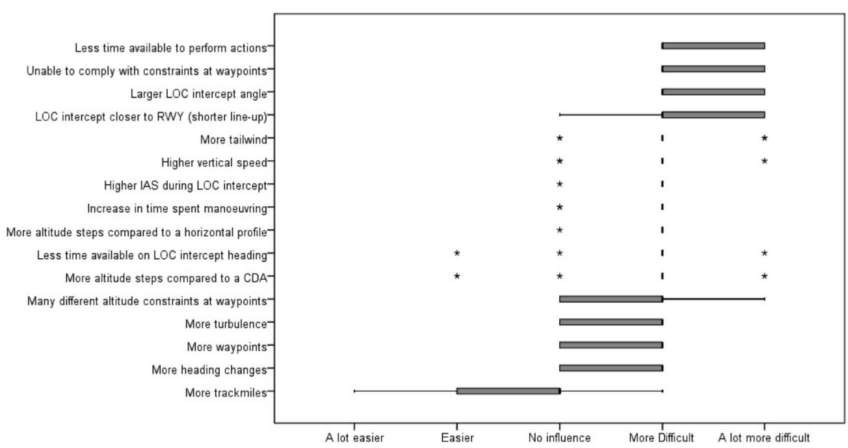


Figure 8 | Pilots' answers (N=9) regarding the influence of the factors mentioned on the left on pilot TDL while flying an approach.

The final two closed format questions regarding factors that influence an RNAV approach were formulated in a different way: "If you *do not* meet the altitude and speed constraints at the waypoints during an RNAV approach, but you *do* meet all the requirements for a stabilized approach at 1,000', would you classify this approach as 'difficult'?", response options: "yes" or "no". Of all pilots, seven pilots answered *yes* two replied *no*. The second question: "If you *do* meet the altitude and speed constraints at the waypoints during an RNAV approach, but you *do not* meet the requirements for a stabilized approach at 1,000', would you classify this approach as 'difficult'?" was answered *yes* by eight of the pilots, one pilot answered *no*.

Following the closed format questions, the pilots were asked whether there were any other factors that influence the difficulty of flying an RNAV approach (open format). The answers given were: mountainous terrain, high landing weight, experience on aircraft, FMS programming, and fatigue. All these factors were mentioned only once.

5.2.3 | Standard RNAV approach

In the third part of the end of day questionnaire pilots were asked whether to their knowledge a standard or optimal RNAV approach exists. Five out of nine pilots replied 'yes', four pilots replied 'no'. Further explanations about what this standard or optimal approach would look like mainly focused on the fact that an optimal approach does not have too strict speed and altitude constraints at waypoints, but rather has 'at or above' altitude constraints or 'at or below' speed constraints. None of the pilots mentioned the standard RNAV approach as described in the PANS-OPS [10]. This differs from the findings in the Australian survey. This might be due to the fact that pilots participating in the Australian survey were used to flying approaches according to the optimum design criteria in the PANS-OPS, since these design criteria are applied to many approaches in Australia (only 21.5% of the RNAV approaches published in the late 2006 varied from the optimum 5 nm configuration [11]), whereas the pilots participating in our experiment were all Dutch pilots, and were not used to flying approaches according to the optimum design criteria.

5.3 | Flight data

During the experiment flight data were recorded. These data were analyzed to determine how many flights were stabilized at 1,000', to determine how often the constraints were met at the waypoints, to calculate the amount of time speedbrakes were used, and to gain insight into all pilots' actions. The results are discussed below.

5.3.1 | Stabilized at 1,000'

Figure 9 shows for each approach the percentage of pilots that were stabilized at 1,000' according to the flight data. For comparison, the data from Figure 6 are included as well (representing the percentage of pilots indicating on the feedback forms that in their opinion the approach was stabilized) It should be noted that pilots did also take the criterion 'all checklists conducted' into account, whereas this was not incorporated in the criteria to analyze the flight data. This causes a difference between the results from the flight data and the pilots' answers.[4]

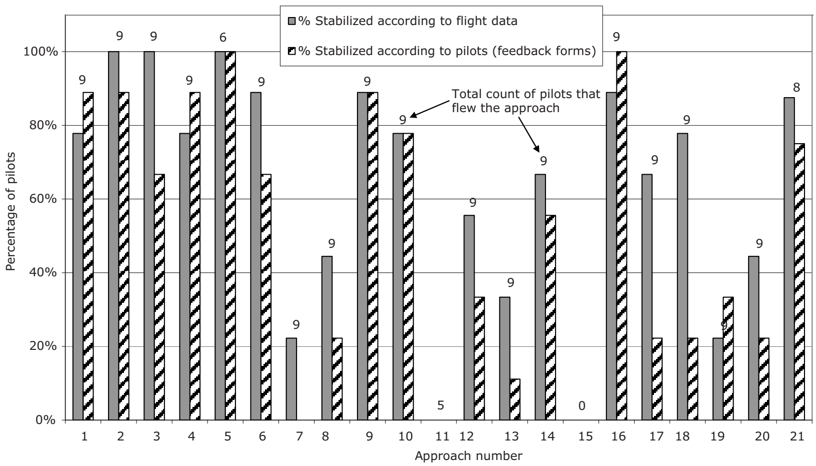


Figure 9 | Percentage of pilots that were stabilized at 1,000' according to the flight data.

4 This difference becomes larger when all other criteria for a stabilized approach (such as correct airspeed, correct flap setting etc.) are met, but the only reason for not being stabilized is the fact that there was not enough time to complete the checklist. This is the case for approaches with a very low FAF (because the final part of the checklist can only be performed after the FAF is reached), and for which the velocity constraints were such that the airspeed at 1,000ft could be lower than VREF+20. This was the case for, for instance, approaches 17 and 18, which indeed show a larger deviation between pilots' perception and flight data.

5.3.2 | Meet constraints at Waypoints

Figure 10 shows how many pilots did, and did not meet the constraints at each waypoint during the experiment, this is depicted for every approach. For each approach the first waypoint (waypoint '1') in this Figure is the IAF, the last waypoint (waypoint with the highest number) is the FAF. For example: approach 7 consisted of four waypoints, where waypoint 4 is the FAF. In total, nine pilots flew approach 7. At the first waypoint all nine pilots met the constraints, which is logical since this was the starting position of the simulation. At waypoint 2 none of the pilots met the constraints, at waypoint 3 only two pilots met the constraints, and at waypoint 4 only one pilot met the constraints.

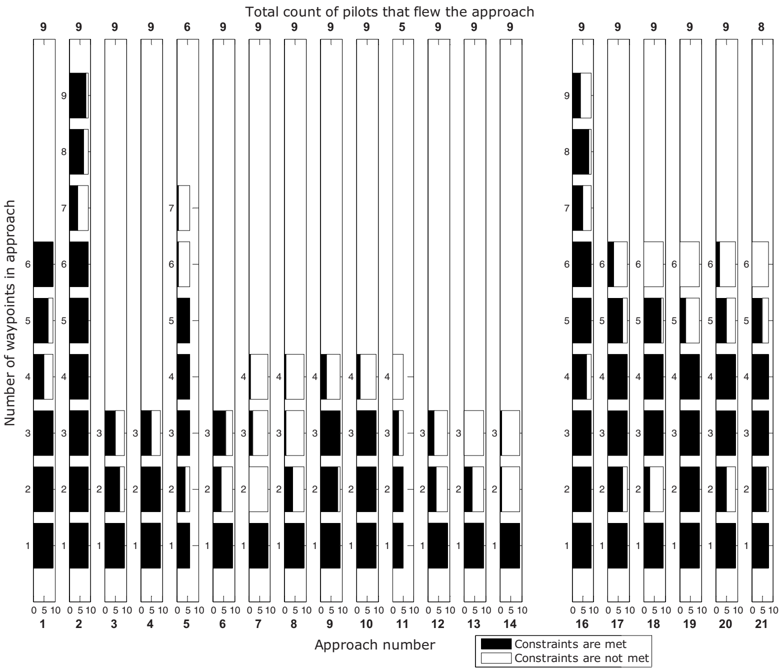


Figure 10 | Count of pilots that met the constraints at the waypoints and did not meet the constraints at the waypoint according to the flight data, for all approaches and all waypoints.

5.3.3 | Speedbrakes

Figure 11 shows the average percentage of time that pilots used the speedbrakes during the approach (percentage calculated with respect to the total flight time during the approach). It can be seen that during some approaches the pilots used the speedbrakes for a substantial amount of time.

5.3.4 | Pilot actions and reaction times

During the experiment all pilot actions (selecting flaps, gear, etc.) were logged. These data were analyzed to verify whether the modeling of the pilots' actions in the Monte Carlo computer simulation of [2] was realistic. In the Monte Carlo simulation the pilot actions were modeled using trigger events (for instance reaching 1,200') and reaction times (for instance 2 seconds after reaching 1,200' flaps 25 are selected). See Table 6 and the accompanying paper [2] for assumed trigger events and reaction times.

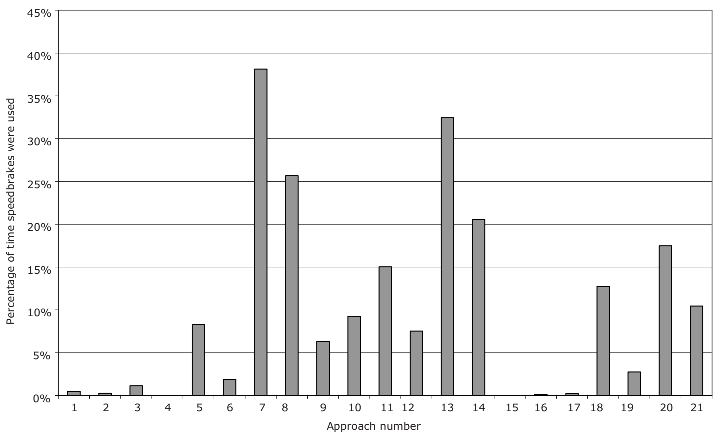


Figure 11 | Average percentage of time (of total approach time) that speedbrakes were used.

Analysis of the flight simulator results showed that the trigger events for the pilots' actions in the Monte Carlo simulation and Table 6 were chosen correctly, except those for the selection of flaps 1 and flaps 5. Experiment results showed that it was better to model the selection of flaps 1 and flaps 5 using the IAS, instead of a predefined location in the approach, see Figures 12 and 13. Better trigger events for selecting flaps 1 and flaps 5 were reaching the Flap Speed Mark for flaps UP (UP mark) at 223 knots (below this speed flaps 1 should be selected) and reaching the Flap Speed Mark for flaps 1 (1 mark) at 203 knots respectively (below this speed flaps 5 should be selected).

3

Table 6 | Trigger events and distributions on which reaction time are based for all pilot actions in Monte Carlo simulation.

Pilot action	**Trigger Event**	**Mean**	**Standard deviation**
Flaps 1	Reaching First waypoint of the leg on which aircraft is 20nm from field	T/3[(1)]	T/3[(1)]
Flaps 5	Reaching first waypoint of the leg before Localizer Intercept Heading	2×T/3[(2)]	T/3[(2)]
Autothrottle / Autopilot Off	Turn to Localizer Intercept Heading	T/2[(3)]	T/2[(3)]
Flaps 10	Autothrottle / Autopilot Off	2 sec	0.5 sec
ARM Approach	Selecting Flaps 10	4 sec	1 sec
Gear Down	Glideslope Intercept / Reaching FAF	2 sec	0.5 sec
Flaps 20	Gear Down	2 sec	0.5 sec
Flaps 25	Reaching 1,200'	2 sec	0.5 sec

[(1)] T = amount of time spent on leg on which the aircraft is 20nm from field
[(2)] T = amount of time spent on leg before Localizer Intercept Heading
[(3)] T = amount of time spent on Localizer Intercept Heading

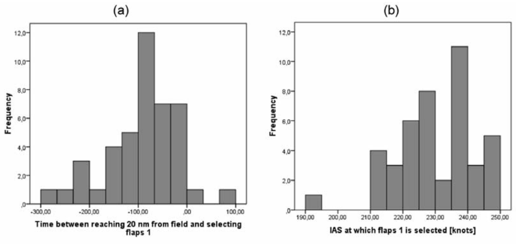


Figure 12 | Histograms for selecting flaps 1, approaches with an initial condition equal to flaps 1 or flaps 5 are not incorporated in the histogram.

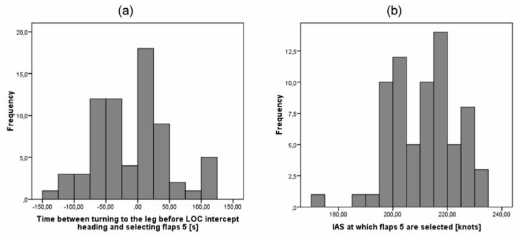


Figure 13 | Histograms for selecting flaps 5, approaches with an initial condition equal to flaps 5 are not incorporated in the histogram.

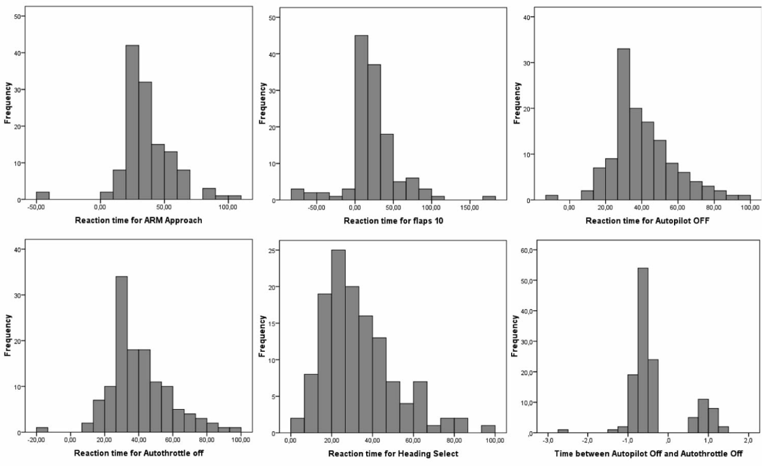


Figure 14 | Histograms for the actions on Localizer Intercept Heading.

Additionally, the trigger events for the actions on Localizer Intercept Heading (see Table 6) implied that these actions would always be performed in the fixed order given in Table 6. The flight data showed that this is not the case. A more realistic trigger event for each of these actions is "Turn to Localizer Intercept Heading", and each of the actions has its own distribution of reaction times with respect to this trigger event, see Figure 14.

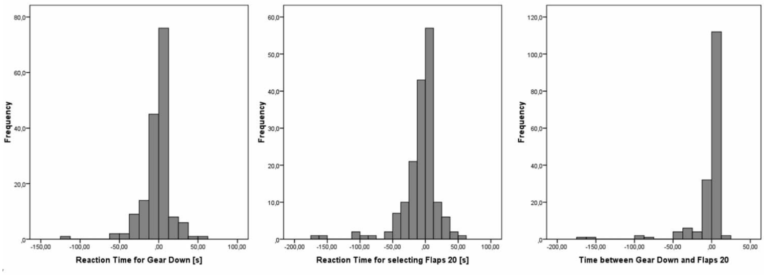


Figure 15 | Histograms for the actions at the FAF.

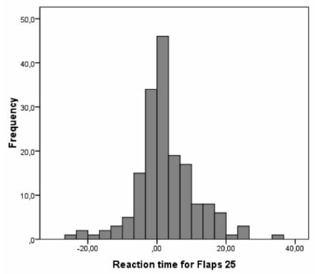


Figure 16 | Histogram for selecting Flaps 25 at 1,200'.

The reaction times for all pilot actions that were assumed in the Monte Carlo simulation were considerably shorter (see Table 6) than the reaction times observed during the experiment, see Figures 12 to 16. Additionally, the flight simulator results showed that the pilot sometimes performed the actions before the trigger event was reached, in the Monte Carlo simulation actions are always modeled to take place after the trigger event.

Regarding the actions performed on Localizer Intercept Heading it can be stated that there was no significant effect of the time available on Localizer Intercept Heading on the reaction times for the actions that were performed on Localizer Intercept Heading. The correlations between time available on Localizer Intercept Heading and all pilot actions are: τ = .008, p = .9 for Arm Approach, τ = .089, p = .155 for flaps 10, τ = .015, p = .811 for Autopilot off, τ = .011, p = .865 for Autothrottle off, and τ = .026, p = .374 for Heading Select. The flight data thus showed that irrespective of the time that was available, pilots tended to start performing these actions the moment they started to turn towards Localizer Intercept Heading, to 'get it over with' as soon as possible.

Additionally, a very high correlation existed between the reaction times for Autothrottle off (significantly non-normal, $D(127) = 0.099$, $p < .01$) and Autopilot off (significantly non-normal, $D(127) = 0.091$, $p < .05$), $r_s = .998$, $p < .001$. This correlation is due to the fact that both actions are asked for in a single call: 'Autothrottle off, Autopilot off', and therefore these two actions were, most of the time, performed very closely together in time, see Figure 14. The same is true for the actions Gear Down and flaps 20, which are performed at the FAF, see Figure 15.

Based on these findings, the trigger events and reaction times in Table 6 are adjusted as given in Table 7. The normal distributions of the reaction times per pilot action given in Table 7 are based on the histograms in Figures 11 to 15. One exception is made: for the time between selecting Gear Down and Flaps 20 the mean and standard deviation of the flight data do not represent the flight data at all when these are applied to a normal distribution. Therefore, based on the recorded reaction times, a different mean and standard deviation are chosen for the approximating normal distribution (M = 2 sec, s = 2 sec). The values in Table 7 will be used to validate the simulation of the pilots' actions in the Monte Carlo simulation.

Table 7 | Adjusted trigger events and reaction times based on flight data.

Pilot action	Trigger Event	Mean	Standard deviation
Flaps 1	IAS	230 knots	11.8 knots
Flaps 5	IAS	212 knots	11.8 knots
Flaps 10	Turn to Localizer Intercept Heading	22.9 sec	33.1 sec
Heading Select	Turn to Localizer Intercept Heading	33.2 sec	17.9 sec
ARM Approach	Turn to Localizer Intercept Heading	35.9 sec	20.3 sec
Autopilot Off	Turn to Localizer Intercept Heading	40.7 sec	17.4 sec
Autothrottle Off	Autopilot Off	-0.35 sec	0.69 sec
Gear Down	Reaching FAF	-1.7 sec	15.9 sec
Flaps 20	Gear Down	2 sec	2 sec
Flaps 25	Reaching 1,200'	2.5 sec	8.6 sec

6 | Discussion

The experiment results that were derived from the feedback forms, end of day questionnaire and flight data are compared in this paragraph, to verify whether these results are consistent.

6.1 | Stabilized at 1,000'

There was a significant relationship (see Figure 9) between the percentage of approaches that were stabilized according to the objective flight data, and the percentage of approaches that were stabilized according to the subjective answers of the pilots on the feedback forms, $\tau = .58$, p(one-tailed) $< .01$ (two times pilots forgot to complete the relevant question in the feedback form, hence $N = 171$). This implies that the judgment of the pilots whether the approach was stabilized at 1,000' correlates well with the objective flight data. An important fact is that in most cases the pilots' judgment is more conservative than the flight data, i.e., although the approach was stabilized at 1,000' according to the flight data, pilots classified the approach as not stabilized. In this respect it should also be noted again that the criterion 'checklists conducted' for a stabilized approach was taken into account by the pilots, but was *not* taken into account when analyzing the flight data.

6.2 | Factors influencing pilot TDL during approach

To check the consistency between the pilots' RSME ratings, factors mentioned on the feedback forms and answers given in the end of day questionnaire, all results are summarized in Table 8. For example: the number of heading changes was an independent variable, and was also changed between approach pairs. The RSME ratings for the approach pair that considered the number of heading changes yielded no significant difference, and the pilots did not mention the number of heading changes on any of the feedback forms as a factor that contributed to the difficulty of flying an approach. However, in the end of day questionnaire, the pilots answered that an approach becomes more difficult when the number of heading changes increases.

The second column thus indicates whether the factor has been changed at all during the experiment (between approach pairs). This is important for example when considering the factor turbulence. The turbulence intensity was not varied during the experiment and was kept at a low setting. Therefore the pilots might not have been triggered to write down any comments regarding turbulence on the feedback forms. The fact that it is not mentioned on the feedback forms does, however, not imply that it does not have an influence on the difficulty.

The factors 'Distance available on Localizer Intercept Heading' and 'Time available for actions on Localizer Intercept Heading' are directly linked. They are mentioned separately to indicate that pilots always commented on 'time' instead of on 'distance' in the Feedback Forms. The factors 'Line-up distance' and 'Time available for actions on final' are also directly linked, for these factors pilots did comment on both factors, the same holds for the factors 'FAF altitude' and 'Time between FAF and 1,000ft'.

Looking at Table 8 there are two factors that require further explanation. The first factor is 'Heading change towards Localizer Intercept Heading'. It was hypothesized that this factor would not influence the difficulty of an approach. The RSME scores however, indicated that pilot TDL increases when the heading change towards Localizer Intercept Heading decreases. On the other hand, none of the pilots mentioned this factor on the feedback forms. The authors were somewhat puzzled by this outcome, especially since if there would be an effect it would be expected that pilot TDL would increase when the heading change *in*creases as well. Therefore, a couple of months after the experiment, four pilots were asked to fly these two approaches again, and were asked to explain which approach they found easier or more difficult to fly. All four pilots commented that there was no difference between the two approaches. This clearly illustrates the limited reliability of the RSME scores, and as stated before, the RSME scores should be considered only an indication of *possible* effects.

3

The second factor that needs further explanation is 'Localizer intercept speed/ground speed', which was linked to a higher Energy rate demand between IF and FAF. Analysis of the RSME scores showed no difference in average RSME scores with increasing Localizer intercept speed and increasing Energy rate demand between IF and FAF. This seems strange when regarding the findings for the factor 'Energy rate demand too high' (linked to a higher Localizer groundspeed), which did show a significantly higher RSME score with increasing Energy rate demand and higher Localizer intercept groundspeed. Also, when looking at the pilots' comments on the feedback forms and the results of the end of day questionnaires, one would expect that the Localizer intercept speed would have an influence on the RSME scores. This will need to be tested again in the next experiment.

Using Table 8 it is now possible to discuss which factors have an influence on pilot TDL. Ideally, factors that influence the difficulty of flying an approach would yield a significant difference in RSME scores, would be mentioned on the Feedback Forms *and* would be rated 'more difficult' or 'a lot more difficult' in the end of day questionnaire. The only factors that fulfill this description are: factor 4 (Energy rate demand too high), factor 8 (Distance (or time) available on Localizer Intercept Heading) and factor 9 (Line-up distance).

On the other hand, factors that do not influence pilot TDL would *not* yield a significant difference in RSME scores, would *not* be mentioned on the Feedback Forms and would *not* be rated 'more difficult' or 'a lot more difficult' in the end of day questionnaire. This applies to none of the factors in Table 8. However, factors 1 (number of heading changes), 2 (CDA compared to horizontal) and 7 (more altitude steps compared to CDA) were tested three

Table 8 | Factors that might influence pilot TDL with corresponding results for RSME ratings, comments on feedback forms and answers given in the end of day questionnaire.

	Factor Nr	Factors	Independent variable	Variable between pairs	RSME scorest-test (1), Wil-coxon signed-rank test(2) or ANOVA(3)	Factors mentioned on Feedback Forms by at least three pilots	Factors indicated to have an effect (#) in the End of day questionnaire	Hypothesis for TDL effect
	1	Number of heading changes	Yes	Yes	- (1)	Not mentioned	#	n.e.
	2	CDA compared to horizontal / Many different altitude constraints	Yes	Yes	-(2)	Not mentioned	#	e.
	3	Heading change towards Localizer Intercept Heading	Yes	Yes	*(1)	Not mentioned	Not asked	n.e.
•	4	Energy rate demand too high / Unable to comply with constraints at waypoints Linked factor: Localizer Groundspeed	Yes	Yes	**(1)	++	#	e.
•	5	LOC intercept speed (IAS)/LOC Groundspeed Linked factor: Energy rate demand IF-FAF	Yes	Yes	-(3)	+	#	e.
•	6	Mass	Yes	No	-(1)	Not mentioned	Not asked	e.
	7	More altitude steps compared to CDA	Yes	Yes	-(1)	Not mentioned	#	n.e.
•	8	Distance available on LOC Intercept Heading Linked factor: Localizer Groundspeed	Yes	Yes	*(1)	Not mentioned	Not asked	e.
		Time available for actions on LOC intercept leg Linked factor: Localizer Groundspeed				++	#	e.
•	9	Line-up distance Linked factor: IF-FAF distance	Yes	Yes	*(1)	++	#	e.
		Time available for actions on final Linked factor: IF-FAF distance				++	#	e.

	Factor Nr	Factors	Independent variable	Variable between pairs	RSME scorest-test (1), Wil-coxon signed-rank test(2) or ANOVA(3)	Factors mentioned on Feedback Forms by at least three pilots	Factors indicated to have an effect (#) in the End of day questionnaire	Hypothesis for TDL effect
•	10	FAF altitude	No	Yes	Not tested	++	Not asked	
		Time available between FAF and 1,000′				++		
•	11	LOC intercept angle	No	Yes	Not tested	++	#	
	12	Tailwind	No	Yes	Not tested	++	#	
•	13	Vertical speed	No	Yes	Not tested	Not mentioned	#	
•	14	Number of waypoints	No	Yes	Not tested	Not mentioned	#	
•	15	Turbulence	No	No	Not tested	Not mentioned	#	
•	16	Trackmiles	No	Yes	Not tested	Not mentioned	#	
•	17	More altitude steps compared to horizontal	No	No	Not tested	Not mentioned	#	
•	18	Airspeed on final	No	Yes	Not tested	++	Not asked	
•	19	Time available during first part of approach	No	Yes	Not tested	++	Not asked	
	20	Increase in time spent manoevring	No	Yes	Not tested	Not mentioned	#	
•	21	Stabilized at 1,000′	No	Yes	Not tested	Not mentioned	#	

- = no significance, * = $p < .05$, ** = $p < .01$
+ = mentioned < 10 times, ++ = mentioned > 10 times
# = factors indicated to have an effect in the end of day questionnaire
n.e. = no effect expected on TDL, e. = effect expected on TDL

3

times (by t-tests or Wilcoxon signed-rank tests for the RSME z-scores, by analyzing whether they were mentioned on the feedback forms, and by asking about these factors in the end of day questionnaire). Only in the end of day questionnaire did the pilots indicate that these factors had an effect on TDL, the other two tests did not show an effect on TDL. Therefore it is assumed that factors 1, 2 and 7 do not have an effect on pilot TDL.

Factor 3 (heading change towards Localizer Intercept Heading) was tested again, as mentioned before, by flying two approaches in the flight simulator, and pilots commented that there was no difference in TDL due to a smaller heading change towards Localizer Intercept Heading. For that reason factor 3 is assumed not to influence pilot TDL.

Factor 12 (tailwind) is by definition directly correlated with the value of the energy rate demand (Energy rate demand, factor 4), and therefore not regarded as an additional factor that influences pilot TDL. Factor 20 (increase in time spent maneuvering) is already covered by factors 1 (number of heading changes), 7 (more altitude steps compared to CDA) and 17 (more altitude steps compared to horizontal), and also not regarded a separate factor.

Given the above, some factors indicated in Table 8 can be removed from the list of factors that influence pilot TDL. The remaining factors in Table 8 are indicated by a black dot in the first column.

7 | Regression analysis for pilot TDL

A regression analysis was performed on the pilots' average RSME z-scores for the approaches. The goal of the regression analysis was to identify the factors that were dominant in affecting the pilots RSME ratings and thereby pilot estimates of TDL during an approach. (The RSME ratings are an indication of the Mental Load experienced by the pilots. However, by choosing pilots with different levels of experience, different ages, different levels of fatigue and by testing the approaches in random order it is assumed that through the RSME ratings of the pilots a good indication of pilot TDL can be obtained.) The predictors used for the regression analysis are based on the factors that are indicated by a black dot in Table 8. For different reasons, not all these factors are incorporated as predictors.

Factors 15 (turbulence) and 17 (more altitude steps compared to horizontal) were not varied among the different approaches, and can therefore, for this exeperiment, not be used as predictors.

Factor 13 (vertical speed) is analytically defined by the IAS, wind conditions and the inertial flight path angle. When pilots commented on the vertical speed, they always referred to the vertical speed on final. On final the inertial flight path angle is equal to 3 degrees for all approaches, implying that the vertical speed on final only depends on the IAS on final and wind conditions on final. Compared to the influence of the IAS on the vertical speed, the influence of the wind is very small for the approaches considered. Therefore the vertical speed can also be represented by the IAS on final, which already is a factor (factor 18). Therefore, for this experiment, vertical speed is not regarded a separate factor for the approaches considered.

Factor 19 (Time available during first part of approach) could only be calculated for a part of the approaches since some approaches did not contain a first part of the approach. It is therefore decided to not use factor 19 as a predictor in the regression analysis for this experiment.

Factor 18 (Airspeed on final) is calculated as the average of Localizer Intercept Speed (which is Factor 5) and IAS at the FAF. Since Factor 5 is already a predictor, the choice is made to include the IAS at the FAF as an additional predictor, instead of the average airspeed on final.

Table 9 | Factors used as predictors in the regression analysis. Correlation coefficients larger than .7 are denoted.

		Energy rate demand LOC intercept HDG	**Energy rate demand Final**	**Localizer Intercept speed (IAS)**	**Aircraft mass**	**Distance available on Localizer Intercept Heading**	**Line-up distance**	**FAF altitude**	**Localizer Intercept angle**	**Number of waypoints**	**Trackmiles**	**IAS at FAF**	**Stabilized at 1000′ (Monte Carlo original simulation)**
4	Energy rate demand LOC int HDG	×	.78	.91	×	×	×	×	×	×	×	.84	-.80
4	Energy rate demand Final	.78	×	.92	×	×	×	×	×	×	×	.87	-.71
5	Localizer Intercept speed (IAS)	.91	.92	×	×	×	×	×	×	×	×	.93	-.80
6	Aircraft mass	×	×	×	×	×	×	×	×	×	×	×	×
8	Distance available on Localizer Intercept Heading	×	×	×	×	×	×	×	×	×	×	×	×
9	Line-up distance	×	×	×	×	×	×	.91	.80	×	×	×	.71
10	FAF altitude	×	×	×	×	×	.91	×	.92	×	×	×	×
11	Localizer Intercept angle	×	×	×	×	×	.80	.92	×	×	×	×	×
14	Number of waypoints	×	×	×	×	×	×	×	×	×	.92	×	×
16	Trackmiles	×	×	×	×	×	×	×	×	.92	×	×	×
18	IAS at FAF	.84	.87	.93	×	×	×	×	×	×	×	×	-.80
21	Stabilized at 1000′ (Monte Carlo original simulation)	-.80	-.71	-.80	×	×	.71	×	×	×	×	-.80	×

The remaining fourteen factors are used as predictors for the regression analysis, these factors and their factor numbers are given in the first column of Table 9. Due to the small amount of different approaches that were tested, some of these predictors are highly correlated, which will most probably pose a problem for the regression analysis, correlations higher than .7 are given in Table 9. It should be noted that for the Localizer Intercept Speed (factor 5), the IAS during Localizer intercept predicted by the original Monte Carlo simulation is used, and for the Airspeed at FAF (factor 22), the IAS at the FAF predicted by the original Monte Carlo simulation is used.

Table 10 | The stepwise linear regression model for the dependent variable RSME z-score.

	B	***SE B***	***β***	***R²***	**Adjusted *R²***
Step 1					
Constant [-]	2.06	.234			
Line-up distance [nm]	-.311	.035	-.597 *	.356	.351
Step 2	×	×	×	×	×
Constant [-]	-2.16	.894			
Line-up distance [nm]	-.226	.037	-.434 *		
Localizer Intercept Speed [knots]	.019	.004	-.345 *	.449	.441
Step 3					
Constant [-]	-1.61	.899			
Line-up distance [nm]	-.185	.039	-.355 *		
Localizer Intercept Speed [knots]	.017	.004	-.301 *		
Total trackmiles [nm]	-.018	.007	.194 *	.475	.464

* $p < .001$

A stepwise regression was performed on these predictors (probability of F to enter <= .05, probability of F to remove >=.01) for the dependent measure average RSME z-score, the resulting model is shown in Table 10. The average Variance Inflation Factor (VIF) for these three predictors is 1.44, which indicates that for these predictors collinearity is not an issue. However, considering the large amount of high correlation coefficients between these three predictors and other factors not included in the regression model in Table 10, other possible regression models are considered by performing a best subsets regression. The best subsets regression revealed, as expected, that several different combinations of the list of predictors in Table 9 provide comparable R squared values to the regression model resulting from the stepwise regression in Table 10.

The authors infer that the following factors should be used as predictors in a regression model for the average RSME z-score:

- Factor 4: Energy rate demand too high, calculated separately for the LOC intercept part and the final part of the approach
- Factor 5: Localizer Intercept Speed, the IAS during Localizer intercept predicted by the original Monte Carlo simulation is used
- Factor 8: Distance available on Localizer Intercept Heading
- Factor 9: Line-up distance
- Factor 10: FAF altitude
- Factor 11: Localizer Intercept angle
- Factor 21: Stabilized at 1,000′, the prediction of the original Monte Carlo simulation is used

This means that of all factors in Table 9 factor 14 (number of waypoints) and factor 16 (total trackmiles) are no longer considered predictors. This choice is made based on observations during the experiment: none of the pilots mentioned the number of waypoints or total trackmiles either in written form or orally, although these factors were varied significantly among approaches. Factor 6 (aircraft mass) is also left out, since from a flight mechanical perspective this only has an influence through a change in Energy rate demand, which is already covered by factor 4. Finally, factor 22 (airspeed at FAF) is no longer considered as a predictor, we think that the speed at the FAF in itself is not a factor for pilot TDL, but that the possibility to reduce from the speed at the FAF to a speed less than V_{REF} + 20 at 1,000′ is a factor for pilot TDL, and this is again covered by the Energy rate demand for the final part of the approach (factor 4).

The resulting regression model for the remaining predictors given in the list above is shown in Table 11. The largest VIF factor for these predictors equals 25, and thus is cause for concern. Cook's distance is smaller than 1 for all cases, and the centered leverage value is smaller than twice the average leverage for all cases.

Table 11 | The linear regression model for the dependent variable RSME z-score chosen as best model from a best subsets regression.

	B	***SE B***	***β***	***p***	***R²***	**Adjusted *R²***
Constant [-]	-5.07	2.489		.044		
Energy rate demand LOC intercept part [-]	-.091	.138	-.109	.512		
Energy rate demand Final part [-]	-.076	.047	-.374	.112		
Localizer Intercept Speed [knots]	.045	.015	.800	.003		
Distance available on LOC int. HDG [nm]	-.128	.068	-.156	.060		
Line-up distance [nm]	-.108	.094	-.208	.251		
FAF altitude [ft]	-.003	.001	-.822	.008		
Localizer Intercept angle [deg]	.053	.022	.548	.018		
Stabilized at 1,000′ [%]	.004	.003	.198	.218	.497	.467

The way forward is to use the predictors of the regression model in Table 11 as a starting point for the design of the approaches that will be tested during the second flight simulator experiment (earlier indicated as step 3 in this research). Special care should be taken to ensure a low correlation between these factors in the next experiment. Additionally, pilots should be asked specifically about these factors in the feedback forms and end of day questionnaire.

8 | Comparison of experiment results and Monte Carlo simulation

Some of the predictors for pilot TDL presented in the previous section are directly evident when looking at the approach charts (for instance, the FAF altitude or the Localizer intercept angle), the flight mechanical factors however (these are the energy rate demand and the fact whether the approach is stabilized at 1,000′) are not directly available and need to be calculated. The Monte Carlo computer simulation aims to do exactly that. As stated in the experiment goals, this experiment aimed to validate the Monte Carlo simulation: with respect to the simulation of the pilot's actions, with respect to the prediction whether or not the constraints at the waypoints are met and with respect to the prediction whether or not the approach is stabilized at 1,000′. For each of these three items the flight simulator results are compared to the Monte Carlo simulation in the following paragraphs.

8.1 | Simulation pilot actions

The simulation of the pilots' actions in the Monte Carlo simulation is adjusted according to the findings presented in section 5.3, and based on the trigger events and reaction times in Table 7. As a first approximation normal distributions are used based on the means and standard deviations given in Table 7.

Additionally, the use of speedbrakes is incorporated in the Monte Carlo simulation. In the original Monte Carlo simulation [2], speedbrakes were assumed not to be used, the experiment results, however, showed that pilots did use the speedbrakes (see Figure 11). Speedbrakes are modeled to be selected when a 'drag required' message would appear on the pilot's displays, indicating that the constraints at the next waypoint would not be met. There is no time delay in applying speedbrakes. Speedbrakes are selected the moment the 'drag required' message appears, and are selected consistently every time the message appears. Due to simulation technical reasons, the speedbrakes could only be used in the Monte Carlo simulation until the Localizer was captured. In reality, pilots could use the speedbrakes for a slightly larger part of the approach: until they reached the FAF, since they are instructed not to use the speedbrakes for flap settings greater than flaps 10 (flaps 20 is selected at the FAF).

The results of this adjusted Monte Carlo simulation are compared to the flight data in the following paragraphs.

8.2 | Prediction stabilized

Figure 17 shows for each approach the percentage of pilots that were stabilized according to the flight data, the percentage of pilots that would be stabilized according to the adjusted Monte Carlo simulation when speedbrakes are applied, and the percentage of pilots that would be stabilized according to the adjusted Monte Carlo simulation when speedbrakes are *not* applied. It can be seen that the fact whether or not speedbrakes are applied (in the Monte Carlo simulation) can have a large influence on the prediction for the percentage of stabilized approaches (for instance, approach 7: 56% is stabilized when using speedbrakes, 0% is stabilized without speedbrakes).

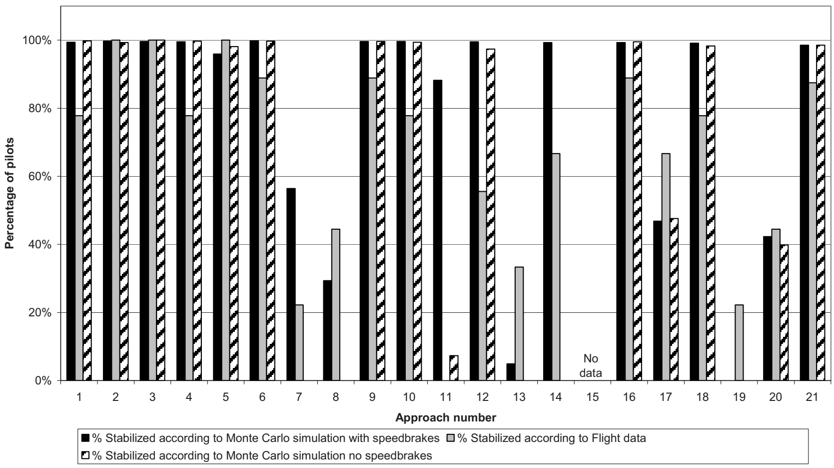


Figure 17 | Percentage of pilots stabilized according to the flight data and stabilized according to the adjusted Monte Carlo simulation (with and without the use of speedbrakes)

There is a significant correlation between the percentage of pilots that were stabilized according to the flight data and according to the adjusted Monte Carlo prediction, $\tau = .53$, p(one-tailed) $< .001$, for the adjusted Monte Carlo prediction when speedbrakes are used, and $\tau = .64$ p(one-tailed) $< .001$ for the adjusted Monte Carlo prediction when speedbrakes are not used.

The discrepancies between the flight data and the adjusted Monte Carlo simulation can be due to the fact that during the experiment, pilots have used the speedbrakes for a longer amount of time (e.g. also after capturing the Localizer) than the amount of time modeled in the adjusted Monte Carlo simulation, or have used the speedbrakes for a shorter amount of time. During the experiment pilots sometimes forgot to retract the speedbrakes due to the absence of buffet cues (while speedbrakes were deployed) in the flight simulator. Another reason for discrepancies could be the fact that, as explained before, the throttle operation differed from reality. Pilots therefore sometimes found it more difficult to regulate the airspeed than they would normally find it during real flight. This sometimes resulted in an airspeed that was too low or too high. This was not incorporated in the Monte Carlo simulation. Additionally, during the experiment pilots sometimes maintained an IAS at the FAF that was lower than defined on the approach charts, as a result they could sometimes achieve a stabilized approach at 1,000′, while the Monte Carlo simulation (maintaining the correct higher IAS at the FAF) predicted an unstabilized approach.

Figure 17 illustrates the large influence that the use of speedbrakes can have on the possibility of achieving a stabilized approach, and hence on the correlation with the predictions of the Monte Carlo simulation. It is therefore better to not allow the use of speedbrakes during the next experiment, and to compare these results to the Monte Carlo simulation without the use of speedbrakes. This is also a better choice regarding the intended use of the method developed during this research: to provide an additional tool for the design of approaches. Approaches should be designed such that they can be flown without speedbrakes.

3

8.3 | Prediction meet constraints

The percentage of pilots that met the constraints according to the flight data (significantly non-normal $D(76) = 0.23$, $p < .05$) was significantly related to the percentage of pilots that met the constraints according to the Monte Carlo simulation. This is true for the adjusted Monte Carlo simulation during which speedbrakes were used τ (one-tailed) = .407, $p < .001$ and for the adjusted Monte Carlo simulation during which speedbrakes were not used t (one-tailed) = .500, $p < .001$. However, in order to be able to use the Monte Carlo simulation to accurately predict whether the constraints at the waypoints will be met, a larger correlation coefficient is desirable and might be achieved by not allowing the use of speedbrakes during the next flight simulator experiment. Since the Monte Carlo simulation will ultimately be used during the design of approaches, and approaches should be designed such that they can be flown without the use of speedbrakes, not allowing the use of speedbrakes corresponds better to the goal of this research.

9 | Conclusions and further research

In this paper a relation has been found between pilot TDL during approaches (measured in terms of the RSME rating scale) on the one hand and flight mechanical factors and factors relating to the approach trajectory on the other hand. Based on the results of flight simulator tests for a B747-100 the authors infer that the following factors influence pilot TDL: the energy rate demand during the last parts of the approach, the Localizer Intercept Speed, the distance available on Localizer intercept heading, the line-up distance, the FAF altitude, the Localizer intercept angle, and the fact whether the approach is stabilized at 1,000′.

Some of these factors are directly evident when looking at the approach charts (for instance, the FAF altitude or the Localizer intercept angle), the flight mechanical factors however (these are the energy rate demand and the fact whether the approach is stabilized at 1,000′) are not directly available and need to be calculated. This paper also presented the validation of a Monte Carlo computer simulation which aims to do exactly that (the Monte Carlo computer simulation itself was presented in the accompanying paper [2]). By modeling the pilot's actions according to the SOPs and applying a distribution in time for all these actions, the Monte Carlo computer simulation predicts the percentage of pilots that will meet the altitude and velocity constraints at the waypoints, and the percentage of pilots that will achieve a stabilized approach at 1,000'. A significant correlation was found between these predictions and the flight data from the flight simulator tests. However, a larger correlation coefficient than achieved in this paper is desirable and might be realized by not allowing the use of speedbrakes during the next experiment.

The factors that were found to influence pilot TDL in this paper are regarded as an initial indication, and further research, based on these findings, is required. We will use the predictors of the regression model as a starting point for the design of the approaches that will be tested during the second flight simulator experiment (indicated as step 3 in this research). Special care should be taken to ensure a low correlation between these factors in the next experiment. Additionally, pilots should be asked specifically about these factors in the feedback forms and end of day questionnaire.

The next flight simulator experiment will also be used to further validate the Monte Carlo simulation. Pilots should not be allowed to use the speedbrakes in the next experiment, in order to cancel out this additional variable and to obtain solid validation data for the Monte Carlo simulation. This is also a better choice regarding the intended use of the method developed during this research: to provide an additional tool for the design of approaches. Approaches should be designed such that they can be flown without the use of speedbrakes.

Acknowledgements

This research was financially supported by the Dutch Technology Foundation STW (DLR7071). We would like to thank all pilots who participated in the flight simulator experiment.

References

1. Stassen, HG, Johannsen, G, & Moray, N. Internal Representation, Internal Model, Human Performance Model and Mental Workload. Automatica, 26(4), 811-820, 1990.
2. Heiligers, M.M., Van Holten, Th., Mulder, Flight mechanical evaluation of approaches, accepted for publication in the Journal of Aircraft (Journal Article 2 in the Dissertation).
3. Heiligers, M.M., Van Holten, Th., Mulder, M, Predicting Pilot Task Demand Load During Final Approach, *The International Journal of Aviation Psychology,* 19(4), 391-416, October-December 2009.
4. Baron, S., Zacharias, G., Muralidharan, R., Lancraft, R. (1980), PROCRU: A model for analyzing flight crew procedures in approach to landing. In NASA-MIT 16th Annual Manual 1980, pp 488-520.
5. Baron, S., Kruser, D.S., Huey, M.B., Quantitative Modeling of Human Performance in Complex, Dynamic Systems,National Academy Press, Washington D.C., 1990.
6. Corker, K.M. (1999), Human Performance Simulation in the Analysis of Advanced Air Traffic Management, In Proceedings of the 1999 Winter Simulation Conference, Farrington, P.A., Nembhard, H.B., Sturrock, D.T., Evans, G.W. eds.
7. Gore, B.F., Corker, K.M. (2000) A Systems Engineering Approach to Behavioral Predictions of an Advanced Air Traffic Management Concept, 19th Digital Avionics Systems Conference, Entering the Second Generation of Powered Flight, Philadelphia, Pennsylvania.

8. Smith, B.R., Tyler, S.W. (1997), The Design and Application of MIDAS: A Constructive Simulation for Human-System Analysis, 2nd Simulation Technology & Training Conference, Canberra, Australia
9. Vicente, K.J., Cognitive Work Analysis, Towards Safe, Productive, and Healthy Computer-Based Word, Lawrence Erlbaum Associates, Publishers, Mahwah, New Jersey, 1999
10. International Civil Aviation Organization, *Procedures for Air Navigation Services Aircraft Operations, Volume II Constructions of Visual and Instrument Flight Procedures,* Fifth edition 2006.
11. Godley, S.T. *Perceived Pilot Workload and Perceived Safety of RNAV (GNSS) Approaches,* Australian Transport Safety Bureau, ATSB Transport Safety Investigation Report, Aviation Safety Research and Analysis Report 20050342, 2006.
12. Vormer, F.J.P., *Theoretical and Operational Aspects of Optimal Airport Arrival Trajectories,* TUD Technische Universiteit Delft, 8 december 2005
13. Hanke, Rodney C. and Nordwall, Donald R., *The Simulation of a Jumbo Jet Transport Aircraft Volume II: Modeling Data*, NASA report D6-30643, September 1970, Wichita, Kansas.
14. Zijlstra, F.R.H., 1993. Efficiency in work behavior. A design approach for modern tools. PhD thesis, Delft University of Technology, Delft University Press, Delft, The Netherlands.
15. Zijlstra, F., Meijman, Th., Het meten van mentale inspanning met behulp van een subjectieve methode, *Mentale belasting en werkstress, een arbeidspsychologische benadering,* Meijman,Th. (Ed.), Van Gorcum, Assen/Maastricht, The Netherlands, 1989.
16. Lodge, M., *Magnitude Scaling, Quantitative Measurement of Opinions,* Sage university paper series on quantitative applications in the social sciences, 07-025, Newbury Park, CA: Sage.
17. Meijman, T., Zijlstra, F., Kompier, M., Mulders, H., Broersen, J., The measurement of perceived effort, *Contemporary Ergonomics 1986,* Oborne, D.J. (Ed.), Taylor & Frances, London, 1986.
18. Hart, S.G., Staveland, L.E., Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research, *Advances in Psychology* Vol. 52, p.139-183, P.A. Hancock, Meskati N. (Eds.), Elsevier Science Publishers B.V., Amsterdam, The Netherlands, 1988.
19. Flight Safety Foundation Approach-and-landing Accident Reduction (ALAR) Task Force. Killers in Aviation: FSF Task Force Presents Facts about Approach-and-landing and Controlled-flight-into-terrain Accidents, *Flight Safety Digest*, November 1998 - February 1999

Appendix A to Journal Article 3

Table 12 shows all relevant data for the approaches that were tested. Most variables in Table 12 are self-explaining; except the energy demand ratio. The maximum value of the energy demand ratio is calculated for each approach part separately. The values in Table 12 are determined using the Monte Carlo simulation, by calculating the average value of the energy demand ratio during a 10 second time interval before a waypoint is reached. The value in Table 12 thus differs from the value that is predicted by the point mass model as explained in the accompanying paper [2], which gives an energy demand ratio that is an *average* value for the entire leg.

If the value for the Energy rate demand is larger than 1 this indicates that the constraints at the next waypoint will not be met, the larger the value for the Energy rate demand, the more the actual IAS will exceed the required IAS at the waypoint. It is important to note that pilots received a 'drag required' message when the Energy rate demand became higher than 1 during the experiment. No message would show up for Energy rate demand values smaller than 1, pilots would thus not be aware of the fact whether the Energy rate demand equaled, for instance, 0.1, 0.6 or 0.9, since in all these cases the constraints would be met. All approach pairs are now briefly described below:

Number of heading changes (approach pair A) - The difference between approaches 1 and 2 was the amount of heading changes within the same amount of trackmiles in the first part of the approach. A difference in Energy rate demand can be observed for the first part of the approach (Table 12), however, the Energy rate demands do no exceed 1.

CDA compared to horizontal (approach pair B) - Approaches 3 and 4 both start on Localizer Intercept Heading. The final part of the approach is identical for approaches 3 and 4, the Localizer Intercept Heading part is a horizontal segment for approach 4 and a CDA segment for approach 3. However, due to keeping the IAS constraints on Localizer Intercept Heading the same, the Energy rate demand will increase for the CDA segment (approach 3) compared to the horizontal segment (approach 4). However, the Energy rate demand does not exceed 1 for either of the approaches.

Heading change towards Localizer Intercept Heading (approach pair C) - Approaches 9 and 10 are identical except for the heading change that is required when turning from the leg before Localizer Intercept Heading to Localizer Intercept Heading. A very small difference in Energy rate demands can be observed, but all are lower than 1.

Energy rate demand too high (approach pair D) - Approaches 17 and 18 are exactly identical. The difference in Energy rate demand is established by applying a wind of 40 knots in approach 18 which resulted in a strong tailwind during the first part of the approach and the Localizer Intercept Heading part. As a result, the Energy rate demand during approach 18 was larger than 1. In approach 17, the wind was only 10 knots and all constraints could be met (Energy rate demand smaller than 1). Since the IAS during Localizer Intercept was identical for both approaches, the groundspeed during Localizer Intercept differed for both approaches due to the wind.

Localizer Intercept Speed (approach pair E) - The IAS during Localizer intercept is different for approaches 12, 13 and 14. This difference in IAS is amplified by adding wind to the experiment resulting in an even larger difference in Groundspeed during Localizer intercept. Since the remainder of the approach is exactly identical for approaches 12, 13 and 14 a higher Localizer Intercept speed (IAS) inevitably means that the Energy rate demand will increase between IF and FAF.

Aircraft Mass (approach pair F) - Approaches 20 and 21 are exactly identical, bur are flown with a different aircraft mass. According to the predictions of the PMM and Monte Carlo simulation it would be easier meet the constraints at the waypoints with the higher aircraft mass, than with the lower aircraft mass (represented by different Energy rate demands in Table 12).

CDA compared to altitude steps (approach pair G) - The first part of the approach for approach 2 was a Continuous Descent Approach, whereas the first part of the approach for approach 16 consisted of three altitude steps (stepped approach). Location of waypoints and IAS restrictions were the same for both approaches. A difference in Energy rate demands can be observed (Table 12, but all remain below 1.

Distance available on Localizer Intercept Heading (approach pair H) - The total amount of trackmiles is equal for both approaches. For approach 3 the Localizer intercept part is 5.1nm and there is no first part of the approach, for approach 9 the Localizer intercept part is 1.6nm long and the remaining 3.5nm constitutes the first part of the approach. Unfortunately, due to an incorrect wind setting during the experiment, there is also a difference in groundspeed during Localizer Intercept within this approach pair: 170kts for approach 3, and 179kts for approach 9. The Energy rate demands are all below 1.

Line-up distance (approach pair I) - Approach 3 has a line-up distance of 7nm, approach 6 has a line-up distance of 10nm. Since the altitude of the FAF is the same for both approaches, this inevitably leads to the fact that the distance between IF and FAF will differ for both approaches, as well as the Energy rate demand between IF and FAF (but remains below 1).

Table 12 | Relevant data for all approaches.

Approach number	# Heading changes	#Altitude steps	Max. Energy rate demand First part approach[-]	Energy rate demand LOC intercept heading[-]	Max. Energy rate demand Final part approach[-]	Heading change towards LOC intercept heading [deg]	Mass{10³ [lbs]	Line-up distance [nm]	Distance on Localizer Intercept Heading [nm]	FAF altitude [ft]	FAF speed [knt]	Localizer intercept angle [deg]	Localizer intercept speed (IAS) [knt]	Localizer intercept speed (Groundspped) [knt]	Distance IAF-IF [nm]	Distance IF-FAF [nm]	Stabilized accordingg to Monte Carlo simulation [%]
1	3	CDA	0,2	0,1	0,1	20	237	9	5	2,000	180	60	180	178	30	3	100
2	6	CDA	0,8	0,1	-0,1	20	237	9	5	2,000	180	60	180	177	30	3	100
3	1	CDA	×	0,6	0,9	×	237	7	5.1	1,590	160	45	180	170	5	2	100
4	1	1	×	-0,2	-0,1	×	237	7	5.1	1,600	160	45	180	176	5	2	100
5	2	CDA	1,6	0,9	1,2	90	237	7	4.1	1,590	160	90	180	180	20	2	100
6	1	CDA	×	0,8	0,3	×	237	10	5	1,600	160	45	180	170	5	5	100
7	2	CDA	1,8	3,7	14,9	45	237	5	3	1,270	160	35	180	181	8	1	0
8	2	CDA	0,6	2,8	5,4	45	237	5	3	1,270	160	35	180	181	8	1	0
9	2	CDA	0,6	0,0	-0,2	30	237	7	1.6	1,590	160	45	180	179	5	2	100
10	2	CDA	0,5	-0,3	-0,1	50	237	7	1.6	1,590	160	45	180	175	5	2	100
11	2	CDA	0,8	0,4	3,7	30	237	7	1.6	1,590	160	45	240	240	5	2	0
12	1	CDA	×	1,3	3,0	×	237	5	4	1,350	160	45	170	150	4	1	0
13	1	CDA	×	2,1	13,0	×	237	5	4	1,350	160	45	190	205	4	1	0
14	1	CDA	×	1,8	9,5	×	237	5	4	1,350	160	45	180	176	4	1	0
16	6	3	0,2	-0,3	-0,3	20	237	9	5	2,000	180	60	180	180	30	3	100
17	4	CDA	1,0	0,1	0,3	30	237	5	3	1,270	160	30	170	161	18	1	100
18	4	CDA	1,4	1,7	2,3	30	237	5	3	1,270	160	30	170	205	18	1	0
19	4	CDA	0,6	0,4	2,9	30	237	5	3	1,270	160	30	200	202	18	1	0
20	4	CDA	1,4	1,2	1,5	30	237	5	3	1,270	160	30	170	166	18	1	3
21	4	CDA	0,7	0,1	0,1	30	295	5	3	1,270	160	30	170	166	18	1	100

3

Appendix B to Journal Article 3

Number of heading changes (approach pair A) - On average, there was no difference in the pilots' RSME rating for approach 1 with only three heading changes (M=-0.95, SE = 0.168) and approach 2 with six heading changes (M=-0.85, SE = 0.162, $t(8)$ = -0.851, p = .4, r = .29). RSME ratings thus indicate that the number of heading changes in an approach most probably does not have an effect on pilot TDL

CDA compared to horizontal (approach pair B) - Using a Wilcoxon signed-rank test it was found that the RSME z-scores for approach 4, which was the horizontal approach, were not significantly lower (Mdn = -0.36) than for approach 5, which was the CDA approach (Mdn = -0.38, T = 15, p =.74, r = .10). A CDA or a horizontal approach would thus result in the same TDL.

Heading change towards Localizer Intercept Heading (approach pair C) - There was a difference in RSME z-scores for approach 9, which required a heading change of 30 degrees when turning towards Localizer Intercept Heading (M = 0.46, SE = 0.181), and approach 10, which required a heading change of 50 degrees (M=-0.01, SE = 0.135), $t(8)$ = 2.57, p < .05, r = .67. This indicates that TDL increased when the heading change towards Localizer Intercept Heading decreased, a surprising result which will be further elaborated upon in paragraph 6.2.

3

Energy rate demand too high (approach pair D) - Approach 18 with a maximum Energy rate demand of 1.5 was rated significantly higher by pilots on the RSME scale (M = 0.72, SE = 0.207) than approach 17 with a maximum Energy rate demand of 1.1 (M=-0.38, SE = 0.199), $t(8)$ = -5.03, p (one-tailed) < .001, r = .87. Table 3 shows that the groundspeed during Localizer Intercept also varied between approach 17 (161kts) and approach 18 (205kts), this was a 'linked factor'. This implies that an Energy rate demand that is too high combined with a higher Localizer Intercept Groundspeed increases pilot TDL.

Localizer Intercept Speed (approach pair E) - The RSME z-scores for approach 12, 13 and 14 with a Localizer Intercept Speed of 170, 190 and 180kts IAS, respectively, were analyzed using a one-way repeated-measures ANOVA. Mauchly's test indicated that the assumption of sphericity was valid ($\chi^2(2)$ = 0.537, p > .05). The results showed that there was no effect of the Localizer Intercept Speed on the average RSME z-scores, $F(2, 16)$ = 2.06, p > .05, ω^2 = .08. The factor that was linked to a higher Localizer Intercept Speed (IAS) was the Energy rate demand between IF and FAF. The results thus seem to indicate that a higher Localizer Intercept Speed (IAS) combined with a higher Energy rate demand between IF and FAF does not influence pilot TDL.

Aircraft Mass (approach pair F) - There was no difference between the average RSME ratings given by pilots for approach 21 (with the higher aircraft mass) (M =0.19, SE = 0.324) and for approach 20 which was performed with a lower aircraft mass (M=-0.105, SE = 0.322), $t(7)$ = -1.14, p >.05, r = .4. This implies that pilot TDL was found to be approximately the same for both aircraft masses.

CDA compared to altitude steps (approach pair G) - On average, there was no difference between the pilots' RSME rating for approach 2 which was the CDA approach (M =-0.85, SE = 0.162) and approach 16 which contained three altitude steps (M=-1.01, SE = 0.139), $t(8)$ = -0.58, p (one-tailed) = .29, r = .2. The RSME ratings thus indicate that there is no significant difference in pilot TDL during a CDA or during a stepped approach.

Distance available on Localizer Intercept Heading (approach pair H) - The pilots' RSME rating for approach 3, designed with 5.1nm available on Localizer Intercept Heading, was significantly lower (M =-0.28, SE = 0.236) than the pilots' RSME rating for approach 9, which was designed with 1.6nm available on Localizer Intercept Heading (M=0.46, SE = 0.181), $t(8) = -2.28$, p (one-tailed) < 0.05, $r = .63$. Keeping in mind that there was also a difference in groundspeed during Localizer Intercept between the two approaches (170kts for approach 3 and 179kts for approach 9) this implies that when the distance available on Localizer Intercept Heading becomes smaller the difficulty of flying an approach increases which might also be influenced by a higher groundspeed during Localizer Intercept.

Line-up distance (approach pair I) - There was a significant difference in RSME rating by the pilots for approach 3 (M =-0.28, SE = 0.236) and approach 6 (M=-0.76, SE = 0.240), $t(8) = 1.89$, p (one-tailed) < 0.05, $r = .55$; implying that approach 3 with a line-up distance of 7nm was rated significantly higher (i.e. more difficult) than approach 6, which was designed with a line-up distance of 10nm. The factors linked to a smaller line-up distance (Table 3) were a higher Energy rate demand between IF and FAF and a smaller distance between IF and FAF. This means that a smaller line-up distance, combined with a higher Energy rate demand between IF and FAF and a smaller distance between IF and FAF increases pilot TDL.

Appendix C to Journal Article 3

Flaps 1
For the analysis of the reaction times and trigger events for selecting flaps 1, only the approaches that were designed to start with an IAS higher than the UP mark have been considered (*N* = 43). For the other approaches (with initial speeds lower than the UP mark) flaps 1 was the initial condition at the start of the simulation, and pilots did not have to select flaps 1 during the approach.

In the Monte Carlo simulation it was assumed that 'reaching the first waypoint of the leg on which the aircraft is 20 trackmiles from field' (from now on referred to as '20 trackmiles from field') was the trigger event for selecting flaps 1. The experiment results, however, showed that this was not a good choice. Figure 12(a) shows a wide spread in reaction times between reaching 20 trackmiles from the field and selecting flaps 1 (a negative reaction time indicates that flaps 1 were selected before reaching the point where the aircraft was 20 trackmiles from the field), therefore this can not be regarded as an actual trigger event. Additionally, during the experiment, none of the pilots mentioned the trackmiles or asked for the amount of trackmiles when selecting flaps 1. Based on the pilots' comments during the approaches and further analysis of the data, a better trigger event for selecting flaps 1 was the IAS, see Figure 12(b). The Flap Speed Mark for flaps UP (UP mark) was at 223 knots (below this speed flaps 1 should be selected), the Placard Speed was 280 knots. Figure 12(b) shows that during the majority of flights, flaps 1 were selected before the IAS reached the UP mark (this means that the IAS was higher than the UP mark).

Flaps 5
For the analysis of the reaction times and trigger events for selecting flaps 5, only the approaches that were designed to start with an IAS higher than the 1 mark have been considered (*N* = 70). For the other approaches (with initial speeds lower than the 1 mark) flaps 5 was the initial condition at the start of the simulation.

Similar to the selection of flaps 1, experiment results showed that flaps 5 also appeared to be selected depending on IAS, instead of at a predetermined location in the approach. The histograms in Figure 13(a) show that there was a large range in reaction times with respect to the trigger event 'turning to the leg before Localizer Intercept Heading' (as assumed in the Monte Carlo simulation), indicating that this can not be regarded an actual trigger event. Selection of flaps 5 as a function of IAS gives better results (see Figure 13(b)), during the majority of flights flaps 5 were selected before the Flaps Speed Mark for flaps 1 was reached (at 203 knots).

Actions on Localizer Intercept Heading
To analyze the pilot actions on Localizer Intercept Heading, only the approaches have been considered during which all five pilot actions were indeed performed on Localizer Intercept Heading. During 31 cases (flight simulator runs), the pilot forgot to switch to Heading Select mode, and therefore these approaches are not taken into account. These 31 cases were equally distributed over the 20 approaches that were designed and tested. During approach number 20 and 21 the IAS profile required the pilots to select flaps 10 before Localizer Intercept Heading was reached, these approaches are also not taken into account. Approaches 20 and 21 were flown by 9 and 8 pilots respectively, during these approaches the pilots forgot to switch to heading select 3 times, this leaves 127 cases to analyze the pilot actions on Localizer Intercept Heading.

Since the pilots were briefed to perform the actions Flaps 10, Arm Approach, Heading Select, Autopilot off and Autothrottle off on Localizer Intercept Heading (as prescribed by the SOPs), the data showed that these actions were indeed performed on Localizer Intercept Heading (a correct trigger event). Regarding the correlation between the reaction times for the actions that were performed on Localizer Intercept Heading and the time available on Localizer Intercept Heading, it was found that there was no significant effect. The correlations between time available on Localizer Intercept Heading and all pilot actions are: $\tau = .008$, $p = .9$ for Arm Approach, $\tau = .089$, $p = .155$ for flaps 10, $\tau = .015$, $p = .811$ for Autopilot off, $\tau = .011$, $p = .865$ for Autothrottle off, and $\tau = .026$, $p = .374$ for Heading Select. The flight data thus showed that irrespective of the time that was available, pilots tended to start performing these actions the moment they started to turn towards Localizer Intercept Heading, to 'get it over with' as soon as possible.

A very high correlation existed between the reaction times for Autothrottle off (significantly non-normal, $D(127) = 0.099$, $p < .01$) and Autopilot off (significantly non-normal, $D(127) = 0.091$, $p < .05$), $r_s = .998$, $p < .001$. This correlation is due to the fact that both actions are asked for in a single call: 'Autothrottle off, Autopilot off', and therefore these two actions were, most of the time, performed very closely together in time. Figure 14 shows the histograms for these two actions, as well as the histogram for the time between these two actions for each flight.

A one-way repeated-measures ANOVA was performed on the reaction times for the actions on Localizer Intercept Heading, to examine whether these actions were performed in a specific order. Mauchly's test indicated that the assumption of sphericity was violated ($\chi^2(9) = 1025$, $p < .01$). The degrees of freedom were therefore corrected using Greenhouse-Geisser estimates of sphericity ($\varepsilon = .402$). The results showed that there was a small effect of the type of action (selecting flaps 10, disconnecting the autopilot, etc.) on the reaction times, $F(1.61, 202.5) = 19.66$, $p < .001$, $\omega^2 = .08$, indicating that there was (at least partially) a preferred order to perform the actions. Post Hoc tests using the Bonferroni correction revealed that selecting flaps 10 was the first action to be performed ($p < .001$) and Autopilot Off was the last action to be performed ($p < .05$) on Localizer Intercept Heading, see Table 6.

Table 6 | Order of reaction times for actions on Localizer Intercept Head

Action	**Mean [sec]**	**Standard deviation [sec]**
Flaps 10	22.9	33.1
Heading Select	33.2	17.9
ARM Approach	35.9	20.3
Autothrottle Off	40.3	17.5
Autopilot Off	40.7	17.4

For the actions in between it was found that Heading Select was on average selected before the Autothrottle was switched off ($p < .001$). There was, however, no significant difference in reaction time between the action Arm Approach and the actions Heading Select and Autothrottle Off, implying that the action Arm Approach could be performed before selecting Heading Select, before switching the Autothrottle off or after switching the Autothrottle off. The last option however, is very unlikely: as explained before the actions Autothrottle off and Autopilot off are asked for in one single call, it is very unlikely that another action will take place in between these two actions. Therefore this last option is not indicated in Table 6. The histograms with reaction times for all the actions on Localizer Intercept Heading are given in Figure 14.

FAF actions

To analyze the pilot's actions at the FAF, all approaches that were flown during the flight simulator experiment were taken into account, except approach 21. During approach 21 the IAS profile required the pilots to select Flaps 20 before the FAF was reached.

The pilots were briefed to select Flaps 20 and Gear Down according to SOPs, which implies that they should be selected when reaching the FAF. As a logical result, the flight data indeed showed that the FAF was the trigger event for these actions. There was a small effect of the FAF altitude (significantly non-normal, $D(330) = 0.231$, $p < .001$) on the reaction times for selecting flaps 20 and gear down (also significantly non-normal, $D(330) = 0.161$, $p < .001$), $\tau = .155$, $p < .001$: when the FAF altitude was lower, the reaction times for the actions at the FAF decreased. The actions flaps 20 and gear down were always asked for in one single call, which is illustrated by a strong correlation between the reaction time for flaps 20 (significantly non-normal $D(165) = 0.146$, $p < .001$) and the reaction time for gear down (significantly non-normal $D(165) = 0.17$, $p < .001$), $r_s = .775$, $p <.001$. Figure 15 also shows that during many flights the actions were performed *before* the FAF was reached.

1,200' actions

All approaches flown during the experiment ($N = 172$) are considered to analyze the pilot's actions at 1,200'. The reaction times for selecting Flaps 25 are given in Figure 16. Again, reaching 1,200' appears to be a correct trigger event for selecting Flaps 25, which could be expected since the pilots were briefed to select Flaps 25 at 1,200' (according to SOPs). Many times Flaps 25 were selected before 1,200' was reached.

Journal Article 4

Factors Influencing Pilot Task Demand Load During RNAV Approaches Based on Theory and Experiments

M.M. Heiligers, Th. van Holten, M. Mulder

Recommended for publication in Journal of Aircraft

Abstract

The goal of this research is to develop a method which can predict pilot task demand load (TDL) during approaches. This paper presents the results of a flight simulator experiment that aimed to identify the factors that influence pilot TDL during an approach with a Boeing 747. Based on the results of this experiment seven guidelines for the design of approaches are presented. When these guidelines are met, pilot TDL during the approach will be acceptable. Additionally, a non-linear Monte Carlo computer simulation as well as a computer simulation based on a point mass model is presented. Both computer simulations can predict, for a given approach, whether the seven guidelines with respect to pilot TDL are met.

1 | Introduction

This research aims to develop a method which predicts the task demand load (TDL) as experienced by the pilot while flying an approach (TDL is defined as the mental workload *imposed* by the system to be controlled or supervised [1]), see also Figure 1. First, this will yield insight in which aspects of an approach actually influence pilot TDL. And second, during the design of approaches this method can be used to rapidly evaluate a potential approach and to 'optimize' an approach with respect to pilot TDL. The TDL is not to be mistaken for the mental workload *experienced* by the human operator, which is referred to as Mental Load.

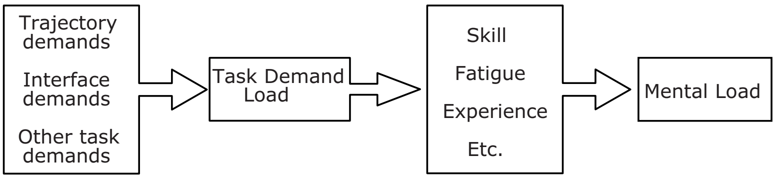


Figure 1 | Difference between Task Demand Load and Workload, adapted from [14].

The rationale within this research is that approaches should be designed such that they can be flown according to Standard Operating Procedures (SOPs) and that a stabilized approach at 1,000ft can be achieved. This is based on the conclusions of the Flight Safety Foundation Approach-and-landing Accident Reduction Task Force [2]. These conclusions read, amongst others, that 'Establishing and adhering to adequate SOPs and flight-crew decision-making processes improve approach-and-landing safety' and that 'Unstabilized and rushed approaches contribute to approach-and-landing accidents'. Therefore, within this research, pilot TDL is predicted for approaches while flying according to SOPs and while aiming to achieve a stabilized approach.

The approaches considered in this research are Area Navigation (RNAV) approaches or, more specifically, RNAV transitions since the final part of the approach is guided by the Instrument Landing System (ILS). The approaches are flown using the Flight Management System (FMS), Autothrottle and Autopilot with Vertical Navigation (VNAV) and Lateral Navigation (LNAV) modes. On Localizer (LOC) intercept heading the autothrottle and autopilot are switched off, and the remainder of the approach is flown using the Flight Director (FD).

Given the level of automation described above, given a certain aircraft with its corresponding SOPs, and given a certain approach, we aim to map pilot TDL and the factors that contribute to pilot TDL. To determine pilot TDL, this research will only take into account the factors that have a direct influence on an approach (see Figure 2). Some of these factors are chosen to be fixed or constant, some factors can be varied and their influence on pilot TDL is investigated. With respect to the meteorological conditions only the wind speed and wind direction are varied. The type of aircraft with its corresponding flight mechanical properties can be varied, but for this paper only a Boeing 747 is considered. Emergency situations such as, for instance, engine fire are not taken into account. The airport infrastructure is defined by assuming an ILS system and a sufficiently long runway, and is therewith a constant factor. The Air Traffic Controller is assumed to be clearly understandable, and is assumed to be giving the standard clearances on time; the variation in this factor lies in the fact that the Air Traffic Controller can order the pilot to fly a different route than originally cleared for. Navigation, communication and Air Traffic Control (ATC) systems are assumed to be working flawlessly, and are therefore not a varying factor. The SOPs that should be adhered to depend on the type of aircraft that is considered, and are regarded a constant factor per aircraft type. The properties of the approach trajectory and its speed and altitude constraints (for instance, the Localizer intercept speed or distance available on Localizer Intercept Heading) are the main focus within this research, and all properties of the approach trajectory are varied.

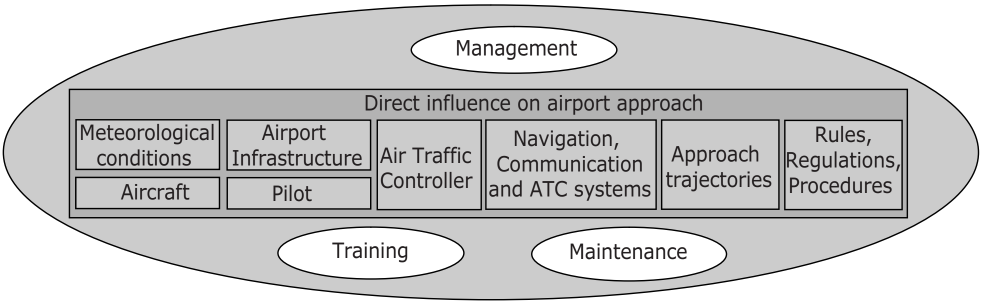


Figure 2 | Direct and indirect factors that influence airport approaches.

To investigate pilot TDL we thus focus on factors that can be described as 'the environment' of the pilot (the direct factors in Figure 2), instead of focusing on the constraints of the pilot himself (like memory capacity, time delay, etc.) [3]. In this respect our work is influenced by the principles of cognitive work analysis [4]. This approach deliberately deviates from the idea behind models such as the Procedure-Oriented Crew model (PROCRU) [5, 6] or the Man-Machine Integrated Design and Analysis System (MIDAS) [7-9] that use human operator models which do focus on the constraints of the human operator. It is anticipated that by focusing on the environment of the pilot instead of on the limitations of the pilot himself much simpler models can be achieved to predict pilot TDL than by using human operator models.

The work by Vormer [10] and the Australian Transport Safety Bureau [11], as well as the guidelines for an optimal RNAV approach [12] were used as a starting point to identify the factors that influence pilot TDL during an RNAV approach. During previous research [13] it was shown through flight simulator experiments and offline computer simulations that some of these factors (based on [10-12]) indeed influenced pilot TDL and that some did not. In [13] also additional factors, not specifically mentioned in [10–12], were found to possibly influence pilot TDL. The results of the research described in [13] form the basis for this paper and are briefly described in the next section.

In this paper a second flight simulator experiment is described, that will study more in depth the factors found in [13] to (possibly) influence pilot TDL in order to arrive at a more complete set of 'environmental' factors that affect pilot TDL relevant for the design of approaches at the end of this paper. Additionally, an extra factor is added to the experiment; this factor is the inclusion of a certain kind of flexibility in an approach in order to give Air Traffic Control (ATC) the possibility to correctly sequence the aircraft. When designing an approach, a choice needs to be made in what way this flexibility will be incorporated. In this research three different options to include such a flexibility are considered, and the effect of each of these three options on pilot TDL is studied.

Finally, this paper presents an offline computer simulation based on the findings of the flight simulator experiments. With this computer simulation it will be possible to rapidly evaluate an approach and to predict the factors that influence pilot TDL according to the results of this experiment; factors such as whether the constraints at the waypoints can be met and whether a stabilized approach can be achieved.

2 | TDL factors considered in this paper

2.1 | Factors related to the approach trajectory

As stated before, this paper can be considered a follow up of the earlier flight simulator experiment described in [13]. The factors influencing pilot TDL during approach relevant for approach design that were studied in [13] are given in the first column of Table 1. For some of these factors it was established in [13] that these factors indeed influenced pilot TDL, these factors are indicated by 'yes' in the third column in Table 1. For some factors it could be established that they did not influence pilot TDL, these factors are denoted by 'no' in Table 1. Then there were factors that did influence pilot TDL, but these factors were already covered by another factor, for instance, the factor "13. Less time available between the Final Approach Fix (FAF, see Figure 1) and 1,000ft" is totally determined by the FAF altitude (factor "12. Lower FAF altitude") when assuming a glideslope of 3 degrees and ignoring the relatively small effect of the airspeed. These factors are indicated by an '-'. Finally, the fourth group of factors, denoted by an 'o' in Table 1, are factors that might possibly have an influence on pilot TDL. These factors were not tested as independent measures in [13], but appeared in the pilots' answers to the questionnaires, and there were no further data to either support or oppose their effect on pilot TDL.

Table 1 shows that some of the factors were also linked to another factor when they were tested in [13]. For example: when testing the effect of factor "8. Less distance available on Localizer (LOC) intercept heading" this was tested as an independent measure between two different approaches, but for these same two approaches there was also a difference in the groundspeed during Localizer capture. The effect on TDL could thus be caused by either the difference in distance on Localizer intercept heading, or by the difference in groundspeed.

The list of factors in Table 1 is considered to be complete; this is based on the observations during the flight tests, the conversations with the pilots, and the fact that (given the type of approaches considered, the level of automation used, the adherence to SOPs, etc.) there are not any other factors relating to the approach trajectory that can be varied or can be taken into account during the design of an approach than those mentioned in Table 1.

Table 1 | Overview of factors that were found to affect pilot TDL in [13] and factors considered in this paper.

#	Factor	Linked factor in [13]	Effect on TDL [13]	Independent measure in this paper
1	More heading changes	Number of waypoints	no	
2	CDA compared to horizontal		no	
3	Larger Heading change towards LOC Intercept Heading		no	
4	Energy rate demand too high	Localizer Groundspeed	yes	*(7)
5	Higher LOC intercept speed	Energy rate demand IF-FAF	yes	*
6	Higher aircraft mass		yes	
7	More altitude steps compared to CDA		no	
8	Less distance available on LOC Intercept Heading	Localizer Groundspeed	yes	*
9	Less time available for actions on LOC intercept leg	Localizer Groundspeed	-(1)	
10	Smaller line-up distance	IF-FAF distance	yes	*
11	Less time available for actions on final	IF-FAF distance	-(2)	
12	Lower FAF altitude		yes	*
13	Less time available between FAF and 1,000'		-(3)	
14	Larger LOC intercept angle		o	*
15	More Tailwind		- (4)	
16	Higher vertical speed during final part of the approach		- (5)	
17	More turbulence		o	
18	Less trackmiles		o	
19	More altitude steps compared to horizontal		o	
20	Higher airspeed during final part of the approach		o	*
21	Less time available during first part of approach		o	
22	Increase in time spent maneuvering		- (6)	-
23	Not stabilized at 1,000ft		o	

yes = proven effect on TDL, no = proven that no effect on TDL, - = effect on TDL already covered by another factor, o = factor that could have an influence on pilot TDL, * = independent measure.

(1) directly related to distance available on LOC intercept heading; (2) directly related to line-up distance; (3) directly related to FAF altitude; (4) already incorporated in energy rate demand; (5) directly related to airspeed on final; (6) directly related to number of heading changes, more altitude steps compared to CDA and more altitude steps compared to horizontal; (7) split into three parts, energy demand ratio for first part of approach, localizer intercept part of approach and final part of approach.

The definitions used in Table 1 are clarified in Figure 1. The analysis of the approach starts at the Initial Approach Fix (IAF) and ends at 1,000ft above airport level. The Intermediate Fix (IF) is the first waypoint on runway heading, and therefore the waypoint where the Localizer is captured. The airspeed at the IF is thus the same as the Localizer intercept speed. The FAF is the waypoint where the aircraft would normally capture the glideslope, when the leg between IF and FAF would be flown horizontally. The entire approach until the FAF can be designed as a horizontal trajectory, or as a trajectory with altitude steps, meaning that the aircraft flies one leg horizontally, then descends, levels off again to fly horizontally, descends again, etcetera. Another option is to design the trajectory as a Continuous Descent Approach (CDA).

When asked to give an opinion about the TDL during the approach, pilots tended to split the approach in multiple parts, giving a separate description for pilot TDL for each of the parts [13]. The same division is adopted for this flight simulator experiment resulting in three parts: 1. the first part of the approach, 2. the Localizer intercept part of the approach, and 3. the final part of the approach (see Figure 3).

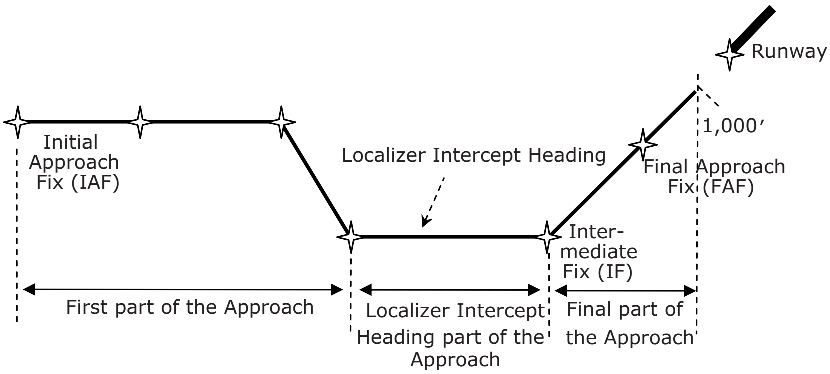


Figure 3 | Division of the approach into three parts

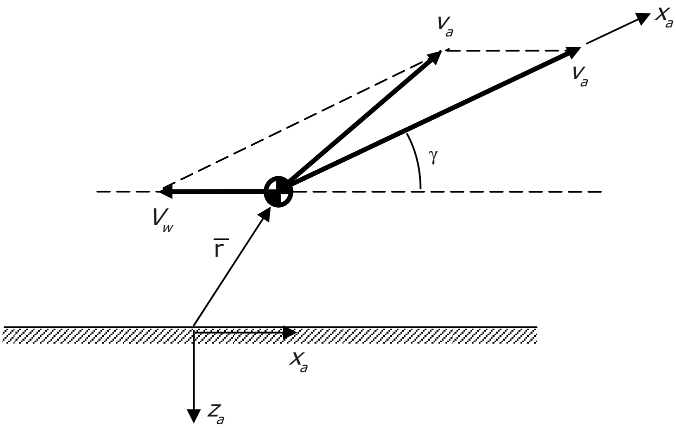


Figure 4 | Definition of axes and velocities (subscript *a* denotes air path reference frame, subscript *g* the geodetic reference frame).

Factor 4 in Table 1 is the "Energy rate demand too high". Assuming small angles, the energy rate demand in the geodetic reference frame can be expressed as (see Figure 4) [13]:

$$\hat{E} = \frac{\frac{(V_a - V_w)}{g}\frac{dV_a}{dx_g} + \frac{(V_a - V_w)}{V_a}\frac{dH}{dx_g}}{\frac{T_0}{W} - \frac{C_D}{C_L}} \tag{1}$$

Where g is the gravitational acceleration, V_a the true airspeed in the air path system, V_w the horizontal wind speed in the geodetic reference frame, T_0 the aircraft flight idle thrust, W the aircraft weight, C_D the aerodynamic drag coefficient, and C_L the aerodynamic lift coefficient. In this equation the numerator represents the energy rate commanded by the trajectory (which will be a negative number in the case of approaches). In the denominator the value for the flight idle thrust T_0 is used, such that the denominator represents the maximum energy rate (decrease) that can be achieved by the aircraft. Note that in the case of approaches this will also be a negative number.

If the value for the energy rate demand becomes larger than 1 this implies that the decrease in energy required by the trajectory can not be met by the energy dissipation of the aircraft, and as a consequence the altitude and/or airspeed constraints at the next waypoint cannot be met. This situation is referred to as 'energy rate demand too high'. The factor 'energy rate demand too high' is, in this paper, also referred to as 'not being able to meet the constraints at the waypoints' since the second expression is clearer when communicating with pilots.

The purpose of the tests presented in this paper now is to get a more comprehensive overview of the factors that influence pilot TDL or, in other words, to 'fill in' the gaps in Table 1 as much as possible. To do so, the factors that were linked to another factor and were proved to have an effect on pilot TDL are tested as separate independent measures in this flight simulator experiment, making sure that these are not linked to any other factor again. The factors that were proved not to have an effect on pilot TDL in Table 1, or the factors that are already represented by another factor are obviously not tested again as independent measures in this flight simulator experiment. Based on the observations during the previous flight simulator tests, and based on conversations with pilots, it is hypothesized that the effect of the energy rate demand will be different for the three approach parts, therefore, factor "4. Energy rate demand too high" is split up for the three approach parts for this flight simulator experiment. The following nine independent measures are now chosen (numbers corresponding to Table 1):

- 4a. Energy rate demand too high in the first part of the approach
- 4b. Energy rate demand too high in the Localizer intercept heading part of the approach
- 4c. Energy rate demand too high in the final part of the approach
- 5. Higher Localizer intercept speed
- 8. Less distance available on Localizer intercept heading
- 12. Lower FAF altitude
- 14. Larger Localizer intercept angle
- 20. Higher airspeed during final part of the approach
- 24. Smaller distance between IF and FAF

In this list factor "24. Smaller distance between IF and FAF", is incorporated as an independent measure instead of factor "10. Smaller line-up distance". The reason for this is the following: the line-up distance is completely determined by the FAF altitude (factor 12) and the distance between IF and FAF, by varying the value of these two underlying factors the value of the line-up distance is also automatically varied. By studying the effect of the FAF altitude and the distance between IF and FAF, the effect of the line-up distance can also be derived.

This choice for the list with independent measures given above implies that the factors 17, 18, 19, 21 and 23 in Table 1 are not tested again as independent measures, although no conclusive decision was obtained about these factors in the previous flight simulator experiment. The reasons for this choice are given below.

The factor "17. More turbulence" can have an effect on pilot TDL during the approach parts considered in this paper, but only in the sense that it can become more difficult to read the flight instruments. This is, however, not something that can be changed by designing an approach differently, and is therefore not considered in this paper.

The factors "18. Less trackmiles", "19. More altitude steps compared to horizontal", and "21. Less time available during first part of approach" all relate to the first part of the approach. Previous research [13] showed that pilots made significantly more comments about the Localizer intercept part of the approach and the final part of the approach than about the first part of the approach. Additionally, the first part of the approach is flown entirely automatically, with very few actions for the pilots to perform. Therefore it is hypothesized that the first part of the approach has the smallest influence on pilot TDL, and the three factors mentioned are not tested. Additionally, the factors "2. CDA compared to horizontal", and "7. More altitude steps compared to CDA" did not influence pilot TDL, therefore the chance that "19. More altitude steps compared to horizontal" would influence pilot TDL while flying automatically is relatively small.

The factor "23. Not stabilized at 1,000ft" is, technically, a result of the factors "4c. Energy rate demand too high in the final part of the approach" and "12. Lower FAF altitude", since these two factors determine whether the airspeed at 1,000ft will be below the value required for a stabilized approach, and whether there is sufficient time available to perform all actions to achieve a stabilized approach. When the values of these two factors change such that a stabilized approach can no longer be achieved, the factor "23. Not stabilized at 1,000ft" will also change. Therefore, for the design of the experiment, the factor "23. Not stabilized at 1,000ft" is not incorporated as a separate independent measure, although its value might vary between approaches.

4

Additionally, the factor "6. Higher aircraft mass" is not included as an independent measure since the effect of aircraft mass is incorporated in the energy rate demand (independent measures 4a, 4b and 4c.

2.2 | Factors related to flexibility in approaches

Next to the factors in Table 1, an additional factor is introduced for the current flight simulator experiment: the inclusion of a certain kind of flexibility in an approach in order to give ATC the possibility to correctly sequence the aircraft. During the design of approaches a choice has to be made as to how to incorporate this flexibility. Based on conversations with ATC, three possible options to include flexibility are considered, see Figure 3.

For each of these options in Figure 5, the aircraft was always initially cleared for the longest route, route C in the examples, this was also the active route in the FMS. During the flight, the pilots could be instructed to follow a different route. For option 1, the pilots were re-cleared for a different, shorter, route before they reached the IAF. This other route (route A or B in Figure 5) was in all cases a published route and already pre-programmed in the FMS.

The Pilot Monitoring (PM) had to select this other route in the FMS and make it the active route. The new route would be flown in VNAV and LNAV mode.

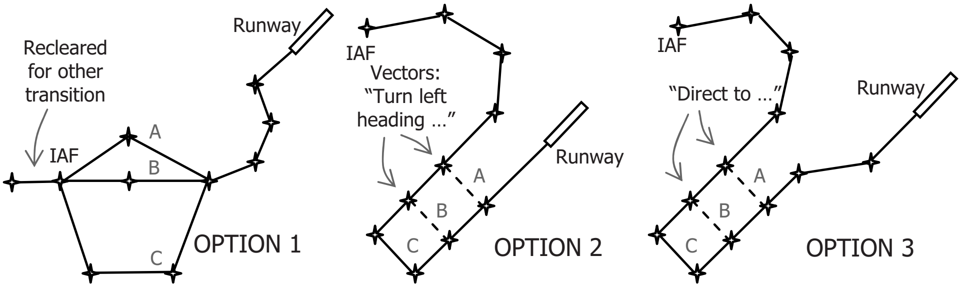


Figure 5 | Three options to include flexibility in approaches

For option 2, the crew was vectored from the downwind leg towards the final leg (via route A or B in Figure 5), on the approach charts there was a note warning to 'expect vectors on final'. When the crew was told, for instance, to 'turn left heading 150, further constraints as published', they had to select the Heading Select mode, thereby deactivating the LNAV mode, and had to switch off the VNAV mode. As a result they could no longer use the route information from the FMS, since the active route in the FMS would still be the initial route C.

In option 3, the crew was told to fly "direct to" a specified waypoint. In all cases, this waypoint was already part of the active route in the FMS, the PM had to select the specified waypoint in the FMS, thereby changing the route in the FMS, and the new route would be flown in VNAV and LNAV mode.

The effects of these three options to include flexibility on pilot TDL are also determined during the current flight simulator experiment. Pilots will fly all three options and will be asked about their preference in a questionnaire. Additionally, the approaches flown during the experiment are designed such that information can be obtained about the factors that increase pilot TDL for each of these approach parts. Factors that are considered are the location of the shortcut (that is, taking route A or B) for flexibility Options 2 and 3, and the fact that due to taking the shortcut the energy rate demand becomes too high. The list with independent measures is therefore augmented with the following independent measures: the location of the shortcut for flexibility Options 2 and 3. Additionally, the independent measures 4a and 4b regarding the energy rate demands are now linked to specific flexibility options, and as a result independent measure 4a is split up in two. Also, the introduction of a shortcut in itself for Option 2 is added to the list of independent measures. For this independent measure, during one approach route B in Option 2 was flown as a shortcut with route C as the initial route, and during another approach, contrary to all other approaches, route B in Option 2 was the published route and the pilot was actually allowed to fly route B. This way it can be checked whether the introduction of a shortcut influences pilot TDL. Due to the limited number of independent measures that could be checked during this experiment this was only done for flexibility Option 2. The above results in a final list of 13 independent measures:

4a1. Energy rate demand too high in the first part of the approach for Option 1
4a2. Energy rate demand too high in the first part of the approach for Option 3
4b. Energy rate demand too high in the Localizer intercept heading part of the approach for Option 2
4c. Energy rate demand too high in the final part of the approach
5. Higher Localizer intercept speed

8. Less distance available on Localizer intercept heading
12. Lower FAF altitude
14. Larger Localizer intercept angle
20. Higher airspeed during final part of the approach
24. Smaller distance between IF and FAF
25. Location of shortcut for flexibility Option 2
26. Location of shortcut for flexibility Option 3
27. Introducing a shortcut for flexibility Option 2

3 | Human-in-the-loop experiment

Nine professional B747 pilots flew 16 custom designed approaches during a human-in-the-loop experiment in a flight simulator.

3.1 | Experiment goal

The experiment was designed to test the influence of the independent measures explained in the previous section on pilot TDL during an approach. Next to testing the influence of these independent measures, additional goals of the experiment were:

- ❖ to test the effect of the three options to include flexibility in an approach on pilot TDL, and to examine the pilots' preference with respect to these three options;
- ❖ to obtain flight data and data about the pilot actions which can later on be used in an offline computer simulation to predict pilot TDL.

3.2 | Method

3.2.1 |Apparatus

The experiment was performed in the six-degree-of-freedom Generic Research Aircraft Cockpit Environment (GRACE) flight simulator at the National Aerospace Laboratory NLR, see Figure 6. During the experiment there was no other traffic, and no emergencies (for instance, engine fire) occurred. All approaches were pre-programmed in the FMS/CDU. The appropriate approach was loaded in the FMS before the start of the run.

Figure 6 | The NLR GRACE Simulator

3.2.2 |Subjects and instructions

Nine B747 pilots participated in the experiment as Pilot Flying (PF), total flight hours ranging from 850 to 16,000 hours (M = 10,205 hours, s = 5,571 hours). Their flight hours on the B747 ranged from 700 to 13,000 hours (M = 5,278 hours, s = 4,309 hours). Seven of these pilots also participated in the earlier flight simulator tests [13] against which the results of this test will be compared. The PF was coupled with a Pilot Monitoring (PM). Four different pilots fulfilled the role of PM, each of them had Multiple Crew Coordination (MCC) experience and held a Commercial Pilot License (CPL). As some of them had no B747 experience they were informed about the B747 SOPs and trained on the simulator in the weeks before the experiment.

The task of the crew was to fly 16 approaches, starting at, or before, the Initial Approach Fix (IAF) and ending at 800' above airport level (AAL). Some crews flew one additional approach. Two weeks before the experiment the pilots (PF) received a briefing by mail. On the day of the experiment they were briefed as well. The pilots were asked to adhere very strictly to SOPs[1], even if they could foresee that by adhering to SOPs they would not meet certain constraints at waypoints or would end up unstabilized at 1,000'. The pilots were briefed that the only situation in which they were allowed to deviate from the flap settings according to SOPs, was when the airspeed required by the approach was lower than the flap setting according to SOPs could accommodate, in that case they were allowed to select the next flap setting. Additionally, they were asked to perform their tasks according to MCC principles and to fly passenger comfort. They were informed that there would be no emergencies during the flight. The pilots were not allowed to use speedbrakes.

3.2.3 | Procedure

Before starting the experiment the pilots could familiarize themselves with GRACE and their task during three to five (depending on the pilot) practice approaches. After that the experiment started. Before every approach the pilots could take as much time as they thought necessary to study the approach and landing charts, to brief the approach and to prepare GRACE for the next approach. The simulation was started when the pilots indicated that they were ready.

After every approach the pilots (PF) were asked to fill in a run questionnaire. Each run questionnaire consisted of three parts: the first part required a rating of the total approach on the Rating Scale Mental Effort (RSME)[15], and required additional RSME ratings for the three individual parts of the approach (see Figure 3), resulting in three RSME sub-ratings per approach. The RSME is constructed according to the 'magnitude estimation' method [16] and the Dutch version of the scale (which was also used for this research) was used and validated in [15, 17]. Though originally intended to measure only one aspect of a task, it is used here to get an indication of the total task because of its simplicity and ease of use when compared to, for example, a NASA TLX rating procedure [18].

When using the RSME pilots obviously base their rating on the mental workload they experienced, see Figure 1. This results in the situation that in order to obtain information about the Task Demand Load, pilots are asked about the Mental Load they experienced during the experiments, unfortunately there is no other way. However, by choosing pilots with different levels of experience, different ages, different levels of fatigue and by testing

1 Pilots were briefed that they were free to select flaps 1 and flaps 5 at a location in the approach they thought appropriate. They were told that they should select flaps 10, Arm the approach, switch to Heading select mode and disconnect the autothrottle and autopilot at Localizer Intercept Heading. Pilots were additionally briefed that they should select flaps 20 and gear down at glideslope intercept, or in the case of a Continous Descent Approach that they should do so at the FAF, and that they should select flaps LAND (flaps 25) at 1200' altitude.

the approaches in random order it is assumed that through the RSME ratings of the pilots a good indication of the TDL can be obtained. Especially, since within this research we are only interested in the *relative* differences in TDL between approaches or approach parts, not in the absolute values of pilot TDL.

In order to compare the pilots' RSME scores for, say, the Localizer intercept part of the approach of approaches 'x' and 'y' (where in approach 'x' there are, for instance, 40 seconds available on Localizer intercept heading, whereas in approach 'y' there are only 20 seconds available) the RSME scores given by a pilot are converted to z-scores per pilot. The reason is the following: one pilot might rate all approaches and approach parts during the experiment high on the RSME scale, whereas another pilot might rate all approaches low on the RSME scale, the absolute values are thus far apart, and difficult to compare. However, we are only interested to find out whether both pilots rated the Localizer intercept part of approach 'x' lower than approach 'y', irrespective of the absolute values. Therefore, the RSME scores of one pilot, are converted to z-scores for this one pilot using all RSME scores given by this pilot.

Table 2 | Independent measures, approaches and hypothesized effect on pilot TDL (+ increase in TDL, = no influence on TDL)

Factor #	Approach pair	Independent measure	Linked factor within relevant approach part	Approaches	Hypothesized TDL effect
4a2	A	Energy rate demand too high in first part of approach for Option 3	·	4 & 7, 3 & 9	=
4b	B	Energy rate demand too high in Localizer int. part for Option 2	·	14 & 15	+
4a1	C	Energy rate demand too high in first part of approach for Option 1	·	1 & 6, 2 & 8	=
26	D	Location of shortcut for flexibility Option 3	·	3 & 4	=
25	E	Location of shortcut for flexibility Option 2	Line-up distance, IF-FAF distance	13 & 14	=
27	F	Introducing a shortcut for flexibility Option 2	·	14 & 19	=
5	G	Higher Localizer intercept speed	·	9 & 10	+
14	H	Larger Localizer intercept angle	·	4 & 9	+
8	I	Less distance available on Localizer Intercept Heading	·	2 & 16	+
20	J	Higher airspeed during final part of approach	·	9 & 10	+
4c	K	Energy rate demand too high in final part of approach	Stabilized at 1,000' actual LOC int speed actual speed at FAF	2 & 6	+
24& 12	L	Smaller distance between IF and FAF and Lower FAF altitude	·	2, 3, 5 & 9	+

4

This is repeated for all pilots participating in the experiment. The resulting RSME z-scores that pilots gave to the Localizer intercept part of approach 'x' and of approach 'y' can now be compared to analyze whether pilots rated the Localizer intercept part of approach 'x' lower than of approach 'y'.

The second part of the run questionnaire contained two questions asking the pilot's opinion on whether the pilot would have adhered to SOPs during real flight, and whether the pilot would

Table 3 | Characteristics of the approaches.

Approach number	**Flexibility option (see Figure 3)**	**Route (see Figure 5)**	**Energy rate demand too high in first part of approach**	**Energy rate demand too high in LOC intercept part**	**Energy rate demand too high in final part of approach**	**Stabilized at 1,000ft[4]**	**IAS loc intercept published [knots]**	**IAS loc intercept actual [knots]**	**LOC intercept angle [deg]**	**Distance available on LOC intercept heading [nm]**	**FAF altitude [ft]**	**Published airspeed during final part of approach [knots][5]**	**Actual airspeed during final part of approach [knots][6]**	**IF-FAF distance [nm][7]**	**lineup distance [nm]**
1	1	B	no	yes	no	yes	190	190	30	4	3,000	190	190	1.3	10.7
2	1	C	yes[1]	no	no	yes	200	200	30	4	1,600	170	190	5.7	10.7
3	3	B	no	no	no	yes	180	180	34	4	1,600	170	170	1.3	6.3
4	3	A	no	no	no	yes	200	200	51	4	3,000	190	190	5.3	14.7
5	1	B	no	no	yes	yes	200	200	30	4	3,000	190	200	1.3	10.7
6	1	B	yes	yes	yes	no	200	220	30	4	1,600	170	220	5.7	10.7
7	3	A	yes	yes	yes	no	200	200	34	4	1,600	190	190	1.3	6.3
8	1	C	yes[2]	no	no	yes	200	200	30	4	1,600	190	190	4.4	9.4
9	3	B	yes	no	no	yes	200	200	38	4	3,000	190	190	5.3	14.7
10	-	-	no	no	no	yes	230	230	38	4	3,000	210	210	5.3	14.7
11	-	-	yes	yes	yes	no	180	210	35	3	1,270	160	200	1	5
12	-	-	no	no	no	yes	180	180	60	5	2,000	180	180	3	9.1
13	2	A	no	no	no	yes	200	200	41	4	2,000	200	200	2	8.3
14	2	B	no	no	no	yes	200	200	41	4	2,000	200	200	6	12.3
15	2	B	no	yes	no	yes	200	210	41	4	2,000	200	200	2	8.3
16	-	-	no	no	no	yes	200	200	30	2	1,600	170	180	5.7	10.7
19	2[1]	B[1]	no	no	no	yes	200	200	41	4	2,000	200	200	2	8.3

[1] Constraints could not be met at three waypoints; [2] Constraints could not be met at one waypoint; [3] For approach 19 route B was the published route (not route C as was the case for all other approaches) and the pilots were also allowed to fly route B; [4] Stabilized at 1,000ft is not an independent measure, but is, as indicated earlier, a result of the factors "4c. Energy rate demand too high in the final part of the approach" and "12. Lower FAF altitude"; [5] Based on the published speed at the FAF; [6] Based on the actual speed at the FAF; [7] The distance between IF and FAF is not an independent measure, but is a result of the factors "12. Lower FAF altitude" and "10. Smaller line-up distance"

have used speedbrakes during real flight. The third part of the run questionnaire consisted of specific closed format questions per approach part regarding the factors hypothesized to influence pilot TDL in that specific approach part.

At the end of the day, after all approaches were flown, the pilots (PF) filled in an end of day questionnaire. The first part of the end of day questionnaire regarded the realism of GRACE and the realism of the experiment as a whole. The second part contained general questions about factors that might possibly influence pilot TDL during approach.

3.2.4 | Independent measures, approaches and hypotheses

The approaches (cases) were designed in pairs to test different independent measures, see Table 2 and Table 3. All approaches were designed for Amsterdam Airport Schiphol Runway 06. Within an approach pair the independent measure was the only changing variable for the relevant approach part, except when the independent measure was inevitably linked to another factor. It was ensured that changing an independent measure in one approach part did not influence the subsequent approach part(s), except when this continuation of an effect in subsequent approach part(s) was deliberately introduced to investigate its effect on pilot TDL. The hypothesized effects of the independent measures on pilot TDL are indicated in the last column of Table 2.

4 | Results

The experiment results consist of subjective data (collected using run questionnaires after each approach and an end of day questionnaire) and of objective flight data, the results are presented below.

4.1 | Run questionnaires

The run questionnaires were filled in after each approach. The number of pilots that flew each approach is given in Table 4. As explained before, each run questionnaire consisted of three parts. The results of each of these parts of the run questionnaires are discussed below.

4

Table 4 | Approaches and number of pilots that flew each approach (total $N = 150$), and number of pilots that gave RSME ratings for each (part of the) approach.

Approach #	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**19**
Nr. of pilots	9	9	9	9	9	9	9	9	8	9	9	9	9	9	9	9	7
Nr. of total RSME ratings	9	9	9	9	9	9	9	9	8	9	9	9	9	9	9	9	7
Nr. of RSME ratings first part of approach	9	9	9	9	9	9	9	9	7	9	9	9	8	9	9	9	7*
Nr. of RSME ratings for LOC Intercept part of approach	9*	9	8	9	9	9	9	8	8	9	9	9	7	8	7	9*	7
Nr. of RSME ratings for final part of approach	9	9	7	8	8	8	9	8	7	9	8	8	8	9	8	9	7

* = RSME z-scores significantly non-normal ($p < .05$) according to Kolmogorov-Smirnov test.

4.1.1 | RSME scores for independent measures

The pilots were asked to rate each approach on an RSME scale. As explained before, each pilot gave four RSME ratings per approach: one rating for the total approach, and three sub-ratings for each of the three approach parts. Unfortunately, pilots sometimes forgot to give a rating for one of the RSME (sub-) ratings, see Table 4.

To explore the relation between the three sub-ratings and the RSME rating for the total approach a regression analysis is performed with the RSME rating for the total approach as dependent measure, and the three separate sub-ratings as predictors, see Table 5. The

Table 5 | The linear regression model for the dependent measure RSME score for total approach, all Variance Inflation Factors (VIFs) below 1.5, average VIF equal to 1.46.

	B	***SE B***	β	R^2	**Adjusted** R^2
Constant	4.954	2.848			
RSME score for first part of approach	0.226	0.052	.261*		
RSME score for Localizer Intercept Heading part of approach	0.361	0.064	.342*		
RSME score for final part of approach	0.374	0.054	.414*	.683	.676

* p < .001

regression model shows that the RSME score for the final part of the approach has the largest influence on the RSME score for the total approach (β = .414), followed by the RSME score for the Localizer Intercept Heading part (β = .342), and that the RMSE score for the first part of the approach has the smallest influence (β = .261). Here it should be noted that all four RSME ratings were given at the end of the approach, and as a result the final part of the approach would be 'fresh in memory', thereby possibly resulting in a larger influence on the RSME rating of the total approach.

The RSME z-scores calculated from the pilots RSME ratings for the relevant approach part were compared for each independent measure between the approaches belonging to this independent measure. To test whether the hypothesized effect in Table 2 actually occurred, t-tests for the normally distributed RSME z-scores, and Wilcoxon signed rank tests for the non-normally distributed RSME z-scores were used.

Due to the total number of tests that is performed on the data set resulting from the flight simulator experiment, when assuming an original p-value of .05 a corrected p-value of .001 should be used (when applying a Bonferroni correction) as criterion for significance in order to control the Type I error rate (this error represents the situation that an effect is found using a statistical test, whereas in reality there is no effect). However, by correcting the p-value to .001 statistical power is lost (meaning that the probability of rejecting an effect that does actually exist is increased (a Type II error)). Therefore, it is decided to use a significance value of .05 for the comparison of RSME z-scores, while keeping in mind that this inflates the Type I error, and that we thus might classify factors as influencing pilot TDL while in reality they do not. For the purpose of this research however, this is deemed more favorable than discarding factors that actually *do* influence pilot TDL. In any regard, due to the small sample size, the results of the tests given below should only be considered as an indication of possible effects. The results of the comparison of RSME z-scores are given in Table 6 and are discussed below.

It was found when analyzing the RSME z-scores for the appropriate part of the approach that not meeting the constraints in the first part of the approach only has a significant effect when the consequences of this high energy rate demand continue into the Localizer intercept heading part of the approach or the final part of the approach, as was the case for approaches 6 and 7 (in approach pairs A and C). When the consequences of the high energy rate demand only remain within the first part of the approach (as was the case for approach 9, in approach pair A) it does not have an effect on the RSME z-scores. On average, the number of waypoints at which the constraints are not met in the first part of the approach does also not influence the RSME z-scores.

Table 6 | Results of the comparisons of RSME z-scores

Factor #	APP-pair	Independent measure	APPR	*df*	*t*	*r*	TDL Effect	Hypoth. TDL effect
4a2	A	Energy rate demand too high in first part of approach for Option 3	4 & 7 3 & 9	8 6	-2.86* -1.14 *ns*	.71 .42	+(8) =	=
4b	B	Energy rate demand too high in Localizer int. part for Option 2	14 & 15	5(3) 6(4)	-1.54 *ns* -3.02*	.56 .78	= +(9)	+
4a1	C	Energy rate demand too high in first part of approach for Option 1	1 & 6(1) 2 & 8(2)	8 8	-2.02* 0.309 *ns*	.58 .11	+ =	=
26	D	Location of shortcut for flexibility Option 3	3 & 4	8	-7.33 *ns*	.26	=	=
25	E	Location of shortcut for flexibility Option 2	13 & 14	4	3.339*	.86	+	=
27	F	Introducing a shortcut for flexibility Option 2	14 & 19	6	-.024 *ns*	.01	=	=
5	G	Higher Localizer intercept speed	9 & 10	7	-3.01*	.75	+	+
14	H	Larger Localizer intercept angle	4 & 9	7	-3.03*	.75	+	+
8	I	Less distance available on Localizer Intercept Heading	2 & 16		*T* = 1.5*(5)	.43	+	+
20	J	Higher airspeed during final part of approach	9 & 10	6	-3.45*	.82	+	+
4c	K	Energy rate demand too high in final part of approach	2 & 6	7	-4.11**	.84	+	+
24&12	L	Smaller distance between IF and FAF and Lower FAF altitude	2, 3, 5 & 9		*F*(1,5) = 4.66*(6) *F*(1,5) = 16.65*(7)	.69 .88	+(6) +(7)	+

* = p < .05, ** = p < .01; (1) For approach 1, the constraints could be met, for approach 6 the constraints could not be met at the waypoints; (2) For approach 2 the constraints could not be met at three waypoints, for approach 8 the constraints could not be met at one waypoint; (3) Test for Localizer intercept part of approach; (4) Test for final part of approach; (5) one-tailed; (6) Effect of lower FAF altitude; (7) Effect of smaller distance between IF and FAF; (8) Only a significant effect when the effect of not meeting the constraints continues into the Localizer intercept part or final part of the approach; (9) Effect on RSME ratings of final part of approach, no effect found in RSME ratings for Localizer intercept part of approach

When the energy rate demand during the Localizer intercept part is larger than 1 (approach pair B), the RSME z-scores for the Localizer intercept part are not significantly higher. For the final part of the approach, however, the RSME z-scores *are* significantly higher, even though the final part of the approach is exactly identical.

When the constraints at the waypoints can be met (the energy rate demand remains below 1) the location of the shortcut (whether the aircraft is guided along route A or B in Figure 3) for flexibility Option 3 does not have a significant effect on the pilots' RSME z-scores (approach pair D). For flexibility Option 2, however, the location of the shortcut *does* influence the RSME z-scores, in the sense that when the shortcut results in a smaller line-up distance, the RSME z-scores increase (approach pair E). The fact that a shortcut is introduced in itself for flexibility Option 2 (approach pair F) does not significantly influence the pilots' RSME z-scores. Vectoring the aircraft off the published route using flexibility Option 2 thus has no effect on the RSME z-scores.

Both a higher Localizer intercept speed (approach pair G) and a larger Localizer intercept angle (approach pair H) resulted, on average, in higher RSME z-scores. The independent measure "Less distance available on Localizer intercept heading" (approach pair I)also had an effect on the RSME z-scores: when the distance available became smaller, the RSME z-scores increased.

A higher airspeed during the final part of the approach resulted in significantly higher RSME z-scores (approach pair J). It should be noted that for both approaches (9 and 10) the energy rate demand was well below 1, and that a stabilized approach at 1,000' could be achieved. When the energy rate demand becomes too high during the final part of the approach and as a consequence the constraints at the waypoints can no longer be met, the pilots' RSME z-scores (also) increase significantly.

Finally, a lower FAF altitude and a smaller distance available between IF and FAF were also found to significantly increase the pilots' RSME z-scores. The final parts of the approach for approaches 2, 3, 5 and 9 were set-up as a two-level factorial design for the independent measures "12. Lower FAF altitude" and "24. Smaller distance available between IF and FAF". The RSME z-scores for these approaches were compared using a factorial repeated measures ANOVA. Since there were only two levels per independent measure, sphericity was not an issue. There was no significant interaction effect, meaning that the distance between IF and FAF had the same effect on the RSME z-scores, independent of the FAF altitude it was coupled to: a smaller distance between IF and FAF would always result in higher RSME z-scores than a larger distance between IF and FAF both when coupled to a high FAF altitude or a low FAF altitude. Since the combination of FAF altitude and distance between IF and FAF completely defines the line-up distance, it can be concluded that as a result a larger line-up distance decreases pilot TDL[2]. Since when the line-up distance increases, this means that either the FAF altitude or the distance between IF and FAF increases, both of which result in lower pilot TDL.

4.1.2 | Questions regarding factors influencing pilot TDL

The second part of the run questionnaire consisted of closed format questions targeting those elements of each approach part that might influence pilot TDL. An example of such a question is: 'The FAF altitude was…', response options: Very low, Low, Normal, High, Very high. The accompanying question then was: 'Because of this I found this part of the approach…', response options: Very difficult, Difficult, No influence, Easy, Very easy. All questions always consisted of the pilot's perception of an objective fact, and the corresponding effect on the difficulty. During the briefing pilots were explained that the second question with response options Very difficult – Very easy, should be answered only in relation to their perception of the objective fact, irrespective of the RSME rat0ing they gave for the entire approach part. For example: for a particular approach a pilot might have rated the final part of the approach high on the RSME scale, indicating that this part as a whole was relatively difficult to fly. The pilot might have rated the FAF altitude as 'high', the question regarding the difficulty should now be read as: "because of this low FAF altitude, I would find this part of the approach (irrespective of what I actually found of this part of the approach)…", answer: 'easy'. The answer to this question might thus be 'easy' whereas the actual RSME rating of this part of the approach indicated that the approach was difficult.

2 For this Ph.D. thesis a more elaborate explanation of the RSME z-scores is added in Appendix A to Journal Article 4.

The questions for each of the approach parts are given in Table 7, for each and everyone of these questions there was an accompanying question 'Because of this I found this part of the approach…', response options: Very difficult - Very easy. For these questions more factors from Table 1 could be incorporated than just the independent measures used during the experiment (which were limited by the number of approaches that could be flown during the experiment). The first column indicates to which factors in Table 1 or which independent measures the questions relate. All independent measures are incorporated in Table 7, except independent measures 23, 25 and 26 since these are not applicable to all approaches.

Table 7 | Closed format questions per approach part (translated from original questions in Dutch).

Factor #	Question	Response options				
	First part of the approach					
1	The number of waypoints was	Very large	Large	Neutral	Small	Very small
1	The number of heading changes was	Very large	Large	Neutral	Small	Very small
18	The amount of trackmiles was	Very small	Small	Neutral	Large	Very Large
21	The time available to perform all actions was	Very short	Short	Neutral	Long	Very long
4a	The altitude and airspeed constraints I could	Not meet at all	Not meet	Just meet	Meet	Easily meet
	Localizer intercept part of the approach					
5	The IAS during LOC intercept was	Very high	High	Normal	Low	Very low
14	The LOC intercept angle was	Very large	Large	Normal	Small	Very small
8	Distance on LOC intercept heading was	Very small	Small	Normal	Large	Very large
9	The time available to perform all actions was	Very short	Short	Neutral	Long	Very long
4b	The altitude and airspeed constraints I could	Not meet at all	Not meet	Just meet	Meet	Easily meet
	Final part of the approach					
12	The FAF altitude was	Very low	Low	Normal	High	Very high
20	The IAS at the FAF was	Very high	High	Normal	Low	Very low
24	The distance between IF and FAF was	Very small	Small	Normal	Large	Very large
10	The line-up distance was	Very small	Small	Normal	Large	Very large
13	The time available to perform all actions was	Very short	Short	Neutral	Long	Very long
4c	The altitude and airspeed constraints I could	Not meet at all	Not meet	Just meet	Meet	Easily meet
23	At 1,000ft I was	Not stabilized at all	Not stabilized	Just stabilized	Stabilized	Easily stabilized
	Coding for correlations and regression:	1	2	3	4	5

4

Four sets of data now result from the run questionnaire (Table 8): first the objective fact (for instance the actual FAF altitude of 3,000'), second there is the pilot's perception of this objective fact (the pilot might classify the actual FAF altitude of 3,000' as 'High'), third there is the effect of this fact on the difficulty experienced by the pilot (for instance: because of this [the 'High' FAF altitude] I found this part of the approach 'Very easy'), and fourth there is the pilot's 'umbrella' RSME rating for this part of the approach which might indicate that this part of the approach was difficult. The question of course is how these four sets of data relate to each other.

Table 8 | Four sets of data resulting from the run questionnaire

Data set		Example for FAF altitude
1	The objective fact	3,000'
2	The pilot's perception of this objective fact	'High' (options Very low – Very high)
3	The effect on the difficulty experienced by the pilot	'Very easy' (options Very difficult – Very easy)
4	The pilot's 'overall RSME rating for this part of the approach	

To find an answer to this question, correlations and regression models were used. For this purpose the response options for all questions have been coded from 1 – 5, and are regarded interval data, see Table 7. Although there is much controversy about whether these response options can be considered ordinal or interval data [19, 20], it is, in this case, deemed appropriate to treat the data as interval scale because the response options were arranged horizontally and were equally spaced apart, and the verbal labels connoted more-or-less evenly-spaced gradations, most of them symmetrical about a neutral middle. Because of the many tied ranks Kendall's tau-b was used for the correlations.

First, the relation between the objective facts[3] (e.g. the FAF altitude of 3,000') and the pilot's perception of this fact (e.g. 'High') was considered (data sets 1 and 2), to check whether pilots would consistently have the same opinion about the same objective fact. The correlation coefficients for all questions in Table 7 are given in the first column of Table 9. For the FAF altitude the correlation is large, for all other factors, however, the effect sizes are merely medium, indicating that pilots were not consistent in their opinions. Analysis of the data showed that these low correlation coefficients were caused by variability in the answers *per pilot*: the same pilot would rate a FAF altitude of 3,000' as 'High' in one run questionnaire, and as 'Low' in another run questionnaire. These inconsistencies might partially be influenced by the design of the approach that was previously flown. For instance: a FAF altitude of 1,600' might be rated 'Low' after an approach with a FAF altitude of 3,000', and might be rated 'High' after an approach with a FAF altitude of 1,300'. Concluding it can be stated that the pilot's perception did not correlate as much as expected with the objective facts.

The next correlation focuses on the relation between the pilot's perception (irrespective of the fact whether this corresponds to reality), and the reported influence on the difficulty (data sets 2 and 3). For instance: when pilots answered that the FAF altitude was 'High', did they then consistently answer that because of this 'High' FAF altitude they found this part of the approach 'Very easy'? These correlations are given in the second column of Table 9 for all factors. It can be seen that the majority of these coefficients represent large effects, indicating that pilots were consistent in coupling the same effect on the difficulty to the

3 The objective fact 'The altitude and airspeed constraints I could…' was coded as follows: 0 if the constraints were not met, 1 if the constraints were met, and 2 if the constraints were easily met. All other objective facts were coded using their appropriate numerical value.

same subjective opinion. A smaller correlation coefficient would indicate that, for instance, a 'High' FAF altitude would be rated 'Easy' in one run questionnaire and 'Difficult' in another run questionnaire, or, that the FAF altitude would be classified as 'Very low' or 'Very high' but that for both classifications pilots would indicate that it had 'No influence' on the difficulty. The only correlation with medium effect in the second column of Table 9 is for the number of waypoints in the first part of the approach, it can therefore be concluded that, of all factors in Table 9, this probably has the smallest effect on the difficulty, or, pilot TDL.

Table 9 | Correlation coefficients (Kendall's tau) for the pilot's perception on the one hand (data set 2) and the objective fact (data set 1) and effect on difficulty (data set 3) on the other hand.

Factor #	Pilot's opinion	Objective fact	Effect on difficulty
	First part of the approach		
1	The number of waypoints	-.42***	.34***
1	The number of heading changes	-.39***	.60***
18	The amount of trackmiles	.21**	.72***
21	The time available to perform all actions	.01 (*n.s.*)	.75***
4a	The altitude and airspeed constraints I could	.54***	.72***
	Localizer intercept part of the approach		
5	The actual IAS during LOC intercept	-.28***	.69***
14	The LOC intercept angle	-.35***	.71***
8	Distance on LOC intercept heading	.32***	.77***
9	The time available to perform all actions	.19**	.71***
4b	The altitude and airspeed constraints I could	.43***	.72***
	Final part of the approach		
12	The FAF altitude	.59***	.70***
20	The actual IAS at the FAF	.28***	-.62***
24	The distance between IF and FAF	.34***	.62***
10	The line-up distance	.45***	.76***
13	The time available to perform all actions	.37***	.85***
4c	The altitude and airspeed constraints I could	.50***	.75***
23	Stabilized/not stabilized at 1,000'	.78***	.71***

** $p < .01$, *** $p < .001$

The final relation that is studied is the relation between data sets 3 and 4, in order to find out which of the factors in Table 7 influence the pilot's overall RSME rating for that part of the approach. For this a linear regression analysis is performed according to the guidelines in [21] for regression analyses of repeated measures data. These guidelines imply that both the subjects (pilots) need to be coded as dummy variables [22, 23], as well as the subject x predictor interactions, and that each predictor variable is tested for significance against the corresponding subject x predictor interaction term [21]. Since three approach parts are considered, three separate regression analyses are performed: one for the set of RSME z-scores for each approach part. The independent measure thus is the RSME z-score for the approach part considered (data set 4), and the predictors for each approach part are the effect on the difficulty of the items given in Table 7 (data set 3), coded 1 – 5, corresponding to very difficult – very easy. The results are described below.

Table 10 | The resulting regression model for the dependent measure RSME z-score for the first part of the approach. Predictors are the effect on the difficulty of the factors (data level 3).

	df	*SS*	*B*	*SE B*	β	*F*	ΔR^2	R^2	*Adj.* R^2
Total	145	132.54							
Between subjects	8	12.01					.091	.091	.037
Within subjects	137	120.53							
Regression	1	66.14							
Meet constraints (4c)	1	66.14	-0.75	0.058	-0.736	123.9*	.499	.590	.562
Subject x Predictor									
Pilot x Meet constraints	8	4.26					.032	.622	.572

* *p* < .0001

Table 11 | The resulting regression model for the dependent measure RSME z-score for the Localizer intercept part of the approach. Predictors are the effect on the difficulty of the factors (data level 3).

	df	*SS*	*B*	*SE B*	β	*F*	ΔR^2	R^2	*Adj.* R^2
Total	136	104.72							
Between subjects	8	10.78					.103	.103	.047
Within subjects	128	93.95							
Regression	3	48.41							
IAS during LOC int. (5)	1	1.58	-0.246	0.118	-0.160	1.66 $^-$			
Time available (13)	1	12.54	-0.655	0.112	-0.438	56.77****			
Meet constraints (4b)	1	3.90	-0.275	0.084	-0.263	3.89 $^-$	.462	.565	.527
Subject x Predictor									
Pilot x IAS during LOC int.	8	7.60							
Pilot x Time available	8	1.77							
Pilot x Meet constraints	8	8.02					.123	.688	.580

$^-$ = *n.s.*, **** *p* < .0001

For the first part of the approach, the best result is obtained when only the indicated effect on the difficulty of the predictor 'The altitude and airspeed constraints I could…' is used (see Table 10). All other four predictors from Table 7 are discarded. For the Localizer intercept part of the approach, the regression model incorporating the effect on the difficulty of the predictors 'The time available to perform all actions', 'The altitude and airspeed constraints I could…' and 'The actual IAS during LOC intercept' yielded the best results, see Table 11. When calculating the significance according to [21], however, it appears that only the effect on difficulty of the factor 'Time available to perform all actions' has a significant effect on the RSME z-score for this part of the approach. To conclude, the regression model for the final part of the approach is given in Table 12. It contains the effect on the difficulty of the predictors 'The actual IAS at the FAF', 'The distance between IF and FAF', 'The altitude and airspeed constraints I could…' and 'Stabilized/not stabilized at 1,000'.

Interesting to note is the fact that whether an approach is stabilized at 1,000' is not a factor in itself but a result of other factors, as explained before. Still, Table 12 shows that the fact whether the approach was stabilized had a larger influence on the RSME z-score, than the underlying factors that actually cause the approach to be unstabilized. For the pilots, the result thus seems to be more important than the cause itself.

Summarizing, it can be stated that the overall RSME z-score for the first part of the approach is mostly influenced by the fact whether or not the constraints can be met in that part of the approach (independent measure 4c). The RSME z-score for the Localizer intercept part of the approach is mainly influenced by the time available on Localizer intercept heading (linked to independent measure 8). And the overall RSME z-score for the final part of the approach is mostly affected by whether or not a stabilized approach can be achieved, and to a lesser extent by the IAS at the FAF (independent measure 20), the distance between the IF and FAF (independent measure 24), and the fact whether or not the constraints can be met (independent measure 4c).

Table 12 | The resulting regression model for the dependent measure RSME z-score for the final part of the approach. Predictors are the effect on the difficulty of the factors (data level 3).

	df	*SS*	*B*	*SE B*	*β*	*F*	*ΔR²*	*R²*	*Adj. R²*
Total	130	148.96							
Between subjects	8	4.35					.029	.029	-.035
Within subjects	122	144.61							
Regression	4	98.59							
Stabilized 1,000′ (23)	1	12.10	-0.544	0.100	-0.444	63.68****			
Meet constraints (4c)	1	3.58	-0.310	0.105	-0.260	8.89*			
IAS at FAF (20)	1	2.44	-0.267	0.109	-0.170	8.27*			
IF-FAF distance (24)	1	1.87	-0.195	0.091	-0.134	7.96*	.633	.662	.629
Subject x Predictor									
Pilot x Stabilized 1,000′	8	1.52							
Pilot x Meet constraints	8	3.22							
Pilot x IAS at FAF	8	2.36							
Pilot x IF-FAF distance	8	1.88					.080	.742	.621

* $p < .05$, **** $p < .0001$

4.1.3 |Stabilized at 1,000′ and adhere to SOPs

To complete the run questionnaire the pilots were asked whether they would have flown this approach according to SOPs in reality (response options “yes” or “no”), and whether they would have used speedbrakes in reality (again, response options “yes” or “no”). In agreement with the results in [13], there was a positive relationship between the pilot indicating the approach was stabilized, and the same pilot indicating that he would adhere to SOPs in reality when flying that approach, $\tau = .36$, $p < .001$. A negative relation was found between the pilot indicating the approach was stabilized, and the same pilot indicating that he would have used speedbrakes during the flight $\tau = -.46$, $p < .001$, indicating that the pilot would use speedbrakes when the approach was *not* stabilized, which would be expected.

4.2 | End of day questionnaire

At the end of the day the pilots were given a questionnaire. The first part of the end of day questionnaire regarded the reality of the GRACE flight simulator and the reality of the experiment as a whole. The second part concerned the factors that influence the difficulty of flying an approach, and the third and final part contained questions about the three options to include flexibility in the approach.

4.2.1 | Reality of GRACE flight simulator and reality of experiment

The pilots were asked their opinion regarding the reality of the flight simulator and experiment as a whole. The Flight Director and the Autothrottle were regarded to be unrealistic by some of the pilots. When asked, three pilots (out of nine) answered that due to this pilot TDL increased. All pilots indicated that ATC provided them with sufficient information regarding the approach, and that the communication between ATC and their flight was "Average" to "Very realistic". Nevertheless, 1 pilot (out of 9) reported that the TDL was influenced by the fact that the contact with ATC differed from reality. The pilots answered that the communication between PF and PM was "Average" to "Realistic", and that the way the PM performed his tasks was "Average" to "Very realistic". All pilots answered that they had sufficient time to prepare for the approach, and reported that the approach charts provided "Sufficient" or "More than sufficient" information. It should be noted that although some of the aspects mentioned above deviated from reality and therefore might have had an influence on the difficulty of flying the approach, all these aspects were the same for all pilots and remained constant during all approaches.

4.2.2 |Factors influencing the difficulty of an approach

In the second part of the end of day questionnaire, pilots were asked to answer questions with respect to factors that might influence pilot TDL. An example of such a question is: "Considering an RNAV approach: when the FAF altitude becomes lower, the approach becomes:" provided response options: A lot easier, Easier, No influence, More difficult, A lot more difficult. The pilots' responses are summarized in Figure 7.

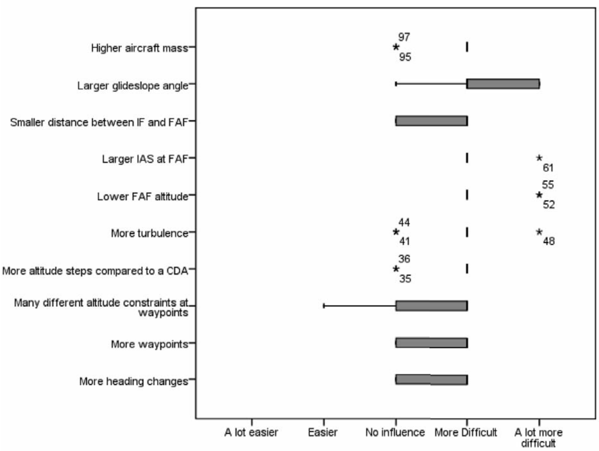


Figure 7 | Pilots answers (N=9) regarding the influence of the factors mentioned on the left on pilot TDL while flying an approach

Pilots were additionally asked to explain why a higher aircraft mass would have an effect on the difficulty. The answers were: larger turning radius (N = 1), due to higher airspeed on final sinkrate might become too high resulting in an unstabilized approach (N = 2), more difficult too dissipate energy (N = 4). This question was asked to verify whether there were any non-flight mechanical factors resulting from a higher aircraft mass that would influence pilot TDL. The answers indicate that this is not the case.

Table 13 | Pilots' answers (N=9) regarding the influence of meeting the constraints at waypoints and achieving a stabilized approach on pilot TDL.

Question	Yes	No
If you *do not* meet the altitude and speed constraints at the waypoints during an RNAV approach, but you *do* meet all the requirements for a stabilized approach at 1,000', would you classify this approach as 'difficult'?"	6	3
If you *do* meet the altitude and speed constraints at the waypoints during an RNAV approach, but you *do not* meet the requirements for a stabilized approach at 1,000', would you classify this approach as 'difficult'?	8	1

The final two closed format questions regarding factors that influence an RNAV approach were formulated in a different way, these are given in Table 13.

Following the closed format questions, the pilots were asked whether there were any other factors that influence the difficulty of flying an RNAV approach (open format). The answers given were: fatigue, time difference, 'at' constraints instead of 'at or above' constraints, tailwind, too many waypoints with speed and altitude constraints (pilot would prefer a Continuous Descent Approach).

To conclude the end of day questionnaire, pilots were asked about their preference regarding the three options to include flexibility in the approach. One pilot preferred Option 1, five pilots preferred Option 2 and three pilots preferred Option 3. The reasons given to choose option 2 are: good overview (w.r.t. other traffic, trackmiles to go and altitude (low or high on glideslope)), resembles the well-known standard circuit, least amount of actions required, lowest workload, lowest level of automation required (using heading select and flight level change), no FMS re-programming required. Reasons to choose option 3 were: least amount of radio communication required, FMS information regarding meeting the constraints is retained, a 'Direct to' does not require major adjustments in the FMS, the pilot is in control of the situation one doesn't have to wait for a vector as in option 2.

Additionally, they were asked what their preference would be if they would only consider the adjustments that are necessary in the FMS. In this case pilots only preferred options 2 and 3 (five pilots preferred option 2, and four pilots chose option 3). As motivation for choosing option 2 pilots reported: attention is not diverted due to re-programming the FMS, more 'in control' when flying heading select and flight level change modes, less heads-down time during critical flight phase, least amount of FMS actions required. Reasons to opt for option 3 were: VNAV and LNAV information is retained, altitude constraints remain active.

Finally, pilots were asked about the effects of a change in the approach on the difficulty of flying that approach for all three flexibility options. A distinction was made between the situation in which, due to the commanded change in the approach, the constraints at the waypoints could no longer be met, and the situation in which, although there was a commanded change in the approach, the constraints could still be met. An example of such a question is: "Considering option 3: if all constraints at the waypoints can be met, even if you are required to take route A or B instead of route C, and you are instructed to take route A, the approach becomes…" response options: A lot easier – A lot more difficult. The pilots' answers are given in Figure 8.

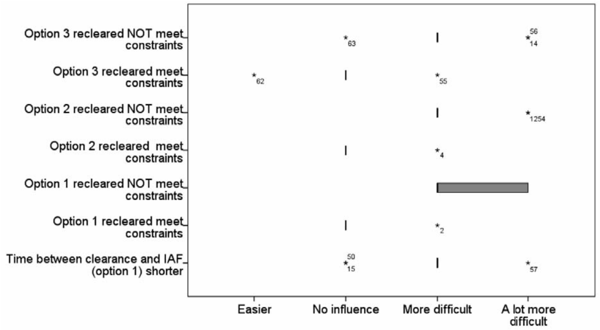


Figure 8 | Pilot responses with respect to the effect on difficulty of the three options to include flexibility in the approach.

Figure 8 clearly shows that, in the pilots' opinion, a change in the approach, whether it concerns options 1, 2, or 3, does not influence the difficulty experienced during that approach as long as the constraints at the waypoints can still be met. On the other hand, the pilots also agree that when, due to the commanded change, the constraints can no longer be met, the approach will become more difficult for all three Options.

4.3 | Flight data

The flight data recorded during the experiment were studied to determine how many flights were stabilized at 1,000', to determine how often the constraints were met at the waypoints, and to gain insight into all pilot actions. The results are discussed below.

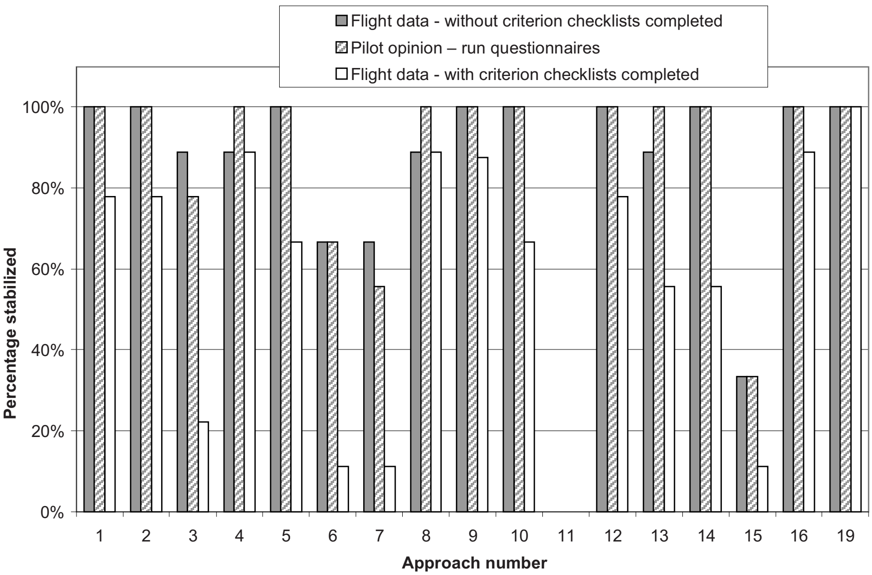


Figure 9 | Percentage of pilots that were stabilized at 1,000' according to the flight data.

4.3.1 | Stabilized at 1,000’

Figure 9 shows for each approach the percentage of pilots that were stabilized at 1,000’ according to the flight data. The following criteria were used (based on [2]):
Heading change and pitch change are within 5 deg/s;

- The IAS is not more than V_{REF} + 20 knots;
- Flaps 25 are selected, landing gear is down;
- Sink rate is not larger than 1,000 feet per minute;
- Localizer and glide slope are within one dot; and
- All checklists are completed

An average value is calculated for the criteria for heading change, pitch change, IAS, Localizer deviation and glide slope deviation: for the time slot starting 5 seconds before reaching 1,000 ft and ending at 1,000 ft. A larger time slot is used to calculate the average sink rate: it starts 1 minute before reaching 1,000 ft, and ends at 1,000 ft. For the other criteria the instantaneous values at 1,000’ are used. Figure 9 shows two results for the flight data: one percentage indicating the flights that were stabilized when taking the criterion ‘all checklists completed’ into account, and one percentage representing the stabilized flights without taking the checklist criterion into account. It should be noted, that during the flight tests, the times pilots were performing checklists were not always consistently logged for all flights. Therefore, due to the way of analyzing the data, the percentage of flights in Figure 10 that represents the stabilized flights with criterion checklists completed might be lower than was actually the case during the experiments. For comparison, the pilots’ answers on the run questionnaires are also included in Figure 9.

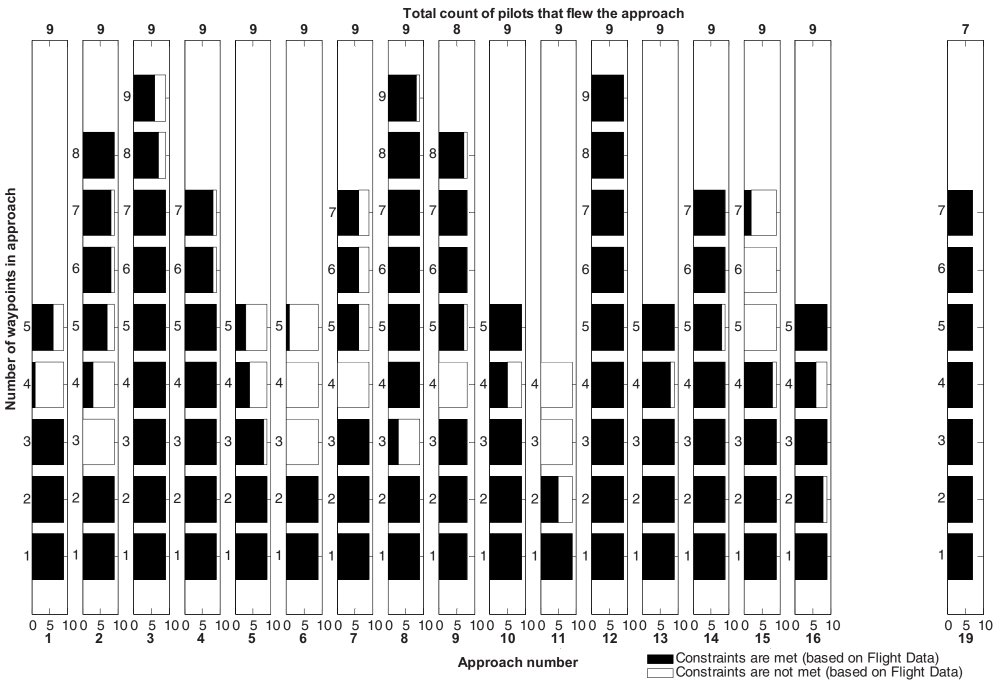


Figure 10 | Count of pilots that met the constraints at the waypoints and did not meet the constraints at the waypoint according to the flight data, for all approaches and all waypoints.

4.3.2 | Meet constraints at Waypoints

Figure 10 shows how many pilots did, and did not meet the constraints at each waypoint during the experiment, this is depicted for every approach. For each approach the last waypoint in this figure (waypoint with the highest number) is the FAF. This figure will be used to check the predictions of an offline computer simulation, which will be presented later on in this paper.

4.3.3 | Pilot actions and reaction times

All pilot actions (selecting flaps, gear, etc.) were logged during the experiment. During an earlier flight simulator experiment [13] it was shown that all pilot actions could be modeled in an offline computer simulation according to the SOPs by using a trigger event (for instance, reaching 1,200’) and a reaction time between reaching the trigger event and performing the action (for instance, selecting flaps 25). The trigger events and reaction times that were derived from this earlier flight simulator experiment are given in Table 14. It can be seen that a trigger event can be represented by reaching a certain location or altitude in the approach, but can also be represented by reaching a certain IAS. In this paragraph it is now analyzed whether the results from this flight simulator tests correspond to the earlier results given in Table 14, and to expand Table 14 with trigger events and reaction times for the performance of the approach and landing checklists. These trigger events and reaction times will eventually be used to model the pilots’ actions in an offline computer simulation, which will be presented later on in this paper.

Table 14 | Trigger events and reaction times for pilot actions resulting from previous flight tests [13].

Pilot action	Trigger Event	Mean	Standard deviation
Flaps 1 (237,000kg)	IAS	230 knots	11.8 knots
Flaps 5 (237,000kg)	IAS	212 knots	11.8 knots
Flaps 10	Start turn to Localizer Intercept Heading	22.9 sec	33.1 sec
Heading Select	Start turn to Localizer Intercept Heading	33.2 sec	17.9 sec
ARM Approach	Start turn to Localizer Intercept Heading	35.9 sec	20.3 sec
Autopilot Off	Start turn to Localizer Intercept Heading	40.7 sec	17.4 sec
Autothrottle Off	Autopilot Off	-0.35 sec	0.69 sec
Gear Down	Reaching FAF	-1.7 sec	15.9 sec
Flaps 20	Gear Down	2 sec	2 sec
Flaps 25	Reaching 1,200’	2.5 sec	8.6 sec

Comparison with previous flight simulator experiment

For the current flight simulator experiment, the same trigger events were found for all pilot actions, except for the actions performed on localizer Intercept heading. The current tests revealed a correlation between the heading change required towards Localizer intercept heading and the reaction times recorded (all significantly non-normal) for selecting flaps 10 (τ = .28, p < .001), ARM approach (τ = .21, p < .001), Autopilot off (τ = .30, p < .001) and Heading select (τ = -.17, p < .01) when using the *start* of the turn towards Localizer intercept heading as a trigger event. Since a larger heading change increases the amount of time the aircraft spends in the turn, thereby consistently increasing the reaction times. A better trigger event appeared to be the *end* of the turn towards Localizer intercept heading, this is the trigger event used for the remainder of this analysis.

To compare the reaction time distributions between the two flight simulator experiments, Mann Whitney tests were used. The results are given in Table 15 and show that for most reaction time distributions there was a significant difference between the two tests. In addition to Table 15, Kolmogorov-Smirnov tests showed that, next to the significant differences found by the Mann Whitney tests, there was also a significant difference between the reaction times for "Autothrottle off" and "Flaps 20". It should be noted that for all comparisons the original distributions for both flight simulator tests are used, not the approximating normal curves in Table 14.

Table 15 | Comparison of reaction time distributions between the current and previous flight simulator experiment for all pilot actions. The last column indicates whether there was a significant difference between the reaction time distributions.

Pilot action	**Trigger Event**	**Mdn previous exp.**	**Mdn. Current exp.**	***U***	***r***	**Difference?**
Flaps 1 (237,000kg)	Same as Table$	230.0 kts	220.7 kts(1)	916 ***	-.38	yes
Flaps 5 (237,000kg)	Same as Table$	212.3 kts	217.9 kts(2)	2948*	-.18	yes
Flaps 10	End turn to LOC int. HDG	-0.50 sec	-4.6 sec	1619 *	-.17	yes
Heading Select	End turn to LOC int. HDG	7.64 sec	-11.64 sec	1081***	-.31	yes
ARM Approach	End turn to LOC int. HDG	6.94 sec	9.12 sec	1889^{-}	-.09	no
Autopilot Off	End turn to LOC int. HDG	12.9 sec	5.12 sec	1152***	-.29	yes
Autothrottle Off	Same as Table 14	-0.50 sec	0.24 sec	1240^{-}	-0.13	no
Gear Down	Same as Table 14	1.70 sec	5.80 sec	8195***	-.29	yes
Flaps 20	Same as Table 14	3.05 sec	3.80 sec	10974^{-}	-.09	no
Flaps 25	Same as Table 14	1.00 sec	-0.20 sec	10182**	-.18	yes

$^{-}$ no significance * $p < .05$, ** $p < .01$, *** $p < .001$; (1) Based on approaches 3, 4, 6-9, 11, 12, 16 which were flown with aircraft mass of 237,000kg *and* started with an initial IAS higher than the UP mark at 223 kts.; (2) Based on approaches 1, 3-9, 11, 12, 16, 16 which were flown with aircraft mass of 237,000kg *and* started with an initial IAS higher than the 1 mark at 203 kts.

4

It is interesting to note that for the actions on Localizer intercept heading the same preferred order was found during these tests, as during the previous tests. A one-way repeated-measures ANOVA revealed that there was (at least partially) a preferred order $F(2.4, 341.9) = 7.32$, $p < .001$, $\omega^2 = .02$. Since Mauchly's test indicated that the assumption of sphericity was violated ($\chi^2(5) = 48.7$, $p < .001$), the degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\varepsilon = .803$). Post Hoc tests using the Bonferroni correction showed that, on average flaps 10 were selected before ARM approach was selected ($p < .01$), and that Heading select was selected before ARM approach was selected ($p < .001$), which agrees with the previous experiment.

Another agreement was that there was no correlation (Kendall's tau) between the FAF altitude (significantly non-normal, $D(150) = 0.264$, $p < .001$) and the reaction times for selecting flaps 20 (significantly non-normal, $D(150) = 0.247$, $p < .001$), τ (150) = .057, $p = .36$, or the reaction times for selecting gear down (significantly non-normal, $D(150) = 0.265$, $p < .001$), τ (150) = .038, $p = .55$. This corresponds to the results from [13].

Next to the comparisons to the previous flight simulator experiment, an additional observation resulting from the current experiment provides a very useful result. Within the current experiment some approaches (approaches 13, 14 and 19) were flown with a higher aircraft mass of 295,000 kg (resulting in an UP mark at 241 kts, and 1 mark at 221 kts). The UP mark is the airspeed below which flaps 1 should be selected, the 1 mark the airspeed below which flaps 5 should be selected. Although the UP mark has a higher value it was found (using a Wilcoxon signed-rank test) that the distribution of speeds *relative* to the UP mark for the higher aircraft weight (*Mdn* = -0.79) did not seem to differ from the speed relative to the UP mark at which flaps 1 were selected for the lower aircraft mass (*Mdn* = -2.34 knots, $T = 137$, $p = .49$, $r = .1$). The same was true for the selection of flaps 5: it appeared that there was no significant difference in this distribution *relative* to the 1 mark between the high aircraft mass (*Mdn* = 16.06) and the low aircraft mass, *Mdn* = 14.89 knots, $T = 161$, $p = .98$, $r = -.00$. This is a very useful result since it implies that, when the distribution around the UP mark is known for one aircraft weight, it is possible to use the same distribution around different UP marks for different aircraft weights[4].

Approach checklist
The approach checklist contains one item: ‘altimeters’, after which the altimeter settings and altimeter readings are checked. In general, the approach checklist is conducted when passing transition level. During the current experiment, the transition level was fixed at FL40, which was equal to 4,000’ altitude, and there were 115 logged recordings of the crew performing the approach checklist. The altitude at which the approach checklist is started and the duration of the approach checklist are given in figure 11.

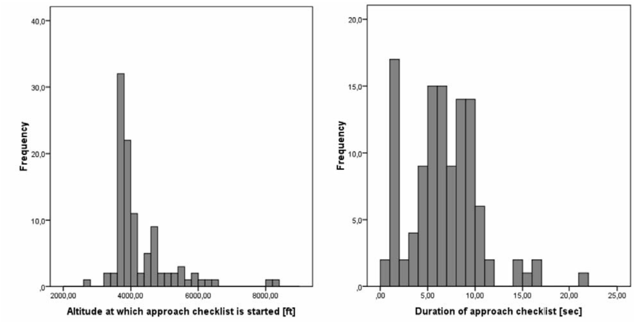


Figure 11 | Altitude at which approach checklist is started and duration of the approach checklist for current flight simulator experiment, with transition level at 4,000’ ($N = 115$).

Landing checklist
The landing checklist consists of three items: ‘speedbrakes’, ‘gear’, and ‘flaps’. This checklist can be performed in two different ways: either the entire checklist is performed after selecting flaps LAND (flaps 25 in this case) at 1,200’, or, the checklist is split into two parts where the first two items of the checklist are performed after selecting Gear down/flaps 20 with the annotation ‘flaps to go’, and the last item (‘flaps’) is performed after selecting flaps LAND.

During the experiment, 74 times a checklist was logged after selecting gear down/flaps 20 but before selecting flaps LANDS. This would always be only the first part of the landing checklist containing the items ‘speedbrakes’ and ‘gear’. The duration of and reaction times for this part of the checklist is given in Figure 12. After selecting flaps LAND a checklist was logged 113 times, this checklist could either be the second part of the landing checklist (with

4 For this Ph.D. thesis a more elaborate explanation of the pilot reaction times is added in Appendix B to Journal Article 4.

item ‘flaps’) or the entire landing checklist. The duration of and reaction times for this (part of the) checklist is given in Figure 13.

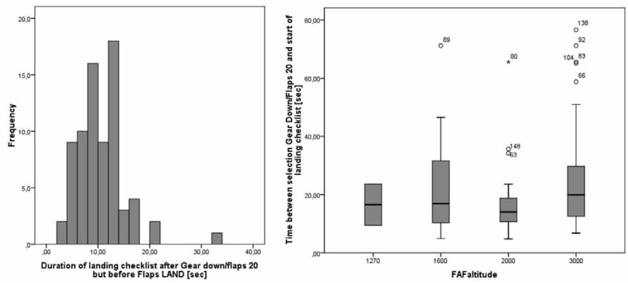


Figure 12 | Duration of and reaction time for the landing checklist performed after selecting Gear down/flaps 20 and before selecting flaps LAND.

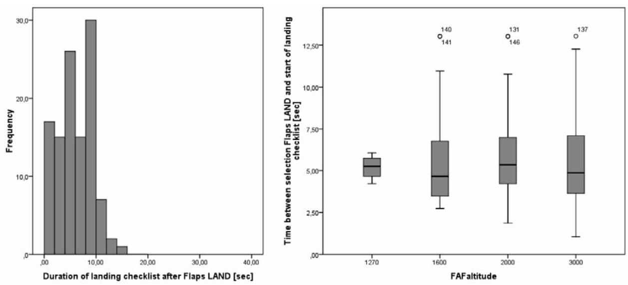


Figure 13 | Duration of and reaction time for the landing checklist performed after selecting flaps LAND.

5 | Discussion

The experiment results that were derived from the run questionnaires, end of day questionnaire and flight data are compared in this paragraph, to verify whether these results are consistent.

5.1 | Stabilized at 1,000’

There was a significant relationship (see Figure 9) between the percentage of approaches that were stabilized according to the objective flight data (without criterion “checklists completed”), and the percentage of approaches that were stabilized according to the subjective answers of the pilots on the run questionnaires, $\tau = .78$, $p < .001$. This implies that the judgment of the pilots whether the approach was stabilized at 1,000’ correlates very well with what the objective flight data indicate, which can also be observed in Figure 9.

5.2 | Three options to incorporate flexibility in an approach

In the end of day questionnaire the pilots’ answers indicated that any of the three options (see Figure 5) became more difficult when, due to a change in route, the energy rate demand became too high and as a result constraints could no longer be met. When looking at the RSME ratings, one can more specifically say that the approach only becomes more difficult when the effect of the high energy rate demand continues into the Localizer intercept part or the final part of the approach.

Table 16 | Summary of the results obtained from the pilots' RSME ratings, answers in the run questionnaires and answers given in the end of day questionnaire

Factor #	Factors	Independent measure	Variable between pairs	RSME scores	Run questionnaires (correlation between data levels 2 and 3)	Regression RSME part ratings and factors	Factors indicated to have an effect (#) in the End of day questionnaire	Hypothesis for TDL effect
4a, 4a1, 4a2	Energy rate demand too high in first part of approach	Yes	Yes	*(1), (4), (5)	***	****	#	=
4b	Energy rate demand too high in Localizer intercept part of approach	Yes	Yes	*(1), (6)	***	*n.s.*	#	+
4c	Energy rate demand too high during final part of approach	Yes	Yes	**(1)	***	*	#	+
5	Higher LOC intercept speed	Yes	Yes	*(1)	***	*n.s.*	# [13]	+
14	Larger LOC intercept angle	Yes	Yes	*(1)	***	*n.s.*	# [13]	+
8	Less distance available on LOC Intercept Heading	Yes	Yes	*(2)	***	*n.s.*	Not asked	+
	Less time available during LOC intercept part of approach				***	****	# [13]	
10	Smaller line-up distance	No	Yes	*(3)	***	*n.s.*	# [13]	
	Time available during final part of approach				***	*n.s.*	# [13]	
12	Lower FAF altitude	Yes	Yes	*(3)	***	*n.s.*	#	+
24	Less distance available between IF and FAF	Yes	Yes	*(3)	***	*	#	+
20	Higher airspeed on final	Yes	Yes	*(1)	***	*	#	+
1	More waypoints	No	Yes	*n.s.* [13]	***	*n.s.*	#	
1	More heading changes	No	Yes	*n.s.* [13]	***	*n.s.*	#	
18	Less trackmiles	No	Yes	Not tested	***	*n.s.*	# [13]	
21	Less time available during first part of approach	No	Yes	Not tested	***	*n.s.*	Not asked	
23	Not stabilized at 1,000'	No	Yes	Not tested	***	****	#	

n.s. = no significance, * = $p < .05$, ** = $p < .01$, *** = $p < .001$, **** = $p < .0001$;# = factors indicated to have an effect in the end of day questionnaire; n.e. = no effect expected on TDL, e. = effect expected on TDL; (1) t-test; (2) Wilcoxon signed-rank test; (3) ANOVA; (4) only a significant effect when the consequences continued into the Localizer intercept part or the final part of the approach; (5) only tested for flexibility options 1 and 3; (6) only tested for flexibility option 2

Pilots also indicated in the questionnaire, that a commanded change in route does *not* influence the difficulty when the constraints at the waypoints can still be met. Based on the RSME scores and conversations with pilots it can be stated that this is partly true: for Options 1 and 3 a change in route does indeed not seem to influence pilot TDL. It was also demonstrated that the location of the shortcut for flexibility Option 3 does not affect pilot TDL. For Option 2, the fact that a shortcut is introduced in itself (when compared to flying the same 'shortcut' route as a published route) does no influence pilot TDL. However, when comparing different shortcuts for option 2, the *location* of the shortcut *does* influence the difficulty, in the sense that a shortcut resulting in a shorter line-up distance increases pilot TDL.

When asked in the questionnaire the majority of the pilots indicated that they, in general, preferred either Option 2 or Option 3, whereas all pilots preferred Option 2 or 3 when asked to only consider the adjustments needed in the FMS.

5.3 | Factors influencing pilot TDL during approach

To check the consistency between the pilots' RSME ratings, answers on the run questionnaires and answers given in the end of day questionnaire, all results are summarized in Table 16. For example: the factor "5. Higher Localizer intercept speed" was an independent measure, and was also changed between approach pairs. The RSME ratings for the approach pair that considered the Localizer intercept speed (approach pair G) yielded a significant difference, and there was a highly significant correlation between the pilots perception of the Localizer intercept speed (data set 2) and the effect on the difficulty as indicated by the pilot (data set 3). However, the indicated difficulty due to the Localizer intercept speed did not significantly influence the pilots RSME sub-rating for that part of the approach (data set 4). But in the end of day questionnaire for the previous flight simulator experiment [13], the pilots answered that an approach becomes more difficult when the Localizer intercept speed increases.

In Table 16 only the independent measures are included that are applicable to all approaches, and are not only valid for one type of flexibility Option. Therefore the independent measures "25. Location of shortcut for flexibility Option 2", "26. Location of shortcut for flexibility Option 3", and "27. Introducing a shortcut for flexibility Option 2" are not included, these were already discussed in the previous paragraph.

4

Based on the summary in Table 16 and the findings in [13], combined with the experience gained from attending all flight simulator experiments, the following is, for now, concluded with respect to the factors that influence pilot TDL during each of the approach parts (it is noted that this is all based on a small data set):

For the first part of the approach:

- ❖ The major contributor to pilot TDL seems to be the energy rate demand, or, the fact whether or not constraints can be met at the waypoints. This is only true when the effect of not meeting the constraints continues into the Localizer intercept part of the approach or final part of the approach. If the consequences of the energy rate demand remain within the first part of the approach this does not influence pilot TDL. In this respect it should be noted that this might be different when other traffic is present or when ATC is urging the pilots to meet the constraints. Although, when asked, most pilots stated that this would have no influence on pilot TDL, because in their opinion ATC would already know upfront that they would not be able to meet the constraints, and would therefore not expect them to meet the constraints.
- ❖ The number of waypoints, number of heading changes and the altitude profile (horizontal approach, CDA, stepped approach) [13] seem not to influence pilot TDL. This is due to

the fact that this part of the approach is flown in LNAV and VNAV modes with autopilot and autothrottle.
- ❖ The time available to perform all actions during this part of the approach is important when flexibility (see Figure 5) is introduced, pilots should then have sufficient time to make all necessary adjustments.

For the Localizer intercept part of the approach:
- ❖ The time available to perform all actions (which is directly related to the distance available on Localizer intercept heading) seems to be the most important factor for pilot TDL.
- ❖ Next to this, pilot TDL is also influenced by the Localizer intercept speed, the Localizer intercept angle, and whether the constraints at the waypoints can be met (the energy rate demand).

For the final part of the approach:
- ❖ The most important factors influencing pilot TDL seem to be whether or not a stabilized approach can be achieved at 1,000′, the distance between IF and FAF and the airspeed on final. Whether an approach is stabilized can, for a B747, be determined from: 1. the energy rate demand during the final part of the approach, 2. the value of the vertical speed (should be below the sink rate warning) which in itself is a function of airspeed on final and glideslope angle, and 3. the FAF altitude and distance between IF and FAF (resulting in the line-up distance) since these two factors together determine whether there is enough time available to perform all actions required for a stabilized approach. All these factors thus influence pilot TDL during the final part of the approach.

5.4 | Guidelines for the design of approaches

The guidelines for the design of approaches with respect to pilot TDL are based on the results presented in the previous section. Starting point for the guidelines is that pilots should fly the approach according to SOPs and that they should aim to achieve a stabilized approach at 1,000′. The guidelines for the contributors to pilot TDL for the B747 then are that:
1. aircraft should be able to meet the altitude and airspeed constraints throughout the approach, especially during the final part of the approach, and during the first part of the approach if this has consequences for the subsequent parts of the approach;
2. there should be sufficient time to perform all actions on Localizer intercept heading;
3. it should be possible to achieve a stabilized approach. Whether a stabilized approach can be achieved depends, for the B747, on 1. the energy rate demand during the final part of the approach, 2. the value of the vertical speed (should be below the sink rate warning) which in itself is a function of airspeed on final and glideslope angle, and 3. the FAF altitude and distance between IF and FAF (resulting in the line-up distance);
4. the distance between IF and FAF should be sufficient;
5. the vertical speed should be below the sink rate warning;
6. the Localizer intercept speed should not be too high, and that
7. the Localizer intercept angle should not be too large.

These seven guidelines for the B747 should be met for the most common prevailing wind conditions at the airport, and for the majority of pilots. The guidelines are based on a small sample of approaches and pilots. Therefore, the decision that some factors from Table 1 do not influence pilot TDL was also based on a small data set, and might not be conclusive for all possible circumstances. In that respect, when also wanting to incorporate these factors in the design of the approach, an additional guideline can be specified that: the number of waypoints and heading changes should not be too high, that there should be sufficient time or trackmiles during the first part of the approach, and that the altitude profile preferably is a Continuous Descent Approach.

6 | An offline computer simulation to predict factors influencing pilot TDL

This section presents an offline computer simulation which predicts whether the guidelines specified in the previous section are met for a particular approach. First the basics of the simulation are explained, which is followed by a validation of the computer simulation by using the results of the flight simulator tests. After that the computer simulation is used to simulate approaches beyond the conditions flown during the flight simulator tests. Finally, a simplified version of the computer simulation is presented and it is checked whether this simplified version can indeed provide predictions concerning all guidelines previously derived.

6.1 | An offline computer simulation

The input of the computer simulation exists of a list of waypoints, defined by their lat-lon coordinates, and the altitude and speed constraints at these waypoints. The computer simulation itself consists of an aircraft model, autothrottle model, autopilot and flight director models, an FMS model and a pilot model. The non-linear aircraft model is based on the Boeing 747 documentation by Hanke and Nordwall [24]. Autopilot, Autothrottle and Flight Director (FD) models are also derived from [24]. The hierarchy in meeting the constraints at the waypoints is as follows: the Autopilot and FD modes will always aim to meet the altitude constraints at the waypoints, second to this, the Autothrottle controls the airspeed. This results in the situation that the altitude constraint at the next waypoint will always be met, while the speed constraint might not be met (airspeed might be higher than required). To these highly detailed, non-linear models a relatively simple pilot manual control model for the flight director task is added, consisting of only a time delay (equal to 0.3 seconds) and pure gain. All other pilot actions such as selecting flaps are modeled according to the SOPs, using the trigger events and reaction time distributions discussed in the previous paragraphs. For a more detailed explanation of the basic principles of the offline computer simulation the reader is referred to [3].

In the GRACE flight simulator the same aerodynamic model for the aircraft is used as in the offline computer simulation. However, within GRACE different engines are incorporated which differ slightly from the ones used in the offline computer simulation. Additionally, the lateral autopilot in the offline computer simulation tends to ‘cut the corners’ when compared to the tracks resulting from the GRACE flight simulator.

6.2 | Validation of the offline computer simulation

When using exactly the same reaction times for all pilot actions for each approach in the offline computer simulation as recorded during the current flight simulator experiment, it is expected that the predictions of the offline computer simulation should closely resemble the logged flight data during the flight simulator experiment. Note, however, that although exactly the same reaction times are used, differences still exist between the computer simulation and the flight simulator experiment: 1. a difference in engines, 2. a slight difference in lateral track, 3. manual throttle control in the Localizer intercept and Final part of the approach during the flight simulator tests opposed to autothrottle control in the computer simulation, 4. a computer simulation that always rigidly flies from one waypoint to the next without anticipation of what comes next and always controls the aircraft towards the exact altitude and airspeed constraints sometimes resulting in unfavorable autothrottle action when considering the *next* waypoint, whereas pilots during the flight simulator experiment (especially during the Localizer intercept part and Final part of the approach) anticipate and sometimes forget altitude and airspeed constraints or control the aircraft towards incorrect constraints, and 5. the groundtracks resulting from ATC intervention (the flexibility options in Figure 5) showed more variability during the flight simulator tests than assumed in the computer simulation. These five differences might cause deviations between the predictions of the computer simulation and the results of the flight simulator experiment.

Figure 14 shows the groundtrack and pilot actions as recorded during the flight simulator experiment, as well as the groundtrack and pilot actions as predicted by the offline computer simulation. Each approach was simulated as many times as it was flown during the experiment, with for each simulated approach exactly the same reaction times as recorded during that experiment run. It can be seen that both plots compare very well, and that modeling the pilots' actions according to trigger events and reaction times actually works.

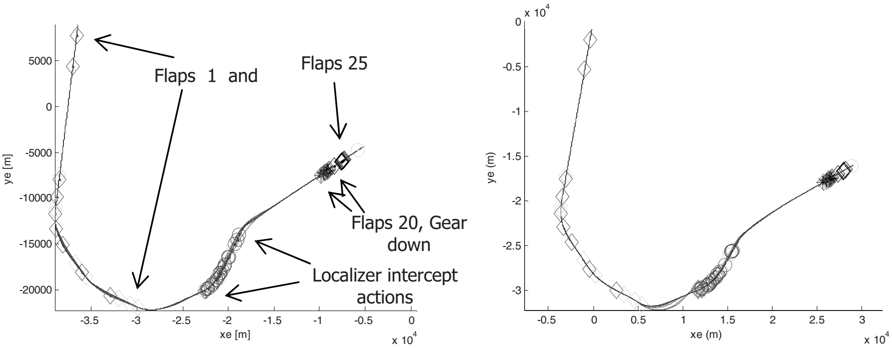


Figure 14 | Groundtracks and pilot actions according to the flight data (left) and according to the offline computer simulation (right)

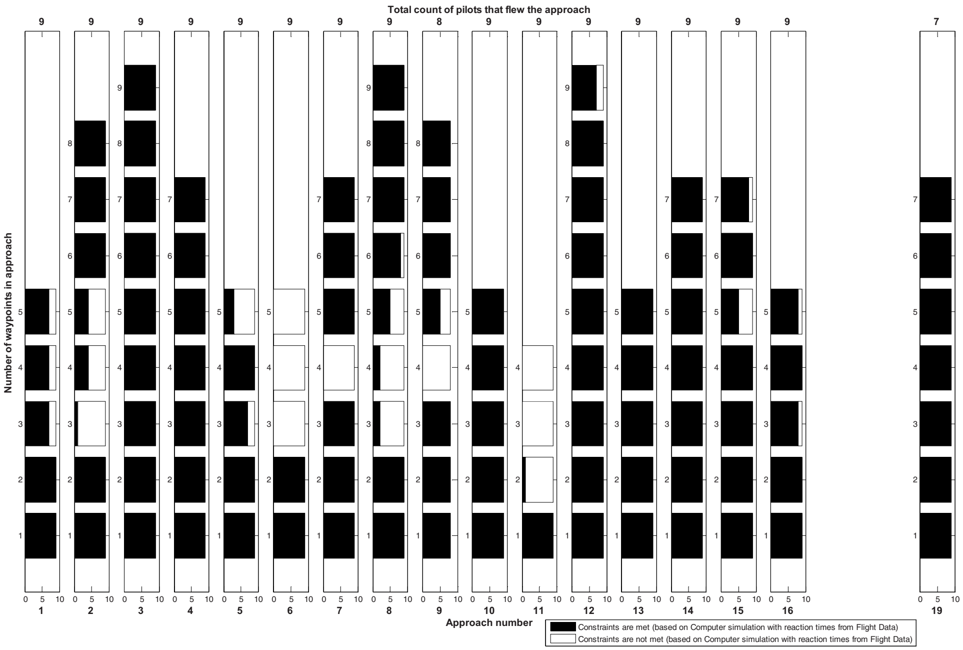


Figure 15 | Count of simulated runs that met the constraints at the waypoints and did not meet the constraints at the waypoint using exactly the same reaction times as recorded during the flight simulator experiment, for all approaches and all waypoints.

Figure 15 shows the predictions of the offline computer simulation with respect to meeting the constraints at all waypoints. Comparing Figure 15 to Figure 10 one can conclude that the results of the computer simulation agree quite well with the flight data. Most notable differences are waypoints 4 and 5 for approach 8 (caused by the difference in engines), waypoint 4 for approach 1 (this is the Localizer intercept waypoint, during the flight tests pilots were more focused on capturing the Localizer than they were on controlling the airspeed) and waypoints 5, 6 and 7 for approach 15. For approach 15 pilots were instructed to follow the shortcut according to flexibility option 2, during this shortcut they also had to increase their descend path. However, when changing to Heading Select mode, they forgot to switch off the VNAV mode, which maintained a shallower descend path than required, and therefore they did not meet the constraints at the subsequent waypoints, although (flight mechanically, according to the computer simulation) they could have been met.

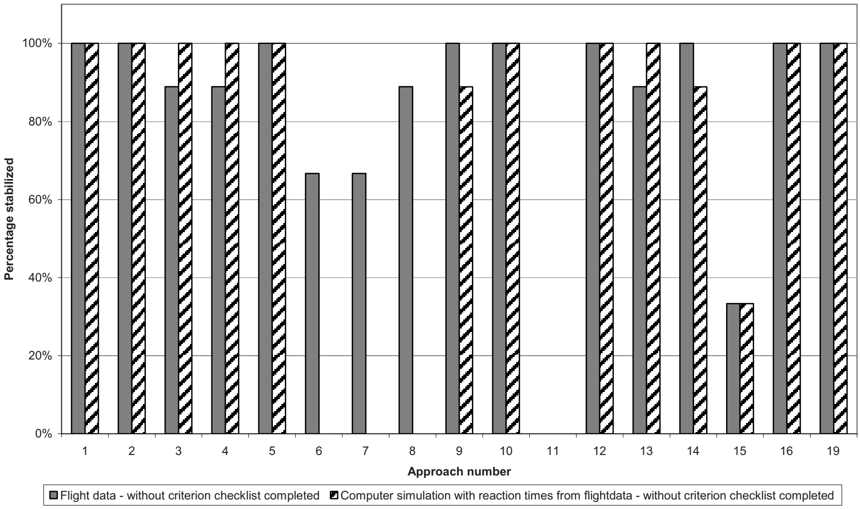


Figure 16 | Percentage of pilots that were stabilized at 1,000' according to the flight data and the predictions of the computer simulation when using exactly the same reaction times as recorded during the flight simulator experiment.

4

Figure 16 shows the percentage of flights that were stabilized at 1,000' according to the predictions of the offline computer simulation, and according to the flight data. The results agree very well, except for approaches 6, 7 and 8. According to the computer simulation all flights for these approaches were unstabilized whereas during the flight simulator experiment a large percentage of the flights were *just* stabilized. This difference is caused by two factors: the difference in flight idle thrust (due tot the difference in engine models) which became very pronounced at these altitudes and airspeeds, and the anticipation (with respect to the constraints at the next waypoint) of the pilots during the flight simulator experiment. These results emphasize the need for a very accurate idle thrust model in order to obtain adequate results with the computer simulation.

6.3 | Monte Carlo runs with the computer simulation

The computer simulation can now be run many times, using all recorded reaction times both from the previous and the current flight simulator experiment, and applying different wind conditions. The computer simulation can then be used to predict under what wind conditions the constraints at the waypoints can be met and a stabilized approach can be achieved.

As an example see the results for approach 06 in Figure 17. The Figure clearly shows that the possibility of achieving a stabilized approach depends on the wind direction (a strong headwind on final results in a stabilized approach). Other reasons for ending stabilized or unstabilized are the moment in time at which flaps 20 and/or gear down are selected. A similar plot as in Figure 17 can be generated for meeting the constraints at each waypoint. The results of the computer simulation also provide insight into the locations in the approach where pilots are performing many actions, or where they are performing checklists, thereby providing approach designers with an indication of the 'busy' parts of an approach, see Figure 18 for an example.

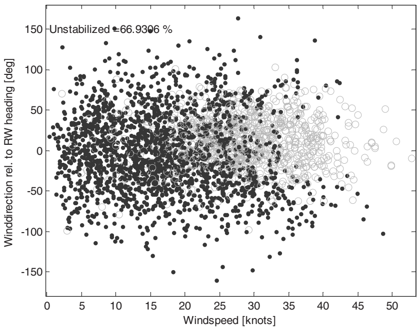


Figure 17 | Example of the results of the computer simulation (3,000 runs) with respect to the possibility of achieving a stabilized (grey circle) or unstabilized (black dot) approach at 1,000'.

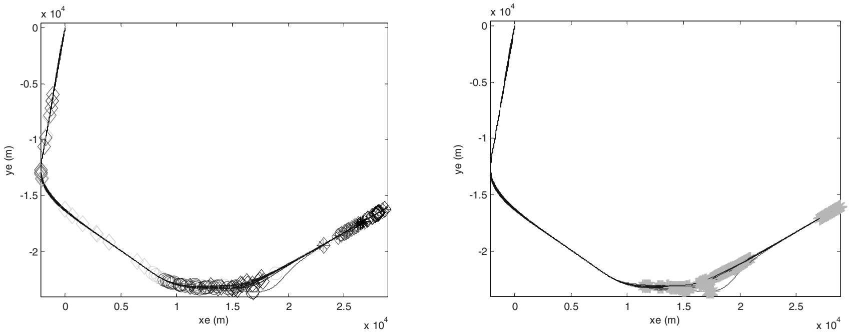


Figure 18 | Example of the results of the computer simulation, providing insight into the locations in the approach where pilots are performing actions such as selecting flaps, gear, etc. (left) or are performing checklists (right).

6.4 | A computer simulation based on a point mass model

Instead of using a complex, highly detailed, non-linear computer simulation with exact replications of autopilots, flight director and autothrottle, it was explored whether a relatively simple point mass model could provide the same predictions that are relevant to predict pilot TDL. This might very well be possible since the factors influencing pilot TDL (see Table 16) all appear to be long time scale flight mechanical factors, not short time scale flight dynamical factors (for which a highly detailed model would be imperative). It was found that a point mass model can indeed accurately predict whether the constraints at the waypoints can be met, whether a stabilized approach can be achieved and whether sufficient time is available to perform all actions. This is true as long as the point mass model contains: 1. a detailed lift-drag polar for all flap settings and gear up/down setting, 2. a detailed model of the flight idle thrust, 3. an accurate model to simulate the lateral track, specifically the distance of turn anticipation since this influences the amount of trackmiles available between two waypoints, and 4. a model to simulate the pilots' actions according to the trigger events and reaction time distributions presented in this paper.

Additionally, a further simplification is possible by not simulating all possible pilot reaction times (resulting in many runs) and afterwards checking the results to identify how many pilots met the constraints and achieved a stabilized approach; but by setting an upper limit to the percentage of pilots that should be able to fly the approach, meet the constraints, achieve a stabilized approach and have enough time available to perform all actions. This way the simulation needs to be run only once for one specific set of reactions times. For example: if the approach designer decides that 95% of the pilots should be able to fly an approach, then for each pilot action we order the reaction times according to their magnitude and select the reaction time with a magnitude higher than 95% of the reaction times. This is the reaction time to use in the computer simulation, and represents a worst case scenario for that percentage. This process is repeated for all pilot actions, and the computer simulation is run only once. If, according to the point mass model computer simulation, the approach can be flown (meet constraints and stabilized) with this set of reaction times, the 95% of pilots with shorter reaction times will also be able to fly the approach. As an indication, the pilot reaction times for different percentages of pilots are given in Table 18, these reaction times are based on the combined distributions of the previous and current flight simulator experiment.

For example: 90% of the pilots selected flaps 5 before they reached an IAS of VREF + 56.2 knots, and selected flaps 25 before they were 9.6 seconds past 1,200'. For the actions Gear Down and Flaps 20 two maximum reaction times are mentioned: the first indicates the latest moment in time at which Gear down was selected (for instance, 90% of the pilots selected gear down before they were 11.3 seconds passed the FAF, to this reaction time a selection of flaps 20 corresponds of 6.9 seconds before selecting GD (-6.9 s in the table)). The second reaction time indicates the latest moment in time at which flaps 20 were selected (for instance, 90% of the pilots selected flaps 20 8.3 seconds after selecting Gear down, which in its turn was 1.3 seconds after reaching the FAF).

6.5 | Are all design guidelines predicted by the computer simulation?

The question now is, whether the output of this computer simulation provides sufficient information in order to assess whether the approach meets all guidelines for the contributors to pilot TDL as given earlier in this chapter. The simulation obviously predicts whether a stabilized approach can be achieved (guideline 3), and whether the constraints at the waypoints can be met (guideline 1). It also provides insight whether there is sufficient time on Localizer intercept heading since the moments in time at which all actions are performed is predicted (guideline 2), and it can easily predict the sink rate (guideline 5).

However, although the simulation can predict or calculate the numerical values for the (actual) Localizer intercept speed (guideline 6), the Localizer intercept angle (guideline 7), and the distance between IF and FAF (guideline 4), it does not give a qualitative indication of whether these numerical values are sufficiently high or low. Fortunately, the minimum or maximum values for these factors are very accurately prescribed in the Procedures for Air Navigation Services Aircraft Operations (PANS-OPS) [12]. The PANS-OPS prescribe a minimum straight distance between IF and FAF of 2nm with an additional turning distance (which depends on the airspeed and intercept angle), and recommend an interception angle at the Localizer not

Table 18 | Maximum pilot reaction times for different percentages..

Pilot action	Trigger Event	85 %	90 %	95 %	99 %
Flaps 1	IAS	VREF+75.5kts	VREF+73kts	VREF+70.2kts	VREF+50kts
Flaps 5	IAS	VREF+57.3kts	VREF+56.2kts	VREF+54.2kts	VREF+43.3kts
Flaps 10	End turn to LOC int. HDG	33.1 s	38.7 s	62.9 s	86.1 s
Heading Select	End turn to LOC int. HDG	24.0 s	30.8 s	45.4 s	147.0 s
ARM Approach	End turn to LOC int. HDG	37.1 s	44.8 s	57.6 s	130.0 s
AP & AT Off	End turn to LOC int. HDG	30.8 s	38.8 s	47.3 s	65.2 s
Gear Down	Reaching FAF	9.6 / -8.5 s	11.3 / 1.3 s	15 / 2.5 s	27.8 / -7 s
Flaps 20	Gear Down	-9.6 / 14.9 s	-6.9 / 8.3 s	-15 / 12.8 s	-26.8 / 30.4 s
Flaps 25	Reaching 1,200’	6.4 s	9.6 s	15.3 s	23.4 s
Approach CL	Transition Level	TL – 333ft	TL – 350ft	TL – 400ft	TL – 1,220ft
Duration		9.8 s	10.4 s	11.5 s	16.5 s
CL after GD/flaps20	Latest of GD or Flaps 20	23.6 s	34.2 s	35.7 s	65.6 s
Duration		13.9 s	15.3 s	16.7 s	21.8 s
CL after flaps 25	Flaps 25	8.4 s	9 s	12.3 s	13.0 s
Duration		9 s	9.3 s	11.5 s	12.1 s

exceeding 30 degrees. The nine pilots that participated in the experiment were also asked to give their opinion about these three factors. Their answers for the minimum distance that should be available between IF and FAF ranged between 2 and 3nm, their answers for the maximum Localizer intercept speed ranged from 170kts to 210kts, and the values given for the maximum Localizer intercept angle ranged between 30 and 60 degrees. These answers correspond to the values found in the PANS-OPS [12]. Actually, the PANS-OPS and the predictions of the computer simulation complement each other very nicely regarding factors contributing to pilot workload, since what is not prescribed in the PANS-OPS is predicted by the computer simulation and vice versa.

The conclusion thus is that the computer simulation, combined with the regulations in the PANS-OP, provides sufficient information to assess whether the guidelines for the contributors to pilot TDL for the B747 are met.

For clarity, the exact quantification of the guidelines as used during this research is given below (numbers correspond to numbered guidelines):

1. The constraints at the waypoints are considered to be met when the airspeed is within 10 knots of the required airspeed and the actual altitude is within 100 feet of the required altitude.
2. Using the reaction times for the actions performed on Localizer intercept heading as given in Table 18 it can be determined (for a percentage of pilots defined by the approach designer) whether all actions that should be performed on Localizer intercept heading are actually performed on Localizer intercept heading. When this is the case this is regarded "sufficient time" to perform all actions.
3. An approach is considered stabilized at 1,000ft when:
   - ❖ Heading change and pitch change are within 5 deg/s;
   - ❖ The IAS is not more than V_{REF} + 20 knots;
   - ❖ Flaps 25 are selected, landing gear is down (whether this is the case for the defined percentage of pilots can be derived from the values in Table 18);
   - ❖ Sink rate is not larger than 1,000 feet per minute;
   - ❖ Localizer and glide slope are within one dot[5]; and
   - ❖ All checklists are completed (whether this is the case can again be derived from the values in Table 18)
4. the distance between IF and FAF is considered sufficient when it meets the requirements in the PANS-OPS [12]
5. the vertical speed should be below the sink rate warning.
6. the Localizer intercept speed is considered "not too high" when it meets the requirements in the PANS-OPS [12]
7. the Localizer intercept angle is considered "not too large" when it meets the requirements in the PANS-OPS [12].

4

7 | Conclusions and further research

In this paper seven approach design guidelines were derived for the B747, resulting from flight simulator experiments. When these seven guidelines are met pilot TDL during the approach will remain at an acceptable level. It was demonstrated that whether these guidelines will be met can be predicted by an offline computer simulation in combination with the requirements in the PANS-OPS [12]. This computer simulation might consist of a simple point mass model as long as it contains: 1. a detailed lift-drag polar for all flap settings and gear up/down setting, 2. a detailed model of the flight idle thrust, 3. an accurate model to

5 One dot deviation on the glide slope equals 0.7° beam error, one dot deviation on the Localizer equals 2.5° beam error.

simulate the lateral track, and 4. a model to simulate the pilots' actions according to the trigger events and reaction time distributions presented in this paper.

The computer simulation model and flight simulator experiments with resulting conclusions are all based on the SOPS and flight mechanical models for a B747. To check the general applicability of the factors which were found to influence pilot TDL and to check the general applicability of the set-up of the computer simulation, a different aircraft needs to be considered. Therefore, the next step in this research is to perform flight simulator experiments with a Cessna Citation, and to adjust the computer simulation such that it contains accurate models for the SOPs and flight mechanical models of the Cessna Citation. Additionally, real flight tests will be performed with the Cessna Citation in order to check whether in reality the same factors are found to influence pilot TDL as during the flight simulator tests.

Acknowledgements

This research was financially supported by the Dutch Technology Foundation STW (DLR7071). We wish to thank the pilots participating in the experiment and the National Aerospace Laboratory NLR for offering the simulation time in, and technical support for, the flight simulator experiment in GRACE.

Literature

1. Stassen, HG, Johannsen, G, & Moray, N. Internal Representation, Internal Model, Human Performance Model and Mental Workload. *Automatica, 26(4)*, 811-820, 1990.
2. Flight Safety Foundation Approach-and-landing Accident Reduction (ALAR) Task Force. Killers in Aviation: FSF Task Force Presents Facts about Approach-and-landing and Controlled-flight-into-terrain Accidents, *Flight Safety Digest*, November 1998 - February 1999
3. Heiligers, M.M., Holten, Th. Van, Mulder, M, *Flight mechanical evaluation of approaches*, submitted to Journal of Aircraft
4. Vicente, K.J., *Cognitive Work Analysis, Towards Safe, Productive, and Healthy Computer-Based Work*, Lawrence Erlbaum Associates, Publishers, Mahwah, New Jersey, 1999
5. Baron, S., Zacharias, G., Muralidharan, R., Lancraft, R. (1980), PROCRU: A model for analyzing flight crew procedures in approach to landing, *NASA-MIT 16th Annual Manual 1980*, pp 488-520.
6. Baron, S., Kruser, D.S., Huey, M.B., *Quantitative Modeling of Human Performance in Complex, Dynamic Systems*,National Academy Press, Washington D.C., 1990.
7. Corker, K.M., Human Performance Simulation in the Analysis of Advanced Air Traffic Management, *Proceedings of the 1999 Winter Simulation Conference*, Farrington, P.A., Nembhard, H.B., Sturrock, D.T., Evans, G.W. (Eds.), 1999.
8. Gore, B.F., Corker, K.M. A Systems Engineering Approach to Behavioral Predictions of an Advanced Air Traffic Management Concept, *19th Digital Avionics Systems Conference, Entering the Second Generation of Powered Flight*, Philadelphia, Pennsylvania, 2000.
9. Smith, B.R., Tyler, S.W., The Design and Application of MIDAS: A Constructive Simulation for Human-System Analysis, *2nd Simulation Technology & Training Conference*, Canberra, Australia, 1997.
10. Vormer, F.J.P., *Theoretical and Operational Aspects of Optimal Airport Arrival Trajectories,* TUD Technische Universiteit Delft, 8 December 2005

11. Godley, Stuart T. *Perceived Pilot Workload and Perceived Safety of RNAV (GNSS) Approaches,* Australian Transport Safety Bureau, ATSB Transport Safety Investigation Report, Aviation Safety Research and Analysis Report 20050342, 2006.
12. International Civil Aviation Organization, *Procedures for Air Navigation Services Aircraft Operations, Volume II Constructions of Visual and Instrument Flight Procedures,* Fifth edition 2006.
13. Heiligers, M.M., Holten, Th. Van, Mulder, M, *Factors that influence pilot task demand load during RNAV approaches,* submitted to Journal of Aircraft
14. Hilburn, B. and Jorna, P., Workload and Air Traffic Control. In PA Hancock and PA Desmond (Eds.) *Stress, Workload and Fatigue: Theory, Research and Practice*, Hillsdale, New Jersey, USA: Erlbaum (2001)
15. Zijlstra, F., Meijman, Th., Het meten van mentale inspanning met behulp van een subjectieve methode, *Mentale belasting en werkstress, een arbeidspsychologische benadering,* Meijman,Th. (Ed.), Van Gorcum, Assen/Maastricht, The Netherlands, 1989.
16. Lodge, M., *Magnitude Scaling, Quantitative Measurement of Opinions,* Sage university paper series on quantitative applications in the social sciences, 07-025, Newbury Park, CA: Sage.
17. Meijman, T., Zijlstra, F., Kompier, M., Mulders, H., Broersen, J., The measurement of perceived effort, *Contemporary Ergonomics 1986,* Oborne, D.J. (Ed.), Taylor & Frances, London, 1986.
18. Hart, S.G., Staveland, L.E., Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research, *Advances in Psychology* Vol. 52, p.139-183, P.A. Hancock, Meskati N. (Eds.), Elsevier Science Publishers B.V., Amsterdam, The Netherlands, 1988.
19. Jamieson, S., Likert Scales: how to (ab)use them, *Medical Education 2004,* 38: 1217-1218, Blackwell Publishing, 2004.
20. Knapp, T.R., Treating Ordinal Scales as Interval Scales: An Attempt To Resolve the Controversy, *Nursing Research,* March/April 1990, Vol.39 No.2. 121-123.
21. Lorch, R.F., Myers, J.L., Regression Analyses of Repeated Measures Data in Cognitive Research, *Journal of Experimental Psychology: Learning, Memory, and Cognition,* Vol. 16, No. 1, 1990, p149-157.
22. Hardy, M. A., *Regression with Dummy Variables,* Sage university paper series on quantitative applications in the social sciences, 07-093, Newbury Park, CA: Sage.
23. Field, A., *Discovering Statistics Using SPSS,* second edition, Sage publication Ltd, London, 2005.
24. Hanke, Rodney C. and Nordwall, Donald R., *The Simulation of a Jumbo Jet Transport Aircraft Volume II: Modeling Data*, NASA report D6-30643, September 1970, Wichita, Kansas.

Appendix A to Journal Article 4

Energy rate demand too high in first part of approach for Option 3 (Approach pair A)

The pilots' RSME z-scores for the first part of the approach for approach 7, which contained a leg with an energy rate demand larger than 1 due to the 'direct to' command by ATC (Option 1 in Figure 3, route A), were significantly higher (M=0.617, SE = 0.220) than the RSME z-scores for the first part of approach 4, which incorporated the same 'direct to' command at the same location but in this approach all energy rate demands remained below 1, (M=-0.093, SE = 0.196, $t(8) = -2.86$, $p < .05$, $r = .71$). However, when comparing the pilots' RSME ratings for approaches 3 and 9, which also contained a 'direct to' command by ATC but now further away from the runway (taking route B instead of route A in Figure 3, Option 3), showed that there was *no* difference in RSME ratings when the energy rate demand in the leg following the 'direct to' command was smaller than 1 (approach 3, M=-0.27, SE = 0.390) or larger than 1 (approach 9, M=0.44, SE = 0.296, $t(6) = -1.14$, $p = .3$, $r = .42$). It thus seems (and this will be further underpinned when considering approach pair C) that an energy rate demand larger than 1 in the first part of the approach only has an effect on pilot TDL, when the consequences of this high energy rate demand continue into the Localizer intercept heading part of the approach or the final part of the approach. This was the case for approach 7 which followed route 'A', for approach 9 there was sufficient time and distance available due to taking route 'B' to dissipate the excessive energy before reaching Localizer intercept heading.

Location of shortcut for flexibility Option 3 (Approach pair D)

The location of the shortcut for flexibility Option 3 (a 'direct to' via either route A or B), did not seem to have an effect on pilot TDL: there was no significant difference between the pilots' RSME scores when guided via route A with an energy rate demand smaller than 1 (approach 4) (M=-0.09, SE = 0.196) or when guided via route B with an energy rate demand smaller than 1 (approach 3) (M=-0.37, SE = 0.305, $t(8) = -.733$, $p = .49$, $r = .26$). The location of the shortcut for flexibility option 3 thus does not seem to have an effect on pilot TDL.

Location of shortcut for flexibility Option 2 (Approach pair E)

On average, the pilots' RMSE z-scores were higher for the Localizer intercept part of approach 13, which comprised a vectored shortcut via route A, with a resulting line-up distance of 8.3nm, (M=0.35, SE = 0.327) than for the Localizer intercept part of approach 14, comprising a vectored shortcut via route B, with a resulting line-up distance of 12.3nm (M=-0.416, SE = 0.216, $t(4) = 3.339$, $p < .05$, $r = .86$). This indicates that pilot TDL increases when the location of the shortcut for flexibility option 2 is closer to the runway (resulting in a shorter line-up distance), it should be noted, however, that this is based on only 5 datapoints.

Introducing a shortcut for flexibility Option 2 (Approach pair F)

There was no significant difference between the RSME z-scores for the total approach for approach 14 (M=-0.24, SE = 0.209) and approach 19 (M=-0.23, SE = 0.137, $t(6) = -0.024$, $p = .98$, $r = .01$).The difference between the two approaches was, that for approach 14 route C was the published route, and during the flight the crew was vectored via route B. For approach 19, route B was the published route, and the crew was also allowed to follow route B. The two routes that were actually flown during approaches 14 and 19 were therefore exactly identical. It thus seems that vectoring the aircraft off the published route using flexibility Option 2 has no effect on pilot TDL.

Energy rate demand too high in Localizer intercept part for Option 2 (Approach pair B)

Approach 14 is a continuous descent approach, and when the crew is vectored towards route 'B' according to flexibility Option 2 the energy rate demand for the Localizer intercept heading part becomes larger than 1. Approach 15 has the same groundtrack as approach 14, but is a horizontal approach, and when vectored along route 'B' the energy rate demand remains below 1. Remarkably, there was no significant difference between the RSME z-scores for the Localizer Intercept heading part between approach 14 (M = -0.33, SE = 0.220) and approach 15 (M = 0.39, SE = 0.463, $t(5)$ = -1.54, p = .19, r = .56), despite the difference in energy rate demand. A significant difference, however, was found for the RSME z-scores for the final part of the approach between approach 14 (M = -0.21, SE = 0.311) and approach 15 (M = 0.94, SE = 0.396, $t(6)$ = -3.02, p < .05, r = .78). Naturally, the high energy rate demand on Localizer intercept heading also has an influence on the final part of the approach, the results show that pilot TDL during the final part of the approach increases due to a high energy rate demand on Localizer intercept heading.

Energy rate demand too high in first part of approach for Option 1 (Approach pair C)

For approaches 1 and 6 the crew was initially cleared for the longest route (route C) of flexibility option 1, prior to reaching the IAF they were re-cleared for route B. Due to this re-clearance the energy rate demand during the first part of the approach for approach 6 became larger than 1, and constraints at the waypoints could not be met. This influence of not meeting the constraints continued into the Localizer intercept part of the approach. For approach 1, the energy rate demand remained below 1 after re-clearance, and the constraints could be met. A significant difference in the pilots RSME z-scores for the first part of the approach was found: the scores for approach 6 (M= 0.53, SE = 0.398) were higher than the RSME z-scores for approach 1, M = -0.19, SE = 0.249, $t(8)$ = -2.02, p (1-tailed) < .05, r = .58. Again this shows that an energy rate demand larger than 1 in the first part of the approach with consequences for the Localizer intercept heading part of the approach increases pilot TDL.

For approaches 2 and 8 the crew was also initially cleared for the longest route (route C) and was now allowed to actually fly route C. For approach 2 the constraints could not be met at three waypoints, for approach 8 the constraints could not be met at only one waypoint. For both approaches, the higher energy rate demand did not have an effect on the Localizer intercept heading part, i.e. before turning to Localizer intercept heading the aircraft was able to meet the constraints again. It was found that there was no significant difference between the pilots' RSME z-scores (for the first part of the approach) for both approaches: approach 2 (M= -0.23, SE = 0.202) and approach 8 (M=-0.32, SE = 0.282, $t(8)$ = 0.309, p = .77, r = .11). Pilot TDL is thus not influenced by the number of waypoints at which the constraints can not be met.

When comparing the two results given above: the pilots' RSME z-scores for the first part of approach 1, which incorporated the shortcut but during which the energy rate demand remained below 1, on the one hand, and the results for approach 8 which did not incorporate a shortcut but followed the original route with energy rate demands larger than 1 on the other hand, it can be seen that there is no significant difference in the RSME z-scores between the two approaches ($t(8)$ = 0.336, p = .75, r = .12). This comparison is slightly ambiguous, because for approach 1 the route is complicated by a commanded change (route "B" instead of "C"), and for approach 8 the route is complicated by higher energy rate demands, these two effects might cancel each other, and therefore the approaches might have the same RSME z-scores. However, based on observations during the experiment, and based on conversations with pilots the following conclusion is presented: both the high energy rate demands that do not influence the Localizer intercept part of the approach, and the change in route (provided the pilot is given sufficient time to reprogram the FMS) do not increase pilot TDL.

Less distance available on Localizer Intercept Heading (Approach pair I)
On average, there was a difference in the RSME z-scores for the Localizer Intercept Heading part of the approach for approach 2, which had 4 nm available on Localizer Intercept Heading (M=-0.28), and the RSME z-scores for the Localizer Intercept Heading part of the approach for approach 16, which had 2 nm available (Mdn=0.30, T = 1.5, p (1-tailed) < .05, r = .43). This implies that pilot TDL increases with decreasing distance available on Localizer Intercept Heading.

Larger Localizer intercept angle (Approach pair H)
The pilots' average RSME z-scores for the Localizer Intercept Part of the approach for approach 4, with a Localizer intercept angle equal to 50 degrees (M=0.16, SE = 0.203), were significantly higher than the average RSME z-scores for the Localizer Intercept Part of the approach for approach 9, with a Localizer intercept angle equal to 38 degrees (M=-0.67, SE = 0.134, t(7) = -3.03, p <.05, r = .75). Indicating that pilot TDL increases with increasing Localizer intercept angle. Although the final part of the approach for approaches 4 and 9 was exactly identical, a difference in RSME z-scores could also be observed for this part: approach 4 was rated significantly higher (M=-0.12, SE = 0.344) than approach 9 (M=-1.13, SE = 0.202, t(5) = -4.42, p <.01, r = .89). It therefore seems that the larger Localizer intercept angle also had an effect on the TDL the pilots experienced during the final part of the approach.

Higher Localizer intercept speed (Approach pair G)
The Localizer intercept part of approach 10, which required a Localizer intercept speed of 230 knots, was rated significantly higher as the RSME z-scores indicate (M=-0.19, SE = 0.152) than the Localizer intercept part of approach 9, requiring a Localizer intercept speed of 200 knots (M=-0.67, SE = 0.134, t(7) = -3.01, p <.05, r = .75). A higher Localizer intercept speed thus seems to increase pilot TDL.

Higher airspeed during final part of approach (Approach pair J)
Due to the difference in Localizer intercept speed, approaches 9 and 10 also have a difference in speed on final and speed at the FAF. The pilots' RSME z-scores for the final part of these approaches show that approach 10, with a initial speed on final equal to 230 knots and 210 knots at the FAF, was rated significantly higher (M=-0.44, SE = 0.266) than approach 9, with an initial speed of 200 knots and 190 knots at the FAF (M=-1.07, SE = 0.179, t(6) = -3.45, p <.05, r = .82). This indicates that a higher initial speed on final coupled to a higher speed at the FAF increases pilot TDL. It should be noted that for both approaches the energy rate demand was well below 1, and that a stabilized approach at 1,000' could be achieved.

Energy rate demand too high during final part of approach (Approach pair K)
The final part of the approach for approaches 2 and 6 was identical and was published with the same speed constraints at the IF and FAF. For approach 2 these published (required) speed constraints could indeed be met, for approach 6, however, the preceding part of the approach caused the aircraft to arrive at the IF with a speed higher than the published speed, as a result the speed constraints at the IF and FAF could not be met. Or, in other words, the energy rate demand for approach 6 was larger than 1. As a result, the pilots' RSME z-scores for approach 6 were significantly higher (M=0.86, SE = 0.289) than for approach 3, M=-0.48, SE = 0.231, t(7) = -4.11, p <.01, r = .84, indicating that pilot TDL increases when the energy rate demand during the final part of the approach is larger than 1.

Smaller distance between IF and FAF and lower FAF altitude (Approach pair L)
The final parts of the approach for approaches 2, 3, 5 and 9 were set-up as a two-level factorial design for the independent measures FAF altitude (1,600' and 3,000') and distance available between IF and FAF (1.3nm and 5.5nm). The energy rate demand was smaller than 1 for all approaches, and the Localizer intercept speed was equal to 200 knots for all approaches but approach 3. For approach 3 a Localizer intercept speed of 180 knots was used, due to the short line-up distance resulting from the combination of low FAF and short distance between IF and FAF. Localizer intercept angles were comparable for all four approaches.

The RSME z-scores for these approaches were compared using a factorial repeated measures ANOVA. Since there were only two levels per independent measure, sphericity was not an issue. There was a significant effect on the pilots' RSME z-scores of both the distance between IF and FAF, $F(1,5) = 16.65$, $p < .05$, $r = .88$, as well as the FAF altitude, $F(1,5) = 4.66$, $p(\text{one-tailed}) < .05$, $r = .69$, in the sense that a higher FAF altitude or a larger distance between IF and FAF resulted in lower RSME z-scores, and hence decreased pilot TDL. There was no significant interaction effect, meaning that the distance between IF and FAF had the same effect on the RSME z-scores, independent of the FAF altitude it was coupled to: a smaller distance between IF and FAF would always result in higher RSME z-scores than a larger distance between IF and FAF both when coupled to a high FAF altitude or a low FAF altitude. Since the combination of FAF altitude and distance between IF and FAF completely defines the line-up distance, it can be concluded that as a result a larger line-up distance decreases pilot TDL.

Appendix B to Journal Article 4

Flaps 1 and flaps 5

When comparing the approaches in the current flight simulator experiment that were flown with an aircraft mass of 237,000kg (resulting in an UP mark at 223 kts, the speed below which flaps 1 should be selected, and a 1 mark at 203 kts, the speed below which flaps 5 should be selected) to the approaches with an aircraft mass of 237,000kg in the previous experiment, it was found, using Mann Whitney tests, that the reaction time distributions for selecting flaps 1 and flaps 5 differed significantly between the two experiments. On average (see Table 15) flaps 1 were selected at a lower airspeed, and flaps 5 were selected at a higher airspeed during the current experiment than during the previous experiment. It should be noted that for all comparisons in this paragraph the original distributions for both flight simulator tests are used, not the approximating normal curves in Table 15.

Within the current experiment some approaches (approaches 13, 14 and 19) were flown with a higher aircraft mass of 295,000 kg (resulting in an UP mark at 241 kts, and 1 mark at 221 kts). Although the UP mark has a higher value it was found (using a Wilcoxon signed-rank test) that the distribution of speeds *relative* to the UP mark for the higher aircraft weight *Mdn* = -0.79) did not seem to differ from the speed relative to the UP mark at which flaps 1 were selected for the lower aircraft mass (*Mdn* = -2.34 knots, T = 137, p = .49, r = .1). The same was true for the selection of flaps 5: it appeared that there was no significant difference in this distribution *relative* to the 1 mark between the high aircraft mass (*Mdn* = 16.06) and the low aircraft mass, *Mdn* = 14.89 knots, T = 161, p = .98, r = -.00.This is a very useful result since it implies that, when the distribution around the UP mark is known for one aircraft weight, it is possible to use the same distribution around different UP marks for different aircraft weights.

Actions on Localizer Intercept Heading

Results of the current flight simulator experiment revealed a correlation between the heading change required towards Localizer intercept heading and the reaction times recorded (all significantly non-normal) for selecting flaps 10 (τ = .28, p < .001), ARM approach (τ = .21, p < .001), Autopilot off (τ = .30, p < .001) and Heading select (τ = -.17, p < .01) when using the *start* of the turn towards Localizer intercept heading as a trigger event. Since a larger heading change increases the amount of time the aircraft spends in the turn, thereby consistently increasing the reaction times. A better trigger event appeared to be the *end* of the turn towards Localizer intercept heading, this is the trigger event used for the remainder of this analysis.

4

A small to medium effect was found of the amount of time available on Localizer intercept heading on the reaction times for the actions selecting flaps 10 (τ = .16, p < .05), ARM approach (τ = .17, p < .01) and Autopilot off (τ = .17, p < .01), indicating that the reaction times decreased when the distance available decreased. This correlation was not found during the earlier experiment [13].

A one-way repeated-measures ANOVA revealed that there was (at least partially) a preferred order in which the actions on Localizer Intercept heading were performed, F(2.4, 341.9) = 7.32, p < .001, ω^2 = .02. Since Mauchly's test indicated that the assumption of sphericity was violated ($\chi^2(5)$ = 48.7, p < .001), the degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity (ε = .803). Post Hoc tests using the Bonferroni correction showed that, on average, flaps 10 were selected before ARM approach was selected (p < .01), and that Heading select was selected before ARM approach was selected (p < .001), this agrees with the earlier experiment [13].

To compare the distributions for the reaction times during the current experiment to those found during the earlier experiment, the reaction times for the approaches with 4nm available on Localizer intercept heading are compared. For both experiments the *end* of the turn towards Localizer intercept heading is used as trigger events. The reaction times for selecting flaps 10, autopilot off and heading select appeared to be significantly shorter during the current experiment than during the earlier experiment, see Table 15. No significant difference was found for the reaction times for selecting ARM approach. Additionally, no significant difference was found between the two experiments for the reaction times for Autothrottle off with respect to the trigger event Autopilot off.

FAF actions

The trigger event for selecting flaps 20 and Gear down is reaching the FAF [13]. It was found that the FAF altitude (significantly non-normal, $D(150) = 0.264$, $p < .001$) did not have an effect on the reaction times for selecting flaps 20 (significantly non-normal, $D(150) = 0.247$, $p < .001$), t (150) = .057, $p = .36$, or the reaction times for selecting gear down (significantly non-normal, $D(150) = 0.265$, $p < .001$), τ (150) = .038, $p = .55$, neither did the tails of the distributions seem to differ with FAF altitude. This corresponds to the results from [13]. When comparing the values of the reaction times of the current flight simulator experiment to the earlier experiment, it can be stated that the reaction times for selecting flaps 20 in the current experiment did not seem to differ from the reaction times in the earlier experiment (Table 15). The reaction times of the current experiment for selecting gear down, however, were significantly longer than found during the earlier experiment, see Table 15.

1,200′ actions

Flaps 25 are selected when reaching 1,200′ [13]. During the current flight simulator experiment, flaps 25 were not selected during one approach, hence $N = 149$. The reaction times for selecting flaps 25 recorded during the current experiment (significantly non-normal, $D(149) = 0.165$, $p < .001$), are significantly shorter than the reaction times for selecting flaps 25 during the earlier flight simulator experiment (significantly non-normal, $D(172) = 0.107$, $p < .001$), see Table 15.

Journal Article 5

Pilot Task Demand Load during RNAV approaches with a Cessna Citation

K.T.P. van Bennekom, T.J. van Tuinen, M.M. Heiligers, Th. van Holten and M. Mulder

Submitted to the Aeronautical Journal

Abstract

This research aims to develop a method which predicts the task demand load as experienced by pilots while flying an Area Navigation (RNAV) approach. First, this will yield insight in which aspects of an approach actually influence pilot task demand load. And second, during the design of approaches this method can be used to rapidly evaluate a potential approach and to 'optimize' an approach with respect to pilot task demand load. During previous research, focusing on approaches flown with a B747, a list of factors that influence pilot task demand load has been obtained, as well as a method to keep pilot task demand load at an acceptable level. The method consists of seven guidelines to be adhered to during approach design. This paper shows that the list of factors and the method do not only apply to a B747 aircraft but are generally applicable to other aircraft as well. This is underpinned by results from both flight simulator tests and real flight tests with TU Delft's Cessna Citation laboratory aircraft. Additionally, it is shown that there are no discrepancies between the list of factors influencing pilot task demand load resulting from the flight simulator tests and the list of factors resulting from the real flight tests.

1 | Introduction

This research aims to develop a method which predicts the task demand load (TDL) as experienced by the pilot while flying an approach. TDL is defined as the mental workload *imposed* by the system to be controlled or supervised [1]. As opposed by mental load, the workload experienced by a particular operator. First, this will yield insight in which aspects of an approach actually influence pilot TDL. And second, during the design of approaches this method can be used to rapidly evaluate a potential approach and to 'optimize' an approach with respect to pilot TDL.

The rationale within this research is that approaches should be designed such that they can be flown according to Standard Operating Procedures (SOPs) and that a stabilized approach at 1,000ft can be achieved. This is based on the conclusions of the Flight Safety Foundation Approach-and-landing Accident Reduction Task Force [2]. These conclusions read, amongst others, that 'Establishing and adhering to adequate SOPs and flight-crew decision-making processes improve approach-and-landing safety' and that 'Unstabilized and rushed approaches contribute to approach-and-landing accidents'. Therefore, within this research, pilot TDL is predicted for approaches while flying according to SOPs and while aiming to achieve a stabilized approach.

The approaches considered in this research are Area Navigation (RNAV) approaches or, more specifically, RNAV transitions since the final part of the approach is guided by the Instrument Landing System (ILS). The approaches are flown using the Flight Management System (FMS), Autothrottle and Autopilot with Vertical Navigation (VNAV) and Lateral Navigation (LNAV) modes. On Localizer intercept heading the autothrottle and autopilot are switched off, and the remainder of the approach is flown using the Flight Director (FD).

Given the level of automation described above, given a certain aircraft with its corresponding SOPs, and given a certain approach, we aim to map pilot TDL and the factors that contribute to pilot TDL. The factors contributing to pilot TDL considered in this research are the properties of the approach trajectory and its speed and altitude constraints (for instance, the Localizer intercept speed or distance available on Localizer Intercept Heading), the meteorological conditions (wind direction and wind speed) and the flight mechanical properties of the aircraft (for instance, how easy it is to dissipate energy) [3]. To investigate pilot TDL we thus focus on factors that can be described as 'the environment' of the pilot, instead of focusing on the constraints of the pilot himself (like memory capacity, time delay, etc.) [3]. In this respect our work is influenced by the principles of cognitive work analysis [4]. This approach deliberately deviates from the idea behind models such as the Procedure-Oriented Crew model (PROCRU) [5, 6] or the Man-Machine Integrated Design and Analysis System (MIDAS) [7-9] that use human operator models which do focus on the constraints of the human operator. It is anticipated that by focusing on the environment of the pilot instead of on the limitations of the pilot himself much simpler models can be achieved to predict pilot TDL than by using human operator models.

During previous research [10-14], factors have been identified that influence pilot TDL for pilots flying an RNAV approach with a B747. These factors were identified based on flight simulator tests. In this paper it will be demonstrated that the same factors also influence pilot TDL when flying an approach with a Cessna Citation. This indicates that the set of factors that has been identified is a generally applicable set of factors, and not only valid for the B747. Additionally, it is investigated whether the same set of factors influences pilot TDL during flight simulator tests and during real flight. To this end, in this paper, both the results of a flight simulator experiment for a Cessna Citation and the results of real flight tests with a Cessna Citation aircraft are compared. It will be demonstrated that the same set of factors results from both experiments. Finally, the simulation tool that was developed for the B747 in order to analyze an approach with respect to the factors that were proven to influence pilot TDL is adjusted in order to include the Cessna Citation. It will be shown that the simulation tool also works and provides reliable results for the Cessna Citation.

This paper is structured as follows. First, the basic principles of this research are introduced as well as the scope of the research. After that, the results of previous research [11-13] which focused on the B747 are briefly explained. Subsequently, the human in the loop experiments are presented, these experiments are conducted for the Cessna Citation aircraft both in a flight simulator and during real flight tests. To conclude, the simulation tool that is adjusted to include the Cessna Citation is explained, and its predictions are illustrated by a case study.

2 | Basic principles of this research

At the heart of the project lies the development of a method that will provide guidelines that can be used during approach design in order to keep pilot TDL during the approach at an acceptable level. In order to analyze whether a newly designed approach actually meets all the guidelines, a computer simulation program is developed. This simulation program incorporates the aspects that affect pilot TDL during approach, including standard operating procedures, altitude-profiles, velocity-profiles, etcetera. It should also be possible to enter different types of aircraft and to change the meteorological conditions (turbulence intensity, amount of wind). These properties are the descriptors of the environment that form the "input" of the computer program as they constitute the specific characteristics of the approach to be evaluated. The "output" of the simulation program is an indication whether the guidelines to keep pilot TDL at an acceptable level are met. This section will explain the basic principles of the method and computer simulation, the assumptions and the choices that have been made as to what is incorporated within this research, and also what is considered to be beyond the scope of this research.

2.1 | Factors of the air transport system included

Many different factors and the interactions between those factors have an influence on the execution of an approach, see Figure 1. This research concentrates on the "pilot" box in Figure 1. It will, e.g., not consider the Air Traffic Controller's TDL. To determine pilot TDL, this research will only take into account the factors that have a direct influence on an approach (see Figure 1), most importantly the characteristics of the trajectory, the type of aircraft and the meteorological conditions.

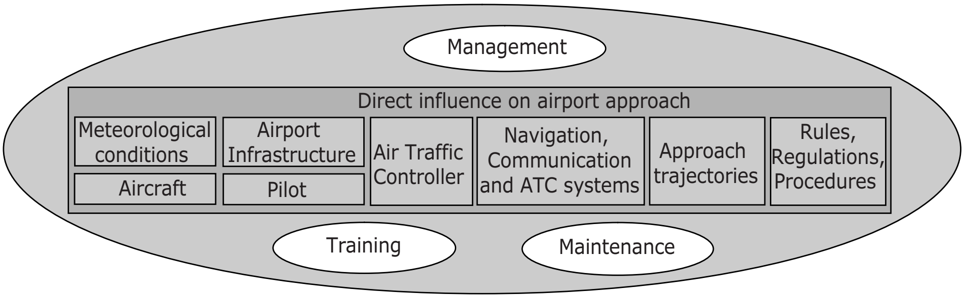


Figure 1 | Direct and indirect factors that influence the safety of airport approaches.

2.2 | Task Demand Load

This research aims to develop a method to predict pilot TDL. Task Demand Load is defined as the mental workload *imposed* by the system to be controlled or supervised [1], see also Figure 2 The TDL is not to be mistaken for the mental workload *experienced* by the human operator, which is referred to as Mental Load (ML). Many of the well-known methods to measure workload, like the NASA Task Load indeX, measure ML, not TDL.

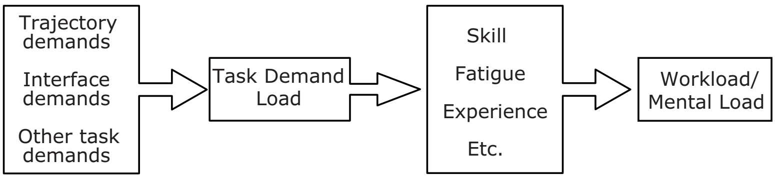


Figure 2 | Difference between Task Demand Load and Workload, adapted from [14].

Within this research several experiments are performed during which pilots are asked to comment on approaches regarding the amount of effort these approaches require, or their effect on the difficulty as experienced by the pilot. When pilots give their opinion on these matters, they obviously base their opinion on the mental workload they experienced. This results in the situation that in order to obtain information about the Task Demand Load, pilots are asked about the mental workload they experienced during the experiments, unfortunately there is no other way. However, by choosing pilots with different levels of experience etc., by testing the approaches in random order and by converting the pilots' ratings to z-scores it is assumed that through the comments of the pilots a good indication of the Task Demand Load can be obtained.

2.3 | Approaches considered and automation used

Obviously, the TDL depends directly on the type of approach that is considered. This research focuses on Area Navigation (RNAV) Approaches. Although it is appreciated that non-precision approaches such as Non-Directional Beacon (NDB) approaches are, in general, more difficult for a pilot to fly than RNAV approaches [16], a deliberate choice is made to focus on RNAV approaches only, since these are expected to become more and more frequently used in the future. The last part of the RNAV approach is assumed to be flown using the Instrument Landing System (ILS).

The part of the flight that is considered starts at the Initial Approach Fix (IAF) and comprises the entire approach (Initial Approach, Intermediate Approach and Final Approach) until 1,000 feet above airport level, see Figure 3. Based on interviews with pilots it was decided to use two different levels of automation during the approach: until Localizer Intercept Heading the approach is flown using the FMS, Autopilots and Autothrottle. At Localizer Intercept Heading (but before Localizer capture) the pilot switches to Flight Director (FD) mode and disconnects the Autothrottle, the remainder of the approach is thus flown using the FMS and FD, which implies manual control by the pilot.

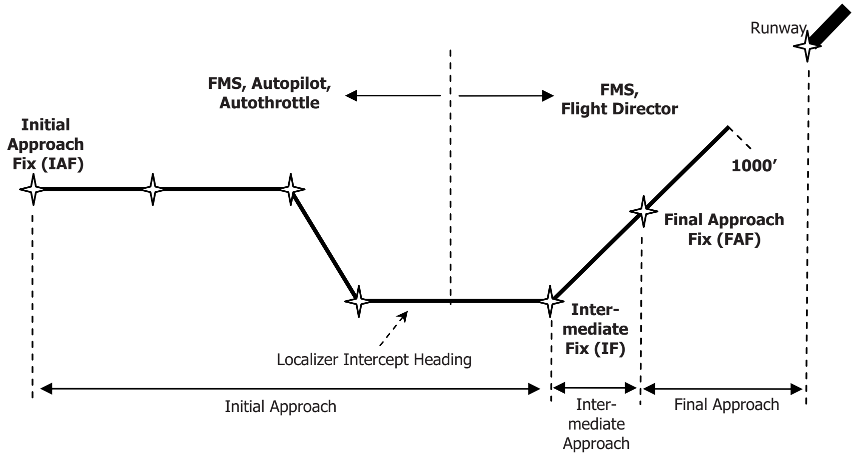


Figure 3 | Part of flight considered (top view) and automation used

2.4 | Non-nominal conditions and emergencies

Non-nominal conditions and emergencies such as engine failure are not considered in this research. The goal is to determine pilot TDL for published RNAV approaches under nominal conditions. When any emergencies such as engine failure occur, the crew will most likely not be required to follow the RNAV approach anyway, but will be vectored to the runway in the most convenient way.

Additionally, the assumption for less severe non-nominal situations is that when flying under nominal conditions, the RNAV approach should provide enough ‘margin’ with respect to pilot TDL, such that the pilot has enough spare capacity and time to deal with non-nominal conditions. This implies that the TDL that is predicted by this research for nominal conditions should be well below the absolute maximum TDL a pilot can cope with in order to guarantee this margin.

2.5 | Boundary conditions: Stabilized approach and Standard Operating Procedures

The TDL experienced by the pilot also depends on the boundary conditions that are set, e.g. the accuracy with which the approach needs to be flown. The boundary conditions chosen for this research are that the approach should be performed according to Standard Operating

Procedures and that pilots should aim to achieve a stabilized approach at 1,000 feet above airport elevation. This decision is based on the conclusions of the ALAR Task Force [2].

To determine whether a stabilized approach is achieved at 1,000 feet, the following nine criteria [2] are used:
1. The aircraft is on the correct flight path;
2. Only small changes in heading/pitch are required to maintain the correct flight path
3. The aircraft speed is not more than VREF + 20 knots Indicated Airspeed (IAS) and not less than VREF[1]*;
4. The aircraft is in the correct landing configuration;
5. Sink rate is no greater than 1,000 feet per minute; if an approach requires a sink rate greater than 1,000 feet per minute, a special briefing should be conducted;
6. Power setting is appropriate for the aircraft configuration and is not below the minimum power for approach as defined by the aircraft operating manual;
7. All briefings and checklists have been conducted;
8. Specific types of approaches are stabilized if they also fulfill the following: instrument landing system (ILS) approaches must be flown within one dot[2] of the glide slope and localizer; a Category II or Category III ILS approach must be flown within the expanded localizer band; during a circling approach, wings should be level on final when the aircraft reaches 300 feet above airport elevation; and
9. Unique approach procedures or abnormal conditions requiring a deviation from the above elements of a stabilized approach require a special briefing.

2.6 | Level of detail of computer simulation models

As briefly explained in the introduction, it is the goal to incorporate very detailed models of the *environment* of the pilot in the computer simulation, and to add to this a rather simple model for the pilot. Therefore, the aircraft with its kinematic and dynamic constraints, the 3-D properties of the trajectory, the velocity profile, turbulence, wind, etcetera, in other words: the factors that have a direct influence on an approach as given in Figure 1, are modeled as detailed and accurate as possible. Whereas the pilot model is kept as simple as possible. This simple pilot model consists of a manual control model (which in effect only contains a pure gain plus time delay) and a model for performing actions such as selecting flaps and gear according to the SOPs.

3 | Results of previous research for B747

Using these basic principles and assumptions, a method (consisting of guidelines to keep pilot TDL at an acceptable level) and two computer simulations have been developed for the B747. The results are briefly explained in this section.

Based on two sets of B747 flight simulator experiments with nine B747 pilots participating in each experiment [12, 13], a list of factors has been identified that influence pilot TDL during approach. This list of factors is considered to be complete, which means that there are not any other factors that fall within the previously defined scope of this research that influence pilot TDL. The factors that influence pilot TDL are grouped per approach part (see Figure 4) and can be summarized as follows:

1 The Reference Speed (VREF) is defined as 1.3 times the stall speed.

2 One dot deviation on the glide slope equals 0.7° beam error, one dot deviation on the Localizer equals 2.5° beam error.

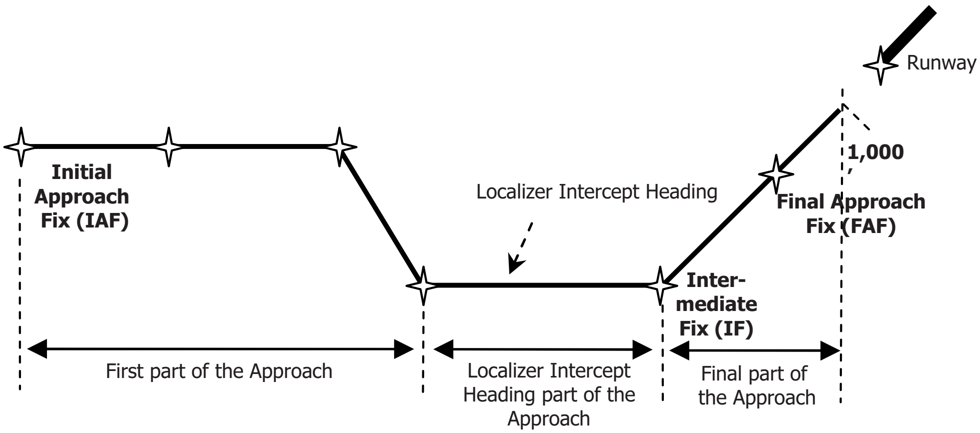


Figure 4 | Division of the approach into three parts

For the first part of the approach:

❖ The major contributor to pilot TDL is the fact whether or not the altitude and velocity constraints can be met at the waypoints. This is only true when the effect of not meeting the constraints continues into the Localizer part or final part of the approach. If the consequences of not meeting the constraints remain within the first part of the approach this does not influence pilot TDL.

❖ The number of waypoints, number of heading changes and the altitude profile (horizontal approach, CDA, stepped approach) do *not* influence pilot TDL. This is due to the fact that this part of the approach is flown in LNAV and VNAV modes with autopilot and autothrottle.

For the Localizer intercept part of the approach:

❖ The time available to perform all actions (which is directly related to the distance available on Localizer intercept heading) is the most important factor for pilot TDL. Actions that need to be performed for the B747 on Localizer intercept heading are: select flaps 10, select heading select, arm the approach, and (due to the choices made for this research, see Figure 2) switch off the autopilot and autothrottle.

❖ Next to this, pilot TDL is also influenced by the Localizer intercept speed, the Localizer intercept angle, and whether the constraints at the waypoints can be met.

For the final part of the approach:

❖ The most important factors influencing pilot TDL seem to be whether or not a stabilized approach can be achieved at 1,000′, the distance between IF and FAF and the airspeed on final. Whether an approach is stabilized can, for a B747, be determined from: 1. whether the constraints at the waypoints can be met during the final part of the approach, 2. the value of the vertical speed (should be below the sink rate warning) which in itself is a function of airspeed on final and glideslope angle, and 3. the FAF altitude and distance between IF and FAF (resulting in the line-up distance) since these two factors together determine whether there is enough time available to perform all actions required for a stabilized approach. All these factors thus influence pilot TDL during the final part of the approach.

The method to predict pilot TDL during approach for a B747 is based on the above factors. The method basically consists of seven guidelines for the design of approaches. When these guidelines are met, pilot TDL during the approach will be acceptable. Starting point for the guidelines is that pilots should fly the approach according to SOPs and that they should aim to achieve a stabilized approach at 1,000'.

Concluding, the guidelines for the contributors to pilot TDL for the B747 then are that:
1. aircraft should be able to meet the altitude and airspeed constraints throughout the approach, especially during the final part of the approach, and during the first part of the approach if this has consequences for the subsequent parts of the approach;
2. there should be sufficient time to perform all actions on Localizer intercept heading;
3. it should be possible to achieve a stabilized approach. Whether a stabilized approach can be achieved for a B747 depends on 1. whether the aircraft can dissipate enough energy during the final part of the approach, 2. the value of the vertical speed (should be below the sink rate warning) which in itself is a function of airspeed on final and glideslope angle, and 3. the FAF altitude and distance between IF and FAF (resulting in the line-up distance);
4. the distance between IF and FAF should be sufficient;
5. the vertical speed should be below the sink rate warning;
6. the Localizer intercept speed should not be too high, and that
7. the Localizer intercept angle should not be too large.

A quantification of these guidelines is given in [13].

It is hypothesized that the same factors will also determine pilot TDL for other aircraft types. In this respect it should be noted that the factors 'FAF altitude' and 'time available on Localizer intercept heading' are factors that originate from the fact that for the B747 the SOPs require pilots to perform a number of actions on Localizer intercept heading and between the FAF and 1,000ft, and that they should have sufficient time to do so. If, for another aircraft type, these actions are required to be performed in another part of the approach, then care should be taken that sufficient time is available in that particular part of the approach. In that case, the factors 'time available on Localizer intercept heading' and 'FAF altitude' might not influence pilot TDL for that particular aircraft type. All other factors in the list above are assumed to be valid for all aircraft types.

In order to obtain a prediction whether these guidelines are met for an approach, a comprehensive Monte Carlo computer simulation was developed for the B747, together with a relatively simple point mass model computer simulation [11-13]. By running either of these computer simulations for a specific approach and analyzing their output combined with the requirements in Procedures for Air Navigation Services Aircraft Operations (PANS-OPS) [17], an indication can be obtained of whether or not the guidelines as described above are met [11-13].

Questions arising from this previous research are whether the guidelines outlined above are indeed valid for other aircraft types, and whether the fact that they were obtained from flight simulator tests might cause discrepancies with guidelines for real flight. To find an answer to these questions, this paper describes flight simulator experiments and real flight tests with a Cessna Citation aircraft. Given the goal of this research (to develop a method to keep pilot TDL at an acceptable level during RNAV approaches) the choice for a Cessna Citation aircraft might not be obvious, since it is not among the aircraft types that will most frequently fly these kinds of approaches. However, it is the only aircraft type for which non-linear aerodynamic models were available for the simulations, and the only aircraft that was available for flight tests. Therefore this paper concentrates on the Cessna Citation. It should be noted in this respect that the particular aircraft that was used does not have a VNAV mode or an autothrottle. When comparing the results found for the Citation to the results found for the B747 this should be kept in mind.

4 | Human in the loop experiment

Two separate human-in-the-loop experiments were performed for the Cessna Citation aircraft which involved pilots flying different approaches under varying conditions. The first experiment was conducted in a six-degree-of-freedom flight simulator, while the second was conducted in a Cessna Citation II aircraft.

4.1 | Experiment goal

The first goal of the experiment was to test whether the same factors that were found to influence pilot TDL for the B747 [14] are general factors, and also influence pilot TDL for other aircraft, in this case a Cessna Citation. The second goal was to investigate whether the factors that influence pilot TDL during flight simulator tests also influence pilot TDL during real flight.

4.2 | Method of experiment 1, flight simulator tests

4.2.1 | Apparatus and Citation model

The experiment was performed in the six-degree-of-freedom TU Delft SIMONA Research simulator (SRS), see Figure 5. The Cessna Citation aerodynamic models as well as the yaw damper are based on the Cessna Citation I [18]. Autopilot modes available during the experiment were: LNAV, Heading Select, Altitude Hold, Vertical Speed and Indicated Airspeed (IAS). In Flight Director operation additional modes available were Glide slope mode and Localizer mode.

The autopilot and flight director models are based on the autopilots developed for the model of the B747 [13], which were derived from [19]. The LNAV mode is based on the VOR modes described in [19].

All approaches that were flown, were pre-programmed in the FMS/CDU. The appropriate approach was loaded in the FMS before the start of the approach, and during the experiment pilots could switch between the 'Progress' and 'Legs' pages, but could not use the CDU interactively, or modify the approach.

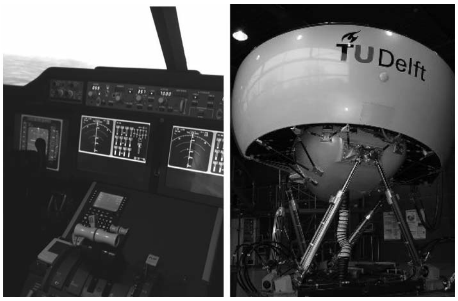


Figure 5 | The SIMONA Research Simulator

There were some discrepancies between the SRS and a Citation that are of importance to the experiment. First, the cockpit lay-out in the SRS differed from reality (see Figure 5). Second, the aircraft was not trimmed when the pilot switched from autopilot to flight director. Third, the altitude capture was slightly abrupt as compared to reality. This resulted in a minor 'bubble feeling' when an altitude was captured. Fourth, with respect to the flight director, at LOC intercept the flight director over exaggerated the bank angle required for a correct intercept. In practice this meant that during LOC intercept the pilots rolled the aircraft to a correct bank angle and waited for the FD bars to return to the centre position. All pilots were briefed about these discrepancies before the tests commenced. Also, all these aspects were the same for all pilots and constant during all runs.

4.2.2 | Subjects and instructions

Six Citation pilots participated in the experiment. The pilots were paired up to form a crew of pilot flying (PF) and pilot monitoring (PM). The task of each pilot was to fly 10 different approaches as PF and the same 10 approaches as PM, starting at the Initial Approach Fix (IAF) and ending at around 800’ above airport level (AAL).

Pilots were instructed to fly the approaches according to SOPs. The SOPs are used as stated for the Aircraft Operations Manual of the Cessna Citation II [20], see Figure 6. It is assumed that before every run the approach checklist has been completed. On Localizer intercept heading pilots should switch to heading select mode, arm the approach and switch from AP to FD (the latter requirement is not prescribed by SOPs but is based on the choices made for this research). The pilots are then required to select Flaps APP between 2 nm and 0.5 nm from the FAF. On the FAF the Gear is selected, followed immediately by the landing checklist. At 1,200’ above the airfield level Flaps LAND are selected.

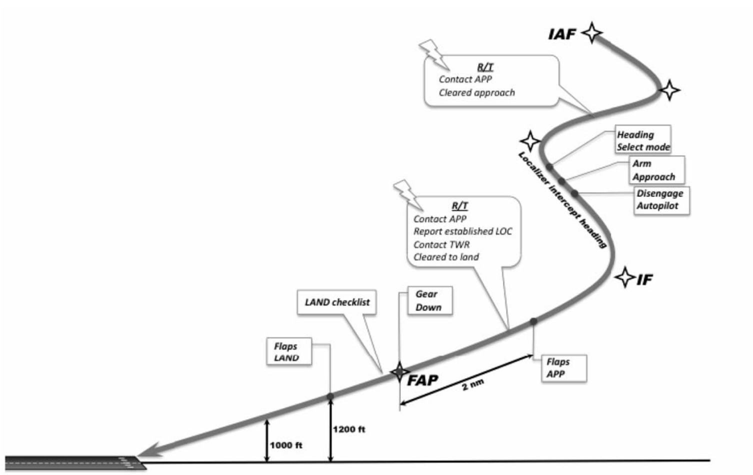


Figure 6 | Standard Operating Procedures for Cessna Citation

Two weeks before the experiment the pilots received a briefing by mail. On the day of the experiment they were briefed as well. The pilots were asked to adhere very strictly to SOPs, even if they could foresee that by adhering to SOPs they would not meet certain constraints at waypoints or would end up unstabilized at 1,000’. Additionally they were asked to perform their tasks according to the principles of Multiple Crew Coordination (MCC) and to fly passenger comfort. They were briefed about the discrepancies between the SRS and the Citation (as explained in the previous paragraph), and were informed that there would be no emergencies (e.g. engine failure) during the flight. They were told that they could fly the approach as published on the approach and landing charts, implying that ATC would not interfere. The pilots were not allowed to use speedbrakes, since approaches should be designed such that they can be flown without the use of speedbrakes.

4.2.3| Procedure

Before starting the experiment the pilots could familiarize themselves with the SRS and their task during three to five (depending on the pilot) practice approaches. After that the experiment started. Before every approach the pilots could take as much time as they thought necessary to study the approach and landing charts, to brief the approach and to prepare the SRS for the next approach. The simulation was started when the pilots indicated they were ready.

After every approach the pilots (PF) were asked to fill in a run questionnaire. Each run questionnaire consisted of three parts: the first part required a rating of the total approach on the Rating Scale Mental Effort (RSME)[21], see Figure 7 and required additional RSME ratings for the three individual parts of the approach, resulting in three RSME sub-ratings per approach. The RSME is constructed according to the 'magnitude estimation' method [22] and can therefore be regarded as interval data. The Dutch version of the scale (which was also used for this research) was used and validated in [21, 23]. Though originally intended to measure only one aspect of a task, it is used here to get an indication of the total task because of its simplicity and ease of use when compared to, for example, a NASA TLX rating procedure [24].

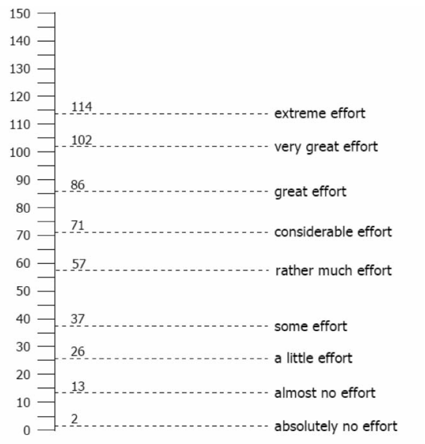


Figure 7 | Rating Scale Mental Effort, during the experiment the Dutch translation was used.

The second part of the run questionnaire contained two questions asking the pilot's opinion on whether the pilot would have adhered to SOPs during real flight, and whether the pilot would have used speedbrakes during real flight. The third part of the run questionnaire consisted of specific closed format questions per approach part regarding the factors hypothesized to influence pilot TDL in that specific approach part. For an example, see Figure 8. To analyze the pilots' answers, the response options for all questions have been coded from 1 – 5, and are regarded interval data. Although there is much controversy about whether these response options can be considered ordinal or interval data [25, 26], it is, in this case, deemed appropriate to treat the data as interval scale because the response options were arranged horizontally and were equally spaced apart, and the verbal labels connoted more-or-less evenly-spaced gradations, most of them symmetrical about a neutral middle.

The number of waypoints was	Very large	Large	Neutral	Small	Very small
As a result I found this part of the approach	Very difficult	Difficult	No influence	Easy	Very easy

Figure 8 | Example of question from the run questionnaire for the flight simulator tests.

At the end of the day, after all approaches were flown, the pilots (PF) filled in an end of day questionnaire. The first part of the end of day questionnaire regarded the realism of the flight simulator and the realism of the experiment as a whole. The second part contained general questions about factors that might possibly influence pilot TDL during approach.

4.3 | Method of experiment 2, Cessna Citation flight tests

4.3.1 | Apparatus

The second experiment was performed in the Cessna Citation II laboratory aircraft (PH-LAB) see Figure 9. The PH-LAB is jointly owned by Delft University of Technology and the National Aerospace Laboratory (NLR). As with the SRS experiments, the same autopilot modes were available to the pilots and the FMS/CDU was used in the same manner.

The flight tests were performed at Malta International Airport. All tests were performed in either CAVOK (officially this means no clouds under 5,000ft, in practice it was a clear blue sky) or in FEW012 to FEW033, which means that ⅛ to ¼ of the sky is covered with clouds with a cloud base at 1,200ft to 3,300ft, respectively. The few small clouds that were present during the tests did not have an effect on the visibility. Maximum windspeeds encountered during the tests were 12 knots, with a maximum tailwind of 7 knots. During the test week the QNH only varied between 1014 and 1017hPa, which means that there was not much variability in air density, and therefore only a very small effect on the flight mechanics due to this change in air density. All in all it can be stated that the meteorological conditions during the real flight tests were very similar to the conditions during the flight simulator experiment.

Figure 9 | Cessna Citation II laboratory aircraft

4.3.2 | Subjects and instructions

For this experiment the same six pilots participated and the same 10 approaches were flown. All pilots were given the same instructions as were given during the SRS experiments.

4.3.3 | Procedure

The procedure used during the Citation flight tests was similar to the SRS experiments. Before starting the experiment the pilots could familiarize themselves with the Citation and their task during one or two (depending on the pilot) practice approaches. After which the experiment started. Before every approach the pilots could study the approach and landing charts, to brief the approach. The experiment started when the aircraft crossed the IAF and ended around 800’ above airport level (AAL). At this altitude a go-around was initiated and the aircraft was maneuvered to the IAF of the next approach.

5

Obviously the aircraft could not be paused after each approach, so there was considerably less time available to fill in a run questionnaire. Instead, during the go-around, the PM was given control over the aircraft while the PF completed a single page questionnaire. This consisted of an RSME scale for the entire approach and three RSME scales for the different parts. Using the headsets the researchers could then ask some specific questions about the approach for additional information.

When the PF had completed all 10 approaches an end of day questionnaire was filled out. This questionnaire resembles the end of day questionnaire used during the SRS experiments and can be used for extra information later in the analysis.

4.4 | Independent variables and approaches

Considering the time and resources available for this research, 10 custom approaches were flown during the human-in-the-loop experiments. It was chosen to design one benchmark approach and nine approaches for each of which a separate independent variable is changed with respect to the benchmark approach, see Table 1.

Table 1 | Independent variable

	Independent variable	**Linked factor**	**TDL effect**
APP01	Benchmark	Benchmark	
APP02	Short LOC intercept heading	-	+
APP03	Large LOC intercept angle	-	+
APP04	Low FAF (normal line-up distance)	Lower FAF speed Lower LOC int. speed Long distance IF-FAF	+
APP05	Short line-up distance	Lower LOC int. speed Lower IF altitude Short distance IF-FAF Lower FAF altitude Lower FAF speed	+
APP06	Short leg IF-FAF	Lower LOC int. speed Lower IF altitude	+
APP07	High LOC intercept speed	Large line-up distance Higher IF altitude Large distance IF-FAF	+
APP08	Not meeting constraints at WP’s, but stabilized at 1000’	Horizontal GS intercept	+
APP09	Not stabilized due to high speed at low FAF (Eratio>1)	Higher LOC int. speed (Lower FAF altitude Higher FAF speed)	+
APP10	Not stabilized due to high energy at start of approach (Eratio>1)	Higher LOC int. speed Low IF altitude Low FAF altitude Higher FAF speed Short distance IF-FAF Short line-up distance	+

The factor "Eratio" in the second column needs explanation: this is the energy rate demand, which is the ratio between the rate at which the trajectory *requires* the aircraft to dissipate energy, and the rate at which the aircraft *can* dissipate energy. Once this ratio becomes larger than 1 this means that the altitude and velocity constraints at the next waypoint will not be met.

If possible, the independent variable was the only changing factor between the benchmark and the respective approach. However, sometimes, due to changing the independent variable another aspect of the approach also had to be changed. These aspects, if there were any, are listed in the column labeled 'Linked factor'. For example, for APP07 a higher Localizer intercept speed was applied. There are then two options: 1. if the FAF altitude, line-up distance and required airspeed at the FAF are maintained equal to the benchmark approach, this will result in an energy rate demand larger than 1 between IF and FAF, 2. the other option is to keep the energy rate demand the same as in the benchmark approach, but in order to do this the line-up distance (and thus IF altitude and distance between IF and FAF) needs to be increased. While designing the approaches in Table 1 the objective was to keep the energy rate demand for all approaches equal to the benchmark approach (except when the energy rate demand was the independent variable), since previous research [12, 13] showed that when the value of the energy rate demand becomes larger than 1, this has a large influence on pilot TDL. Therefore, for APP07 the second option is chosen for the design.

4.5 | Hypotheses

Regarding the influence of the independent variables the following was hypothesized (see also last column of Table 1):

- A short LOC intercept heading increases pilot TDL
- A large LOC intercept heading angle increases pilot TDL
- A low FAF increases pilot TDL
- A short line-up distance increases pilot TDL
- A short leg IF-FAF increases pilot TDL
- A high LOC intercept speed increases pilot TDL
- Not meeting constraints increases pilot TDL
- Not being stabilized at 1000' increases pilot TDL

5 | Results

Due to the total number of tests that is performed on the data set resulting from the flight simulator experiment and real flight tests, when assuming an original p-value of .05 a corrected p-value smaller than .05 should be used (when applying a Bonferroni correction) as criterion for significance in order to control the Type I error rate (this error represents the situation that an effect is found using a statistical test, whereas in reality there is no effect). However, by correcting the p-value to smaller than .05, statistical power is lost (meaning that the probability of rejecting an effect that does actually exist is increased (a Type II error Therefore, it is decided to use a significance value of .05 for all comparisons, while keeping in mind that this inflates the Type I error, and that we thus might classify factors as influencing pilot TDL while in reality they do not. For the purpose of this research however, this is deemed more favorable than discarding factors that actually *do* influence pilot TDL. In any regard, due to the small sample size, the results of the tests given below should only be considered as an indication of possible effects.

5.1 | Comparison between flight simulator tests and real flight tests

The first goal of the flight simulator tests and real flight tests is to identify whether pilots classify the same factors as increasing or decreasing pilot TDL during both experiments. If there would be discrepancies between the results of the two experiments, for instance, pilots would identify a large Localizer angle as increasing pilot TDL during the simulator tests, and would identify this same factor as decreasing pilot TDL during the real flight tests, this would be apparent from the RSME scores for the approach in which this factor was tested.

To test whether this is the case the RSME z-scores for all pilots and all approaches are calculated for each of the two experiments. The RSME z-scores for the entire approach are plotted in Figure 10 for both experiments, the RSME z-scores for the three approach parts are also calculated but not given in Figure 10. If the same factors influence pilot TDL during both experiments, one would expect to see the same trend relative to the benchmark approach (approach 1) for both experiments.

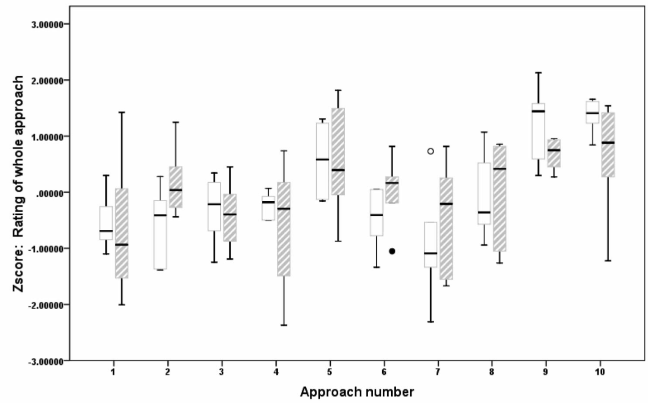


Figure 10 | Boxplot of standardized ratings of entire approach, white boxplots are for flight simulator tests, striped boxplots for real flight tests.

To analyze whether the same trend can be observed, the RSME z-scores for both experiments are compared for *each approach.* If there is no difference between the RSME z-scores per approach this means that the factor that was tested during that approach had the same influence both during the simulator experiment and during the real flight tests. The RSME z-scores are compared using the paired samples t-test for the parametric cases, and using the Wilcoxon matched pairs test for the non-parametric cases. In total 40 comparisons were made (four RSME ratings per approach for 10 approaches).

The results for the RSME ratings for the entire approach are given in Table 2, there was no significant difference ($p<.05$) for any of the approaches, and hence for any of the TDL influencing factors tested during these approaches. Comparison of the RSME z-scores for the three approach parts proved that three comparisons out of the thirty possible comparisons were significant ($p<.05$), which indicates that there was, in some way, a difference in effect for the same approach parts between the two experiments.

For approach 6 (short leg IF-FAF) there was a difference in the RSME z-scores for the first part of the approach. Compared to the benchmark approach, the RSME z-scores were lower for both experiments, but the RSME z-scores for the real flight tests were even lower when compared to the z-scores for the flight simulator tests. What caused this difference is unclear, since compared to the benchmark approach nothing was adjusted in this part (segment) of the approach.

Table 2 | Results of paired samples t-test (comparison of SIMONA research simulator test and Citation) of the z-scores of the RSME ratings of the entire approach

		Paired Differences							
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df	Sig. (2-tailed)
Pair 1	SIM1 - Cit1	.10609	1.32690	.54170	-1.28640	1.49859	.196	5	.852
Pair 2	SIM2 - Cit2	-.75148	1.12667	.45996	-1.93385	.43089	-1.634	5	.163
Pair 3	SIM3 - Cit3	.09902	1.15279	.47062	-1.11075	1.30880	.210	5	.842
Pair 4	SIM4 - Cit4	.36378	1.29521	.52877	-.99546	1.72302	.688	5	.522
Pair 5	SIM5 - Cit5	.03972	1.61328	.65862	-1.65331	1.73275	.060	5	.954
Pair 6	SIM6 - Cit6	-.50136	1.07567	.43914	-1.63021	.62748	-1.142	5	.305
Pair 7	SIM7 - Cit7	-.51029	1.51476	.61840	-2.09993	1.07936	-.825	5	.447
Pair 8	SIM8 - Cit8	-.13877	1.65862	.67713	-1.87938	1.60185	-.205	5	.846
Pair 9	SIM9 - Cit9	.56194	.63780	.26038	-.10739	1.23127	2.158	5	.083
Pair 10	SIM10 - Cit10	.73134	1.12780	.46042	-.45221	1.91489	1.588	5	.173

For approach 2 (Short Localizer intercept heading) there was a difference in RSME z-scores for the final part of the approach. The RSME z-scores were both higher than for the benchmark, but for the real flight tests even more so. Approach 2 was the approach with the short Localizer intercept heading. Therefore, this might indicate that this factor had more effect on pilot TDL during the real flight tests than during the flight simulator tests.

For approach 9 (Not stabilized due to high speed at low FAF) the RSME z-scores for the final part of the approach were significantly different. Again, both RSME z-scores were higher than for the benchmark, but this time this effect was more pronounced for the flight simulator tests. Approach 9 was designed such that it was very difficult to achieve a stabilized approach at 1,000'. This factor thus appears to have more effect on pilot TDL during the simulator tests than during the real flight tests.

Although some differences were found in the pilots' ratings of the approach during both experiments it can be concluded that the 'direction' of the effect, i.e., the trend, was always the same. That is, for both experiments the same factor always either increased or decreased pilot TDL, there was never an opposite effect between the two experiments for the same factor. The difference in RSME z-scores was caused by the fact that the effect was more pronounced in one of the two experiments. Additionally, this difference in effect was only found in the RSME sub-ratings (the rating for an approach part), it never affected the 'overall' RSME rating for the entire approach. Therefore it can be concluded that pilots classify the same factors as increasing or decreasing pilot TDL during both experiments.

Some differences, although they did not result in different factors for TDL, could be observed between the two experiments. It appeared that the duration of the landing checklist is longer during the real flight tests, than during the flight simulator experiment. The mean for the simulator data being 15.7 seconds and the mean for the Citation real flight tests is 20.9

seconds. During the flight simulator tests it was already observed that the landing checklist was often hastily completed. In some cases certain actions that needed to be performed by the PF during the checklist (physically checking the brake pressure by pressing the pedals, actually flipping the ignition switches, etc) were in fact not done. During the real flights in the Citation aircraft the checklist was taken more seriously for obvious reasons.

Another difference that was observed concerns the communication with Air Traffic Control (ATC). During the flight simulator tests all communication was standard and the same for all approaches (in order not to add yet another variable to the test). As a result pilots would know, after having flown a couple of approaches, what ATC was going to say. Consequently, after a while, they would continue performing checklists even if ATC was giving them instructions. During the real flights this was not the case: once the pilot monitoring received a call from ATC all attention was diverted to ATC contact and all other activities (such as performing checklists, selecting flaps, corresponding to calls from the pilot flying) stopped.

5.2 | Factors that influence pilot TDL

Six sets of subjective data are available to determine which factors influence pilot TDL

These six sets are:
1. The RSME z-scores for the entire approach, and
2. the RSME z-scores for the three approach parts. Since the effect of the different approaches on the RSME z-scores was the same for both the flight simulator tests and the real flight tests, the RSME z-scores for both experiments are combined, in order to yield one larger data set. As a result each approach now has 12 RSME ratings. The boxplots of this combined set of ratings can be found in Figure 11. To analyze whether there was a difference between the RSME z-scores for any approach (part) and the benchmark approach (approach 1), paired samples t-test were used when the RSME z-scores were parametric, and Wilcoxon matched pairs tests were used when they were non-parametric.
3. The run questionnaires from the flight simulator tests. The pilots' answers to the questions regarding the difficulty (see Figure 8 for an example) are converted to z-scores. To determine whether there was an effect of a factor, the answers given for one specific approach are compared to the answers given for the benchmark approach. The comparison is again performed by using paired samples t-tests and Wilcoxon matched pairs test.
4. The answers from pilots to the brief in-flight interview during the real flight tests. To analyze these answers the majority rule is used: when more than three of the pilots have the same opinion on a matter, this opinion is used for further analysis.
5. The answers to the end-of-day questionnaires for the flight simulator tests, and
6. The answers to the end-of-day questionnaires for the real flight tests. These two end-of-day questionnaires are mostly identical. It is interesting to see whether the general opinion of the pilots changes between the two test series. Again, the majority rule is used to analyze the answers.

5.2.1 | Overview per approach

Table 3 shows the results when comparing each approach to the benchmark approach, when using the results from the first four datasets.

To explain the idea behind Table 3 an example for APP02 is given here. The data for each approach are compared to the data of the benchmark approach (APP01). The first comparison is between the RSME z-scores for the entire approach for APP02 and APP01. The result of the t-test is that there is no significant difference ($t(11) = -1.649$, $p > .1$). For the second test the RSME z-scores of part 2 of the approach are used, since the independent measure for

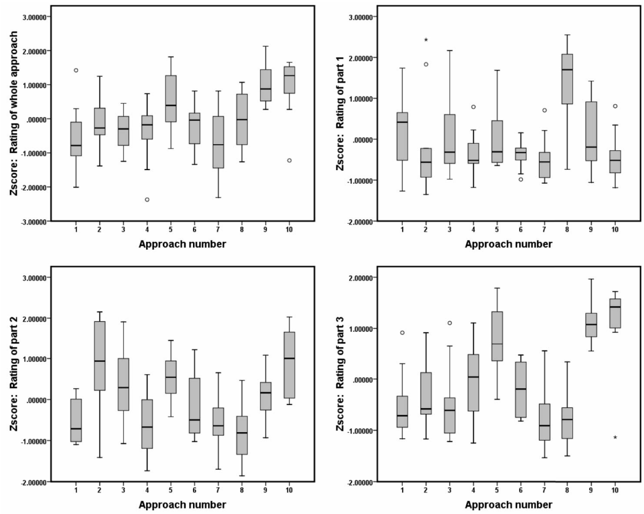


Figure 11 | Boxplots of ratings approaches (SIMONA research simulator test and Citation combined)

Table 3 | Analysis of differences in effort between the approaches and APP01.

	Independent measure	RSME z-scores entire approach	RSME z-scores part of approach	Run questionnaire flight sim tests	In-flight interview real flight tests	Conclusion
APP01	-	-	-	-	-	-
APP02	Short LOC intercept heading	No difference	More effort	More effort	More effort	More effort
APP03	Large LOC intercept angle	No difference	More effort	More effort	More effort	More effort
APP04	Low FAF, normal line-up distance	No difference	More effort	More effort	More effort	More effort
APP05	Short line-up distance	More effort	More effort	More effort	More effort	More effort
APP06	Short leg IF-FAF, normal line-up distance	No difference	No difference	No difference	No difference	No difference
APP07	High LOC intercept speed	No difference	No difference	No difference	No difference	No difference
APP08	Not meeting constraints at WP's, but stabilized at 1000'	No difference	More effort	More effort	More effort	More effort
APP09	Not stabilized due to high speed at low FAF	More effort	More effort	More effort	More effort	More effort
APP10	Not stabilized due to high energy at start of approach	More effort	More effort	More effort	More effort	More effort

5

APP02 is a short LOC intercept heading. The result of the t- test is that there is a significant difference between the two ratings ($t(11) = -3.003$, $p < .05$). On average the ratings of part 2 of APP02 are significantly higher than the ratings of part 2 of APP01. Hence the table states "more effort". The next column is the flight simulator run-questionnaire. In this particular case only the question about the length of the LOC intercept heading is important. The standardized data for this question are normally distributed, so again a paired samples t-test is used. The result is that a short LOC intercept heading significantly increases the amount of effort needed ($t(5) = 2.965$, $p < .05$). Finally the in-flight interview from the real flight tests is reviewed on this issue. All six pilots stated that this short LOC intercept heading increased the amount of effort needed.

Since 3 out of 4 sets of data indicate "more effort" the conclusion can be drawn that APP02 requires more effort than the benchmark approach, APP01. For this reason the final column reads "more effort" for approach 2. Using this method for each approach gives the results in Table 3.

Table 4 | Overview of factors that increase the TDL during RNAV approaches

Factor	**Approach**	**Run-question-naire flight sim test**	**In-flight interview real flight test**	**End-of-day question-naire flight sim tests**	**End-of-day question-naire real flight tests**	**RSME z-scores part of approach**	**Conclusion**
Eratio more than 1, part 1 approach	APP08	More effort	More effort	More effort	More effort	More effort	More effort
Eratio more than 1, part 2 approach	APP10	More effort	-	More effort	More effort	-	More effort
Eratio more than 1, part 3 approach	APP09 & APP10	More effort	-	More effort	More effort	-	More effort
Short LOC intercept head-ing	APP02	More effort	More effort	-	-	More effort	More effort
Large LOC intercept angle	APP03	More effort	More effort	More effort	More effort	More effort	More effort
High LOC intercept speed	APP07, 09 & 10	More effort	Neutral effort	More effort	More effort	-	More effort*
Short leg IF-FAF	APP05, 06 & 10	More effort	-	More effort	More effort	-	More effort*
Short Line up distance	APP05 & 10	More effort	-	More effort	More effort	-	More effort
High Speed on FAF	APP09 & 10	More effort	More effort	More effort	More effort	More effort	More effort
Low FAF	APP04, 05 & 10	More effort	More effort	More effort	More effort	-	More effort
Not stabilized at 1000'	APP09 & 10	More effort	-	More effort	More effort	-	More effort
Horizontal intercept of GS	APP08	Less effort	Less effort	-	-	-	Less effort

(* = under specific circumstances - = No information available)

5.2.2 | Overview per factor

Table 3 thus gives the overview per approach. However, per approach there were sometimes more factors that were changed relative to the benchmark approach than just the independent variable, see Table 1. Therefore it is interesting to also consider the effects per factor. For this reason Table 4 is created. This table lists all the specific factors that are investigated in this research in the first column.

The second column lists the approach numbers where the factor mentioned in the first column occurred. The next 4 columns list the results of the flight simulator test run-questionnaires, the real flight test in-flight interviews, the end-of-day questionnaire for the flight simulator tests and the end-of-day questionnaire for the real flight tests respectively. The fifth column indicates the results according to the RSME z-scores for that particular part of the approach that is connected to the factor. There is only an entry in this column when there was no other linked factor (see Table 1) that was also changing in that part of the approach. The same procedure is used as before: when a certain factor is identified to increase the effort at least in 3 of the data sets, it is concluded that this factor increases the effort.

An example: the first entry in the table is "Eratio more than 1, in part 1 of the approach". This only occurs in APP08. Using the data from APP08, statistical tests are performed on the answers to the flight simulator run questionnaires. A paired samples t-test is conducted on the answers of the question in the run questionnaire (regarding the ability to meet constraints at waypoints in the first part of the approach, see Figure 8 for an example of the format). The results indicate that this significantly increases the effort (t = 2.939, p < .05). During the in-flight interview of the Citation tests 4 pilots answered that this aspect increased the amount of effort. The same answers are found from the two end-of-day questionnaires. Additionally, in APP08 the only factor that was changed in the first part of the approach relative to the benchmark approach was the Eratio in the first part of the approach, there were no linked factors. Therefore in the fifth column the result of the RSME z-scores for the first part of the approach is also incorporated (taken from Table 3). So, in the case of "Eratio more than 1, in part 1 approach" all the data concur, this factor increases the effort.

In several parts of the table it states "no information available". This occurs frequently in the column of the Citation in-flight interview. These interviews were so short that only a limited amount of information could be gathered. In some cases the end of day questionnaires do not contain a question that specifically handles a factor. So these cases are also noted as "no information available". When this occurs twice in one row and the other two columns do state "increase in effort", the conclusion is that that particular factor does increase the effort. An example is "short LOC intercept heading". Unfortunately there was no question incorporated in the end of day questionnaires that specifically targeted this aspect. But from the flight simulator test run questionnaire and the real flight test in-flight interview it was found that this factor does increase the effort. In this case only 2 out of 4 are needed for a positive conclusion.

Very interesting to note in Table 4 are the factors: "High LOC intercept speed" and "short leg IF-FAF". From the comparisons of the RSME ratings of APP06 (which has a short leg IF-FAF) it was concluded that this approach does not require significantly more effort to fly. However APP06 is not the only approach with a short IF-FAF leg. As can be seen in the second column of Table 4, APP05 and APP10 also have a short leg IF-FAF (among other factors). The results of the Wilcoxon test on the answers of the question in the run questionnaire regarding the length of leg IF-FAF of APP05 and APP10 showed a significant difference compared to the benchmark approach. This difference is not found on APP06. From this it seems that a short leg IF-FAF only increases the effort when an approach also has a short line-up distance (like APP05 and APP10). In APP06 the line-up distance is normal, so pilots have enough time to ‘recover’ from the short leg IF-FAF. However in the end of day questionnaires 5 out the 6 pilots answered “more effort” on the question regarding the effect of a short IF-FAF distance. Taking this information into consideration it is concluded that in the researchers’ opinion a short leg IF-FAF does increase the TDL, but the increase in TDL is limited and only really occurs when the line-up distance is short as well.

The same conclusion is made regarding the high LOC intercept speed. No increase in TDL is found when investigating APP07 (which has a high LOC intercept speed, but with a long line-up distance). But analysis of APP09 and APP10 shows that a high LOC intercept speed accompanied by a short line-up distance (little time to recover on final), does indeed increase the TDL.

Interesting to note is the last row, where the factor “horizontal intercept of the GS” is stated. This aspect occurs in APP08. This approach is designed in such a way that instead of a CDA, this approach has several “step-down’s” (legs where the aircraft descends, followed by legs where the aircraft flies level). In APP08 the leg before the GS is flown level and as a result the GS is intercepted horizontally. From the flight simulator test run questionnaire and the Citation real flight tests in-flight interviews it can be concluded that this in fact *decreases* the effort during an RNAV approach.

5.3 | Conclusions TDL factors

Analyzing the factors from Table 4 and taking the observations made by the authors during all the tests (both flight simulator and real flight tests) into consideration, it can be concluded that especially the final part of the approach (the glideslope) has great influence on the TDL of pilots. An Eratio of more than 1 in the beginning of an approach (and not being able to meet constraints on waypoints as a result) increases the TDL to some extent. However, when this occurs in the final part of the approach the increase in TDL is much more significant.

The factors concerning the LOC intercept (length of localizer intercept heading, angle and speed) increase the TDL slightly. When pilots have a long line-up distance to “recover”, the increase in TDL is even less.

A short line-up distance and low FAF are of great influence to the TDL. Even when the Eratio on the glideslope is less than 1 (and a stabilized approach is possible), a short line-up distance (and a low FAF) results in very limited time available to perform the necessary pilot actions. This “limited time available” (which is always a consequence of one of the above stated factors), is indeed the driving force on increasing the TDL. It was observed during both experiments that due to limited time, pilots were often late with SOPs and occasionally even forgot certain actions altogether.

To summarize the analysis of this section, the following list of factors increase the TDL during an RNAV approach:

Limited increase in TDL:
❖ Eratio more than 1 (before LOC intercept heading)
❖ Short LOC intercept heading
❖ Large LOC intercept angle
❖ High LOC intercept speed (in combination with other factors, e.g. short line-up distance)
❖ Non-horizontal intercept of the Glideslope

High increase in TDL:
❖ Short leg IF-FAF (in combination with other factors, e.g. short line-up distance)

- Short line-up distance
- High speed on FAF
- Low FAF
- Eratio more than 1 (after LOC intercept heading)
- Not stabilized at 1000’

The factors that are found to influence pilot TDL for the Cessna Citation were found to be identical to those found for the B747. Except for the fact that for the Citation experiments pilots indicated that a horizontal intercept of the Glideslope resulted in a decrease in effort, whereas for the B747 experiments the results showed that this did not have an influence on the effort. This difference can be explained by the fact that during the B747 experiments pilots had a VNAV mode available, guiding them correctly towards the Glideslope intercept independent of the altitude profile, whereas pilots during the Citation experiment had not. It is therefore stated that for aircraft with VNAV mode the Glideslope intercept does not influence pilot TDL.

The hypothesis that the list of factors that influence pilot TDL and the guidelines to keep pilot TDL at an acceptable level as found for the B747 are also valid for other aircraft types, thus is a reasonable one. It is noted again that the factors ‘FAF altitude’ and ‘time available on Localizer intercept heading’ are factors that originate from the fact that both for the B747 and for the Citation a number of actions need to be performed on Localizer intercept heading and between the FAF and 1,000ft. If, for other aircraft, these actions are performed at a different location in the approach, then there should be sufficient time at that particular location.

6 | Flight mechanical tool

For the B747 a comprehensive Monte Carlo simulation and a Point Mass Model (PMM) simulation were developed which could, when combined with the regulations in the PANS-OPS [17], for a given approach predict whether the guidelines to keep pilot TDL at an acceptable level are met. It is now investigated whether these two simulation models can also provide predictions for another aircraft, in this case the Cessna Citation. To this end, first the general idea behind the Monte Carlo computer simulation [11] is repeated here, and the necessary adjustments to incorporate the Cessna Citation are explained. After that, a case study is considered for the Cessna Citation. To conclude the PMM simulation is briefly mentioned.

6.1 | Monte Carlo Computer simulation

When a (newly designed) approach is entered into the Monte Carlo computer simulation, the simulation predicts, amongst others, the percentage of flights that will meet the constraints at the waypoints, and the percentage of flights that will result in a stabilized approach at 1,000 feet, both factors proved to have a significant influence on pilot TDL. It also predicts under what circumstances (e.g., wind conditions) this can be achieved. This section will describe the aircraft model, pilot model, SOPs, wind model and turbulence model that are used within the Monte Carlo computer simulation.

6.1.1 | Computer simulation input

The input of the Monte Carlo computer simulation exists of a list of waypoints of a (newly designed) approach, defined by their lat-lon coordinates, and the altitude and speed constraints at these waypoints. Additionally, the user has to define which waypoint in the list is the Final Approach Fix (FAF).

6.1.2 | Aircraft (Cessna Citation) and autopilot models

The Monte Carlo computer simulation [11] was set-up in a modular way. The B747 aircraft is replaced by the model for the Cessna Citation I (500) [18] which is exactly similar to the model used in the SIMONA flight simulator. Just as in the flight simulator models (and identical to the flight simulator models), the autopilot models are based on the autopilots developed for the model of the B747 [11], which were derived from [19]. Autopilot modes included are: LNAV, Altitude hold, Vertical Speed Select, Glideslope, Heading Select and Localizer modes. The LNAV mode is based on the VOR modes described in [19]. As the Cessna Citation II Laboratory aircraft does not contain a VNAV mode the aircraft is modeled to descend from waypoint to waypoint in the vertical speed mode where the selected (calculated) vertical speed depends on the constraints at the waypoints and the wind conditions.

The hierarchy in meeting the constraints at the waypoints is as follows: the Autopilot models will always aim to meet the altitude constraints at the waypoints, second to this, the Autothrottle controls the airspeed. This results in the situation that the altitude constraint at the next waypoint will always be met, while the speed constraint might not be met (airspeed might be higher than required).

6.1.3 | Pilot model and Standard Operating Procedures for the Cessna Citation Laboratory Aircraft

All pilot actions such as selecting flaps are modeled according to the SOPs for the Cessna Citation (see Figure 6). Each of the pilot actions prescribed by the SOPs is modeled using a 'trigger' event (e.g. reaching 1,200 feet) and a reaction time representing the time between reaching the trigger event and actually performing the action (e.g., 2 seconds after reaching 1,200 feet, flaps LAND are selected). These reaction times are modeled as normal distributions and are based on the distributions of the reaction times as obtained from the flight simulator tests and real flight tests. The trigger events and corresponding reaction time normal distributions as used in the computer simulation are given in Table 5.

Table 5 | Trigger events and reaction time distributions for pilot actions in Monte Carlo simulation

Pilot action	Trigger Event	Mean	Standard deviation
Switch to heading select mode	Waypoint capture at start Localizer Intercept Heading	16.4 sec	9.4 sec
Arm approach	Waypoint capture at start Localizer Intercept Heading	22.2 sec	10.1 sec
Disengage autopilot	Waypoint capture at start Localizer Intercept Heading	23.5 sec	10.6 sec
Flaps APPROACH	2.0 nm before reaching FAF	10.3 sec	13.7 sec
Gear Down	Reaching FAF	0.5 sec	6.9 sec
LAND checklist start (Checklist duration	Gear Down -	14.3 sec 20.9 sec	12.8 sec 8.7 sec)
Flaps LAND	Reaching 1,200' above airport elevation	1.5 sec	4.1 sec

It is important to note that if the airspeed constraints at the waypoints required an airspeed lower than the instantaneous flap speed mark (the IAS below which the next flap setting needs to be selected), the next flap setting is selected in the Monte Carlo simulation irrespective of SOPs. Also, when the airspeed exceeds the placard speed of a certain flap setting, this flap setting cannot be selected. Gear Down selection has no upper speed limit.

The Cessna Citation II Laboratory aircraft does not contain an autothrottle. In the model the airspeed is regulated using a simple proportional controller to simulate the pilots' manual throttle control. Additionally, a relatively simple pilot manual control model for the flight director task is added, consisting of only a time delay (equal to 0.3 seconds) and pure gain.

6.1.4 | Turbulence and Wind models

A Patchy turbulence model is used within the Monte Carlo simulation. The intensity of the turbulence can be adjusted with a gain. During one Monte Carlo simulation run the turbulence intensity, wind direction and wind speed are constant throughout the entire approach, between different Monte Carlo runs these are varied.

6.1.5 | Outputs of the computer simulation

The constraints at a waypoint were considered to be met when the actual Indicated Airspeed (IAS) at that waypoint (to be more specific: the point in the trajectory closest to that waypoint) was less than the required IAS plus 10 knots, and the actual altitude at that waypoint was less than the required altitude plus 100 feet. A lower boundary for these constraints is not necessary since the Monte Carlo simulation always regulates the airspeed and altitude towards the constraints at the next waypoint, when the required airspeed and altitude are attained, the Monte Carlo simulation maintains the required airspeed and altitude until the waypoint is reached. Therefore, in the Monte Carlo simulation, the altitude and airspeed will never be too low at a waypoint.

To determine whether a Monte Carlo simulation run of the approach resulted in a stabilized approach the criteria given earlier were quantified as follows:

- ❖ Heading change, and roll rate are within 5 deg/s
- ❖ The IAS is not more than V_{REF} + 20 knots;
- ❖ Flaps LAND are selected, landing gear is down;
- ❖ Sink rate is not larger than 1,000 feet per minute; and
- ❖ Localizer and glide slope are within one dot;

These criteria are evaluated exactly at 1,000 ft above airport elevation.

6.2 | Validation of Monte Carlo simulation with experiment data

To properly validate the predictions of the Monte Carlo simulation, a comparison is made with the results obtained during the two experiments (flight simulator test and real flight test). During these experiments flight data were recorded and analyzed to evaluate how many flights were stabilized at 1,000', to determine how often the constraints were met at the waypoints, and to gain insight into all pilots' actions. The Monte Carlo simulation will be run with exactly the same conditions as encountered during each of the runs during the two experiments, and while simulating the pilots' actions at exactly the same moments in time at which the pilots performed their actions during the experiments. This implies that the RNAV routes, aircraft weight, wind speed and wind direction, QNH pressure levels, airport elevation, runway heading and pilot action times (the moment that SOPs are performed) are simulated exactly identical to the experiment situation. This results in a total of 12 simulated Monte Carlo runs per approach (since each approach was flown only once by each pilot during each experiment).

The results of the Monte Carlo simulation are presented in Figure 12 as the percentage of approaches that are predicted to be stabilized at 1,000ft.

As Figure 12 shows, there is still a discrepancy between the predictions of the Monte Carlo computer simulation and the actual flight simulator test and real flight test results. Excessive airspeed was always one of the violated conditions in the unstabilized approaches in the experiments. Even though APP01, APP02, APP03 and APP05 are no high energy approaches, the airspeed at 1000′ still exceeded the V_{REF} + 20 kts limit during the flight simulator tests and real flight tests resulting in unstabilized approaches. These large variations in airspeed can all be traced back to human performance in throttle control. Figure 13 shows the airspeed profiles during APP01 as an example to indicate the difference in variation between a throttle controlled by a human pilot and by the simple throttle model in the Monte Carlo simulation. It can be seen that the simple throttle model in the Monte Carlo computer simulation regulates the airspeed more strictly than the pilots did during the experiment.

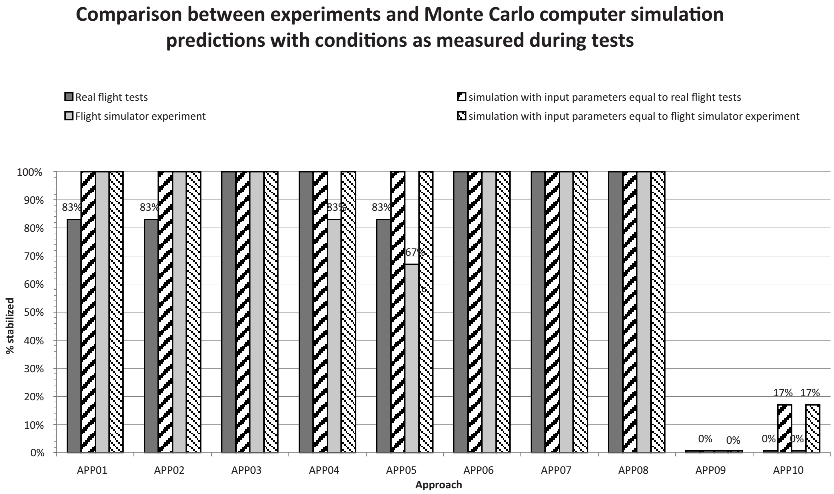


Figure 12 | Stabilized approaches during tests compared to the simulations with environment input and reaction times equal to every test performed in the SRS and during the Citation flight tests

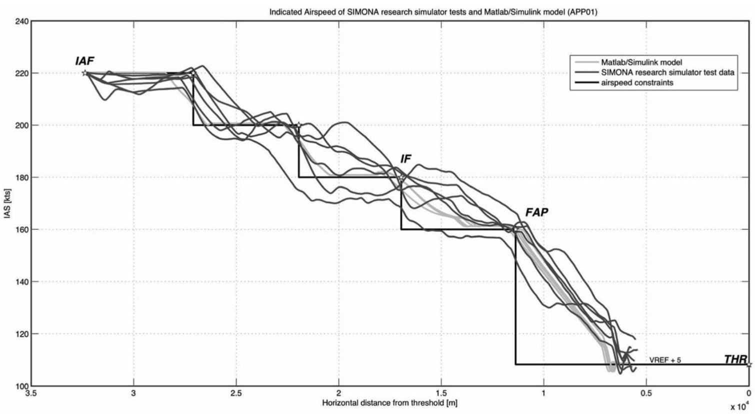


Figure 13 | Airspeed profile of APP01

In the case of APP10, a high-energy approach, the model simulates a stabilized approach, even though none of the approaches during the flight simulator tests and real flight tests were stabilized approaches (all due to an excessive airspeed at 1000′ above airport elevation). This difference is related to the slightly larger variation in manual altitude control by the pilots compared to the altitude control of the model, resulting in slightly different altitude profiles. For this particular case the altitude profile during the flight simulator test was steeper than the three degree glideslope. If, for this case, flaps LAND were selected at 1,500ft this occurred at a later moment in time for the steep approach during the flight simulator test than during the three degree glideslope approach simulated in the Monte Carlo computer simulation. This results in a stabilized approach in the computer simulation and an unstabilized approach during the flight simulator tests. This is the reason for the discrepancy for APP10 in

Table 6 | Waypoint constraints met during tests compared to the simulations with input parameters as measured during test runs (*N* = 6)

	2nd WP (APP08 only)		3rd WP (APP08 only)		2nd WP or 4th WP (APP08 only)		Begin LOC Intercept hdg		IF		FAF	
	SRS tests	**Monte Carlo simulation**	**SRS tests**	**Monte Carlo simulation**	**SRS tests**	**Monte Carlo simulation**	**SRS tests**	**Monte Carlo simulation**	**SRS tests**	**Monte Carlo simulation**	**SRS tests**	**Monte Carlo simulation**
APP01					100%	100%	100%	100%	100%	100%	100%	100%
APP02					100%	100%	**83%**	**100%**	**67%**	**100%**	**83%**	100%
APP03					100%	100%	100%	100%	**83%**	**100%**	**67%**	**100%**
APP04					100%	100%	100%	100%	100%	100%	**83%**	100%
APP05					100%	100%	83%	100%	100%	100%	**83%**	**100%**
APP06					100%	100%	100%	100%	100%	100%	100%	100%
APP07					100%	100%	**83%**	**100%**	100%	100%	**67%**	**100%**
APP08	100%	100%	0%	0%	100%	100%	0%	0%	100%	100%	100%	100%
APP09					100%	100%	**83%**	**100%**	100%	100%	**83%**	**100%**
APP10					100%	100%	100%	100%	0%	0%	0%	0%
	Citation flight tests	**Monte Carlo simulation**	**Citation flight tests**	**Monte Carlo simulation**	**Citation flight tests**	**Monte Carlo simulation**	**Citation flight tests**	**Monte Carlo simulation**	**Citation flight tests**	**Monte Carlo simulation**	**Citation flight tests**	**Monte Carlo simulation**
APP01					100%	100%	100%	100%	100%	100%	**83%**	**100%**
APP02					100%	100%	100%	100%	100%	100%	**67%**	**100%**
APP03					100%	100%	100%	100%	100%	100%	100%	100%
APP04					100%	100%	100%	100%	100%	100%	**83%**	**67%**
APP05					100%	100%	100%	100%	100%	100%	**67%**	**100%**
APP06					100%	100%	100%	100%	100%	98%	**83%**	**100%**
APP07					100%	100%	100%	100%	100%	100%	100%	100%
APP08	100%	%	0%	0%	100%	100%	**50%**	**0%**	100%	100%	100%	100%
APP09					100%	100%	100%	100%	**83%**	**100%**	100%	100%
APP10					100%	100%	100%	100%	**17%**	**0%**	0%	0%

Table 6 shows how many times the constraints at the waypoints have been met. The grey cells indicate the fact that a large amount of energy (altitude and/or airspeed) before that waypoint needed to be dissipated in order to meet the constraints. Again, there are discrepancies to be found in how many times waypoint constraints have been met during the experiments compared to the output of the simulation. However, as with the prediction of stabilized approaches, the reason for the differences lies in the unpredictability of the manual throttle control by the pilots.

It can thus be concluded that the predictions of the Monte Carlo computer simulation model are not always fully in agreement with reality, but that the trends are predicted rather well. The major cause for discrepancies is the lack of correct simulation of the pilots' manual control of the throttle. It should be noted that the method and computer simulation developed within this research will be used to give an indication of pilot TDL during RNAV approaches at airports. The majority of these aircraft will have an autothrottle, and therefore it is expected that better results will be obtained for these aircraft.

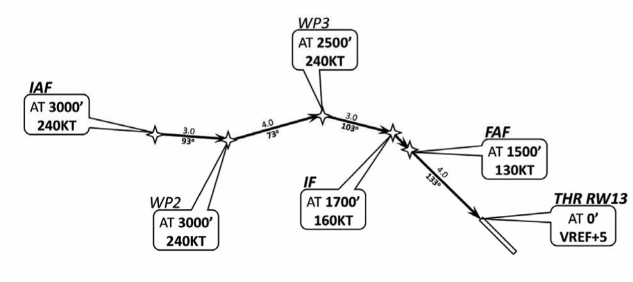


Figure 14 | Example approach towards runway 13.

Table 7 | Waypoints with altitude and airspeed constraints for example approach towards runway 1

Waypoint	Altitude [ft]	IAS [knots]
IAF	3000	240
WP2	3000	240
WP3	2500	240
IF	1700	160
FAF	1500	130
THR RW13	0	VREF + 5

6.3 | Case study

The Monte Carlo computer simulation can now be run many times, using all recorded reaction times, and applying different wind conditions. The computer simulation can then be used to predict under what wind conditions the constraints at the waypoints can be met and a stabilized approach can be achieved. As an example an approach towards runway 13 is considered, see Figure 14 and Table 7. The results are depicted in Figure 15 and Figure 16. Figure 15 clearly shows that the possibility of achieving a stabilized approach depends on the wind direction (a headwind on final results in a stabilized approach). Other reasons for ending stabilized or unstabilized are the moment in time at which flaps LAND and/or gear down are selected. A similar plot as in Figure 15 can be generated for meeting the constraints at each waypoint. The results of the computer simulation also provide insight into the locations in the approach where pilots are performing many actions, thereby providing approach designers with an indication of the 'busy' parts of an approach, see Figure 16 for an example.

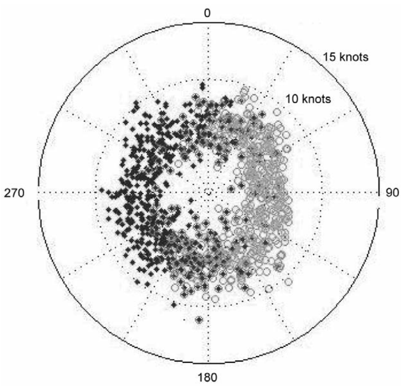


Figure 15 | Example of the results of the Monte Carlo computer simulation with respect to the possibility of achieving a stabilized (grey circle) or unstabilized (black dot) approach, as a function of wind direction and wind speed.

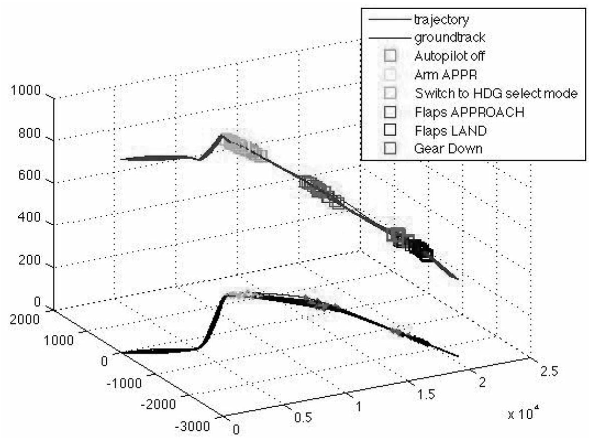


Figure 16 | Example of the results of the computer simulation, providing insight into the locations in the approach where pilots are performing actions such as selecting flaps, gear, etc.

6.4 | Point mass Model

The Monte Carlo simulation model generates reliable results regarding the percentage of flights that can achieve a stabilized approach and factors that influence pilot TDL, but takes a long time to produce these results (in the order of several hours per approach that is analyzed). This is not very practical when the computer simulation is intended to be used as a tool during the design of approaches. Therefore it was investigated [13] for the B747 whether a much simpler model based on a point mass model, with a considerably shorter

calculation time, can generate results as reliable as the highly detailed computer simulation. It was found for the B747 that a point mass model could indeed generate the same results as the more detailed computer simulation. It can now be stated that this is also the case for the Cessna Citation. This is true as long as the point mass model contains:
1. a detailed lift-drag polar for all flap settings and gear up/down setting,
2. a detailed model of the flight idle thrust,
3. an accurate model to simulate the lateral track, specifically the distance of turn anticipation since this influences the amount of trackmiles available between two waypoints, and
4. a model to simulate the pilots' actions according to the trigger events and reaction time distributions found in this research.

6.5 | Are all design guidelines predicted by the computer simulation?

The question now is, whether the output of this computer simulation provides sufficient information in order to assess whether the approach meets all guidelines for the contributors to pilot TDL as given earlier in this chapter. The simulation obviously predicts whether a stabilized approach can be achieved (guideline 3), and whether the constraints at the waypoints can be met (guideline 1). It also provides insight whether there is sufficient time on Localizer intercept heading since the moments in time at which all actions are performed is predicted (guideline 2), and it can easily predict the sink rate (guideline 5).

However, although the simulation can predict or calculate the numerical values for the (actual) Localizer intercept speed (guideline 6), the Localizer intercept angle (guideline 7), and the distance between IF and FAF (guideline 4), it does not give a qualitative indication of whether these numerical values are sufficiently high or low. Fortunately, the minimum or maximum values for these factors are very accurately prescribed in the Procedures for Air Navigation Services Aircraft Operations (PANS-OPS) [17]. The PANS-OPS prescribe a minimum straight distance between IF and FAF of 2nm with an additional turning distance (which depends on the airspeed and intercept angle), and recommend an interception angle at the Localizer not exceeding 30 degrees. Actually, the PANS-OPS and the predictions of the computer simulation complement each other very nicely regarding factors contributing to pilot workload, since what is not prescribed in the PANS-OPS is predicted by the computer simulation and vice versa.

The conclusion thus is that the computer simulation, combined with the regulations in the PANS-OP, provides sufficient information to assess whether the guidelines for the contributors to pilot TDL for the Citation are met.

6.6 | Quantification of guidelines

In order to be able to assess whether, for instance, sufficient time is available on Localizer Intercept heading, the time required in this part of the approach needs to be quantified. Table 8 gives an indication of the reaction times for all pilot actions for the Citation based on the reaction times recorded during the two experiments. As an example: 41 seconds after starting the turn towards Localizer intercept heading, pilots had switched off the autopilot during 95% of all runs. From this it can be concluded that, when it is assumed that 95% of the pilots should be able to fly the approach, the time available on Localizer intercept heading should at least be 41 seconds. These values can also be used in the point mass model.

It is now interesting to compare the values found for the Citation (Table 8) to the values found for the B747 [13], see Table 9. The only actions that are based on the same trigger event for both aircraft are the actions Flaps LAND and Gear down (the reaction time for the B747 stated to the left of the forward slash is of importance here). It is important to note that the reaction times given for each percentage of pilots in both tables are based on the *actual recorded reaction times*, not on the approximating normal distribution curves.

Table 8 | Maximum pilot reaction times in seconds for different percentages of pilots based on 120 samples for the Cessna Citation

Pilot action	Trigger Event	85 %	90 %	95 %	99 %
Heading Select	start turn to LOC int. HDG	23.1	25.3	29.6	55.2
ARM Approach	start turn to LOC int. HDG	30.5	33.6	38.3	57.0
AP OFF	start turn to LOC int. HDG	31.7	35.4	41.0	58.8
Flaps APP	2 NM before FAF	21.5	29.6	38.1	44.3
Gear Down	Reaching FAF	5.6	7.3	9.2	19.8
Flaps LAND	1200ft	5.5	7.5	10.6	14.5
Landing Checklist	Gear Down	55.2*	60.2*	70.2*	75.7*

* Time in seconds until the checklist was completed

Table 9 | Maximum pilot reaction times for different percentages of pilots for the B747 [13].

Pilot action	Trigger Event	85 %	90 %	95 %	99 %
Flaps 1	IAS	VREF+75.5kts	VREF+73kts	VREF+70.2kts	VREF+50kts
Flaps 5	IAS	VREF+57.3kts	VREF+56.2kts	VREF+54.2kts	VREF+43.3kts
Flaps 10	End turn to LOC int. HDG	33.1 s	38.7 s	62.9 s	86.1 s
Heading Select	End turn to LOC int. HDG	24.0 s	30.8 s	45.4 s	147.0 s
ARM Approach	End turn to LOC int. HDG	37.1 s	44.8 s	57.6 s	130.0 s
AP & AT Off	End turn to LOC int. HDG	30.8 s	38.8 s	47.3 s	65.2 s
Gear Down	Reaching FAF	9.6 / -8.5 s	11.3 / 1.3 s	15 / 2.5 s	27.8 / -7 s
Flaps 20	Gear Down	-9.6 / 14.9 s	-6.9 / 8.3 s	-15 / 12.8 s	-26.8 / 30.4 s
Flaps LAND	Reaching 1,200’	6.4 s	9.6 s	15.3 s	23.4 s
Approach CL	Transition Level	TL – 333ft	TL – 350ft	TL – 400ft	TL – 1,220ft
Duration		9.8 s	10.4 s	11.5 s	16.5 s
CL after GD/flaps20	Latest of GD or Flaps 20	23.6 s	34.2 s	35.7 s	65.6 s
Duration		13.9 s	15.3 s	16.7 s	21.8 s
CL after flaps 25	Flaps 25	8.4 s	9 s	12.3 s	13.0 s
Duration		9 s	9.3 s	11.5 s	12.1 s

It can be seen that the reaction times for the B747 are in all cases larger than for the Citation. This might be caused by the following two facts: first, the sample size for the B747 was much larger (see also the histograms in Figure 17 as an example), and second, the pilots participating in the Citation experiments were pilots that were used to participate in scientific research and flight tests, the pilots participating in the B747 experiment were commercial airline pilots. For these reasons, the reaction times for the B747 given in Table 9 are regarded more reliable and more realistic.

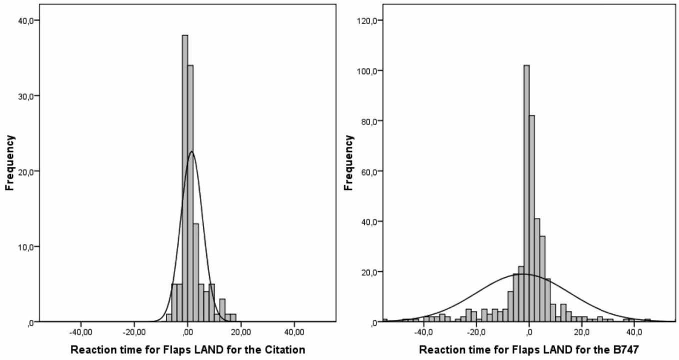


Figure 17 | Histograms for the reaction times for selecting Flaps LAND for the B747 (left, *N* = 116) and Citation (right, *N* = 421)

To illustrate the difference between the reaction times for the B747 and Citation, Figure 17 is presented. Mann-Whitney, Wald-Wolfowitz and Kolmogorov-Smirnov tests all indicated that the distributions for the B747 and the Citation were significantly different.

For clarity, the exact quantification of the guidelines as used during this research is given below (numbers correspond to numbered guidelines):

1. The constraints at the waypoints are considered to be met when the airspeed is within 10 knots of the required airspeed and the actual altitude is within 100 feet of the required altitude.
2. Using the reaction times for the actions performed on Localizer intercept heading as given in Table 9 it can be determined (for a percentage of pilots defined by the approach designer) whether all actions that should be performed on Localizer intercept heading are actually performed on Localizer intercept heading. When this is the case this is regarded "sufficient time" to perform all actions.
3. An approach is considered stabilized at 1,000ft when:
   - ❖ Heading change and pitch change are within 5 deg/s;
   - ❖ The IAS is not more than V_{REF} + 20 knots;
   - ❖ Flaps 25 are selected, landing gear is down (whether this is the case for the defined percentage of pilots can be derived from the values in Table 9);
   - ❖ Sink rate is not larger than 1,000 feet per minute;
   - ❖ Localizer and glide slope are within one dot[3]; and
   - ❖ All checklists are completed (whether this is the case can again be derived from the values in Table 9)
4. the distance between IF and FAF is considered sufficient when it meets the requirements in the PANS-OPS [17]
5. the vertical speed should be below the sink rate warning.
6. the Localizer intercept speed is considered "not too high" when it meets the requirements in the PANS-OPS [17]
7. the Localizer intercept angle is considered "not too large" when it meets the requirements in the PANS-OPS [17].

3 One dot deviation on the glide slope equals 0.7° beam error, one dot deviation on the Localizer equals 2.5° beam error.

7 | Conclusions

The list of factors that influence pilot TDL during RNAV approaches as found for the B747 are also applicable to the Cessna Citation. In other words: the same factors influence pilot TDL while flying an RNAV approach both for the B747 and for the Citation. This is concluded from the results from flight simulator tests and real flight tests with the Cessna Citation. Consequently, the guidelines to keep pilot TDL at an acceptable level as defined for the B747 also apply to the Cessna Citation. From this it is concluded that these guidelines are valid for aircraft types that fly approaches according to the assumptions set forth in this research, and that these guidelines can be used as such during the design of approaches.

It was also found that there were no discrepancies between the list of factors influencing pilot TDL resulting from the flight simulator tests and the list of factors resulting from the real flight tests.

Finally, the computer simulations that were developed for the B747 in order to predict whether the guidelines were met for a certain approach, were successfully adjusted to incorporate the Cessna Citation and produced reliable results. The same simulation technique can thus be used for different types of aircraft.

Acknowledgements

This research was financially supported by the Dutch Technology Foundation STW (DLR7071). We wish to thank the technical staff and pilots participating in the experiments, as well as Air Traffic Control at Malta International Airport for accommodating the test flights.

Literature

1. Stassen, HG, Johannsen, G, & Moray, N. Internal Representation, Internal Model, Human Performance Model and Mental Workload. *Automatica, 26(4)*, 811-820, 1990.
2. Flight Safety Foundation Approach-and-landing Accident Reduction (ALAR) Task Force. Killers in Aviation: FSF Task Force Presents Facts about Approach-and-landing and Controlled-flight-into-terrain Accidents, *Flight Safety Digest*, November 1998 - February 1999
3. Heiligers, M.M., van Holten, Th., Mulder, M., Flight Mechanical Evaluation of Approaches, *Journal of Aircraft*, to be printed.
4. Vicente, K.J., *Cognitive Work Analysis, Towards Safe, Productive, and Healthy Computer-Based Work*, Lawrence Erlbaum Associates, Publishers, Mahwah, New Jersey, 1999
5. Baron, S., Zacharias, G., Muralidharan, R., Lancraft, R. (1980), PROCRU: A model for analyzing flight crew procedures in approach to landing, *NASA-MIT 16th Annual Manual 1980*, pp 488-520.
6. Baron, S., Kruser, D.S., Huey, M.B., *Quantitative Modeling of Human Performance in Complex, Dynamic Systems*,National Academy Press, Washington D.C., 1990.
7. Corker, K.M., Human Performance Simulation in the Analysis of Advanced Air Traffic Management, *Proceedings of the 1999 Winter Simulation Conference*, Farrington, P.A., Nembhard, H.B., Sturrock, D.T., Evans, G.W. (Eds.), 1999.
8. Gore, B.F., Corker, K.M. A Systems Engineering Approach to Behavioral Predictions of an Advanced Air Traffic Management Concept, *19th Digital Avionics Systems Conference, Entering the Second Generation of Powered Flight*, Philadelphia, Pennsylvania, 2000.

9. Smith, B.R., Tyler, S.W., The Design and Application of MIDAS: A Constructive Simulation for Human-System Analysis, *2nd Simulation Technology & Training Conference*, Canberra, Australia, 1997.
10. Heiligers, M.M., van Holten, Th., Mulder, M., *Feasibility analysis of Achieving a Stabilized Approach,* 26th International Congress of the Aeronautical Sciences, 2008.
11. Heiligers, M.M., van Holten, Th., Mulder, M., Flight Mechanical Evaluation of Approaches, submitted to Journal of Aircraft
12. Heiligers, M.M., van Holten, Th., Mulder, M., Factors that influence Pilot Task Demand Load during RNAV Approaches, submitted to Journal of Aircraft
13. Heiligers, M.M., van Holten, Th., Mulder, M., Factors influencing pilot task demand load during RNAV approaches based on theory and experiments, submitted to Journal of Aircraft.
14. Heiligers, M.M., *Pilot Task Demand Load During RNAV Approaches,* …., The Netherlands, December 2010.
15. Hilburn, B. and Jorna, P., Workload and Air Traffic Control. In PA Hancock and PA Desmond (Eds.) *Stress, Workload and Fatigue: Theory, Research and Practice*, Hillsdale, New Jersey, USA: Erlbaum (2001)
16. Godley, Stuart T. *Perceived Pilot Workload and Perceived Safety of RNAV (GNSS) Approaches,* Australian Transport Safety Bureau, ATSB Transport Safety Investigation Report, Aviation Safety Research and Analysis Report 20050342, 2006.
17. International Civil Aviation Organization, *Procedures for Air Navigation Services Aircraft Operations, Volume II Constructions of Visual and Instrument Flight Procedures,* Fifth edition 2006.
18. Borst, C., *Citast: Citation Analysis and Simulation Toolkit,* Delft University of Technology, The Netherlands, 2004.
19. Hanke, Rodney C. and Nordwall, Donald R., *The Simulation of a Jumbo Jet Transport Aircraft Volume II: Modeling Data*, NASA report D6-30643, September 1970, Wichita, Kansas.
20. Aircraft Operations Manual TU Delft Laboratory Aircraft
21. Zijlstra, F., Meijman, Th., Het meten van mentale inspanning met behulp van een subjectieve methode, *Mentale belasting en werkstress, een arbeidspsychologische benadering,* Meijman,Th. (Ed.), Van Gorcum, Assen/Maastricht, The Netherlands, 1989.
22. Lodge, M., *Magnitude Scaling, Quantitative Measurement of Opinions,* Sage university paper series on quantitative applications in the social sciences, 07-025, Newbury Park, CA: Sage.
23. Meijman, T., Zijlstra, F., Kompier, M., Mulders, H., Broersen, J., The measurement of perceived effort, *Contemporary Ergonomics 1986,* Oborne, D.J. (Ed.), Taylor & Frances, London, 1986.
24. Hart, S.G., Staveland, L.E., Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research, *Advances in Psychology* Vol. 52, p.139-183, P.A. Hancock, Meskati N. (Eds.), Elsevier Science Publishers B.V., Amsterdam, The Netherlands, 1988.
25. Jamieson, S., Likert Scales: how to (ab)use them, *Medical Education 2004,* 38: 1217-1218, Blackwell Publishing, 2004.
26. Knapp, T.R., Treating Ordinal Scales as Interval Scales: An Attempt To Resolve the Controversy, *Nursing Research,* March/April 1990, Vol.39 No.2. 121-123.

Discussion

The overall goal of this research was to identify the factors that influence pilot task demand load during approach, and to develop a method that can predict this task demand load for any (existing or to-be-designed) approach.

As first possible contributing factors to pilot TDL, the pilot's scanning and manual control behaviour were studied. A scanning model was developed that was much simpler and easier to implement than existing scanning models such as, for instance, the queuing model of visual sampling by Carbonell (1966). The control model was also kept as simple as possible, deliberately deviating from more complex models such as the optimal control model (Baron, Kleinman & Levison, 1970). These simple models were easy to use, easy to understand and generated interesting results. For the simple task of piloting an aircraft along a horizontal trajectory in the presence of turbulence, an elegant relation was found between the pilot's scanning behaviour, control workload and task performance when plotted as a function of turbulence intensity. When the pilot's task was expanded to flying an entire approach, however, this relation was no longer apparent when analyzing the data of the flight simulator experiment. It appeared that the relation could no longer be recognized due to the other additional tasks that were added to the simple task of scanning the instruments and manually controlling the aircraft. Additional tasks such as selecting flaps and gear, managing the aircraft's energy, communicating with air traffic control, etc.

From the flight simulator test and conversations with pilots it appeared that it were, actually, these additional tasks that constitute the largest contribution to pilot TDL during approaches. This insight resulted in a shift in focus within the research, from the continuous tasks (scanning and manually controlling the aircraft) to the more discrete tasks. As a starting point to identify these discrete factors that influence pilot task demand load, the work of Vormer (2005) and Godley (2006) was used. Although these two researches were not entirely based on the same assumptions and same types as flight as considered in this Ph.D. research, their results provided a good basis. Based on the results of the flight simulator tests performed during this Ph.D. research, it could be concluded that the factors that proved to influence pilot TDL were partly the same as found by Vormer or Godley, but additional factors, not mentioned by Vormer or Godley, also proved to influence pilot TDL. Additionally, some factors that were indicated by Vormer or Godley, were found *not* to influence pilot TDL for the approaches and boundary conditions considered within this research.

The factors that were identified to influence pilot TDL were found to be consistent throughout three separate flight simulator experiments, and one real test flight experiment. Moreover, the same factors were found for two different aircraft types: the Cessna Citation and the Boeing 747. For the Cessna Citation, the same factors were found both during a flight simulator experiment as well as during real flight. This proved that the factors are generally applicable, although it has to be noted that some factors might change when different SOPs are required for different types of aircraft.

The conclusions about which factors influence pilot TDL are based on relatively small data sets, with 6-10 pilots participating per experiment, flying 10-16 different approaches. Due to the high costs of these flight simulator experiments or real test flights, it is very difficult to obtain larger data sets. Although it is appreciated that statistical tests based on these small sample sizes do not justify definite conclusions, nor that statements can be made about the general applicability of the factors found to influence pilot TDL, the perception nevertheless is that the set of factors derived during this research provides a good basis for the guidelines for the design of approaches. The list of factors found to influence pilot TDL is considered to be complete, this is based on the observations during the flight tests, the conversations with the pilots, and the fact that there are not any other factors relating to the approach trajectory that can be varied or can be taken into account during the design of an approach than those mentioned in Table 1 at the start of journal article 4. The factors in Table 1 were used as a starting point, and based on observations, and pilots' answers some of these factors were proved to influence pilot TDL, whilst others were proved to not influence pilot TDL.

It should, however, be noted that the list of factors and conclusions are only valid for flights that are performed according to the assumptions set forth in this thesis. When, for instance, manual flight during the first part of the approach is assumed, there might be additional factors influencing pilot TDL. Another example: when pilots are not required to strictly adhere to SOPs, some factors that were found to influence pilot TDL might no longer have an influence.

Also, it should be mentioned that there are other factors that influence pilot TDL, which are beyond the scope of this research. This research concentrated on factors that influence pilot TDL that are relevant for the design of approaches, with a focus on the properties of the approach trajectory. Other factors, such as the way the approaches are presented to the pilots on the approach charts, or the naming convention of waypoints (Godley, 2006) also influence pilot TDL.

Based on the factors that were found to influence pilot TDL a method was developed that can predict pilot TDL during the design of approaches. The method consists of seven guidelines, when these guidelines are met, pilot TDL during the approach will be acceptable. Whether these guidelines are met can be determined by using the requirements in the PANS-OPS (ICAO 2006) together with an offline computer simulation of the approach. These guidelines were the same for both the B747 and the Cessna Citation.

The offline computer simulation was initially based on a Monte Carlo simulation, with detailed models of the pilot's environment, and relatively simple models for the pilot's actions. This approach to modeling deliberately deviated from existing models such as MIDAS or PROCRU, which intend to model the pilot's behaviour as accurately as possible. Both for the B747 and the Cessna Citation, the simple pilot models proved to generate sufficiently accurate results for the specific purpose of this computer simulation, which was to predict whether a stabilized approach could be achieved and whether the constraints at the waypoints could be met. More detailed pilot models, including, for instance, models for the pilots memory capacity are not needed.

It was found that detailed models of the pilot's environment, that is, a non-linear model of the aircraft, were also not needed for the specific purpose of the computer simulation. A computer simulation based on a point mass model representation of the aircraft can generate the same results. This was again found for both the B747 and the Cessna Citation.

Conclusions and Recommendations

For the conclusions and recommendations the reader is referred to chapters 12 and 13 of Part I of this thesis.

Samenvatting

De taakbelasting van vliegers tijdens RNAV naderingen

Inleiding

De vraag waar dit onderzoek mee begonnen is luidt: "Waarom is de ene naderingsroute nu moeilijker te vliegen voor een vlieger dan de andere?". De naderingsroute is het laatste gedeelte van een vlucht. Het is een gepubliceerde driedimensionale route die vliegtuigen moeten volgen bij het naderen van een luchthaven. Om te kunnen bepalen waarom de ene naderingsroute moeilijker te vliegen is dan de andere, moet allereerst uitgezocht worden welke factoren nu eigenlijk bijdragen aan deze moeilijkheid. Wanneer deze factoren bekend zijn, kan hier tijdens het ontwerp van naderingsroutes rekening mee gehouden worden, zodat de moeilijkheid van het vliegen van de nieuw ontworpen nadering laag kan worden gehouden.

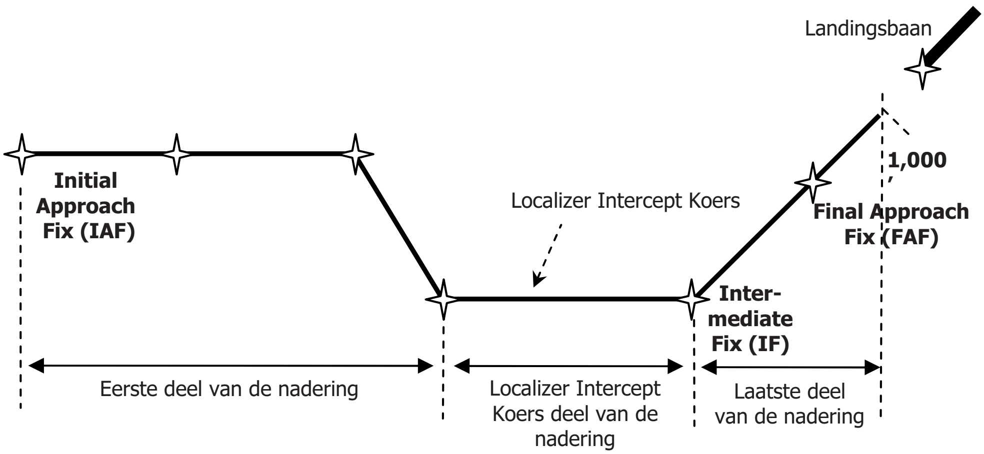
Er zijn twee redenen waarom het belangrijk is om te weten welke factoren bijdragen aan de moeilijkheid van het vliegen van een naderingsroute. Allereerst gebeurt een groot percentage van de ongelukken tijdens de naderings- en landingsfase van de vlucht. De hypothese is dat wanneer een nadering makkelijker te vliegen is, de factoren die bijdragen aan deze ongelukken een kleinere kans van voorkomen hebben. Daarnaast wordt voorspeld dat het aantal vliegtuigbewegingen de komende jaren aanzienlijk zal toenemen. Tegelijkertijd is er ook de ambitie om de overlast veroorzaakt door vliegtuiggeluid en de emissies veroorzaakt door vliegtuigen terug te dringen. Een deel van de oplossing om deze tegenstrijdige intenties te verenigen is het ontwerpen van nieuwe naderingen. Wanneer de factoren bekend zijn die bijdragen aan de moeilijkheid, dan kunnen deze nieuwe naderingen ontworpen worden op basis van deze kennis.

Gebaseerd op het bovenstaande zijn de drie doelen van dit onderzoek als volgt gedefinieerd:

Doel 1: De factoren, die relevant zijn voor wat betreft het ontwerp van naderingen, en die invloed hebben op de moeilijkheid van het vliegen van een nadering moeten geïdentificeerd worden.
Doel 2: Er moet een methode worden ontwikkeld waarmee voorspeld kan worden hoe moeilijk het voor een vlieger zal zijn om een bepaalde nadering te vliegen.
Doel 3: Deze methode moet gevat worden in een computer simulatie, zodat deze gebruikt kan worden tijdens het ontwerp van naderingsroutes.

Uitgangspunten en aannames

Omdat er wereldwijd een overgang plaatsvindt naar Area Navigation (RNAV) operaties, wordt er in dit onderzoek alleen gekeken naar RNAV naderingen. Een RNAV nadering is gedefinieerd door een lijst met waypoints. Voor elk van deze waypoints kan een vereiste hoogte of snelheid opgegeven zijn, deze worden de hoogte en snelheids constraints genoemd. Tijdens het vliegen van een nadering bestuurt de vlieger het vliegtuig zodanig dat het opgegeven driedimensionale traject wordt gevolgd en dat aan alle snelheids constraints wordt voldaan. Een nadering begint bij de Initial Approach Fix (IAF), zie Figuur S.1, en is voor dit onderzoek opgedeeld in drie delen. De Intermediate Fix (IF) is het eerste waypoint dat in het verlengde van de landingsbaan ligt, hier wordt de Localizer opgepakt. Het laatste karakteristieke waypoint in een nadering is de Final Approach Fix (FAF).



Figuur S.1 | Een RNAV nadering met karakteristieke waypoints.

De moeilijkheid die een vlieger ervaart tijdens een nadering wordt uitgedrukt in Task Demand Load (TDL). De TDL geeft aan hoe moeilijk een *taak* is, en moet niet verward worden met Mental Load. Mental Load is de werkbelasting die een vlieger ervaart tijdens het uitvoeren van een taak wanneer factoren zoals vermoeidheid, training en ervaring worden meegenomen. Voor een gegeven approach, en gegeven omstandigheden, is de TDL dus altijd hetzelfde, maar kan de Mental Load anders zijn voor verschillende vliegers, of zelfs voor dezelfde vlieger wanneer deze de ene keer vermoeid is, en de andere keer niet.

Er wordt aangenomen dat de nadering gevlogen wordt volgens de standaard procedures, ofwel Standard Operating Procedures, en dat de vlieger ernaar streeft om op 1.000 voet hoogte stabiel uit te komen. Een nadering is stabiel op 1.000 voet wanneer aan een aantal voorwaarden wordt voldaan, zoals correcte stand van de flaps, het landingsgestel moet naar beneden zijn, de snelheid moet een bepaalde waarde hebben, etc. Wanneer de nadering niet stabiel is moet de vlieger de nadering afbreken en een doorstart uitvoeren.

De TDL van de vlieger wordt in dit onderzoek bepaald als een functie van: 1. de naderingsroute en bijbehorende hoogte en snelheids constraints, 2. de windsnelheid en windrichting, 3. het type vliegtuig (Boeing 747 of Cessna Citation), en 4. de massa van het vliegtuig.

Resultaten

Omdat verwacht werd dat het aflezen van de vluchtinstrumenten, ofwel scannen, en de continue stuuracties van de vlieger invloed zouden hebben op de TDL, zijn een theoretisch scanmodel en stuurmodel ontwikkeld. Deze modellen zijn gevalideerd op basis van twee vluchtsimulator testen waarbij de oogbewegingen van de vliegers werden geregistreerd. De conclusie was dat het scannen en continu besturen van het vliegtuig niet in betekenende mate bijdroegen aan de TDL. Het bleek dat de TDL van de vlieger veel meer werd beïnvloed door andere taken zoals het selecteren van flaps en landingsgestel, het uitvoeren van checklists, etc. De vliegers leken ook veel meer bezig te zijn met eigenschappen van de naderingsroute die betrekking hadden op een grotere tijdschaal, zoals de tijd beschikbaar tussen bepaalde waypoints, dan met korte tijdschaal acties als het wegregelen van verstoringen in de stand van het vliegtuig die optreden door het selecteren van flaps.

Daarom is een volgend vluchtsimulator experiment uitgevoerd dat tot doel had om inzicht te krijgen in deze andere factoren die een invloed hebben op TDL. Het experiment is uitgevoerd voor een Boeing 747. Een uitgebreide lijst met factoren die mogelijk invloed hadden op TDL is voortgekomen uit dit experiment. Een tweede experiment (ook voor de Boeing 747) is uitgevoerd, waardoor deze lijst kon worden ingekort. Gebaseerd op de resultaten van deze beide experimenten is de volgende lijst samengesteld met factoren die een invloed hebben op TDL:

- ❖ Voor het eerste deel van de nadering (zie Figuur S.1): het feit of de hoogte en snelheids constraints op de waypoints gehaald kunnen worden.
- ❖ Voor het Localizer Intercept Koers deel van de nadering (zie Figuur S.1): de beschikbare tijd om alle acties uit te voeren. En daarnaast, in mindere mate: de snelheid waarmee de Localizer wordt opgepakt, de hoek waaronder de Localizer wordt opgepakt, en het feit of de hoogte en snelheids constraints kunnen worden gehaald.
- ❖ Voor het laatste deel van de nadering (zie Figuur S.1): het feit of het mogelijk is om op 1.000 voet stabiel uit te komen, de afstand tussen IF en FAF en de snelheid tijdens het laatste deel van de nadering.

Gebaseerd op deze lijst met factoren is een methode ontwikkeld voor het voorspellen van TDL. De methode bestaat uit zeven richtlijnen. Als, voor een gegeven nadering, aan deze richtlijnen wordt voldaan, is de voorspelling dat de TDL die de vlieger ervaart acceptabel is. De richtlijnen voor de Boeing 747 zijn dat:

- ❖ Het mogelijk moet zijn om gedurende de gehele nadering te kunnen voldoen aan de hoogte en snelheids constraints;
- ❖ Er voldoende tijd moet zijn om de acties uit te voeren op Localizer Intercept Koers;
- ❖ Het mogelijk moet zijn om stabiel uit te komen op 1.000 voet hoogte;
- ❖ De afstand tussen IF en FAF voldoende groot is;
- ❖ De verticale snelheid lager is dan de "sink rate warning";
- ❖ De snelheid waarmee de Localizer wordt opgepakt niet te hoog is; en dat
- ❖ De hoek waaronder de Localizer wordt opgepakt niet te groot is.

Vervolgens is getest of de factoren en richtlijnen die gevonden zijn voor de Boeing 747 ook gelden voor een ander vliegtuig type. Daarom is ook een vluchtsimulator experiment uitgevoerd voor de Cessna Citation, gevolgd door echte vluchten met de Citation. Gebaseerd op de uitkomsten van deze testen, kan geconcludeerd worden dat voor de Citation inderdaad dezelfde factoren en richtlijnen gelden. Ook was er geen verschil tussen de factoren die invloed hebben op TDL tussen de simulator testen en de echte vluchten.

Als onderdeel van dit onderzoek is ook een zeer gedetailleerde, niet lineaire computer simulatie ontwikkeld, zowel voor de Boeing 747 als voor de Citation. Met deze computer simulatie kan voorspeld worden of, voor een gegeven nadering, voldaan wordt aan de zeven richtlijnen. De computer simulatie bestaat uit een vliegtuigmodel, een windmodel en een model voor de vlieger. De naderingsroute met de bijbehorende hoogte en snelheids constraints vormt de input voor de simulatie. Er is gekozen om de modellen voor de vlieger zo simpel mogelijk te houden in de simulatie, en de omgeving van de vlieger (vliegtuig en wind) zo exact mogelijk te simuleren. Dit omdat het niet de intentie is de vlieger exact na te bootsen, maar het alleen de bedoeling is een indicatie te krijgen van hoe "hard" de vlieger moet werken als gevolg van de eisen die de omgeving aan de vlieger stelt. Het model voor de vlieger is gebaseerd op de data die verkregen zijn uit de simulator testen. De acties van de vlieger zijn gemodelleerd op basis van "trigger" gebeurtenissen, en reactie tijden ten opzichte van deze trigger gebeurtenissen. Deze zeer gedetailleerde, niet-lineaire computer simulatie, gebaseerd op de Monte Carlo simulatie techniek, kon voor de Boeing 747:

- ❖ Het percentage vluchten voorspellen dat niet aan de hoogte en snelheids constraints kan voldoen, en dat niet stabiel uit kan komen op 1.000 voet, als functie van windsnelheid, windrichting, type vliegtuig en de massa van het vliegtuig;
- ❖ Inzicht geven in de delen van de nadering waar de vlieger "druk" is; en
- ❖ De beweging van het vliegtuig voorspellen in het longitudinale (verticale) vlak.

Daarnaast kon de computer simulatie, wanneer deze gecombineerd werd met de regelgeving uit de Procedures for Air Navigation Services - Aircraft Operations (PANS-OPS), voorspellen of voor een gegeven nadering voldaan wordt aan de zeven richtlijnen.

Tot slot is ook een vereenvoudigde computer simulatie ontwikkeld, gebaseerd op een punt massa representatie van het vliegtuig. Deze vereenvoudigde computer simulatie kon dezelfde resultaten genereren als de zeer gedetailleerde, niet lineaire computer simulatie, maar in een veel kortere rekentijd. Dit puntmassa model kan gebruikt worden als hulpmiddel tijdens het ontwerpen van naderingsroutes.

Aanbevelingen

Voor verder onderzoek naar de TDL van vliegers tijdens naderingsroutes wordt aangeraden om verder te werken met het vereenvoudigde computer simulatie model dat gebaseerd is op een puntmassa representatie van een vliegtuig. Om een nadering te kunnen evalueren voor alle typen vliegtuigen die gebruik zullen maken van de naderingsroute is het nodig dat de computer simulatie met deze vliegtuigtypen wordt uitgebreid. Wanneer er een vliegtuigtype aan de simulatie wordt toegevoegd dat gebruikt maakt van SOPs die sterk afwijken van de SOPs van de Boeing 747 of Cessna Citation, is het advies om voor deze vliegtuigtypen extra simulatie testen uit te voeren. Dit vanwege twee redenen: ten eerste om na te gaan of ook voor afwijkende SOPs gebruik gemaakt kan worden van de simulatie filosofie die in dit onderzoek is voorgesteld (dat is, het gebruik van trigger gebeurtenissen en reactietijden), en, ten tweede, om na te gaan of dezelfde factoren een invloed hebben op TDL voor deze typen vliegtuigen.

De onderliggende ideeën die gebruikt zijn tijdens dit onderzoek kunnen ook toegepast worden op andere vluchtfases, bijvoorbeeld op het vliegen van vertrekroutes.

About the author

Monique Heiligers was born on the 16th of May 1977 in Geldrop, The Netherlands. From 1989 until 1995 she attended the "Scholengemeenschap Augustinianum" in Eindhoven, from which she obtained her "Gymnasium" diploma. She then moved to Delft where she enrolled as a student at the Faculty of Aerospace Engineering at Delft University of Technology in September 1995. During her studies she worked as a student assistant: she taught mechanics to first year students, supervised the rocket propulsion practical, and composed the Study Guide for the student administration office. Driven by her interest in aviation safety, she arranged an internship at the National Transportation Safety Board in Washington D.C.. She obtained her M.Sc. degree (Cum Laude) in June 2002, and was granted the award for best graduate student of the Faculty of Aerospace Engineering during the academic year 2001/2002.

After her graduation she moved with her husband to Saudi Arabia, where she lived for two years. During this period she worked for Delft University of Technology as a researcher on interesting topics such as the Ornicopter and the Condi-cyclone. She also wrote her research proposal for this Ph.D. research and applied for funding at the Dutch Technology Foundation STW.

In 2004 she returned to the Netherlands and continued to work for Delft University of Technology, now shifting her focus from researching to teaching, which she enjoyed very much. She taught the first and second year courses in Airplane Performance, and participated as a teacher in numerous (laboratory) practicals such as the Flight Tests with the Cessna Citation aircraft and flight path simulation computer practicals.

In 2006 the Dutch Technology Foundation STW decided to finance her Ph.D. research, which implied that her focus shifted back from teaching to researching again. The research was supervised by prof dr ir. Th. van Holten and prof dr ir. M. Mulder, both professors at the Faculty of Aerospace Engineering at Delft University of Technology. In order to become more familiar with the topic of her Ph.D. research, Monique obtained her Private Pilot License in 2007. The Zonta International Foundation supported her research by granting her the Zonta International Amelia Earhart Fellowship Award.

Acknowledgements

This dissertation is the result of four years of hard work, which I enjoyed very much, and would not have been possible without the help, advice and encouragement of many people. I want to thank everybody who has in some way contributed to this research!

First of all, the Dutch Technology Foundation STW for providing the financial support for this project. It is fair to say that, without this support, this research would not have taken place. For each project STW requires the researchers to install a so-called users committee. This users committee consists of people from the industry or government, and their task is to make sure that the results of each project are also of value to parties outside the university. My committee did much more than that. I want to thank all committee members for the time and effort they invested in this research, and for all valuable discussions, tips, suggestions and ideas that came up during our meetings. Often the meetings were too short to finish our valuable discussions. The committee consisted of:

- Gerard van Es and Henk Blom, Air Transport Safety Institute, National Aerospace Laboratory NLR;
- Rob Ruigrok, National Aerospace Laboratory NLR;
- Sjoerd de Groot and Jeroen Timmers, To70;
- Gerard Temme, Dutch National Aviation Authority (CAA-NL);
- Pieter de Bruijne, PecoTrans;
- Juriaan Hoekstra, Air Traffic Control the Netherlands (LVNL);
- Jurriaan Breedeveld, Dutch Air Line Pilots Association (VNV), and
- Tony Joustra, Amsterdam Airport Schiphol.

During the last four years this Ph.D. research was supervised by my two promotors: prof. dr ir Theo van Holten and prof. dr ir Max Mulder.

Theo, I would like to thank you, first of all, for allowing me to start the work when I was still living abroad, and for giving me total freedom to write a research proposal on a subject I found interesting. These two factors together resulted in my starting this Ph.D. research. I am also grateful for the opportunity to obtain my Private Pilot License, which, apart from the fact that it is amazing to fly, was very valuable for this research, and also came in very handy when teaching Aircraft Performance classes. I very much enjoyed our flights, during which we experienced different levels of pilot TDL and workload ourselves ("Where is that Hercules??!"). During discussions about this research you always urged me to look at the bigger picture, and to not get lost in details. One of your main goals was to prevent me becoming one of "them". "Them" being people who can only talk in incomprehensible jargon and who are no longer able to explain what they are doing to "normal people". I hope I managed to remain "me", which I tried to proof with the first part of this thesis. Thank you for all your support and advice!

Max, since the topic of this research was actually a merge of your area of expertise and Theo's area of expertise, you kindly agreed to become my supervisor and promotor. You also gave me the opportunity to set-up my own research and gave me the opportunity to explore my own ideas, which I very much appreciated. In the beginning of this research you advised me to focus on pilot TDL instead of on pilot workload, a very valuable advice, which greatly reduced the number of variables. During our conversations you always urged me to pay sufficient attention to those details that *are* important for the research. Thank you for you support and your time!

The eye-track equipment that was used during the first two test series (explained in Chapter 6) was kindly made available by the Faculty of Industrial Engineering. I would like to thank Theo Boersema and Cees Jorens for their help and support in using the eye-track equipment and analysis of the data. The first flight simulator experiment of these two (Chapter 6) was performed in the Frasca simulator of IST Flight Training, and was made possible through the help of Paul van Spijker. The second flight simulator experiment was performed in the SIMONA Research Simulator. It focused on the pilots' scanning and control behavior during approach (Chapter 6) and was organized and analyzed by Gijs Blok and Kirsten Nijsten as part of their master theses. Gijs and Kirsten, you work contributed very much during the early stages of this research.

The flight simulator experiment in SIMONA that focused on a first indication of factors that influence pilot TDL (Chapter 7) required a lot of preparation. I want to thank everybody who contributed. Harm Boschloo, for converting the Matlab/Simulink code I had written to simulate an FMS in my offline simulation to C++ such that it could be used in the SIMONA flight simulator. Joost Ellerbroek, for programming the Primary Flight Display and Navigation Display required for the tests and connecting these to the FMS. Ferdinand Postema, for helping me model the (incomprehensible) autothrottle. Bas Lansdorp, for explaining basic control issues when I was modeling the autopilots and flight director (I know I'm not allowed to reveal that you know a lot about control, but, hey, who reads a Ph.D. thesis anyway?). And Henk Lindenburg, for scheduling the tests and acquiring the hardware for the FMS. But most of all, I am greatly indebted to Olaf Stroosma. Olaf, thank you for making all these separate components WORK. Thanks for the many hours spent debugging the simulator ("Why is it turning right??!"), and solving all technical problems. Thanks to Alwin Damman for his help during the experiment. And last but not least, I am grateful to Justin Konings and Tobie van den Berg who acted as co-pilots during the experiment.

I would like to thank Fred Abbink, Michel Peters and Rob Ruigrok of the National Aerospace Laboratory NLR for offering me the opportunity to perform the flight simulator experiment in GRACE (Chapter 8). Many people from NLR helped with the preparation of this experiment: Paul Blok, Chris Pols, Jaap Kuiper, Wilfred Rouwhorst and Hafid Smaili, your contribution is very much appreciated. A special thanks goes to Bart Heesbeen and Michiel Valens: Bart for his work on the B747 model, and Michiel for his help in the GRACE control room during the entire experiment (and of course for updating my knowledge about Michael Jackson). Paul Hille, Dane vd Meer, Hans Mulder and Xander in 't Veld completed the experiment by assisting as Pilot Monitoring.

The flight simulator test and real flight test with the Cessna Citation aircraft (Chapter 10) were the subjects of the Master theses of Thierry van Bennekom and Tiemen van Tuinen. Thierry and Tiemen, you arranged, performed and analyzed these tests, I am very grateful for all the effort and time you devoted to this research. Andries Muis, Menno Klaassen, and Ferdinand Postema, I very much appreciate the time you invested in adjusting the Flight Test Instrumentation System in the Cessna Citation aircraft such that all pilot actions relevant for our research could be logged. Part of the success of the flight tests was due to the outstanding services of Malta ATC and the handling agent Globe Ground at Malta International Airport, we could not have wished for a better cooperation. Finally, I would like to thank everybody of the crew for the good time we had at Malta.

Many pilots participated in the experiments performed within this research: Jorg, Alwin, Ramses, Herman, Sander, Evert-Jan, Pascal, Maki, Wiebe, Bart, Daan, Mike, Pieter, Hessel, Gerard, Metske, Cees, Tjipke, Alwin, Bob, Xander and Hans, your participation in these experiments is of vital importance for research topics as discussed in this thesis, and is very much appreciated.

The computer simulation program based on a point mass model that provides the output as given in Appendix II is developed by Jelmer Steen during his internship at To70. I would like to thank both Jelmer and To70 for their willingness to make this part of my Ph.D. research the topic of an internship.

As already indicated, many students contributed to this research, either as part of their fourth year flight mechanics exercise, or as part of their Master Thesis. Arnout van Veen, Jeroen van der Zee, Kirsten Nijsten, Koen Brinkman, Marnix Wouters, Pascal Geusgens, Menno Braakenburg, Justin Konings and Tobie van den Berg, devoted their flight mechanics exercise to this subject. Next to the Master theses contributions already mentioned, Eelco Bakker and Rafik Essayah also worked the final year of their studies on subjects related to this dissertation. To all of you: I very much enjoyed being your supervisor, and want to thank you for your enthusiasm, time and effort. You were always willing to put in that extra bit of time to arrive at better results. Not all of you will find your own results incorporated in this thesis, in that case, know that your work was very valuable as background information! Lynn de Proft, Haley Watson and Jenny van der Pols, although your Master thesis subjects were not directly part of this Ph.D. research, your findings and work also helped my research indirectly.

I performed my research within the Design of Aircraft and Rotorcraft (DAR) group at the faculty of Aerospace Engineering. I like to thank all (former) DARlings for the great time I had, especially: Dries Visser, Marilena Pavel, Mark Voskuijl, Roelof Vos, Gianfranco la Rocca, Sander Hartjes, Egbert Torenbeek, John Stoop, Lin Pijpaert, Michel van Tooren, Michiel Haanschoten, Arvind Gangoli Rao, Stefan van der Elst, Dennis van Gerwen, Benjamin Marrant, Joost Schut, Jochem Berends and Hans Mulder. Dirk and Ronald, my thoughts are with you.

Lin, thanks for your help with all administrative challenges.

Michiel, thank you for the computer and IT support, especially when my laptop was quarantined because of an alleged virus a couple of weeks before the deadline of my thesis.

Michel, three years ago, you urged me to quit my teaching activities in order to fully concentrate on my Ph.D. research. At first I did not really agree. Looking back, I am glad I did follow your advice, otherwise I would never have been able to finish in time!

Joost, Jochem and Hans, I really enjoyed the two years all four of us were working at DAR. Thanks for all the practical jokes, the trips we organized (Reno, Alaska) and the discussions (sometimes even relating to work). I, for sure, had a lot of fun.

Hans, you were the best roommate. As a pilot you answered many of my questions, and as a friend you encouraged me when things became really difficult and shared my small "victories". I am very grateful for all your support, thanks for being my paranimf.

Tom van Tol, my other paranimf. Tom, we sat through many, many, lectures together: you, always late with a cup of coffee, me, on time, providing you with pens and paper. Ever since, you have been there for me, during good times but also during difficult times. Thank you for being my paranimf and such a good friend.

I want to thank my parents for giving me a wonderful childhood, for supporting me during my studies, and for raising me with the motto "*Als* je iets doet, doe het dan goed" which I kept in the back of my mind during this Ph.D. work: I did my best.

Dear Friso, home is where you are. Thank you so much.

Monique