Lehrstuhl für Ergonomie
Technische Universität München

An Integrative Evaluation of Airline Pilots' Manual High-Precision Flying Skills in the Age of Automation

Andreas Haslbeck

Vollständiger Abdruck der von der Fakultät für Maschinenwesen
der Technischen Universität München
zur Erlangung des akademischen Grades eines
Doktor-Ingenieurs (Dr.-Ing.) genehmigten Dissertation.

Vorsitzender:		Prof. Dr.-Ing. Manfred Hajek
Prüfer der Dissertation:	1.	Prof. Dr. phil. Klaus Bengler
	2.	Prof. Dr. phil. Alf C. Zimmer

Die Dissertation wurde am 05.07.2017 bei der Technischen Universität eingereicht und durch die Fakultät für Maschinenwesen am 23.10.2017 angenommen.

Kurzfassung / Abstract

In der vorliegenden Arbeit werden in mehreren Flugsimulatorstudien Faktoren untersucht, welche die manuelle Flugfertigkeit von Linienpilotenbeeinflussen können. Kurz- und Langstreckenpiloten, die sich in ihrer Flugpraxis unterscheiden, müssen einen manuellen Endanflug mit Landung absolvieren, während ihr Blick- und Steuerverhalten aufgezeichnet wird. Im ersten Schritt der Auswertungen stellt sich die aktuelle Flugpraxis als der wichtigste Einflussfaktor auf die manuelle Flugleistung heraus, während die Erfahrung der Piloten nur einen marginalen Einfluss auf deren Leistung ausübt. Im zweiten Schritt der Analyse werden Verhaltensmuster bei der Informationsaufnahme sowie der Flugzeugsteuerung untersucht. Hierbei zeigen die Kurzstreckenpiloten wiederum vorteilhaftere Muster und Strategien bei der Ausführung der Flugaufgabe, während Langstreckenpiloten teilweise deutlich nachteiligere Muster verwenden, die mitunter den Flugaufgaben nicht angemessen sind. Zusammengefasst zeigt das zweite Ergebnis, dass die Untersuchung von Verhaltensmustern notwendig ist, um das, auf den ersten Blick paradoxe Ergebnis manueller Flugfertigkeiten, zu erklären.

This thesis reports the analysis of manual flying skills of airline pilots. Short and long-haul pilots differing in recent flight practice have had to complete a manual approach and landing scenario. Their visual scanning and fine-motor control of the aircraft has been recorded and analyzed. In the first step it is found that recency of practice has the most important effect on manual flight performance, while over-all experience only has a marginal effect. In the second step behavioral patterns have been analyzed. Short-haul pilots show superior patterns and strategies for manual high-precision aircraft control, while—at least partially—long-haul pilots partly have not been able to adapt their patterns to the given situation and flying tasks. This thesis shows that an analysis of behavioral patterns is a necessary step to explain pilots' psycho-motor performance.

Publications Included in this Work

This thesis is based on the work presented in the following papers (order in respect of content):

Haslbeck, A., & Bengler, K. (2016). Pilots' gaze strategies and manual control performance using occlusion as a measurement technique during a simulated manual flight task. Cognition, Technology & Work, 18(3), 529–540. doi:10.1007/s10111-016-0382-2

Haslbeck is the main author of the manuscript. Bengler supported this manuscript in the theoretical and methodological foundation, the introduction, applications for industry and revisions.

Haslbeck, A., & Zhang, B. (2017). I spy with my little eye: Analysis of airline pilots' gaze patterns in a manual instrument flight scenario. Applied Ergonomics, 63(September 2017), 62–71. doi:10.1016/j.apergo.2017.03.015

Both authors have equally contributed to this paper.

Haslbeck, A., Kirchner, P., Schubert, E., & Bengler, K. (2014). A Flight Simulator Study to Evaluate Manual Flying Skills of Airline Pilots. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 58(1), 11–15. doi:10.1177/1541931214581003

Haslbeck is the main author of the manuscript. Kirchner supported this manuscript in data analysis and artworks. Schubert supported this manuscript in data analysis and the introduction. Bengler supported this manuscript in defining the rationale and revisions.

Haslbeck, A., & Hoermann, H.-J. (2016). Flying the Needles: Flight Deck Automation Erodes Fine-Motor Flying Skills among Airline Pilots. Human Factors. doi:10.1177/0018720816640394

Haslbeck is the main author of the manuscript. Hoermann supported this manuscript in deriving the motivation, data analysis and the discussion.

Haslbeck, A., Hoermann, H.-J., & Gontar, P. (2017). Stirring the Pot: Comparing Outer and Inner Loop Control Behavior in Airline Pilots. (Manuscript under review)

Haslbeck is the main author of the manuscript. Hoermann supported this manuscript in statistical analysis, and the discussion. Gontar supported this manuscript in deriving the motivation, data analysis and method development.

Additional Publications Resulting from the Corresponding Research

Haslbeck, A., Schmidt-Moll, C., & Schubert, E. (2015). Pilots' Willingness to Report Aviation Incidents. In *Proceedings of the 18th International Symposium on Aviation Psychology* (pp. 596–601).

Hoermann, H.-J., Gontar, P., & **Haslbeck, A.** (2015). Effects of Workload on Measures of Sustained Attention During a Flight Simulator Night Mission. In *Proceedings of the 18th International Symposium on Aviation Psychology* (pp. 602–607).

Haslbeck, A., & Gontar, P. (2014). How Pilots Believe to Act - and how They Really Do: Insights from Flight Simulator Studies. In A. Droog (Ed.), *Aviation Psychology: faciliating change(s): Proceedings of the 31st EAAP Conference* (pp. 317–324).

Haslbeck, A., Schneider, A., & Gontar, P. (2014). Analyse menschlicher Informationsverarbeitung am Beispiel manuellen Fliegens. In Gesellschaft für Arbeitswissenschaft (Ed.), *Gestaltung der Arbeitswelt der Zukunft. Bericht zum 60. Kongress der Gesellschaft für Arbeitswissenschaft vom 12. - 14. März 2014* (pp. 43–45). Dortmund: GFA-Press.

Schubert, E., & **Haslbeck, A.** (2014). Gestaltungskriterien für Szenarien in Flugsimulatoren zur Untersuchung von Verhalten und Leistung von Verkehrspiloten. In M. Grandt & S. Schmerwitz (Eds.): *Vol. 2014-01. DGLR-Bericht, Der Mensch zwischen Automatisierung, Kompetenz und Verantwortung. 56. Fachausschusssitzung Anthropotechnik* (pp. 125–137). Bonn.

Haslbeck, A., Drees, L., Rehmann, K., & Tüshaus, A. T. (2013). Anforderungen an das Training manueller Flugfertigkeiten. In M. Grandt & S. Schmerwitz (Eds.): *Vol. 2013-01. DGLR-Bericht, Ausbildung und Training in der Fahrzeug- und Prozessführung. 55. Fachausschusssitzung Anthropotechnik* (pp. 159–167). Bonn.

Haslbeck, A., Eichinger, A., & Bengler, K. (2013). Pilot Decision Making: Modeling Choices in a Go-Around Situation. In *Proceedings of the 17th International Symposium on Aviation Psychology* (pp. 548–553).

Gontar, P., Thoma, O., & **Haslbeck, A.** (2013). Adaption hochgradig geübter psychomotorischer Muster an ungewohnte Simulationsumgebungen am Beispiel von Piloten. In M. Grandt & S. Schmerwitz (Eds.): *Vol. 2013-01. DGLR-Bericht, Ausbildung und Training in der Fahrzeug- und Prozessführung. 55. Fachausschusssitzung Anthropotechnik* (pp. 169–181). Bonn.

Haslbeck, A., & Bengler, K. (2012). Entwicklung einer Fragebogenmethodik zur Erhebung leistungsbeeinflussender Faktoren in unterschiedlichen Berufsfeldern.

In Gesellschaft für Arbeitswissenschaft (Ed.), *Gestaltung nachhaltiger Arbeitssysteme. 58. Kongress der Gesellschaft für Arbeitswissenschaft* (pp. 615–618). Dortmund: GfA Press.

Haslbeck, A., Gontar, P., & Schubert, E. (2012). The way pilots handle their control stick - effects shown in a flight simulator study. In Deutsches Zentrum für Luft und Raumfahrt e.V. (Ed.): *Vol. 19. Reports of the DLR-Institute of Transportation Systems, 30th European Annual Conference on Human Decision-Making and Manual Control. Proceedings* (pp. 21–26). Braunschweig.

Haslbeck, A., Schubert, E., Gontar, P., & Bengler, K. (2012). The relationship between pilots' manual flying skills and their visual behavior: a flight simulator study using eye tracking. In S. Laundry, G. Salvendy, & W. Karwowski (Eds.): *Advances in Human Factors and Ergonomics, Advances in Human Aspects of Aviation* (pp. 561–568). Boca Raton: CRC Press.

Haslbeck, A., Schubert, E., Onnasch, L., Hüttig, G., Bubb, H., & Bengler, K. (2012). Manual flying skills under the influence of performance shaping factors. *Work: A Journal of Prevention, Assessment and Rehabilitation*, *41*(Supplement 1/2012), 178–183. Retrieved from http://dx.doi.org/10.3233/WOR-2012-0153-178

Gontar, P., & **Haslbeck, A.** (2012). Untersuchung der Nutzung des Primary Flight Displays durch Piloten unterschiedlicher Trainiertheit mittels Blickerfassung. In M. Grandt & S. Schmerwitz (Eds.): *Vol. 2012-01. DGLR-Bericht, Fortschrittliche Anzeigesysteme für die Fahrzeug- und Prozessführung. 54. Fachausschusssitzung Anthropotechnik* (pp. 263–271). Bonn.

Schubert, E., **Haslbeck, A.**, Hüttig, G., & Bengler, K. (2011). Evaluation manueller fliegerischer Leistung von Piloten anhand erfasster technischer Parameter in Flugsimulatoren hochautomatisierter Flugzeuge. In M. Grandt & S. Schmerwitz (Eds.): *Vol. 2011-01. DGLR-Bericht, Ergonomie im interdisziplinären Gestaltungsprozess. 53. Fachausschusssitzung Anthropotechnik* (pp. 163–173). Bonn.

Acknowledgements

First of all I would like to thank several persons who supported this work by their scientific advice and supervision. The professors Klaus Bengler, Heiner Bubb and Alf Zimmer were my academic supervisors while Hans-Juergen Hoermann contributed as a supervisor from aviation practice. Professor Armin Eichinger supported this work in terms of methods and statistics advice. There is also special thanks to several anonymous reviewers who gave worthwhile feedback on the included and further papers.

Furthermore, I would like to thank all persons who have contributed in the SaM-Sys research project, especially the flight simulator team, Manfred Müller, Carsten Schmidt-Moll, Peter Croeniger, Manfred Binder and Ekkehart Schubert.

I also would like to express my thanks to large numbers of students who supported this work by semester, diploma, bachelor and master theses.

Finally, I would like to thank Patrick Gontar who started to contribute to this work several years ago with his own semester thesis and ended up as a colleague and good friend.

All included papers and the thesis itself were corrected by native English speaking persons.

The approach of Saccade Grouping sets was motivated by Visual Grouping Sets (Kang & Landry, 2015) and initially developed by Hang Yin in his semester's thesis.

List of Abbreviations

2C	captain on Airbus A320
2F	first officer on Airbus A320
4C	captain on Airbus A340
4F	first officer on Airbus A340
ALT	altitude indicator
AOI	area of interest
ATP	airline transport pilot
ATT	attitude indicator / artificial horizon
CPT	captain
ECAM	electronic centralized aircraft monitoring
FO	first officer
GS	glideslope
HDG	heading indicator
IATA	International Air Transport Association
ILS	instrument landing system
LOC	localizer
PF	pilot flying
PFD	primary flight display
RMSE	root mean squared error
SPD	speed indicator

Table of Contents

1. Introduction

The aviation industry's challenge Aviation is a major domain in transportation with a forecasted passenger growth for the next 20 years aiming at a duplication in this period from 3.3 billion passengers up to 7 billion in 2034 (Boeing, 2015, International Air Transport Association [IATA], 2015). However, this trend challenges flight safety: the number of passengers will double, but an increase of accidents would be an unwanted side effect. Consequently, to retain the absolute number of accidents their frequency needs to be halved. In spite of this demand, annual accident rates have remained stable in the last years (Boeing, 2015; IATA, 2016).

A human factors approach This outline highlights the aviation industry's challenge. There must be a reason why accident rates do no longer decline compared to earlier years of aviation. This thesis sheds light on factors affecting pilots' performance, and thus, on aviation safety from a human factors perspective. Automation was suspected to negatively affect pilots' skills (Sarter & Woods, 1994) even though automation led to great safety improvements decades ago. To understand this apparent contradiction and to propose mitigation strategies an ergonomics approach is necessary. In the advent of automation, human limitations were covered by this new technical approach to enable longer flights and higher accuracy in navigation and flightpath management. However, after a long time of highly automated aircraft this initially blessed concept also turns out to be a curse: pilots seem to have lost their skills first without any reason. When considering this issue in more detail, a very logical and simple conclusion comes in one's mind: skill comes with practice – but skill also diminishes without practice.

Ironies of automation Automation took over a big portion of pilots' work(load), but facilitated (Billings, 1991a) and ironically complicated (Bainbridge, 1983) it at the same time. The use of automation frequently comes with a reduction in the manual performance of an active task such as manual control of an aircraft or monitoring system states. This thesis focusses on manual skills during approach and landing. Three flight simulator experiments investigated manual flying skills in terms of visual scanning and fine-motor control of the aircraft.

A research project to foster safety management Two out of three experiments reported in this thesis took place in the SaMSys research project (Deutsche Lufthansa, 2016; Lehrstuhl für Ergonomie, 2016), which was conducted in two phases: SaMSys I (2009–2012) and SaMSys II (2013–2015). The aim of this research project was to develop and identify metrics (safety performance indicators) to measure an airline's recent level of safety. This project was funded by the German Federal Ministry for Economic Affairs and Energy via the Federal Aeronautical Research Program LuFo (Grant 20V0803).

Work program of this thesis This publication based (cumulative) thesis includes peer-reviewed articles as core chapters reporting the research conducted and follows the structure of a research paper. Introductory chapters briefly set the frame for this study (motivation, theory, method) and a summarizing discussion including recommendations reports the outcome of this work. The primary goal of this thesis was to evaluate manual flying skills of airline pilots experimentally, due to the fact that there was no equivalent study to the present date. To facilitate the primary aim there were several secondary goals:

- Developing an integrative definition of *manual flying* and derive an essential part of these skills as subject for this thesis.
- Defining and developing metrics and concepts to cover all relevant characteristics of manual flying to analyze and to evaluate these skills.
- Defining a relevant scenario for the evaluation of manual flying skills.
- Concluding appropriate recommendations for the aviation industry to address the retention of manual skills.

2. Human Information Processing in the Age of Automation

Human information processing A classical approach to describe human work from an ergonomics perspective is to consider a human operator, the utilized technical system and the environment when performing a given task (Bubb & Schmidtke, 1993, p. 308). With a more thorough focus on the human operator, Wickens' model for human information processing (Wickens, Hollands, Banbury, & Parasuraman, 2013, p. 4) furtherly decomposed these cognitive functions into three stages (Figure 1): information acquisition through gaze patterns (I), cognition (II) and performance execution through motor behavior (III). This model provides a valid framework for neuro-cognitive investigations of human performance. However, all three stages vary in their hardness to investigate.

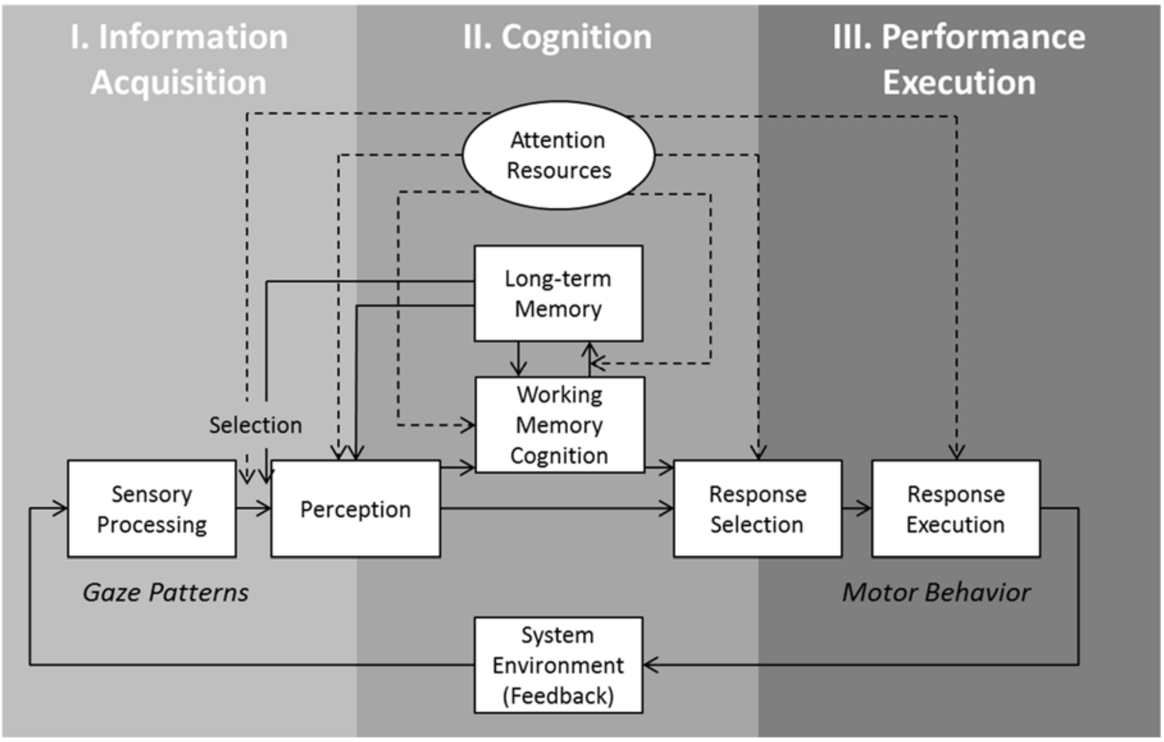


Figure 1: Model of human information processing (Wickens et al., 2013, p. 4) adopted to show three main stages.

Hypotheses-based assumptions for perception In terms of visual perception, there is only a certain probability depending on the context that the human perceives an information, e.g., in a controlled experimental environment (Just & Carpenter, 1976). Eye-tracking approaches (visual scanning) suppose the validity of the *eye-mind-hypothesis* (Goldberg & Wichansky, 2003). This assumption is valid in professional environments when skilled performance is applied and top-down processes prevail (Einhäuser, Rutishauser, & Koch, 2008; Schütz, Braun, & Gegenfurtner, 2011; Foulsham et al., 2012; Allsop & Gray, 2014).

Cognitive processing is hard to observe Linking information from the environment with performed actions using previously gathered information happens in the stage of cognition. Mental models are a common way to give a formalized description of knowledge about facts, relations and processes in the real world (for aviation examples, see Sarter et al., 2003; Hamblin, Gilmore, & Chaparro, 2006). Kahneman (2011) contributed two different mechanisms of problem solving and decision making (*system 1 and system 2*) to cognitive ergonomics, to explain different mechanisms of cognitive processes. The analysis of cognition is not pursued in this thesis.

Skilled performance depends on practice The third stage, motor behavior, exhibits human skills being observed, measured and analyzed in scientific experiments (Flach, 1990). Rasmussen (1983) proposed a classical tripartite scheme how to develop these skills, emphasizing the need of frequent rehearsals to gain *stimulus-response* skills. Many manual tasks in transportation (Langewiesche, 1944) or in other areas of daily life like sports (Chapman, 1968) depend on highly skilled psychomotor processes. Two components are important for the acquisition of motor skills: (1) guided training with sufficient repetitions (Lintern & Gopher, 1978; Buckley & Caple, 2009) and (2) practice such as on-the-job-training (Fleishman, 1966; Savion-Lemieux & Penhune, 2005).

3. Manual Flying

From round dials to the glass cockpit Especially civil aviation has experienced significant changes in the last decades and so have pilots' tasks. Forty years ago most aircraft had analogous round dial flight instruments, denoted as conventional flightdecks, with only few automated systems. Glass cockpits (beginning in the late 1970s) with synthetic displays (*information automation*), flight management systems (*management automation*) and fly-by-wire control systems (*control automation*) extensively changes a pilot's task from an active to a monitoring task (Billings, 1991b). Even when a modern, highly automated aircraft can be manually controlled, today's airliners are predominantly operated under high levels of automation for safety and economic reasons (see Parasuraman, Sheridan, & Wickens, 2000). Nowadays, the aviator turned into a system manager and is endangered to lose situation awareness (Endsley & Kiris, 1995; Sarter & Woods, 1991) and to suffer from mode confusion (Sarter & Woods, 1995). Manual flying has suffered from reduced repetitions due to increasing system management and is, therefore, subject to this thesis. In addition, a pilot with distinct manual flying skills is considered as the *last line of defense*, i.e., if all automated systems fail, the pilot must be able to land the aircraft safely.

A pilot's manual flying task at a glance Manual instrument flying, denoted as *raw data* flying, is a psychomotor process of continuously gathering information by referencing the primary flight instruments (Casner, Geven, Recker, & Schooler, 2014) on the primary flight display (PFD; Figure 2) and performing manual control by primary and secondary flight controls (Appendix, Table 6). These tasks can be structured into primary and secondary flight controls (SKYbrary, 2016a) and analogously to primary and secondary driving tasks (Bubb & Wohlfarter, 2013). Two control loops can be considered for manual flying. The inner control loop comprises the pilot interacting with the inceptor (flightstick or yoke)—a typical corresponding analysis focuses on flightstick inputs. The outer control loop additionally comprises the flight control system manipulating control surfaces and environmental influences resulting in a measurable flightpath.

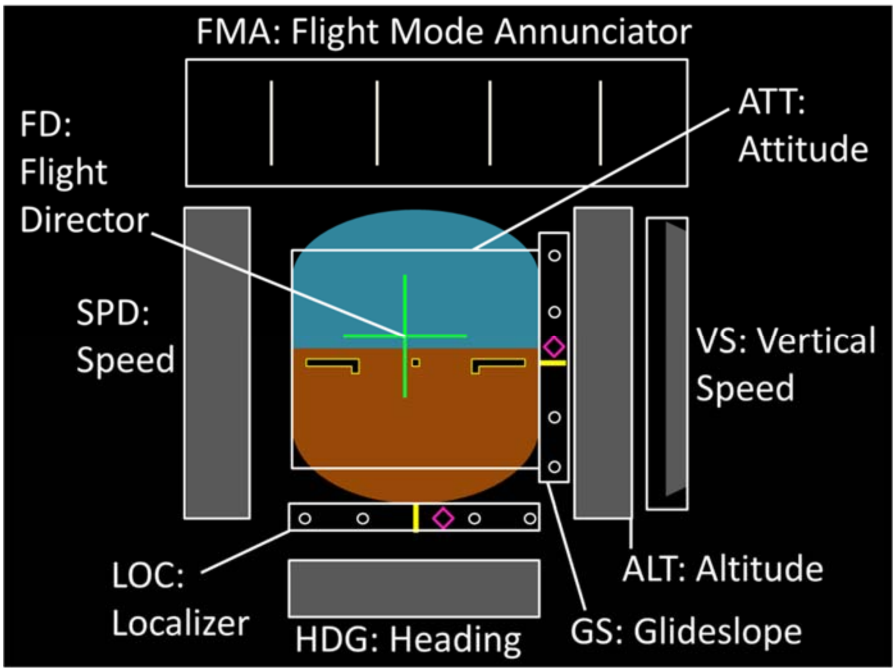


Figure 2: Primary flight display (PFD) showing the primary flight instruments (most important indicators for manual flying): artificial horizon / attitude (ATT), speed (SPD), vertical speed (VS), altitude (ALT), heading (HDG), glideslope (GS) and localizer (LOC). Further displays for flying with automation: flight mode annunciator (FMA) and flight director (FD).

A pilot flying's role Independently from the hierarchy on the flight deck (i.e., the captain and the first officer) only one pilot, denoted as pilot flying (PF), conducts the primary flying task, while the other pilot is denoted as pilot monitoring (PM), or recently, as pilot supporting (PS; Popp & Kemény, 2016). According to Billings (1991b), flying can be manually performed with or without supportive systems such as flight director (information automation), autopilot (control automation), autothrust, or autotrim (control automation), depending on the aircraft type. The highest level of automation is provided by the flight management system (management automation) and only cedes a supervisory role to the pilot flying—the very opposite of *raw data* flying.

Definition of manual flying skills SKYbrary (2016b) describes manual flying skills and pilot handling skills:

> *Manual Flying Skills are typically thought of as pure core flying skills, where maneuvers are flown solely by reference to raw data obtained from the heading, airspeed, attitude, altitude and vertical speed instruments, and without the use of technology such as auto-throttles, auto-pilot, flight*

> *director or any other flight management system. This might extend as far as requiring manual trim inputs and navigation using basic systems.*

This refers to a complete manual control of an aircraft as mentioned above and includes all stages of human information processing. It sets the agenda how to advance in an analysis of manual flying skills: a pilot's performance has to be measured and evaluated in terms of visual scanning and flightpath tracking (see Dick, 1980; Yang, Kennedy, Sullivan, & Fricker, 2013; Allsop & Gray, 2014).

Manual flying skills are considered as the last line of defense If highly automated systems fail, the human pilot must be capable to take over control of an aircraft and to land the aircraft safely, postulated by the *aviate–navigate–communicate–administrate* prioritization (Schutte & Trujillo, 1996). Pilot candidates start their training in small aircraft to develop basic flying skills (Langewiesche, 1944). They follow a typical learning process from knowledge-based to skill-based behavioral patterns founded on frequent repetitions (Rasmussen, 1983). However, in line operation of modern aircraft only few opportunities are available for manual control: the takeoff (~ 30 s of manual control after liftoff) is always performed manually and in many cases the landing including final approach, flare and touchdown is also this way (< 5 min). In short-haul service, pilots frequently conduct shorter flights due to legal flight and rest time regulations, while on long-haul service few but long flights are typical. Pilots also practice their manual flying skills in recurrent flight simulator training depending on the training curricula and on the number of trainings per year. The legal minimum for recurrent training and checking is twice per year, since an operator (pilot) proficiency check is valid for six months (European Union, 2008, p. 170).

Manual skills decay due to the use of automation Manual skills are obtained with repetition, but also decay without practice (Mengelkoch, Adams, & Gainer, 1971; Childs & Spears, 1986; Arthur, Bennett, Stanush, & McNelly, 1998). Warnings (Wiener & Curry, 1980; Arnold, 2015) and empirical evidence for diminishing skills are reported (Puentes, 2011). Different studies found evidence for declining manual flight proficiency due to a rising exhibition to automation (Veillette, 1995; Young, Fanjoy, & Suckow, 2006), insufficient experience (Taylor, Kennedy, Noda, & Yesavage, 2007; Yang et al., 2013) and infrequent practice (Gillen, 2008; Ebbatson, 2009). This loss of manual skills threatens airline pilots differently depending on their personal type of operation (short-haul vs. long-haul) and additional factors, especially part-time schedules due to parental leaves or management duties. This thesis focuses on measurement and evaluation of two opposing areas of manual skills according to human information: visual perception and manual performance.

Evaluation of visual scanning by gaze patterns Visual attention is influenced by salience, effort, expectancy and value of an information (Wickens et al., 2013).The observation of the human eye can reveal which information is available for cognitive processes (Goldberg & Wichansky, 2003). The analysis of visual scanning in the tradition of eye-tracking can be done by several different ways (Milton, Jones, & Fitts, 1950). Gaze-based metrics according to the automotive standard ISO 15007-1 (2015) are taken for basic performance evaluation, e.g., to identify the ratio of performed checks for important indicators (Björklund, Alfredson, & Dekker, 2006; Sarter, Mumaw, & Wickens, 2007) based on the lengths of scanning intervals (Allen, Clement, & Jex, 1970). Gaze-based analyses answer the question *how performant was the scanning*, but does not explain the *structure of the scanning process* and *why the objects were scanned*. Theory suggests models emphasizing on salience, value (Schütz et al., 2011), effort and expectancy (Wickens et al., 2013). They lead to distinct behavioral patterns (see van Leeuwen, Happee, & de Winter, 2015) to be analyzed with sequence-based methods such as the analysis of gaze patterns (Dick, 1980; Colvin, Dodhia, & Dismukes, 2005) or concrete scanpaths (Noton & Stark, 1971; Underwood, Chapman, Brocklehurst, Underwood, & Crundall, 2003; Holmqvist, Nyström, Andersson, Dewenhurst, Jarodzka, & van de Weijar, 2011, p. 254).

Evaluation of manual control Measurement and evaluation of the accuracy of a manually controlled aircraft according to a desired flightpath (the outer control loop) is straightforward. A common approach is to compare an ideal flightpath to the observed deviations. The root mean squared error (RMSE) is a common averaging metric (Rantanen, Johnson, & Talleur, 2004) when the direction information is not of interest (see Hubbard, 1987). The mean error and the standard deviation are further metrics to consider (Ebbatson, 2009) or put into relationship to each other (McClernon & Miller, 2011). Pilots’ control inputs are smoothened by the flight control system in normal flight operation. A pilot’s control behavior on the inceptor (the inner control loop) is therefore rarely analyzed in civil aviation, but can evaluate raw behavioral patterns. Ebbatson, Harris, Huddlestone and Sears (2010) proposed a frequency-domain approach (Baron, 1988), however, the corresponding metrics (e.g., power spectral density) lack of practical relevance for aviation. Time-domain measures seem more promising for civil aviation (training) due to their easy applicability and clarity.

4. Accident Statistics

The landing phase bears the highest risk The Boeing (2015, p. 20) statistical summary compares the occurrence of aviation accidents for the last ten years with the exposition to different flight phases during a flight (Table 1). It highlights takeoff, initial climb, final approach and landing as high-risk phases. The highest risks are found at the final approach and the landing: these two flight phases are very short (~ 5-10 min) but feature a high number of accidents. These highlighted phases are burdened with high relative accident rates and have several characteristics in common:

- high amount of manually flown maneuvers
- high amount of flight management / avionics computers inputs
- high amount of flight deck communication and decisions
- high amount of radio communication with air traffic control (ATC)
- low time budget in regard to a low altitude which potentially means a low time to ground collision

Table 1: Distribution of accidents to different flight phases compared to their duration (exposition) derived from the Boeing Statistical Summary (Boeing, 2015, p. 20)

	fatal accidents	exposure	ratio
taxi / load / parked / tow	10%		
takeoff	7%	1%	7
initial climb	6%	1%	6
climb	7%	14%	0.5
cruise	13%	57%	0.2
descent	3%	11%	0.3
initial approach	8%	12%	0.7
final approach	24%	3%	8
landing	24%	1%	24

Strong statistical evidence for lacking manual skills In the past five years, flight crew errors were among the top three contributing factors for aircraft accidents with a share of 30% (IATA, 2016, p. 46). Manual handling and flight control errors were the most important errors in this category, with clear relationships towards *loss of control* (also denoted as *upset*, Newman, 2012), *runway excur-*

sion, *undershoot*, *hard landing* and *tailstrike*. The IATA (2016, p. 119) recommends (1) to land in the touchdown zone, (2) to define the aiming point of the runway as the target for pilots, (3) to adhere to parameters defining a stabilized approach, (4) to actively call out deviations from a desired landing course, (5) to actively measure the safety management system and (6) to implement flight data monitoring. These two statistical sources (Boeing, 2015; IATA, 2016) clearly suggest improving safety in (1) the final approach and landing phase as the flight phase to represent the highest risk and (2) manual handling and flying skills as an element of airmanship to reveal large deficits.

5. Research Goal

Research gap The preceding chapters presented the concept of pilots' information processing to manually fly an aircraft, how automation counteracts with practicing these skills and indicates a landing scenario as the most relevant scenario. However, to date, there was no holistic approach to evaluate manual flying skills under consideration of the following requirements:

(1) Addressing different levels of practice
(2) Addressing different levels of experience
(3) Investigating under preferably realistic conditions
(4) Measuring and in-depth analyzing visual scanning as well as manual control performance in parallel
(5) Deriving holistic and practice-oriented opportunity areas such as training improvements or organizational changes for application in the airline industry

The goal is to quantitatively evaluate pilots' performance This thesis aims to address the abovementioned research gap by flight simulator experiments with airline pilots representing different levels of practice and experience in a realistic, manual and fine-motor (final) approach and landing scenario analyzing pilots' visual behavior and control performance. Pilots' performance ratings are frequently unsoundly rated by flight instructors showing poor inter-rater reliability (Gontar & Hoermann, 2016). In this work pilots' manual flight performance, therefore, is evaluated against objectively measurable performance standards in the first step and behavioral patterns are identified and compared in the second step.

Research conducted and content structure Due to organizational reasons there were three complementing experiments (Table 2). The main research to analyze manual flight performance was split into two periods: two out of four groups of pilots participated in experiment 1 while the other two groups of pilots took part in experiment 3. Experiment 2 was additionally conducted to furtherly investigate visual scanning without interrupting the process of the main experiment and to evaluate methods to analyze gaze patterns in a continuous tracking task. For this thesis experiment 2 aiming at a special case of visual scanning is reported first followed by general results on visual scanning derived from experiment 3. Finally, consecutive analyses of manual control performance derived from experiment 1 and 3 are reported.

Table 2: Overview of research conducted

Experiment	Year	Aim	Participants	Publications
1	2011	Effect of practice and training on fine-motor flying skills	30 FOs A320, 30 CPTs A340	Haslbeck, Kirchner, Schubert, & Bengler (2014); Haslbeck & Hoermann (2016); Haslbeck, Hoermann, & Gontar (2017)
2	2013	Effect of restricted vision (occlusion) on manual aircraft control and effect of wind interferences on visual scanning	11 FOs A320	Haslbeck & Bengler (2016)
3	2013	Effect of practice on fine-motor flying skills	30 CPTs A320, 30 FOs A340	Haslbeck & Zhang (2017); Haslbeck & Hoermann (2016); Haslbeck, Hoermann, & Gontar (2017)

Note: The numbering of the experiments follows an order in respect of time, while the sequence of the research reported follows a consecutive order in respect of content.

Challenges for objectivity, reliability, validity Safety research aims at safety improvements by investigating the interplay of human performance and human errors denoted as performance variability (Hollnagel, 2012) and depends on the observation of natural and realistic behavior of operators. This thesis focuses on experimental studies with scientific standards in terms of objectivity, reliability, and validity. To reach objectivity, adequate metrics are necessary to evaluate human performance and avoid subjective ratings by flight instructors (see Gontar & Hoermann, 2016) which are typically taken for pilots' legal licensing tests. For reliability, the use of highly standardized and repeatable experimental conditions ensured by a programmed simulator scenario, a scripted procedure of the flight scenario and a clear definition of the roles of all experimenters have to be ensured (see Schubert & Haslbeck, 2014). Pilots representing different levels of practice need to be considered and other potential effects need to be controlled or eliminated to ensure validity in measuring the influence of practice on manual flight performance. In addition, the experiments shall maintain a possibly high degree of realism to observe pilots' unbiased professional behavior (see Rosenthal & Rosnow, 2007) by the use of a random sample of pilots.

Practice and training The independent variable of the main experiment (1 and 3) was the *level of practice and training* (following the German term *Trainiertheit*) describing the manual flight proficiency level of a pilot. This construct frequently addressed in training research is not a single measurable variable but depends on at least three influencing effects (Table 3). In flight school the foundation for manual skills is established by intense and highly repetitive training of manual flying tasks, also denoted as *stick and rudder skills* (Langewiesche, 1944). Daily flight practice represents the amount of recent iterations for psycho-motor tasks such as fine-motor flying (see Mengelkoch, Adams, & Gainer, 1971; Ebbatson, 2009). Based on legal flight time regulations, the kind of flight operation has a strong effect on the recent number of flights and, therefore, on daily flight practice. This number could be high in short-haul operation, while being low in long-haul operation. A third and not less important influence on flying skills is the recurring flight simulator training offering short but intense opportunities to train previously learned control patters under an instructor's supervision (Buckley & Caple, 2009). However, flight simulator training does not only focus on manual aircraft handling, but also on operational tasks such as standard operating procedures or the conduct of abnormal situations. As a consequence, flight simulator training might not sufficiently support manual flight proficiency alone. However, simulator training can hardly be operationalized and expressed in standardized numbers. This leads to the conclusion that flight simulator training cannot (easily) be taken as predictor for manual flying skills although it is an influencing factor.

Table 3: Three main components of the concept ‘level of practice and training’ having effects on manual flying skills.

Source	Operationalization [measurable index]	Relation to a high and low level of manual flying skills
initial flight training / school	time dated back to flight school [years]	first officers (high) captains (low)
daily / recent flight practice	number of recent flight operations including manual aircraft handling [manual landings as pilot flying within past 30 days]	short-haul pilots (high) long-haul pilots (low)
flight simulator training	time dated back to simulator training addressing manual skills [months] duration of simulator training [hours] content of simulator training [-]	

Visual interruptions Pilots have to scan several concurring areas of interest (AOI) and manage a trade-off between visual resources (Wickens et al., 2013). Scanning one display results in a visual interruption for another one. To evaluate the influence of these interruptions on visual scanning (experiment 2) the occlusion paradigm (Gelau & Krems, 2004; Bengler, 2014) simulating a non-cognitive task was selected.

Age and experience In all experiments there were several potential covariates mediating manual flight performance and have, therefore, been recorded in some cases. A covariate being recorded and analyzed was age (see Table 4). Due to the strong effect of daily flight practice the time since flight school was also considered as a covariate but highly correlates to age. For pilots’ experience partly behaves inversely proportional to practice (Table 4). Consequently, experience was subsequently varied like an independent variable but can also be considered as a covariate.

Table 4: Generic demographics of pilots showing typical ranges for different variables

Fleet	Rank	Practice	Experience	Age	Time since Flight School [years]
short-haul	first officer	high	low	20-30	0-10
short-haul	captain	high	high	40-50	20-30
long-haul	first officer	low	low	30-40	10-20
long-haul	captain	low	high	50-60	30-40

Conduct and design of experiments According to the variation of practice and a large spread for age and experience, four groups of pilots (ATP licensed) were selected to take part for the experiments of this thesis (Table 4). Experiment 1 considered an extreme groups design investigating manual skills of short-haul first officers and long-haul captains. Experiment 3 complemented these two groups by the missing two groups, short-haul captains and long-haul first officers. For a realistic spread of manual flight performance and to avoid a self-selection bias (Rosenthal & Rosnow, 2007) pilots in these four groups were randomly selected. Experiment 2 concentrated on a homogenous high-level group according to practice and selected volunteering short-haul first officers. The basic scenario for all included experiments was a manual raw data based approach to a familiar and safe airport (Munich and Frankfurt). The experimental task was a raw data based fine-motor approach and landing from an altitude of 3,000 ft above ground level (AGL) until touchdown. The following chapter introduces the research question for all experiments.

6. Research Questions

The main research question for this thesis and the superordinate research questions for all included papers was:

- Main RQ: What is the effect of *the level of practice and training* on manual flying skills?

Besides this, the included papers have additional, focused research questions to investigate human information processing during manual flight more in detail.

Paper on scanning under restricted conditions (Haslbeck & Bengler, 2016):

- RQ1: How do visual interruptions influence a pilot’s visual scanning?
- RQ2: Which information has the highest importance for a pilot’s attention under visual interruptions?
- RQ3: Which displays are primarily scanned when a pilot initiates visual scanning during manual flight?

Paper on pilots’ gaze patterns (Haslbeck & Zhang, 2017)

- RQ4: Do pilots exhibit recurring gaze patterns in visual scanning?
- RQ5: Is there a superior gaze pattern in manual flying?

Papers on pilot’s manual flight performance (Haslbeck, Kirchner, Schubert, & Bengler, 2014; Haslbeck & Hoermann, 2016):

- RQ6: Is there a strongest predictor among all variables for manual flight performance?
- RQ7: Are there additional predictors for manual flight performance?

Paper on outer and inner loop control (Haslbeck, Hoermann, & Gontar, 2017):

- RQ8: Do pilots exhibit recurring outer loop control patterns in manual flightpath tracking?
- RQ9: Is there a superior outer loop control pattern in manual flying?
- RQ10: Do pilots exhibit recurring inner loop control patterns in manual flightpath tracking?

- RQ11: Can pilots access different outer and inner loop control patterns depending on the context?

7. Results

7.1. Pilots‘ Visual Scanning and Scanpaths Under Restricted Conditions (Haslbeck & Bengler, 2016)

Background Airline pilots experience a high density of different tasks, decisions and information in the landing phase. Besides the primary flying task, the pilot flying has some other tasks and checks which occasionally interrupt the visual scanning. This experiment is as a pre-study to the main experiment.

Rationale and method The occlusion paradigm was used to analyze the effect of interruptions on pilots’ manual flight performance. Certified airline pilots representing a high and homogenous level of manual skills had to perform a fine-motor landing scenario in a fixed-base Airbus-like flight simulator. The manual flight scenario was a fine-motor ILS approach (Frankfurt airport) starting at an altitude of 4,500 ft AGL until touchdown and was flown with and without occlusion. Gaze-based metrics such as the percent time on AOI, the number of glances and the mean single glance duration were taken to describe pilots’ visual scanning. To identify the first three single glances, a sequence-based approach was selected and 3-elements scanpaths were analyzed.

Results The primary flight display accounted for about 95% of all glances—a finding which might reflect the simplified experimental task. The attitude indicator proved its importance for manual flying by all analyses and thereby confirmed previous research.

RQ1 Even if occlusion showed a significant effect on the attention distribution, it did not result in relevant changes. Consequently, the influence of visual interruptions was considered as minor and pilots seemed not to have experienced a general breakdown in visual scanning—they were able to keep a stable tracking.

RQ2 However, precursors for attentional narrowing were found for the speed and heading indicators in conjunction with visual interruptions. Occlusion has indicated a prioritization among different information (attitude > altitude > heading > speed) when visual resources get severely restricted.

RQ3 Pilots showed a measurable link between their initial scanpaths and the predominant flight control task, both focusing on vertical tracking.

Conclusion Well practiced pilots demonstrated certain resilience in situations of limited resources to safely handle their flying tasks. They also showed potential areas for breakdowns of visual scanning. In addition, gaze based metrics contributed in determining pilots' performance. The sequence-based approach analyzing scanpaths added valuable information describing pilots' behavioral patterns.

7.2. Pilots' Gaze Patterns in Manual Flight (Haslbeck & Zhang, 2017)

Background Pilots need an efficient panelscan strategy to facilitate accurate flight maneuvers, but low practice might dampen performance of visual perception: previous research in road traffic identified behavioral deficits in driver's information acquisition due to lacking experience.

Rationale and method This paper addressed the influence of practice on visual scanning, attempted to identify different gaze patterns, and evaluated them against each other under regard of the corresponding manual control performance. Participating pilots had to perform a fine-motor instrument approach and landing scenario (at the Munich airport). The pilots were assigned into two groups representing different levels of practice: short-haul pilots represented a high level of practice, while long-haul pilots represented a low level. The subsequent gaze patterns analysis applied a new method based on saccade grouping sets to classify different visual patterns.

Results The well-practiced short-haul pilots showed a significantly better manual performance on the ILS flightpath. The gaze-based analysis revealed: (1) long-haul pilots had a larger percent time on AOI on the attitude indicator, (2) glances towards the heading indicator had a significant longer duration compared to other AOIs and (3) the attitude indicator corresponded to the highest glance rates. However, univariate tests indicated that (4) the factor AOI had significantly more influence on gaze-based metrics compared to practice.

RQ4 The structure in pilots' visual scanning confirmed that recurring gaze patterns exist. Two basically differing saccade grouping sets were found: an attitude-centered pattern denoted as *spokes-of-a-wheel* and a *triangular* pattern. The spokes types were mainly observed among long-haul pilots, while short-haul pilots showed a more balanced scanning using both types at equal shares.

RQ5 Pilots predominantly using the triangular types showed significantly better manual flight performance compared to their colleagues, who mainly used the spokes types. Since the spokes types are considered to support monitoring, the triangular patterns can be judged as superior for manual flight.

Conclusion This paper showed again, that gaze-based analyses alone hardly identify a successful visual scanning. Only with the more detailed analysis of dominant saccades, correlations with fine-motor flight performance were found. Short-haul pilots showed other dominant saccades and partly a broader repertoire of scanning strategies in combination with a superior flight performance. The findings of this paper (using a new method) are unique: concrete gaze patterns helped to identify behavioral deficits of lesser practiced pilots. These patterns showed that unpracticed pilots apply an inappropriate monitoring strategy in situations when active control instead of sole monitoring needs to be performed.

7.3. Investigation on Pilots‘ Fine-Motor Control of an Aircraft (Haslbeck, Kirchner, Schubert, & Bengler, 2014)

Background Lacking manual flying skills came into the focus of aviation industry due to prominent aviation accidents like Air France flight 447 or Asiana Airlines flight 214. However, there is a research gap concerning studies to investigate the effect of practice and training on fine-motor flying skills.

Rationale and method The effect of the level of practice and training on fine-motor flying skills was investigated by an extreme-groups design: 57 randomly selected professional airline pilots had to complete a flight simulator scenario with a fine-motor ILS approach. All participants were assigned to two groups which extremely differed in their level of practice and training: short-haul (Airbus A320) first officers represented a high level, while long-haul (Airbus A340) captains represented a low level. Deviations from an ideal ILS flightpath (maximum deviations and root mean square errors) were taken as measures for the fine-motor flight performance.

Results Highly practiced short-haul first officers showed better fine-motor flight performance compared to the less practiced long-haul captains. According to legal licensing standards, 25.9% of the captains and 6.7% of the first officers would have failed a licensing test with their manual flight performance.

RQ6 Practice was found to be a predictor for manual flight performance differentiating two groups of pilots.

RQ7 The aircraft type, age and time since flight school (see Table 3) were also identified as potential factors to influence manual flight performance. However, this experimental design could not rule out this potential confound.

Conclusion This experiment showed that well practiced pilots could easily perform the experimental task and stay within legal limits for ILS approaches. In contrast, hardly practiced pilots performed the task with way more variance. In short, practice defeated experience for manual flight performance. However, there were several possible confounds with the different levels of practice: different ages and the two different aircraft types. While the effect of age was considered as minor this experimental design couldn't rule out the other confounds. Consequently, more pilot groups needed to be considered.

7.4. Effects on Pilots‘ Outer Loop Control Performance (Haslbeck & Hoermann, 2016)

Background Experiment 1 showed that practice had a significant effect on manual flying skills, but several confounding factors limited the expressiveness. The experimental design needed to additionally address more groups of pilots to rule out the abovementioned confounds and to get more detailed insights into the effects on manual flight performance.

Rationale and method Experiment 3 presented the same scenario and the same experimental task to short-haul (Airbus A320) captains and long-haul (Airbus A340) first officers compared to experiment 1. The main hypothesis of this paper was that daily flight practice has a stronger effect on manual skills compared to the time period since flight school leading to the following expected order of manual flight performance: 2F > 2C > 4F > 4C. The factor *fleet* was taken to express the differences between pilots on the Airbus A320 and A340 fleet; the factor *rank* was taken to distinguish between first officers and captains. Three different altitude levels representing different task difficulties were considered for the approach, denoted as the factor *altitude*. The covariates *time since initial training* and *age* were also analyzed.

Results The data of experiment 1 and 3 were merged, refined and analyzed by several multivariate analyses of variance to investigate the effect of fleet, rank, altitude and time since initial training on the fine-motor flight performance. The factor fleet was found to have a significant effect on fine-motor flight performance showing large effect sizes. The factors rank and altitude had also a significant effect on flight performance with smaller effect sizes. The analyses of time since initial training gave evidence that recent flight practice is more important for manual flight proficiency compared to long past flight school. The resulting order of manual flight performance was: 2C > 2F > 4F > 4C, indicating that short-haul captains showed the best fine-motor flying skills.

RQ6 Practice improves flight performance–it was found to be the strongest predictor for manual flight performance showing a large effect.

RQ7 The time since flight school only showed a minor effect on manual flight performance. Altitude addressing different taskloads also showed a large effect. The covariates age respectively time since flight school only showed minor effects.

Conclusion This study emphasized the importance of different effects on fine-motor flight performance, identifying daily flight practice as the most important one. It also indicated that manual skills can be regained by a high amount of daily practice after several years of only few practice opportunities (2C > 4F). The confounding effect of fleet and age / time since flight school could be ruled out by statistical analyses, as well as the confounds of fleet and aircraft type was considered as minor due to the design of the fly-by-wire systems.

7.5. Outer and Inner Loop Control Performance (Haslbeck, Hoermann, & Gontar, 2017)

Background The preceding investigation found an erosion of manual flying due to an extensive use of automation and identified several factors to have an effect on fine-motor flying skills. However, it still remained unclear which behavioral patterns or strategies pilots applied to perform fine-motor flying in detail. Other researchers analyzed manual flight data in the frequency-domain, but there are no in-depth analyses of manual flight data in the time-domain addressing effects such as recent practice and experience.

Rationale and method Two different areas were considered to investigate strategies and patterns in pilots' control behavior: outer and inner loop control performance. All valid datasets from experiments 1 and 3 were analyzed. A new method to analyze outer loop flight performance by differentiating between constant and variable flightpath errors was proposed. This method facilitates the identification of different control behaviors depending on pilots' effort. Time-domain measures were taken into account for the evaluation of flight control inputs. The dimensionality of sidestick inputs—combining (two-dimensional) or separating (one-dimensional) the pitch and roll axis—was used to describe how the pilots handled the inceptor.

Results The sidestick input data confirmed the previously identified effect of daily flight practice. In addition, the pilots who predominantly used one-dimensional sidestick inputs and, thus, separated both input axes also had smaller deviations from the ideal flightpath.

RQ8 Two different outer loop performance strategies were found, denoted as *optimizer* and *steady path* strategy.

RQ9 The effortful strategy denoted as optimizer turned out to be superior against an effortless strategy (steady path) in terms of flightpath deviations. However, workload was not considered in this analysis, which also might have an effect on *success* in manual flying.

RQ10 There was a significant effect of practice on the way how pilots controlled their aircraft in the inner loop: well-practiced pilots did less two-dimensional sidestick inputs and in general did fewer inputs on the sidestick.

RQ11 The factor altitude, representing different taskloads and contexts, showed large effects on manual flight performance in all analyses.

Conclusion Behavioral patterns play an important role to explain human performance in detail. In manual flight performance data, different behavioral patterns were identified and evaluated against each other. Factors showing an effect on these patterns have been identified and analyzed, first of all practice. These findings are noteworthy for pilots' training: there are different strategies how to control an aircraft and one of those facilitates better manual flight performance.

8. Effects on Visual Scanning and Manual Control

Effect sizes on manual flight performance The different experiments applied analyses of variance to investigate different effects on manual flight performance. The according effect sizes were quantified by η_p^2 (partial eta squared) as a common measure. In spite of commonly overestimating effects (Levine & Hullett, 2002), some conclusions are drawn from the comparison of effect sizes (Table 5):

- For single glances, the context (condition) and the kind of information (AOI) have the most important effects, while practice (fleet) seems to be negligible.
- For manual control performance, the context (altitude) and practice (fleet) have the most important effects, while experience (rank) only plays a minor role.
- For performance on the outer control loop, the difference of effect sizes between practice (fleet) and experience (rank) is larger compared to the difference on performance on the inner control loop.

Table 5: Analyzed effects on manual flight performance

Paper	Dependent Variable	Factor	η_p^2
Haslbeck & Bengler, 2016	percent time on AOI	condition	.59
		AOI	.53
		condition * AOI	.25
	number of glances	condition	.52
	mean glance duration	condition	.34
		AOI	.26
		condition * AOI	.19
Haslbeck & Zhang, 2017	percent time on AOI	fleet	< .01
		AOI	.37
		fleet * AOI	.06
	mean glance duration	fleet	< .01
		AOI	.45
		fleet * AOI	< .01
	glance rate	fleet	.03
		AOI	.46

		fleet * AOI	.09
Haslbeck & Hoermann, 2016	ILS flightpath tracking	fleet	.45
		rank	.08
		fleet * rank	.08
		altitude	.80
		altitude * fleet	.14
		altitude * rank	.02
		altitude * fleet * rank	.03
	touch down point deviations	fleet	.18
		rank	.04
		fleet * rank	.11
Haslbeck, Hoermann, & Gontar, 2017	ILS sidestick inputs	fleet	.29
		rank	.12
		fleet * rank	.01
		altitude	.90
		altitude * fleet	.18
		altitude * rank	.01
		altitude * fleet * rank	.16

Note. Effects which were analyzed and found to be not significant ($\alpha > .05$) are marked in grey letters. Since effect sizes do not add to 1.0, η_p^2 (partial eta squared) was taken as measure for the effect size due to comparability (Levine & Hullett, 2002). The effect sizes for percent time on AOI were analyzed and reported in the manuscript version of Haslbeck & Zhang (2017), however, they were dropped for the final version.

9. Conclusions

From simple performance metrics to behavioral patterns This thesis applied a two-level approach to analyze human performance for both fields of interest, i.e., visual scanning and manual control. The first step was to roughly analyze pilots' performance by means of *simple metrics* (e.g., percent time on AOI, deviations from ideal ILS course). This first step is a common approach for the evaluation of human performance in complex human-machine systems if rating criteria exist. Corresponding result are insights about which (groups of) observed participants showed good or poor performance. The second step was an in-depth analysis of *behavioral patterns* (e.g. gaze patterns, fine-motor patterns in sidestick inputs). The derived results showed which behavioral patterns were applied while performing specific tasks. This second and effortful step is rarely found in applied studies. However, both steps together deliver more worthwhile insights compared to their independent application. With the combination of both steps different levels of manual flight performance were measured in this thesis. Corresponding behavioral patterns were successfully matched to these different performance levels and, consequently, an evaluation of behavioral patterns was possible.

9.1. Effects on manual flight performance

Practice has the most important effect on manual control The fleet affiliation indicating practice was found to have the most important effect on manual control performance, but had only a negligible effect on scanning of single information displays. The effect of practice on manual control performance was one of the most important research questions and corresponding hypotheses were proposed and confirmed. However, the low effect of practice on single glances was unanticipated. To put it another way, performing single glances seems to be quite robust against different and potential insufficient levels of practice. Only when analyzing gaze patterns practice showed an effect on the structure of the sequences. Pilots with insufficient flight practice showed behavioral deficits in manual flying. Increasing the flight practice for long-haul crews is an important goal for future flight safety measures.

Experience has a small effect and frequently it is ill-defined Experience (i.e., the crew rank) only had a minor effect on manual control performance. In many cases (e.g., professional domains) the common understanding of *experience* is misleading (Duncan, Williams, & Brown, 1991). Experience is frequently considered to include recent fine-motor practice as well as theoretical problem solving and operational knowledge (Robinski & Stein, 2013). In the case of senior long-haul pilots, this assumption is not valid any longer: these pilots have accumulated a vast amount of flight hours and recorded a lot of knowledge how to cope with certain situations but they recently lack of practice for important psycho-motor processes. Expertise, therefore, is an ill-defined concept in aviation and should be avoided as an explanatory factor if possible. Only when having a strong theoretical foundation and well measurable metrics this concept may be accepted as an independent variable in according experiments.

Importance of the meaning of an information The kind of information (depicted by the AOI) per se effects a corresponding task. Thus, AOIs frequently affected single glances. However, Haslbeck and Zhang (2017) reported one important detail when the meaning of an information changes due to a change of the pilot’s task. When compensating ILS flightpath deviations the meaning of the indicators for localizer and glideslope change from performance instruments to control instruments. Only well practiced pilots could cope with this change and adjust their gaze patterns while lesser practiced pilots remained in accustomed patterns (see Haslbeck et al., 2012).

The given context has the strongest effect The context (depicted by condition and altitude) in which pilots perform their task has an important effect on their performance. Especially different altitude levels mean a different content and difficulty of the task resulting in various performances. The three final sub-phases of an approach prior to the landing (preparing the stabilized approach, stabilized approach and visual approach below 250 ft AGL in this case) require different tasks and difficulties. Table 5 indicated the highest effect sizes for the experiments reported. However, the context cannot be easily manipulated for real scenarios and flying tasks and is a fixed precondition for flight operations.

9.2. Summarized findings and take home messages

Performance is similarly distributed compared to practice The level of practice and training had a significant and large effect on manual flying skills: a high level of practice and training facilitates distinct and accurate flying skills. Short-haul pilots formed a homogenously performing group, showed superior manual flight proficiency and mastered the experimental scenario. In contrast, participating long-haul pilots formed a heterogeneously performing group: no one performed better than the best short-haul colleagues, some also mastered the scenario to satisfaction and some could only complete the scenario with difficulty. Several long-haul pilots would even have failed a licensing test with the performance shown in this experiment.

Behavioral patterns also reflect the level of practice Both groups of pilots could also be differentiated by the exhibited behavioral patterns in regard of visual scanning and fine-motor control. Well-practiced pilots applied superior patterns and strategies and adapted their behavior according to the context. In contrast, many long-haul pilots partially showed behavioral deficits: they applied less sophisticated strategies and patterns and were not able to adapt their behavior to a situation’s demands.

Less flying pilots are at risk of skill degradation According to the findings of this thesis, long-haul operation experiences a higher exposition to risks derived from manual flying skill deficits due to few practice opportunities. Obviously, long-haul crews could not compensate their skill deficits with a higher experience expressed by flight hours (Ebbatson, 2009). Besides long-haul pilots, there are two further groups at risk of skill degradation. Pilots with management duties in different operations departments at their airlines have reduced flight duties. Even in short-haul operation, a 50% or less part time schedule might remarkably deteriorate manual flight proficiency. The other group is pilots on parental leave for longer periods of time such as one year or longer. Even if they get additional flight simulator training when returning to the flightdeck they experience a substantial decrease in practice and consequently a decrease in skill.

9.3. Recommendations for the airline industry

Organizational recommendations for air transport companies By remembering the three components of *the level of practice and training* (practice, flight school and simulator training), two areas for safety improvements can be identified: recent practice and simulator training. To raise recent practice, i.e., the number of flights of long-haul crews, mixed-fleet-flying with the same (Soo, Mavin, & Roth, 2016) or several (Lyall & Wickens, 2005) type ratings should be applied. Long-haul pilots (e.g., on Airbus A330/340) could obtain a second type rating for the Airbus A320 fleet and quickly raise their level of practice when facing a long and short haul roster. However, the above mentioned authors also pointed out potential safety issued with this practice. An alternate approach is to operate highly frequented short-haul connections with larger long-haul aircraft, which is not a rare practice. A general increase in flight simulation training lessons would be desirable, however, well directed aims have to be addressed such as manual flying tasks (Haslbeck, Drees, Rehmann, & Tüshaus, 2013) or gaze patterns (Shapiro & Raymond, 1989; Wetzel, Anderson, & Barelka, 1998) adapted to individual needs (Lintern & Gopher, 1978).

Design recommendations for airframe manufacturers Another set of recommendations addresses aircraft design. The pure automation of flying tasks while allowing the pilots to get *out of the* loop cannot be considered as a desirable goal. Intelligent or adaptive (Parasuraman, 2000; Geiselman, Johnson, & Buck, 2013) automation was proposed as a dynamic distribution of flying tasks between human pilots and the technical systems. The pilots should be kept *in the loop* to a medium level according to the context, pilots' capacity and the availability of the technical systems. However, it seems that highly automated systems would not reach extreme low failure rates practically excluding technical defects in aviation. The main irony of automation, automation failing in the most difficult situations for a human operator still needs to be expected to happen (Bainbridge, 1983). Consequently, human operators need to maintain their full range of manual skills according to their control tasks.

References

Allen, R. W., Clement, W. F., & Jex, H. R. (1970). *Research on Display Scanning, Sampling, and Reconstruction Using Separate Main and Secondary Tracking Tasks* (NASA Contractor Report No. 1569). Washington, DC.

Allsop, J., & Gray, R. (2014). Flying under pressure: Effects of anxiety on attention and gaze behavior in aviation. *Journal of Applied Research in Memory and Cognition*, *3*(2), 63–71. doi:10.1016/j.jarmac.2014.04.010

Arnold, R. (2015). Role of Pilot Lack of Manual Control Proficiency in Air Transport Aircraft Accidents. *Procedia Manufacturing*, *3*, 3142–3146. doi:10.1016/j.promfg.2015.07.862

Arthur Jr., W., Bennett Jr., W., Stanush, P. L., & McNelly, T. L. (1998). Factors That Influence Skill Decay and Retention: A Quantitative Review and Analysis. *Human Performance*, *11*(1), 57–101. doi:10.1207/s15327043hup1101_3

Bainbridge, L. (1983). Ironies of Automation. *Automatica*, *19*(6), 775–779.

Baron, S. (1988). Pilot control. In E. L. Wiener & D. C. Nagel (Eds.), Human factors in aviation (pp. 347–386). San Diego: Academic Press.

Bengler, K. (2014). Driver Distraction. In D. Crolla, D. E. Foster, T. Kobayashi, & N. Vaughan (Eds.), *Encyclopedia of Automotive Engineering* (pp. 1–8). Chichester, UK: John Wiley & Sons, Ltd.

Billings, C. E. (1991a). Toward a Human-Centered Aircraft Automation Philosophy. *The International Journal of Aviation Psychology*, *1*(4), 267–270.

Billings, C. E. (1991b). *Human-Centered Aircraft Automation: A Concept and Guidelines: NASA-TM-103885*. Moffet Field, California. Retrieved from http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19910022821_1991022821.pdf

Björklund, C. M., Alfredson, J., & Dekker, S. W. A. (2006). Mode Monitoring and Call-Outs: An Eye-Tracking Study of Two-Crew Automated Flight Deck Operations. *The International Journal of Aviation Psychology*, *16*(3), 257. doi:10.1207/s15327108ijap1603_2

Boeing Commercial Airplanes. (2015). *Statistical Summary of Commercial Jet Airplane Accidents*. Worldwide Operations 1959-2014. Retrieved from http://www.boeing.com/resources/boeingdotcom/company/about_bca/pdf/statsum.pdf

Bubb, H., & Schmidtke, H. (1993). Systemstruktur. In H. Schmidtke (Ed.), *Ergonomie* (3rd ed., pp. 305–333). München: Hanser.

Bubb, H., & Wohlfarter, M. (2013). Eye-Tracking Data Analysis and Neuroergonomics. In M. Fafrowicz (Ed.), *Ergonomics design and management. Neuroadaptive systems. Theory and applications.* Boca Raton, Fla: Taylor & Francis.

Buckley, R., & Caple, J. (2009). *The theory & practice of training* (6th ed.). London: Kogan Page.

Casner, S. M., Geven, R. W., Recker, M. P., & Schooler, J. W. (2014). The Retention of Manual Flying Skills in the Automated Cockpit. *Human Factors*, *56*(8), 1506–1516. doi:10.1177/0018720814535628

Chapman, S. (1968). Catching a Baseball. *American Journal of Physics*, *36*(10), 868–870.

Childs, J. M., & Spears, W. D. (1986). Flight-skill decay and recurrent training. *Perceptual and Motor Skills*, *62*(1), 235–242. doi:10.2466/pms.1986.62.1.235

Colvin, K., Dodhia, R., & Dismukes, K. (2005). Is Pilots' Visual Scanning Adequate to Avoid Mid-Air Collisions? In *Proceedings of the 13th International Symposium on Aviation Psychology* (pp. 104–109).

Deutsche Lufthansa. (2016). *Safety Management Systems zur Verbesserung der Flugsicherheit (SaMSsys): Verbundvorhaben im Luftfahrtforschungsprogramm*. Schlussbericht (No. 20V0803). Frankfurt am Main.

Dick, A. O. (1980). *Instrument Scanning and Controlling: Using Eye Movement Data To Understand Pilot Behavior and Strategies* (NASA Contractor Report No. 3306). Rochester.

Duncan, J., Williams, P., & Brown, I. (1991). Components of driving skill: experience does not mean expertise. *Ergonomics*, *34*(7), 919–937. doi:10.1080/00140139108964835

Ebbatson, M. (2009). *The Loss of Manual Flying Skills in Pilots of Highly Automated Airliners* (PhD Thesis). Cranfield University, Cranfield.

Ebbatson, M., Harris, D., Huddlestone, J., & Sears, R. (2010). The relationship between manual handling performance and recent flying experience in air transport pilots. *Ergonomics*, *53*(2), 268–277. doi:10.1080/00140130903342349

Einhäuser, W., Rutishauser, U., & Koch, C. (2008). Task-demands can immediately reverse the effects of sensory-driven saliency in complex visual stimuli. *Journal of Vision*, *8*(2), 2.1-19. doi:10.1167/8.2.2

Endsley, M. R., & Kiris, E. O. (1995). The Out-of-the-Loop Performance Problem and Level of Control in Automation. *Human Factors*, *37*(2), 381–394. doi:10.1518/001872095779064555

Commission Regulation (EC) No 859/2008. Official Journal of the European Union (European Union 2008).

Flach, J. M. (1990). Control With an Eye for Perception: Precursors to an Active Psychophysics. *Ecological Psychology*, *2*(2), 83–111. doi:10.1207/s15326969eco0202_1

Fleishman, E. A. (1966). Human Abilities and the Acquisition of Skill: Comments on Professor Jones' Paper. In E. A. Bilodeau (Ed.), *Acquisition of Skill* (pp. 147–167). New York: Academic Press.

Foulsham, T., Dewenhurst, R., Nyström, M., Jarodzka, H., Johansson, R., Underwood, G., & Holmqvist, K. (2012). Comparing scanpaths during scene encoding and recognition: A multi-dimensional approach. *Journal of Eye Movement Research*, *5*(4), 1–14. Retrieved from www.jemr.org/online/5/4/3

Geiselman, E. E., Johnson, C. M., & Buck, D. R. (2013). Flight deck automation: Invaluable collaborator or insidious enabler? *Ergonomics in Design: The Quarterly of Human Factors Applications*, 21(3), 22–26. doi:10.1177/1064804613491268

Gelau, C., & Krems, J. F. (2004). The occlusion technique: a procedure to assess the HMI of in-vehicle information and communication systems. *Applied Ergonomics*, *35*(3), 185–187. doi:10.1016/j.apergo.2003.11.009

Gillen, M. (2008). *Degradation of Piloting Skills* (Master's Thesis). University of North Dakota, Grand Forks.

Goldberg, J. H., & Wichansky, A. M. (2003). Eye Tracking in Usability Evaluation - A Practitioner's Guide. In J. Hyönä, R. Radach, & H. Deubel (Eds.), *The mind's eye. Cognitive and applied aspects of eye movement research* (pp. 493–516). Amsterdam: North-Holland.

Gontar, P., & Hoermann, H.-J. (2015). Interrater Reliability at the Top End: Measures of Pilots' Nontechnical Performance. *The International Journal of Aviation Psychology*, *25*(3-4), 171–190. doi:10.1080/10508414.2015.1162636

Hamblin, C. J., Gilmore, C., & Chaparro, A. (2006). Learning to Fly Glass Cockpits Requires a New Cognitive Model. In Human Factors and Ergonomics Society (Ed.), *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting* (pp. 1977–1981).

Haslbeck, A., & Bengler, K. (2012). Entwicklung einer Fragebogenmethodik zur Erhebung leistungsbeeinflussender Faktoren in unterschiedlichen Berufsfeldern [Development of a Survey Method to Analyze Performance Shaping Factors in Different Domains]. In Gesellschaft für Arbeitswissenschaft (Ed.), *Gestaltung nachhaltiger Arbeitssysteme. 58. Kongress der Gesellschaft für Arbeitswissenschaft* (pp. 615–618). Dortmund: GfA Press.

Haslbeck, A., & Bengler, K. (2016). Pilots' gaze strategies and manual control performance using occlusion as a measurement technique during a simulated manual flight task. *Cognition, Technology & Work*, *18*(3), 529–540. doi:10.1007/s10111-016-0382-2

Haslbeck, A., Drees, L., Rehmann, K., & Tüshaus, A. T. (2013). Anforderungen an das Training manueller Flugfertigkeiten. In M. Grandt & S. Schmerwitz (Eds.): *Vol. 2013-01. DGLR-Bericht, Ausbildung und Training in der Fahrzeug- und Prozessführung. 55. Fachausschusssitzung Anthropotechnik* (pp. 159–167). Bonn.

Haslbeck, A., Gontar, P., & Schubert, E. (2012). The way pilots handle their control stick - effects shown in a flight simulator study. In C. Lemmer (Ed.): *Vol. 19. Reports of the DLR-Institute of Transportation Systems, 30th European Annual Conference on Human Decision-Making and Manual Control. Proceedings* (pp. 21–26). Braunschweig: DLR-Institut für Verkehrssystemtechnik.

Haslbeck, A., & Hörmann, H.-J. (2016). Flying the Needles: Flight Deck Automation Erodes Fine-Motor Flying Skills among Airline Pilots. *Human Factors*, *58*(4), 533–545. doi:10.1177/0018720816640394

Haslbeck, A., Kirchner, P., Schubert, E., & Bengler, K. (2014). A Flight Simulator Study to Evaluate Manual Flying Skills of Airline Pilots. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *58*(1), 11–15. doi:10.1177/1541931214581003

Haslbeck, A., Schubert, E., Onnasch, L., Hüttig, G., Bubb, H., & Bengler, K. (2012). Manual flying skills under the influence of performance shaping factors. *Work: A Journal of Prevention, Assessment and Rehabilitation*, *41*(Supplement 1/2012), 178–183. doi:10.3233/WOR-2012-0153-178

Haslbeck, A., & Zhang, B. (2017). I spy with my little eye: Analysis of airline pilots' gaze patterns in a manual instrument flight scenario. Applied Ergonomics, 63(September 2017), 62–71. doi:10.1016/j.apergo.2017.03.015

Hollnagel, E. (2012). Coping with complexity: Past, present and future. *Cognition, Technology & Work*, *14*(3), 199–205. doi:10.1007/s10111-011-0202-7

Holmqvist, K., Nyström, M., Andersson, R., Dewenhurst, R., Jarodzka, H., & van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford, New York: Oxford University Press.

Hubbard, D. C. (1987). Inadequacy of root mean square error as a performance measure. In *International Symposium on Aviation Psychology*, 4th (pp. 698–704).

International Air Transport Association. (2015). *IATA Air Passenger Forecast Shows Dip in Long-Term Demand: Total passengers set to double to 7 billion by 2034* (Press Release No. 55). Retrieved from http://www.iata.org/pressroom/pr/Pages/2015-11-26-01.aspx

International Air Transport Association. (2016). *Safety Report 2015*. Montréal.

ISO 15007-1:2015 (2015). *Road vehicles - Measurement of driver visual behaviour with respect to transport information and control systems - Part 1: Definitions and parameters*.

Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive Psychology*, *8*(4), 441–480. doi:10.1016/0010-0285(76)90015-3

Kahneman, D. (2011). *Thinking, Fast and Slow*. London: Allen Lane.

Kang, Z., & Landry, S. J. (2015). An Eye Movement Analysis Algorithm for a Multielement Target Tracking Task: Maximum Transition-Based Agglomerative Hierarchical Clustering. *IEEE Transactions on Human-Machine Systems*, 45(1), 13–24. doi:10.1109/THMS.2014.2363121

Langewiesche, W. (1944). *Stick and rudder: An explanation of the art of flying*. New York: McGraw-Hill.

Lehrstuhl für Ergonomie. (2016). *SaMSys – Safety Management System zur Verbesserung der Flugsicherheit*. Retrieved from https://www.lfe.mw.tum.de/en/projects/samsys-safety-management-system-zur-verbesserung-der-flugsicherheit/

Levine, T. R., & Hullett, C. R. (2002). Eta Squared, Partial Eta Squared, and Misreporting of Effect Size in Communication Research. *Human Communication Research*, *28*(4), 612–625. doi:10.1093/hcr/28.4.612

Lintern, G., & Gopher, D. (1978). Adaptive training of perceptual-motor skills: issues, results, and future directions. *International Journal of Man-Machine Studies*, *10*(5), 521–551. doi:10.1016/S0020-7373(78)80018-2

Lyall, B., & Wickens, C. D. (2005). Mixed Fleet Flying Between Two Commercial Aircraft Types: An Empirical Evaluation of the Role of Negative Transfer. In Human Factors and Ergonomics Society (Ed.), *Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting* (Vol. 49, pp. 45–48).

McClernon, C. K., & Miller, J. C. (2011). Variance as a Measure of Performance in an Aviation Context. *The International Journal of Aviation Psychology*, 21(4), 397–412. doi:10.1080/10508414.2011.606765

Mengelkoch, R. F., Adams, J. A., & Gainer, C. A. (1971). The Forgetting of Instrument Fyling Skills. *Human Factors*, *13*(5), 397–405. doi:10.1177/001872087101300502

Milton, J. L., Jones, R. E., & Fitts, P. M. (1950). *Eye Fixations of Aircraft Pilots* (AF Technical Report No. 6018). Dayton, Ohio.

Newman, R. L. (2012). *Thirty Years of Airline Loss-of-Control Mishaps* (AIAA Modeling and Simulation Technologies Conference and Exhibit No. AIAA 2012-4495). Minneapolis, Minnesota.

Noton, D., & Stark, L. (1971). Scanpaths in Eye Movements during Pattern Perception. *Science*, *171*(3968), 308–311. doi:10.1126/science.171.3968.308

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A Model for Types and Levels of Human Interaction with Automation. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, *30*(3), 286–297. doi:10.1109/3468.844354

Parasuraman, R. (2000). Designing automation for human use: empirical studies and quantitative models. *Ergonomics*, *43*(7), 931–951. doi:10.1080/001401300409125

Popp, C., & Kemény, C. (2016). One Team. *The Journal for Civil Aviation training*, *27*(2), 26–28.

Puentes, A. (2011). *The Manual Flight Skill of Airline Pilots* (Master's Thesis). San Jose State University, San Jose. Retrieved from http://scholarworks.sjsu.edu/etd_theses/4109

Rantanen, E. M., Johnson, N. R., & Talleur, D. A. (2004). *The Effectiveness of Personal Computer Aviation Training Device, a Flight Training Device, and an Airplane in Conducting Instrument Proficiency Checks: Volume 2: Objective Pilot Performance Measures*. Final Technical Report AHFD-04-16/FAA-04-6. Oklahoma City.

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*, SMC-13(3), 257–266. doi:10.1109/TSMC.1983.6313160

Rosenthal, R., & Rosnow, R. L. (2007). *Essentials of behavioral research: Methods and data analysis* (3rd). Boston: McGraw-Hill.

Sarter, N. B., Mumaw, R. J., & Wickens, C. D. (2007). Pilots' Monitoring Strategies and Performance on Automated Flight Decks: An Empirical Study Combining Behavioral and Eye-Tracking Data. *Human Factors*, *49*(3), 347–357. doi:10.1518/001872007X196685

Sarter, N. B., Wickens, C. D., Mumaw, R., Kimball, S., Marsh, R., Nikolic, M., & Xu, W. (2003). Modern Flight Deck Automation: Pilots' Mental Model and Monitoring Patterns and Performance. In *Proceedings of the 12th International Symposium on Aviation Psychology.* Retrieved from http://humanfactors.illinois.edu/research/ConferencePapers/isap03.aspx

Sarter, N. B., & Woods, D. D. (1991). Situation Awareness: A Critical But Ill-Defined Phenomenon. *The International Journal of Aviation Psychology*, *1*(1), 45–57.

Sarter, N. B., & Woods, D. D. (1995). How in the World Did We Get into That Mode? Mode Error and Awareness in Supervisory Control. *Human Factors*, *37*(1), 5–19.

Savion-Lemieux, T., & Penhune, V. B. (2005). The effects of practice and delay on motor skill learning and retention. *Experimental brain research*, *161*(4), 423–431. doi:10.1007/s00221-004-2085-9

Schubert, E., & Haslbeck, A. (2014). Gestaltungskriterien für Szenarien in Flugsimulatoren zur Untersuchung von Verhalten und Leistung von Verkehrspiloten. In M. Grandt & S. Schmerwitz (Eds.): *Vol. 2014-01. DGLR-Bericht, Der Mensch zwischen Automatisierung, Kompetenz und Verantwortung. 56. Fachausschusssitzung Anthropotechnik* (pp. 125–137). Bonn.

Schutte, P. C., & Trujillo, A. C. (1996). Flight Crew Task Management in Non-Normal Situations. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *40*(4), 244–248. doi:10.1177/154193129604000422

Schütz, A. C., Braun, D. I., & Gegenfurtner, K. R. (2011). Eye movements and perception: a selective review. *Journal of Vision*, *11*(5). doi:10.1167/11.5.9

Shapiro, K. L., & Raymond, J. E. (1989). Training of efficient oculomotor strategies enhances skill acquisition. *Acta Psychologica*, *71*(1-3), 217–242. doi:10.1016/0001-6918(89)90010-3

SKYbrary. (2016a). *Flight Controls*. Retrieved from http://www.skybrary.aero/index.php/Flight_Controls

SKYbrary. (2016b). *Pilot Handling Skills*. Retrieved from http://www.skybrary.aero/index.php/Pilot_Handling_Skills

SKYbrary. (2016c). *Airmanship*. Retrieved from http://www.skybrary.aero/index.php/Airmanship

Soo, K., Mavin, T. J., & Roth, W.-M. (2016). Mixed-fleet flying in commercial aviation: A joint cognitive systems perspective. *Cognition, Technology & Work*, *18*(3), 449–463. doi:10.1007/s10111-016-0381-3

Taylor, J. L., Kennedy, Q., Noda, A., & Yesavage, J. A. (2007). Pilot age and expertise predict flight simulator performance: A 3-year longitudinal study. *Neurology*, *68*(9), 648–654. doi:10.1212/01.wnl.0000255943.10045.c0

Underwood, G., Chapman, P., Brocklehurst, N., Underwood, J., & Crundall, D. (2003). Visual attention while driving: sequences of eye fixations made by experienced and novice drivers. *Ergonomics*, *46*(6), 629–646. doi:10.1080/0014013031000090116

van Leeuwen, P. M., Happee, R., & Winter, J. de. (2015). Changes of Driving Performance and Gaze Behavior of Novice drivers During a 30-min Simulator-based Training. *Procedia Manufacturing*, *3*, 3325–3332. doi:10.1016/j.promfg.2015.07.422

Veillette, P. R. (1995). Differences in aircrew manual skills in automated and conventional flight decks. *Transportation Research Record*, (1480), 43–50.

Wetzel, P. A., Anderson, G. M., & Barelka, B. A. (1998). Instructor Use of Eye Position Based Feedback for Pilot Training. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *42*(20), 1388–1392. doi:10.1177/154193129804202005

Wickens, C. D., Hollands, J. G., Banbury, S., & Parasuraman, R. (2013). *Engineering psychology and human performance* (Fourth edition). Boston: Pearson.

Wiener, E. L., & Curry, R. E. (1980). Flight-Deck Automation: Promises and Problems. *Ergonomics*, *23*(10), 995–1011. doi:10.1080/00140138008924809

Yang, J. H., Kennedy, Q., Sullivan, J., & Fricker, R. D. (2013). Pilot Performance: Assessing How Scan Patterns & Navigational Assessments Vary by Flight Expertise. *Aviation, Space, and Environmental Medicine*, *84*(2), 116–124. doi:10.3357/ASEM.3372.2013

Young, J. P., Fanjoy, R. O., & Suckow, M. W. (2006). Impact of Glass Cockpit Experience On Manual Flight Skills. *Journal of Aviation Aerospace Education and Research*, *15*(2), 27–32.

Appendix

Table 6: Manual flying with primary and secondary flight controls

Primary (P) or Secondary (S) Flight Controls	Dimension / Function	Control Element	Actuator	Information Source	Supportive Automation
P	lateral control of the roll axis	inceptor (flight stick, yoke)	ailerons	heading (HDG) indicator	autopilot, flight director
P	°lateral control of the yaw axis	rudder pedals	rudder	heading (HDG) indicator	
P	vertical control of the pitch axis	inceptor (flight stick, yoke)	elevators	altitude (ALT) and vertical speed (VS) indicators	autopilot, flight director
P	°vertical control of the vertical speed	thrust levers	engines	altitude (ALT) and vertical speed (VS) indicators	autothrust
P	longitudinal control of the airspeed	thrust levers	engines	speed (SPD) indicator	autothrust
S	high lift devices	flaps lever	flaps, slats	ECAM flaps indicator	
S	spoilers, speedbrakes	speedbrake lever	spoilers		
S	trim	trim wheel	horizontal stabilizer	ECAM flight controls page	autotrim

Note. Some degrees of freedom (marked by °) can be controlled by more than one control element or are cross-coupled, i.e., a pilot's control action has an effect on more than one dimension.

Appended Papers

Haslbeck, A., & Bengler, K. (2016). Pilots’ gaze strategies and manual control performance using occlusion as a measurement technique during a simulated manual flight task. Cognition, Technology & Work, 18(3), 529–540. doi:10.1007/s10111-016-0382-2

This paper can be originally found at:

https://link.springer.com/article/10.1007/s10111-016-0382-2

Cogn Tech Work (2016) 18:529–540
DOI 10.1007/s10111-016-0382-2

ORIGINAL ARTICLE

Pilots' gaze strategies and manual control performance using occlusion as a measurement technique during a simulated manual flight task

Andreas Haslbeck[1] · **Klaus Bengler**[1]

Received: 10 April 2016 / Accepted: 2 June 2016 / Published online: 13 June 2016

Abstract The aim of this study was to analyze pilots' visual scanning under conditions visually restricted by the occlusion paradigm. During manual flight, pilots experienced interruptions in their panel scan due to concurring tasks and monitoring of distant displays. Eleven volunteer airline pilots performed several manual instrument landing system approaches in a fixed-base flight simulator. Some of these approaches were performed using the paradigm of occlusion with shutter glasses. Under occlusion, the gaze pattern analysis revealed that pilots demonstrated reduced mean glance durations, but did not reduce their attention to lesser information displays. The results also indicated that the attitude indicator (artificial horizon) as a preview instrument was less affected by occlusion compared to other areas of interest. A subsequent scanpath analysis revealed that vertical tracking was the predominant information acquisition strategy and corresponded to larger deviations on the glideslope. These results imply the need to optimize information even for short glances, and to be very cautious with adaptive layouts of free programmable or dynamic displays, and not to overburden the pilot flying with parallel tasks.

Keywords Occlusion · Visual scanning · Gaze pattern · Scanpath · Manual flying

1 Introduction

Manual flying is a psychomotor process and an active task involving predominantly visual scanning and manual performance (Childs and Spears 1986; Flach 1990; Sarter and Woods 1994). Especially in raw data conditions, i.e., without supporting systems such as autothrust, autopilot, or the flight director, this task becomes challenging for pilots. The pilot has to cognitively process information about altitude, speed, and the flightpath by reference to the corresponding indicators on the primary flight display (PFD). We have used the occlusion paradigm in a flight simulator study to gain a better understanding of the information acquisition processes.

Occlusion is a well-established experimental method in automotive as well as in usability research, and it is standardized by ISO 16673 (ISO 2007) ('Occlusion method to assess visual demand due to the use of in-vehicle systems'). It is designed primarily to investigate visual information processing and is at the foundation of tachistoscope-based research (Milgram 1987). Sleight (1948) evaluated cockpit instrumentation using tachistoscopic presentations of different dial types. Senders et al. (1967) introduced occlusion to road traffic safety research. In contrast to eye tracking, the occlusion method is an experimental paradigm that allows for the sequential observation of gaze patterns using highly standardized and restricted information presentation. It facilitates research questions different to eye-tracking approaches such as an easy identification of the starting and ending eye fixation points (see Kang and Landry 2015). Furthermore, occlusion can simulate task interruptions and investigates the extent to which an operator can handle these disruptions by adding visual but not cognitive demand. In contrast, occlusion does not facilitate direct measurement of eye glances. Therefore, it

---

✉ Andreas Haslbeck
haslbeck@tum.de

[1] Institute of Ergonomics, Technische Universität München, Boltzmannstrasse 15, 85747 Garching, Germany

may be an ideal complement for eye tracking to measure and evaluate visual strategies under restricted conditions. Occlusion was successfully used to evaluate operator performance and estimate attentional demand not only in the automotive domain, but also in other fields involving highly skilled performance such as aviation (Milgram 1987). Gray et al. (2008) have taken the occlusion paradigm in a different way: in a manual fine-motor flying task to maintain a constant altitude, they occluded the ground surface by 3D objects. In the current study, the benefit compared to more intense techniques such as eye tracking is that occlusion allows for a very selective and controlled manipulation of an operator's visual perception, which is a very important channel for many psychomotor control tasks (Helleberg and Wickens 2003; Gelau and Krems 2004) with a clear distinction between the information a subject should receive and the information to be ignored.

Another purpose of the method is to measure the extent to which the processing of rapidly changing information suffers from interrupted presentation of information (Chen and Milgram 2011). Under occlusion, the gaze of a person is periodically interrupted and reestablished either by the subject, i.e., *subject paced* (van der Horst 2004), or by using a shutter mechanism, i.e., *system paced*. By experimentally restricting visibility, the minimum time required for the perception of visualized information can be determined for a given experimental task. There are two common methods for occluding visual displays. The more common approach is performed using shutter glasses, which can be switched between a transparent and opaque view, thus maintaining a continuous environmental luminous density. An alternative method is called display blanking, where information displays are periodically switched off (Krause et al. 2015). Occlusion was standardized as a paradigm to evaluate the visual demand of secondary tasks while driving a car (e.g., manipulating a navigation system or a radio): an investigator measures how much visual attention is required for an individual task, while the glance is interrupted at 1.5-s intervals (ISO 16673). The shutter open (or a display visible) represents the time available for the task of interest, while the shutter closed (blanked display) represents the time needed for a concurring task, which is the driving task/scene with an automotive background (Krems et al. 2000). Such studies aim to measure human performance to develop and evaluate technical systems.

In this study, performed according to the human information processing model by Wickens et al. (2013) and the approach by Senders et al. (1967), we assumed that trained and experienced pilots will use strategies to optimize visual behavior in coping with limited resources and time restrictions to maintain performance of a continuous control task. The use of occlusion gives standardized conditions to limit the access to visual information and to let pilots apply strategies to focus attention and visual behavior.

1.1 Research questions

We assumed that there is a lower limit of visual information to successfully perform the manual flying task. One aim of this paper was to determine whether occlusion affects pilots' gaze patterns while performing a manual instrument landing system (ILS) task and, subsequently, whether occlusion can be applied for pilots' performance research. Under occlusion, two different mitigation strategies are possible due to the reduced time budget for visual perception. One strategy is a reduction in attention due to a smaller number of observed areas of interest (AOIs). An alternative could be a decline in the duration of single glances to cope with the reduced time budget without reducing the number of observed AOIs. A second aim of this paper was to investigate the first sequence of glances when pilots initiate a panel scan. Three research questions were developed for this study:

RQ 1: How does occlusion influence a pilot's visual behavior?
RQ 2: To which AOI does occlusion reduce a pilot's attention?
RQ 3: Which AOIs are scanned first at the beginning of a pilot's visual strategy?

The research questions were all based on the assumption that pilots attempt to keep their flight performance stable/constant even when their vision is interrupted (as per the instructions). The first research question refers to an exploratory approach for measuring pilots' behavior and performance with and without interrupted vision due to occlusion. This paradigm is rarely found in the aviation context, and the method is not familiar to commercial pilots. However, pilots experience intermittent vision interruption as a result of intramodal (visual) resource competition (Helleberg and Wickens 2003) due to concurring visual demands from parallel tasks. Besides the visual scanning of the PFD in a manual approach, pilots also have to check other near and distant displays in parallel such as the wind indicator on the navigation display, the percent thrust on the engines display, and superordinate tasks such as performing the landing checklist. Harris and Christhilf (1980) believe that a glance duration of 500 ms is essential for gathering information from most cockpit instruments. It is worth noting that pilots can maintain such glance durations while under the influence of occlusion. When applying the occlusion technique, we hypothesized that mean single glance durations decline because of the reduced time budget for visual perception.

The second research question addresses the assumption that pilots must reduce their attention to relevant information under occlusion, thus showing strategies for management of visual attention resources (see Wickens and Alexander 2009). A typical occlusion setting of 1500 ms shutter open and 1500 ms shutter closed (ISO 16673) cuts the available time for information acquisition in half. The hypothesis behind this research question assumes a concentration on the primary task (flying the aircraft) and thus a reduction in viewed areas of interest (reduced monitoring), mainly focusing on the (PFD) applying the aviate–navigate–communicate–manage systems (ANCS) prioritization (Schutte and Trujillo 1996; Morris and Leung 2006). In addition, the question arises how difficult would the ISO-suggested shutter open–close procedure be for pilots. We hypothesized that the system-paced 1500 ms open/1500 ms closed scheme is a manageable restriction for them. Thus, a slightly more difficult occlusion condition should also be covered by the experiment: a more severe occlusion condition with 1500 ms open/2000 ms closed was added. We did not intend to disrupt pilots' visual scanning with a much harder occlusion condition possibly leading to failure in the landing scenario.

The third research question aims to analyze the initial sequence of AOI reflecting the core elements of the visual strategy to maintain the scanning task. With longer sequences, recurrent gaze patterns can be identified, as Dick (1980, p. 12) and Jones (1985, p. 17) mentioned in their work. Typical glance durations in pilots' visual scanning are 500 ms (Harris and Christhilf 1980). Thus, a sequence of three AOIs will be considered for this analysis, as this is the most likely length of a sequence during 1500 ms shutter open time and represents the information most prioritized by the pilots. If the predominant patterns show a higher frequency than the product of the independent likelihood of the single AOIs, these patterns can be considered meaningful and regarded as implicit visual strategies (Ellis and Stark 1986). Sequence analyses were applied by Underwood et al. (2003) in road traffic. They analyzed gaze patterns of novice and expert drivers considering single fixations, transition probabilities between two AOIs (2-element scanpaths), and 3-element scanpaths. Single fixations revealed no group differences in the mean single glance durations but an effect of different road types and traffic situations. Transition probabilities calculated by the first-order Markov matrix (see Liu 1998) were visualized and depicted in several figures. These single transitions delivered first indications for differences between novices' and experts' visual scanning. However, only the 3-element scanpaths allowed a deeper insight into drivers' visual behavior, denoted as scanning strategies. The most dominant one was the *preview strategy* starting and ending at the road several seconds ahead.

2 Method

In this study, we analyzed eye-tracking data derived from a flight simulator study using the occlusion paradigm based on the experiments by Senders et al. (1967) to measure human performance in a manual flying task. Further results of this study have already been reported (Gontar et al. 2013; Haslbeck et al. 2014a). The main research question in the original study addressed the influence of pilots' information acquisition on their performance execution and vice versa. The basic idea of one part of the original study was to manipulate (disturb) a pilot's visual perception and to measure execution in terms of flightpath deviations using a manual ILS approach. The occlusion technique by shutter glasses was selected for the manipulation of the visual channel. In the second part of this study, the aim was to disturb the aircraft on its flightpath by cross- and tailwind while measuring changes in the (non-occluded) visual behavior.

2.1 Participants

All eleven volunteer participants (9 men and 2 women) holding a valid airline transport pilot license met the requirements for a very high and homogenously performing group of pilots, when considering manual flying skills. First officers (FOs) scheduled on the Airbus A320, working for the same airline as pilots for at least 2 years, participated (mean age = 29.4 years, SD = 3.4). Their manual flight proficiency was very high due to very frequent flights on short-haul duty [mean individual landings as pilot flying (PF) in past 30 days = 16.4, SD = 6.6 as measure of practice]; longtime operational experience was neither needed for this experiment nor present in participating FOs (mean flight hours on A320 = 2368, SD = 1231 as measure of experience).

2.2 Apparatus

We conducted this experiment in a fixed-base simulator with a generic Airbus A320 cockpit instrumentation system (Fig. 1) and that of Dornier 728-based flight dynamics. For the entire test, the flight dynamics were set to a Direct Law configuration (SKYbrary 2016), which meant that pilots' control inputs were directly transmitted to the control surfaces and no automatic trim or protection modes were available. For the occlusion, a set of PLATO (portable liquid crystal apparatus for tachistoscopy via visual occlusion) visual occlusion spectacles were used. Pilots' eye movements were recorded with a monocular head-mounted Dikablis Essential eye tracker. Both visual research devices can be combined and worn by the participant (Fig. 1); however, eye-tracking data quality was reduced by the shutter glasses.

Fig. 1 Simulator setting showing a participant with mounted shutter glasses and eye tracker

2.3 Scenario, instruction, and procedure

The flight scenario, with a duration of about 4 min, was a manual ILS approach to Frankfurt Airport (EDDF) beginning at an altitude of about 4500 ft. above ground level (AGL). From this altitude down to 500 ft. AGL, different interferences were presented, depending on the experimental conditions. Below 500 ft. AGL, all interferences faded out, to make it possible for participants to safely land and to avoid a missed approach. All participants had the role of the PF, and their task was to land manually as accurately as possible according to glideslope, localizer, and airspeed, equivalent to a real approach. One member of the experimental team operated the landing gear and the flaps and assisted as a rudimentary pilot monitoring. However, there were no other tasks except for the manual approach for the PF such as communication with air traffic control or navigation tasks for the participants, implying a distinctive focus on aviating compared to navigating, communicating, and systems management (see Schutte and Trujillo 1996).

The experiment started with a (1st) set of three baseline approaches with moderate turbulences and light tail- and crosswinds so the pilots could become acclimated with the simulator. These moderate turbulences were present in all subsequent trials to maintain a moderate difficulty level with continuous control activities. The consecutive set of three approaches was dedicated to interfering with the manual control by an oscillating tailwind (5–25 kts.) in one and an oscillating crosswind (10–50 kts.) in another trial. A third trial in this set used an approach without interferences (no wind). The next set of three approaches was dedicated to interferences due to occlusion with an easy and a difficult occlusion condition (Table 1), and one trial was done with the shutter open constantly (no occlusion). The order of the trials within the wind (2nd set) and the occlusion (3rd set) set was counterbalanced, while the order of the sets was kept constant. The experiment ended with a baseline approach equal to one of the first three approaches. Four approaches are considered for this paper: two trials under different occlusion conditions and two without occlusion, all four without any wind interference (Table 1), but all with the aforementioned moderate turbulences.

Table 1 Four experimental conditions considered in this paper

Name	Conditions
no_wind	Moderate turbulences (2nd set)
no_occ	Same as no_wind, participants wear shutter glasses, shutter open all time (3rd set)
occ_easy	Same as no_occ but shutter is 1500 ms open and 1500 ms closed (3rd set)
occ_diff	Same as no_occ but shutter is 1500 ms open and 2000 ms closed (3rd set)

2.4 Dependent measures

Eye-tracking data on the PFD were calculated and analyzed for four different areas of interest within the PFD (Fig. 2): attitude indicator/artificial horizon (ATT), altitude (ALT), speed (SPD), and heading (HDG). The flight mode annunciator (FMA) was not used because of the manual flight scenario, without automation, thus not providing any information. Especially in a manual ILS approach, the indicators for glideslope (GS), localizer (LOC), and vertical speed (VS) were among the most important AOI. However, due to insufficient tracking accuracy, they could not be handled as separate AOI: GS and VS were combined with ALT, and LOC was combined with HDG.

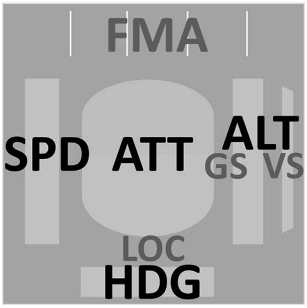


Fig. 2 Areas of interest on the primary flight display considered in this study (*black letters*): speed (*SPD*), attitude (*ATT*), altitude (*ALT*) including the glideslope indicator (*GS*) and vertical speed (*VS*), and heading (*HDG*) including the localizer indicator (*LOC*). Some areas (*GS*, *VS*, *LOC*) could not be separately considered due to eye-tracking accuracy. The flight mode annunciator (*FMA*) was not considered because no automation modes were used

Table 2 Different information acquisition strategies and matching AOI

Information acquisition strategy	Matched AOI
Vertical tracking	ATT, ALT, (VS), (GS)
Horizontal tracking	ATT, HDG, (LOC)
Airspeed tracking	ATT, SPD, (VS), (GS)

AOI in brackets indicates that these AOIs provide meaningful information about the respective strategy; however, they could not be recorded and analyzed separately in this study

Hayashi (2004, p. 17) matched several instruments to three types of instrument groups: vertical, horizontal, and airspeed tracking (see Table 2). Within these groups, different instruments depicting similar information were included, e.g., the vertical-tracking instrument group included the pitch indicator of ATT, ALT, VS, and GS. The same applied for horizontal tracking, including the bank indicator on ATT, HDG, and LOC. Thus, our practice combining AOI seemed valid, matching different information together.

For the evaluation of information acquisition (RQ1 and RQ2), gaze-based metrics according to ISO 15007-1 (ISO 2015) ('Measurement of driver visual behavior with respect to transport information and control systems') were taken into account. Gaze data were only measured when the shutter was open. The *percent time on AOI* informs about the distribution of glances and shows changes comparing the visual behavior under different conditions. In terms of attentional narrowing, we defined 5 % as a threshold for the percent time on AOI: percentages below this value might indicate the potential for attentional narrowing. The *number of glances* provides information regarding the glances performed on a specific AOI during one approach. *The mean single glance duration* is a measure used to express the temporal effort for visually obtaining information on a specific AOI: it indicates how much time is needed to gather certain information on an AOI, depending on the design of an AOI and the time available. To answer RQ 3, a sequence-based procedure was necessary: the sequence of single glances was analyzed. Every time the shutter opened in one of the two occlusion conditions, the first three glances at different AOIs were identified, listed, and counted afterward. These 3-element scanpaths can easily be found under occlusion conditions.

Manual flight performance accuracy in terms of flight-path deviations from glideslope and localizer was measured in *dots* and displayed by the GS and LOC. One dot indicates a half scale deflection from localizer or glideslope (European Union 2011, p. 117) and corresponds to a deviation of ±0.8° on the localizer and ±0.4° on the glideslope for Airbus types. The root-mean-square error (RMSE) was taken as a measure to represent the flight performance (Rantanen et al. 2004) for every single trial.

3 Results

3.1 Analysis of eye fixations

The first exploratory analysis of all data showed that a very high percentage of about 95 % of all gazes were directed to the PFD. Thus, other AOIs such as the navigation display or the engines display were nearly completely ignored (see Edwards et al. 1982). The percent time on AOI is a general look at the distribution of glances at the relevant AOI. Therefore, only gaze data regarding the PFD (Fig. 3) were considered in the further analyses. A 4 × 4 (Condition [no_wind, no_occ, occ_easy, occ_diff] × AOI [ATT, ALT, SPD, HDG]) repeated-measures ANOVA was performed finding significant effects of Condition, AOI, and the interaction Condition × AOI (Table 3 upper panel). These results show the importance of ATT with no percentage below 40 %, while SPD had no percentage higher than 15 %. However, no meaningful differences between occluded and non-occluded conditions were discerned. Besides the analysis of mean values, some additional statistics are reported (Table 4). A Chi-square test indicated a significant difference between occluded and non-occluded conditions concerning the number of cases when the percent time on AOI was 5 % or below, $X^2(1) = 6.30$; $p = .010$.

To analyze the mean number of glances, the glances of a participant were summed up over the four AOIs for every condition and compared in a repeated-measures ANOVA with the factor Condition. Results show that the mean number of glances was significantly affected by the type of condition (Table 2 middle panel). Under occlusion which

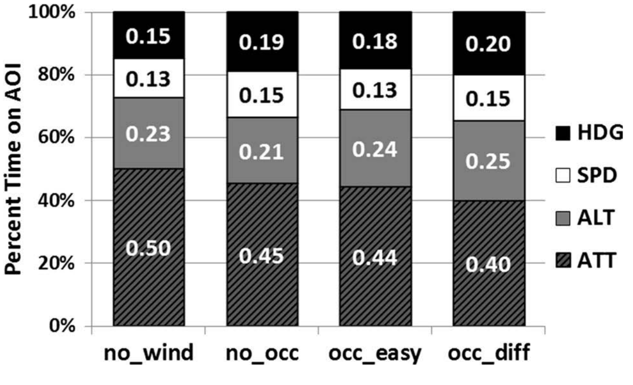


Fig. 3 Percent time on AOI showing the glance distribution depending on the conditions

Table 3 Results of statistical tests for percent time on AOI, mean number of glances, and mean single glance duration

Source	Univariate tests	Bonferroni-corrected post hoc tests
Repeated-measures ANOVA for percent time on AOI[a]		
Condition	$\varepsilon = .42$, $F(1.27, 12.67) = 14.11$, $p = .002$, $\eta_p^2 = .59$	SPD: no_occ-occ_easy ($p = .008$), no_occ-occ_diff ($p = .022$) ALT: no_occ-occ_easy ($p = .034$), no_occ-occ_diff ($p = .024$) ATT: no_occ-occ_easy ($p = .015$), no_occ-occ_diff ($p = .022$)
AOI	$\varepsilon = .41$, $F(1.23, 12.34) = 11.34$, $p = .004$, $\eta_p^2 = .53$	no_wind: HDG-ATT ($p = .026$), SPD-ATT ($p = .004$) no_occ: SPD-ATT ($p = .024$) occ_easy: SPD-ATT ($p = .020$)
Condition × AOI	$\varepsilon = .33$, $F(2.94, 29.36) = 3.36$, $p = .033$, $\eta_p^2 = .25$	
Repeated-measures ANOVA for mean number of glances[b]		
Condition	$\varepsilon = .61$, $F(1.82, 18.17) = 10.72$, $p = .001$, $\eta_p^2 = .52$	no_occ-occ_easy ($p = .011$), no_occ-occ_diff ($p = .009$)
Repeated-measures ANOVA for mean single glance duration[c]		
Condition	$\varepsilon = .67$, $F(2.00, 19.95) = 5.08$, $p < .016$, $\eta_p^2 = .34$	SPD: no_occ-occ_easy ($p = .027$), no_occ-occ_diff ($p = .032$)
AOI	$\varepsilon = .52$, $F(1.57, 15.74) = 3.57$, $p = .062$, $\eta_p^2 = .26$, n.s.	
Condition × AOI	$\varepsilon = .47$, $F(4.27, 42.66) = 2.30$, $p = .071$, $\eta_p^2 = .19$, n.s.	

All effects have been reported as significant at $p \leq .05$ and η_p^2 has been reported as the effect size. If the assumption of sphericity was violated, Greenhouse–Geisser's estimates of sphericity were reported by ε

[a] A Shapiro–Wilk test has revealed that four out of sixteen datasets did not meet the assumption of normality: the glances at HDG for no_occ ($p = .001$), at ALT for occ_diff ($p = .005$) and at ATT for no_wind ($p = .049$) and occ_diff ($p = .002$)

[b] A Shapiro–Wilk test has revealed that two out of four datasets did not meet the assumption of normality: occ_easy ($p < .001$) and occ_diff ($p = .005$)

[c] This analysis is based on log10-transformed gaze data. A Shapiro–Wilk test has revealed that two out of sixteen datasets did not meet the assumption of normality: the glances at ALT for occ_diff ($p = .021$) and at ATT for no_wind ($p = .018$)

Table 4 Number of cases when percent time on AOI was below 5 % for any condition/AOI

ATT	ALT	SPD	HDG	no_wind	no_occ	occ_easy	occ_diff
1 (2 %)	7 (16 %)	15 (34 %)	10 (23 %)	8 (18 %)	2 (5 %)	11 (25 %)	12 (27 %)

A percent time on AOI of 5 % and less was selected as a threshold for the potential of attentional narrowing

halved the available time, the number of glances was reduced by one-third (Fig. 4).

The last metric taken into account was the *mean single glance duration*. Single glance durations are not normally distributed but positively skewed. Thus, a repeated-measures ANOVA with the factors Condition and AOI was performed (Table 2 bottom panel) with log10-transformed data. Figures 5 and 6 show the significant effect of Condition on mean single glance durations. The first and second condition, both without occlusion, showed about 25 % longer mean glance durations than the third and fourth condition with intermittent occluded vision. Neither AOI nor an interaction between both main effects was significant.

Mean values are only of little significance due to positive skewness of glance durations. Thus, distributions of glance durations were plotted (Fig. 6). For ATT, the mode of both types of conditions (occluded and non-occluded) was about 300 ms. Both ATT and HDG had a higher proportion of longer glances in comparison with the SPD and ALT. In contrast to SPD, and similar to HDG, the

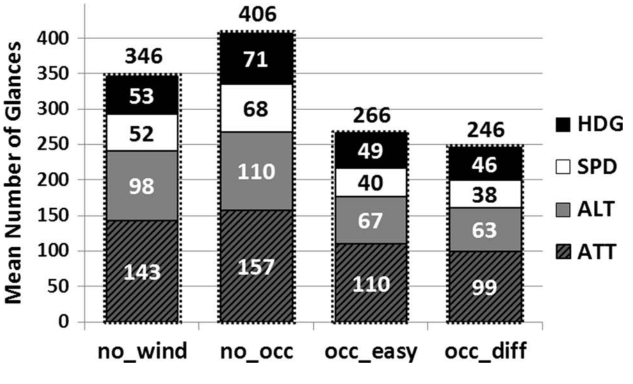


Fig. 4 *Mean number of glances* for four areas of interest on the primary flight display

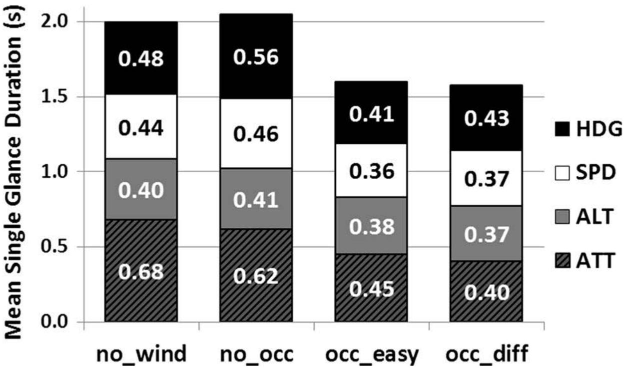


Fig. 5 Mean single glance duration for four areas of interest on the primary flight display

mode of the non-occluded conditions was between 450 and 500 ms, while the occluded conditions show an approximately bimodal distribution with a major mode of 300 ms and minor mode of 500 ms. For ALT, a mode of 300 ms was also found for the occluded conditions. In summary, occlusion changed the modes of ALT, SPD, and HDG, while ATT seemed rather robust against occlusion.

3.2 Analysis of glances to the first three AOIs

All 3-element patterns were listed for every participant and were combined for all participants. It can be seen that ATT is the most frequent AOI (about 43 %, Figs. 3, 4) and found to be the starting AOI for the five most frequent sequences (Table 5), and the starting AOI in 64 % of all patterns. Consequently, ATT is scanned disproportionate high as the first AOI in a panel scan.

To ensure that those 3-element scanpaths were not random, we calculated the independent likelihood of all possible patterns based on the pilots' individual gaze distributions and compared it to the occurring frequencies. If the latter was remarkably higher, we assumed that a 3-element pattern had not occurred randomly, which was the case for at least the five most frequent ones. Table 5 shows the occurrence frequencies for all recorded patterns as well as their independent likelihoods. An illustration derived from eye-based data showing the five most frequent gaze patterns is given in Fig. 7. These five strategies represent nine of eleven participants in the test very accurately, and two showed different visual behaviors.

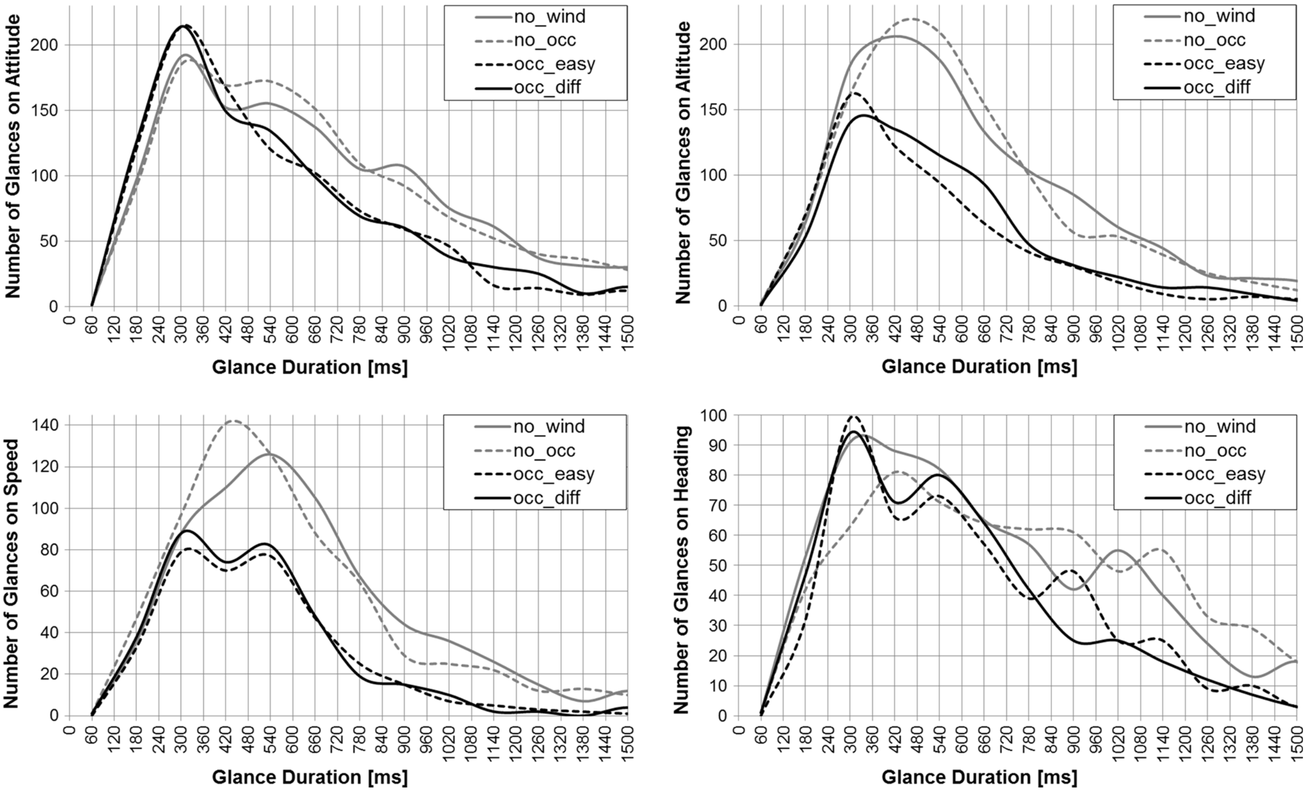


Fig. 6 Distribution of glance durations on four different areas of interest: attitude (*top left*), altitude (*top right*), speed (*bottom left*), and heading (*bottom right*)

Table 5 Absolute number, relative frequency of the 15 most frequent 3-element scanpaths and corresponding independent likelihood

Scanpath	*n*	Relative frequency	Independent likelihood
ATT–ALT–ATT	82	0.14	0.04
ATT–SPD–ATT	46	0.08	0.03
ATT–ALT–HDG	39	0.07	0.02
ATT–ALT–SPD	36	0.06	0.01
ATT–HDG–ALT	33	0.06	0.02
ATT–HDG–ATT	25	0.04	0.03
SPD–ATT–ALT	23	0.04	0.01
ATT–SPD–ALT	22	0.04	0.01
ALT–ATT–HDG	15	0.03	0.02
SPD–ALT–ATT	15	0.03	0.01
ATT–SPD–HDG	13	0.02	0.01
ALT–ATT–SPD	13	0.02	0.01
ALT–HDG–ATT	11	0.02	0.02
HDG–ALT–ATT	10	0.02	0.02
SPD–ALT–HDG	9	0.02	0.01

3.3 Complementary manual flight performance data

Gaze data and corresponding performance data are necessary for the interpretation of psychomotor performance. The pilots' tracking performance on the ILS (localizer and glideslope) also needs to be considered, which was already reported (Haslbeck et al. 2014a). An excerpt of corresponding flight performance data (Table 6) was analyzed: a repeated-measures MANOVA was performed with the factor Condition and two dependent variables GS and LOC. The assumption of normality was violated in two datasets, no_wind_LOC, $D(8) = .3$, $p = .034$, and occ_diff_LOC, $D(8) = .34$, $p = .008$. Univariate tests indicated an effect of Condition on flight performance on the glideslope: $F(3, 21) = 6.62$, $p = .003$, $\eta_p^2 = .49$.

4 Discussion and conclusions

For pilots, it is unfamiliar to wear shutter glasses while performing panel scans in a compelling schedule. However, the usage of occlusion allows for a more systematic analysis of information processing under restricted conditions. The observed flightpath deviations show similar amplitudes compared to other flight simulator studies (Ebbatson et al. 2010; Haslbeck and Hörmann 2016), thus indicating that the occlusion conditions in this study did affect the participants but did not lead to a complete failure in the manual flying task (Table 6). Thus, occlusion appears as an approach to generate valid data to analyze pilots' performance under conditions of restricted visual resources (see Gray et al. 2008).

4.1 Occlusion's influence on pilots' visual behavior (RQ 1)

The results of this study imply that a pilot's visual behavior during a manual approach is influenced by occlusion representing dual-task situations for pilots. Occlusion has exerted significant influence on the percent time on AOI and has significantly reduced the mean single glance durations. One strategy shift revealed here was the reduction in glance duration by about one fourth on average (Fig. 5). The mean number of glances was reduced by one-third under occluded conditions (Fig. 4), which is plausible because of the bisection of the available time for acquisition of visual information. The distribution of glance durations on all AOIs (Fig. 6) delivers additional information. Occlusion has only shown a small influence on the control instrument ATT considering glance durations. However, the performance instruments ALT, SPD, and HDG were read out with reduced glance durations shown by the declined modes. Besides precise values, both indicators also show important trend information. However, to perceive slow trends, longer glances are necessary compared to the perception of a stable value. Insufficient time to track important indicators supporting a stable ILS course (GS on ALT and LOC on HDG) might handicap the manual flight performance. Consequently, occlusion has revealed especially a significant increase in glideslope deviations.

In the non-occluded approaches, the deviations on glideslope are about one-third larger compared to deviations on the localizer (Table 3). The same relative difference of one-third was found in an earlier manual flying experiment without occlusion (Haslbeck et al. 2014b). In contrast to

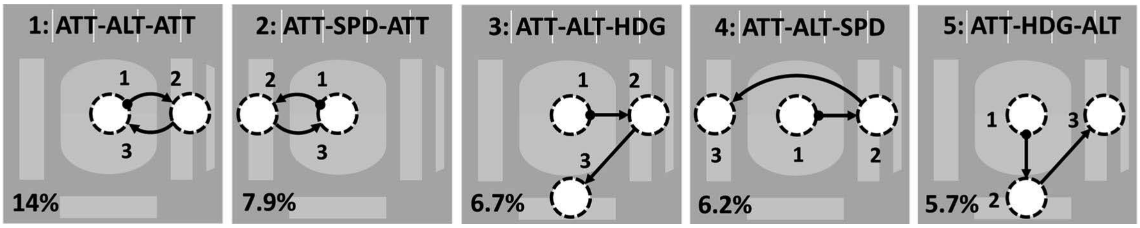


Fig. 7 Five most frequent scanpaths

Table 6 Flightpath deviations on glideslope (GS) and localizer (LOC), excerpt from Haslbeck et al. (2014a)

	no_wind		no_occ		occ_easy		occ_diff	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
RMSE on GS [dot]	.73	.3	.51	.14	.9	.34	1.11	.49
RMSE on LOC [dot]	.54	.25	.39	.15	.65	.43	.54	.36

these approaches with non-restricted vision, the glideslope deviations in the occluded approaches in this study are about two-thirds larger than the localizer deviations. Three supposable reasons for the vulnerability of glideslope performance are (1) the lower angle for of the funnel-shaped glideslope of only ±0.4° compared to ±0.8° on localizer; (2) glideslope is not only manipulated by the sidestick like localizer, but also by the thrust of an aircraft; and (3) in a straight approach, glideslope corresponds to the changing altitude, while localizer refers to the widely stable course.

In summary, occlusion representing secondary tasks (e.g., the landing checklist) has significantly impaired pilots' manual flight performance. These findings show that interruptions of the visual scan of a pilot without any additional cognitive tasks may change a pilot's visual behavior, as well as the accuracy of the manual flying task. Examples for such non- or low cognitive tasks in parallel to the primary flying task are checks of the engines or short glances to the outside while checking the visibility of the runway. Because occlusion interrupts visual perception but not cognition (Monk et al. 2002; Gelau and Krems 2004), the occluded time intervals have presumably been used for further cognitive processing, thus supporting the manual flying task, and there was an interruption of visual contact but no complete interruption of information processing. However, the extent to which an effect of preemption (Helleberg and Wickens 2003) seizes only these cognitive phases remains unanswered. In a real aircraft cockpit, many visual, contextual, and cognitive interruptions, regardless of the secondary tasks occurring simultaneously with the main flying task such as a go-around, do occur. These interruptions might even more negatively influence a pilot's information acquisition (compared to these test conditions) as such tasks also require mental capacity (Morris and Leung 2006) and, therefore, could negatively influence manual flight performance even more. Pilots have significantly reduced the duration of each glance, i.e., accelerated their panel scans; however, many glance durations still have durations similar to glances when reading a digital instrument of about 500 ms (see Senders et al. 1967; Unema and Rötting 1990; Steelman et al. 2011; Chen and Milgram 2011) and less like assessing analog instruments of about 125 ms as established by Harris and Christhilf (1980). This length of time also corresponds to information processing for a manual control task (Zimmer and Stein 2012).

These results give clear implications to the aviation industry: (1) the need to optimize information even for short glances such as with analog instruments but the need to be very careful with adaptive layouts of free programmable or dynamic displays in stressful situations. And (2) it is important not to overload the PF with (standard operating) procedures and secondary tasks and furthermore to consider more support for the PF in high-workload situations by the second pilot. In this regard, the pilot monitoring's role can be discussed and may be advanced to an explicit *pilot supporting* (Popp and Kemény 2016).

4.2 Reduction in pilots' attention under occlusion (RQ 2)

The percent time on AOI was significantly affected by the conditions; however, this statistical effect has not shown meaningful differences between non-occluded and occluded approaches as was hypothesized (regarding the mean values given by Fig. 3). The reduction in the mean single glance duration seems not to generally inhibit thorough information acquisition as discussed above. This leads to the conclusion that pilots' attention distribution did not fundamentally change under occlusion, and no specific information on the PFD was neglected in general. However, individual data (Table 4) implied that there were several cases indicating a tendency of attention narrowing might have been present, supported by a significant Chi-squared test. A trend to neglect SPD can be derived from comparing Figs. 4, 5, and 6 as well as Table 4. The attention on ATT, ALT, and HDG is essential for maintaining the correct flightpath in an ILS approach though they do not account for SPD. SPD is an important parameter in the final phase of a landing and is related to accidents such as runway excursion, long/short landing or tailstrikes (International Air Transport Association 2016, pp. 62–75). The accident of Asiana Airlines flight 214 is a prominent example, in which airspeed was neglected by the pilots and it was one of the most important contributing factors (National Transportation Safety Board 2014, p. 11). However, it is unclear which threshold to select for indicating the beginning of an effect of attention narrowing (see Wickens and Alexander 2009). Thus, a general effect of attention narrowing by occlusion could not be verified. Pilots have shown a certain amount of resilience to interruptions under high workload situations. This experiment cannot document the extent to which this resilience exists:

it can be assumed that longer shutter closed periods or shorter shutter open periods will significantly impair pilots' performance, both representing more interruptions. An implication for the aviation industry is to thoroughly account for pilots' tasks when developing the concept of single-pilot operation without support from a second pilot. The pilot monitoring is an essential backup for the PF when he is working to capacity and single parameters such as SPD fade out of his attention.

A further look at the data has shown that about 95 % of glances were directed to the above-mentioned AOIs on the PFD (see Colvin et al. 2005). This portion is even higher compared to earlier studies (Gontar and Haslbeck 2012). Pilots during a manual flying task focused on their acquisition of information under occlusion clearly to the PFD. This prioritization of selected areas could be due to the instruction and clarification that no other task should be relevant for them besides manual flying. Otherwise, it could be an indication that when there are limited visual resources, pilots demonstrate very efficient visual strategies, when managing the manual flying task (to aviate). The flight data quantify the level of performance that can be achieved under these conditions. This confirms the recommendation to minimize distractions that may disturb the strategy under comparable circumstances.

4.3 First scanned AOI: starting point of a visual strategy (RQ 3)

The analysis of all 3-element scanpaths has also shown the prominence of ATT with about 42 % of percent time on AOI (Fig. 3). Moreover, ATT demonstrated even higher importance as the first scanned AOI in 64 % of all 3-element patterns and the first scanned AOI in seven out of ten of the most frequently shown patterns. This leads to two conclusions: first, it supports the fact that ATT is the most important AOI on a PFD and for a manual flying task (see Spady 1978; Harris and Christhilf 1980). ATT is a window to the outside showing synthetic and important information concerning the exact position of the aircraft and the attitude of flight on one glance. In other words, ATT is a display supporting a preview strategy (see Underwood et al. 2003). Secondly, pilots can memorize the position of the ATT during the shutter closed period and successfully recall it when the shutter opens. A similar habit was described by Anderson et al. (2013) when participants in an observation experiment re-fixated on previously viewed scenes again (see also Foulsham and Kingstone 2013). This finding emphasizes the above-mentioned application to be careful with dynamic displays and not to (dynamically) alter the position of the PFD.

Besides the first scanned AOI, an analysis of the first sequence of three AOIs was performed with emphasis on the five most frequent 3-element patterns. These five prevailing patterns were evaluated as intentional and not random (see Myers 2007, pp. 2–4). However, the question arises how to interpret these different patterns. Hayashi (2004, p. 17) has merged different instruments into three instrument groups based on the tracking of different information for a cockpit layout with analog instruments. For an analysis of a typical PFD (in this case similar to the Airbus layout), we adopted her mapping (Table 2) and this accounted for the fact that some AOI could not be recorded separately from others. When interpreting the five most frequent strategies illustrated in Fig. 7, most transitions between different AOIs occur horizontally on the PFD. Strategy 1 (ATT–ALT–ATT) can be matched to vertical tracking, while strategy 2 (ATT–SPD–ATT) can be matched to airspeed tracking. Strategy 3 (ATT–ALT–HDG) and strategy 5 (ATT–HDG–ALT) both represent a mixed strategy, addressing vertical and horizontal tracking simultaneously. Strategy 4 (ATT–ALT–SPD) refers to combined vertical and airspeed tracking. The question arises whether the temporal order of elements in strategies 3 and 5 may be neglected. Many pilots reported in our studies that they were never told how exactly to perform a panel scan in terms of a direction for an ongoing scan. Under the assumption that temporal order can be neglected, the frequencies of strategies 3 and 5 can be added, resulting in a common frequency of 12.4 %. This approach seems valid at least when focusing on the AOI and neglecting the sequence. This leads to the conclusion that vertical tracking was the most important information acquisition strategy for this manual ILS approach. When comparing this conclusion to the manual flight performance data (Table 6), both correspond to each other: the vertical position of the aircraft was the more important control task, given the larger deviations from an ideal glideslope. And, therefore, the vertical tracking was the prevailing visual strategy. However, within the context of the importance of extrafoveal information for guiding gaze, the following findings may be important for airline training departments: if it is possible to identify dominant gaze patterns and result in highly skilled visual strategies, situation-dependent successful strategies can be developed and taught to pilots to further enhance manual flying skills (see Shapiro and Raymond 1989).

4.4 Limitations to this study

One apparent limitation in this study is that occlusion and interrupted vision is somewhat artificial to airline pilots, which could have been a bias to their visual behavior in the test. However, no explicit evidence for such a bias was found. The lack of aviation occlusion studies is a problem when discussing these results, and comparing them to other domain-related studies and trying to confirm the results. The generic, fixed-base simulator had an unknown characteristic for airline pilots. The results reported in this paper might be confounded with the effects of adaption processes to the unknown flight dynamics (Gontar et al. 2013). A major limitation in the study was the small number of participants ($n = 11$), which accounts for a reduced power in all statistical comparisons.

This experiment was performed using airline pilots holding a valid ATP license and with a type rating for the Airbus A320—a very homogeneous group. We would assume a higher variability in all experimental results when observing pilots with a low level of practice and training, e.g., long-haul crews. Other experiments have shown deteriorated manual flying on long-haul operation (Haslbeck and Hörmann 2016) including decreased visual information acquisition skills (Gontar and Haslbeck 2012). In addition, this study only considered manual approaches and no other flight phases, display concepts, or automation levels (see Edwards et al. 1982).

4.5 Summary

This study analyzed pilots' visual behavior during a manual flying task (ILS approach) under occlusion. Due to these visual interruptions, pilots showed reduced mean glance durations and larger glideslope tracking errors. Hence, occlusion deteriorated manual flight performance. However, pilots did not generally reduce their attention to lesser information displays. Practical implications are the need for optimal information even for short glances, to be very careful with adaptive layouts of free programmable or dynamic displays, and to not to overload the pilot flying with parallel tasks. Afterward, a scanpath analysis revealed that vertical tracking was the predominant information acquisition strategy and this corresponds to larger deviations on the glideslope, fostering visual behavior to be analyzed and improved for training demands.

Acknowledgments The authors acknowledge their thanks to Andreas Jaroš and Ludwig Drees at the Institute of Flight System Dynamics for their technical support in the Research Flight Simulator, as well as Patrick Gontar, Anja Schneider, Olivia Thoma, Moritz Bohnenberger, Hang Yin, and Heiner Bubb for their support with data analysis and interpretation and finally Bastiaan Petermeijer (all Institute of Ergonomics) for his review of the manuscript of this paper.

References

Anderson NC, Bischof WF, Laidlaw K, Risko EF, Kingstone A (2013) Recurrence quantification analysis of eye movements. Behav Res Methods 45(3):842–856. doi:10.3758/s13428-012-0299-5

Chen H-YW, Milgram P (2011) Determining fixed glance duration for visual occlusion research. In: Proceedings of the human factors and ergonomics society annual meeting, vol 55, no. 1, pp 1904–1908. doi:10.1177/1071181311551396

Childs JM, Spears WD (1986) Flight-skill decay and recurrent training. Percept Mot Skills 62(1):235–242. doi:10.2466/pms.1986.62.1.235

Colvin K, Dodhia R, Dismukes K (2005) Is pilots' visual scanning adequate to avoid mid-air collisions? In: Proceedings of the 13th international symposium on aviation psychology, pp 104–109

Dick AO (1980) Instrument scanning and controlling: using eye movement data to understand pilot behavior and strategies. NASA Contractor Report, Rochester

Ebbatson M, Harris D, Huddlestone J, Sears R (2010) The relationship between manual handling performance and recent flying experience in air transport pilots. Ergonomics 53(2):268–277. doi:10.1080/00140130903342349

Edwards RE, Tolin P, Jonsen GL (1982) Pilot visual behavior as a function of navigation and flight control modes in the Boeing 757/767. In: Proceedings of the human factors and ergonomics society annual meeting, vol 26, no. 5, pp 441–445. doi:10.1177/154193128202600513

Ellis SR, Stark L (1986) Statistical dependency in visual scanning. Hum Factors 28(4):421–438

European Union (2011) Commission Regulation (EU) No 1178/2011 of 3 November 2011. In: Official Journal of the European Union, vol 54

Flach JM (1990) Control with an eye for perception: precursors to an active psychophysics. Ecol Psychol 2(2):83–111. doi:10.1207/s15326969eco0202_1

Foulsham T, Kingstone A (2013) Optimal and preferred eye landing positions in objects and scenes. Q J Exp Psychol 66(9):1707–1728. doi:10.1080/17470218.2012.762798

Gelau C, Krems JF (2004) The occlusion technique: a procedure to assess the HMI of in-vehicle information and communication systems. Appl Ergon 35(3):185–187. doi:10.1016/j.apergo.2003.11.009

Gontar P, Haslbeck A (2012) Untersuchung der Nutzung des Primary Flight Displays durch Piloten unterschiedlicher Trainiertheit mittels Blickerfassung [Eye-Tracking Investigation of the Use of the Primary Flight Displays by Pilots of Different Practice]. In: Grandt M, Schmerwitz S (eds) Fortschrittliche Anzeigesysteme für die Fahrzeug- und Prozessführung: 54. Fachausschusssitzung Anthropotechnik, Bonn, pp 263–271

Gontar P, Thoma O, Haslbeck A (2013) Adaption hochgradig geübter psychomotorischer Muster an ungewohnte Simulationsumgebungen am Beispiel von Piloten [Adaptation of highly skilled psychomotor patterns to unfamiliar simulation environments using the example of pilots]. In: Grandt M, Schmerwitz S (eds) Ausbildung und Training in der Fahrzeug- und Prozessführung: 55. Fachausschusssitzung Anthropotechnik, Bonn, pp 169–181

Gray R, Geri GA, Akhtar SC, Covas CM (2008) The role of visual occlusion in altitude maintenance during simulated flight. J Exp Psychol Hum Percept Perform 34(2):475–488. doi:10.1037/0096-1523.34.2.475

Harris RL, Christhilf DM (1980) What do pilots see in displays? In: Proceedings of the human factors and ergonomics society annual meeting, vol 24, no. 1, pp 22–26. doi:10.1177/107118138002400106

Haslbeck A, Hörmann H-J (2016) Flying the needles: flight deck automation erodes fine-motor flying skills among airline pilots. Hum Factors 58(4):533–545. doi:10.1177/0018720816640394

Haslbeck A, Kirchner P, Schubert E, Bengler K (2014) A flight simulator study to evaluate manual flying skills of airline pilots. In: Proceedings of the human factors and ergonomics society annual meeting, vol 58, no. 1, pp 11–15. doi:10.1177/1541931214581003

Haslbeck A, Schneider A, Gontar P (2014) Analyse menschlicher Informationsverarbeitung am Beispiel manuellen Fliegens [Analysis of human information processing using the example of manual flying]. In: Gesellschaft für Arbeitswissenschaft (ed) Gestaltung der Arbeitswelt der Zukunft: Bericht zum 60. Kongress der Gesellschaft für Arbeitswissenschaft vom 12–14 März 2014. GFA-Press, Dortmund, pp 43–45

Hayashi M (2004) Hidden Markov Models for analysis of pilot instrument scanning and attention switching. Massachusetts Institute of Technology

Helleberg JR, Wickens CD (2003) Effects of data-link modality and display redundancy on pilot performance: an attentional perspective. Int J Aviat Psychol 13(3):189–210. doi:10.1207/S15327108IJAP1303_01

International Air Transport Association (2016) Safety Report 2015, 52st edition, Montréal

ISO (2007) Road vehicles—Ergonomic aspects of transport information and control systems—Occlusion method to assess visual demand due to the use of in-vehicle systems 13.180(ISO 16673:2007)

ISO (2015) Road vehicles—measurement of driver visual behaviour with respect to transport information and control systems—part 1: Definitions and parameters 01.040.43(ISO 15007-1:2015)

Jones DH (1985) An error-dependent model of instrument-scanning behavior in commercial airline pilots. NASA Contractor Report

Kang Z, Landry SJ (2015) An eye movement analysis algorithm for a multielement target tracking task: maximum transition-based agglomerative hierarchical clustering. IEEE Trans Hum Mach Syst 45(1):13–24. doi:10.1109/THMS.2014.2363121

Krause M, Donant N, Bengler K (2015) Comparing occlusion method by display blanking to occlusion goggles. Proc Manuf 3:2650–2657. doi:10.1016/j.promfg.2015.07.622

Krems JF, Keinath A, Baumann M, Gelau C, Bengler K (2000) Evaluating visual display designs in vehicles: advantages and disadvantages of the occlusion technique. In: Camarinha-Matos LM, Afsarmanesh H, Erbe H-H (eds) Advances in networked enterprises: virtual organizations, balanced automation, and systems integration, vol 53. Springer, Heilderberg, pp 361–368

Liu A (1998) What the driver's eye tells the car's brain. Chapter 20. In: Underwood G (ed) Eye guidance in reading and scene perception. Elsevier Science Ltd, Amsterdam, pp 431–452

Milgram P (1987) A spectacle-mounted liquid-crystal tachistoscope. Behav Res Methods Instrum Comput 19(5):449–456

Monk CA, Boehm-Davis DA, Trafton JG (2002) The attentional costs of interrupting task performance at various stages. In: Proceedings of the human factors and ergonomics society annual meeting, vol 46, no. 22, pp 1824–1828. doi:10.1177/154193120204602210

Morris CH, Leung YK (2006) Pilot mental workload: how well do pilots really perform? Ergonomics 49(15):1581–1596. doi:10.1080/00140130600857987

Myers CW (2007) Scan pattern adaptation during repeated visual search. PhD Thesis, Rensselaer Polytechnic Institute

National Transportation Safety Board (2014) Descent below visual glidepath and impact with seawall: Asiana Airlines Flight 214, Boeing 777-200ER, HL7742, San Francisco, California July 6, 2013. Accident Report. NTSB/AAR, Washington, DC

Popp C, Kemény C (2016) One team. J Civil Aviat Train 27(2):26–28

Rantanen EM, Johnson NR, Talleur DA (2004) The effectiveness of personal computer aviation training device, a flight training device, and an airplane in conducting instrument proficiency checks: volume 2: objective pilot performance measures. Final Technical Report AHFD-04-16/FAA-04-6, Oklahoma City

Sarter NB, Woods DD (1994) Pilot interaction with cockpit automation II: an experimental study of pilots' model and awareness of the flight management system. Int J Aviat Psychol 4(1):1–28. doi:10.1207/s15327108ijap0401_1

Schutte PC, Trujillo AC (1996) Flight crew task management in non-normal situations. In: Proceedings of the human factors and ergonomics society annual meeting, vol 40, no. 4, pp 244–248. doi:10.1177/154193129604000422

Senders JW, Kristofferson AB, Levison WH, Dietrich CW, Ward JL (1967) The attentional demand of automobile driving. Paper sponsored by committee on highway safety and presented at the 46th annual meeting. Highway Research Record

Shapiro KL, Raymond JE (1989) Training of efficient oculomotor strategies enhances skill acquisition. Acta Psychol 71(1–3):217–242. doi:10.1016/0001-6918(89)90010-3

SKYbrary (2016) Flight control laws. http://www.skybrary.aero/index.php/Flight_Control_Laws. Accessed 24 Feb 2015

Sleight RB (1948) The effect of instrument dial shape on legibility. J Appl Psychol 32(2):170–188. doi:10.1037/h0063435

Spady AA (1978) Airline pilot scan patterns during simulated ILS approaches. NASA Technical Paper, Hampton

Steelman KS, McCarley JS, Wickens CD (2011) Modeling the control of attention in visual workspaces. Hum Factors 53(2):142–153. doi:10.1177/0018720811404026

Underwood G, Chapman P, Brocklehurst N, Underwood J, Crundall D (2003) Visual attention while driving: sequences of eye fixations made by experienced and novice drivers. Ergonomics 46(6):629–646. doi:10.1080/0014013031000090116

Unema P, Rötting M (1990) Differences in eye movements and mental workload between experienced and inexperienced motor-vehicle drivers. In: Brogan D, Gale AG, Carr K (eds) Visual search. Taylor & Francis, London, pp 193–202

van der Horst R (2004) Occlusion as a measure for visual workload: an overview of TNO occlusion research in car driving. Appl Ergon 35(3):189–196. doi:10.1016/j.apergo.2003.11.010

Wickens CD, Alexander AL (2009) Attentional tunneling and task management in synthetic vision displays. Int J Aviat Psychol 19(2):182–199. doi:10.1080/10508410902766549

Wickens CD, Hollands JG, Banbury S, Parasuraman R (2013) Engineering psychology and human performance, 4th edn. Pearson, Boston

Zimmer AC, Stein M (2012) Information Systems in Transportation. In: Stein M, Sandl P (eds) Information ergonomics: a theoretical approach and practical experience in transportation. Springer, Heidelberg, pp 1–22

Haslbeck, A., & Zhang, B. (2017). I spy with my little eye: Analysis of airline pilots' gaze patterns in a manual instrument flight scenario. Applied Ergonomics, 63(September 2017), 62–71. doi:10.1016/j.apergo.2017.03.015

This paper can be originally found at:

http://www.sciencedirect.com/science/article/pii/S000368701730073X

Applied Ergonomics 63 (2017) 62–71

Contents lists available at ScienceDirect

Applied Ergonomics

journal homepage: www.elsevier.com/locate/apergo

I spy with my little eye: Analysis of airline pilots' gaze patterns in a manual instrument flight scenario

Andreas Haslbeck [a, *, 1], Bo Zhang [a, b, 1]

[a] Institute of Ergonomics, Technical University of Munich, Munich, Germany
[b] Centre for Transport Studies, University of Twente, Enschede, The Netherlands

ARTICLE INFO

Article history:
Received 17 March 2016
Received in revised form
20 February 2017
Accepted 29 March 2017

Keywords:
Gaze pattern
Manual flying
Visual scanning

ABSTRACT

The aim of this study was to analyze pilots' visual scanning in a manual approach and landing scenario. Manual flying skills suffer from increasing use of automation. In addition, predominantly long-haul pilots with only a few opportunities to practice these skills experience this decline. Airline pilots representing different levels of practice (short-haul vs. long-haul) had to perform a manual raw data precision approach while their visual scanning was recorded by an eye-tracking device. The analysis of gaze patterns, which are based on predominant saccades, revealed one main group of saccades among long-haul pilots. In contrast, short-haul pilots showed more balanced scanning using two different groups of saccades. Short-haul pilots generally demonstrated better manual flight performance and within this group, one type of scan pattern was found to facilitate the manual landing task more. Long-haul pilots tend to utilize visual scanning behaviors that are inappropriate for the manual ILS landing task. This lack of skills needs to be addressed by providing specific training and more practice.

1. Introduction

Since the 1980s, aircraft have been supplied with highly automated flight control systems. These systems have taken over many processes previously performed by pilots including the setting and supervision of flight performance parameters for the engines, the course, and appropriate speed and altitude. In spite of the accrued advantages of modern aircraft with highly automated systems, there is one large disadvantage: the more automation is used, the more manual skills diminish due to an absence of practice opportunities (Wiener and Curry, 1980; Sarter and Woods, 1994; Veillette, 1995; Ebbatson, 2009; Haslbeck and Hörmann, 2016). In an in-depth analysis of 415 commercial aviation accidents, which occurred between 2010 and 2014 by the International Air Transport Association (2015), evidence can be found that manual handling flight crew errors were involved in nearly one third of these accidents. It is therefore of great importance to enhance flight safety by providing pilots with adequate and effective training programs and help them maintaining sufficient manual flying skills, especially during the most vulnerable flight phases such as approach and landing.

The level of practice, strongly associated with the daily flight practice, is assumed to have the biggest influence on manual flight performance (Ebbatson, 2009; Haslbeck and Hörmann, 2016). Airline transport pilots face the same flying tasks and conditions independently from the type of operation. They also perform very similar and predominantly standardized maneuvers in aircraft families (e.g. Airbus A320 family) offering unified cockpit layouts and highly comparable handling qualities (Brière and Traverse, 1993; Favre, 1994; Joint Aviation Authorities, 2004; Bissonnette and Culet, 2013). In spite of these similarities, crews on short-haul routes perform more than five times as many flights as crews on long-haul ones and show better fine-motor flight performance (Haslbeck and Hörmann, 2016) as well as superior visual skills (Haslbeck et al., 2012). If we can identify different instrument scanning patterns in correlation with good or poor performance, it could be beneficial for future cockpit design and training programs.

1.1. Pilot's instrument scanning and analytical methods

Manual flying has been denoted as a closed-loop control problem (Field and Harris, 1998). This is a psycho-motor and highly skilled task where the pilot needs to continuously control and monitor six variables which are usually cross-coupled (Field and

---

* Corresponding author. Boltzmannstrasse 15, 85747 Garching, Germany.
E-mail address: haslbeck@tum.de (A. Haslbeck).
[1] Joint first authors.

http://dx.doi.org/10.1016/j.apergo.2017.03.015
0003-6870/© 2017 Elsevier Ltd. All rights reserved.

Harris, 1998; Wickens, 2002): pitch, roll, and yaw of the aircraft as well as altitude, lateral, and longitudinal deviation from the desired flight path. Raw data flying specifies the fact when these parameters have to be actively scanned, cognitively processed, and transferred to adequate control inputs by the pilots (instead of the flight computers). Effective instrument scanning strategies can keep the pilots in the loop by continuously updating their memory about the current state of the aircraft. Adequate visual scanning also enables pilots to take effective control in time, which requires adequate cognitive skills and a sophisticated understanding of the relationship between the instruments.

To maintain sufficient spatial awareness, continuous and logical scanning across multiple instruments is required, also known as cross-checking in aviation domain. The two common visual strategies for pilots are the *radial cross-check* technique (Federal Aviation Administration, 2012, pp. 6–24) and the *circular* one (Dick, 1980, p. 12; Jones, 1985, p. 17). In training documentation the FAA describes the *radial cross-check* as a technique where the pilot starts a scan in the center of a primary flight display where the attitude indicator (ATT) is located (Federal Aviation Administration, 2012, pp. 6–24). After that, scans are to be performed left to the airspeed tape (SPD), right to the altitude indicator (ALT), and down to the heading indicator (HDG). However, every scan returns to the center for an intermediate scan – representing a pattern like the spokes of a wheel (Dick, 1980, p. 12; Jones, 1985, p. 17) and comparable to the *basic-T* pattern matching with the instrument layout in conventional cockpits (Fig. 1, left side). A pilot checks control inputs on the control instruments (mainly on attitude) and monitors their effects on the performance instruments such as speed, altitude, and heading (Federal Aviation Administration, 2012, p. 6–18 - 6–19).

Gaze-based metrics such as *mean glance duration*, *glance rate*, and *percent time on areas of interests (AOIs)* are mostly analyzed to gain insight into pilots' visual behavior and attentional allocation, which were standardized for road traffic research by ISO 15007-1 ("*Measurement of driver visual behaviour with respect to transport information and control systems")*. Nevertheless, regardless of pilots' attentional switching across the instruments, these metrics alone are inadequate to understand pilots' scanning strategies and reflect complex information acquisition processes.

To take the gaze sequences into account, the term *gaze pattern* describes the order of a person's scanning behavior (Dorr et al., 2010) while *scanpath* refers to vectors, i.e. the geometric characteristics of subsequent glances (Holmqvist et al., 2011, p. 254; Kang and Landry, 2015). Myers (2007) highlights two main influences on sequences of saccades: exogenous and endogenous. Exogenous saccades account for bottom-up (data/sensory-driven) processes being considered as non-deliberate (Einhäuser et al., 2008b; Schütz et al., 2011; Kowler, 2011). Endogenous saccades were described as deliberate top-down (goal-driven) processes, when the successful conduct of a visual task accounts for specific gaze patterns and scanpaths (Noton and Stark, 1971; Holmqvist et al., 2011, p. 253–254). Endogenously initiated saccades can be assumed when an operator fulfills a specific skilled task (Einhäuser et al., 2008a; Schütz et al., 2011; Foulsham et al., 2012) such as continuous flightpath tracking (Allsop and Gray, 2014).

The sequence of how different displays are looked at reveals how this information is cognitively processed by pilots. The comparison of sequences, however, is complex (Foulsham et al., 2012; Anderson et al., 2013), especially when these sequences become extremely long (Kang and Landry, 2015), for example in a continuous flight tracking task. One solution is to break a longer sequence into smaller segments with easily analyzable lengths (Tole et al., 1983; Simon et al., 1993). Hence, the most frequent sequences may reflect the main scanning strategies. Transition matrices are an economic alternative approach indicating the probability of an AOI being next in the sequence based on the current AOI. The highest transition probabilities give a good representation of the overall scanning pattern (Milton et al., 1950; Spady, 1978; Harris et al., 1986). In a very recent study Kang and Landry (2015) introduced the MTAHC algorithm to analyze eye movements during the tracking of multiple moving targets based on unordered transition matrices. This method was developed to find an adequate representation for large and complex gaze patterns when comparing them among several individuals. AOIs with higher transition values were hierarchically clustered and integrated into *visual grouping sets*. In summary, focusing on transition probabilities facilitates the analysis of gaze patterns and reduces complexity of longer sequences.

1.2. Expertise-related differences in visual scanning

Evidence can be found in many transportation studies that expertise levels have a major influence on visual behavior and task performance. One very early study can be dated back to the 1940s when Fitts et al. (1949) investigated the effect of experience on 40 pilots' eye movement measures in a ground control approach flying scenario, suggesting that experienced pilots had more frequent fixation and correspondingly short fixation duration on flight instruments. These findings are fairly consistent with similar studies. For instance, Bellenkes et al. (1997) indicated that expert pilots' scanning strategies differed from those of novices in several aspects

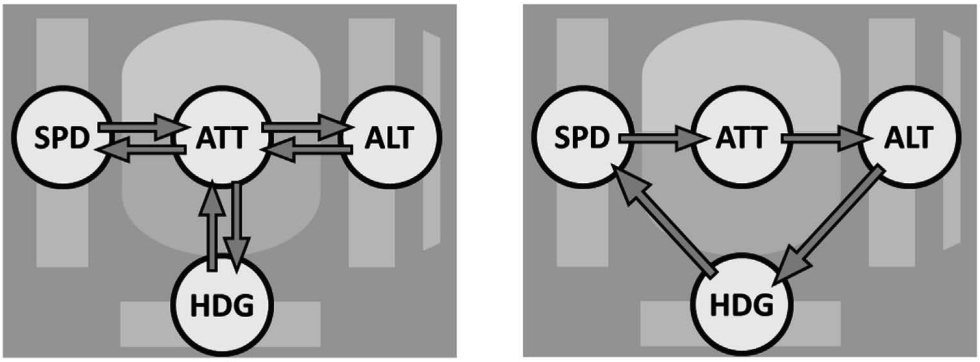


Fig. 1. The radial cross-check technique (spokes-of-a-wheel) showing a center-bound pattern (left side) and the circular scanning technique showing a clockwise pattern (right side). There is no distinct starting point per se, however, a typical starting point is the attitude indicator (ATT), while speed (SPD), altitude (ALT), and heading (HDG) are scanned frequently.

including (1) shorter fixation times and more active visits to most instruments, (2) distinct attentional flexibility in response to changing task demands and (3) more checks on the flight variables that were not being manipulated. In the studies of Ottati et al. (1999) and Kasarskis et al. (2001) experts showed again differences in attention allocation, total number of fixations and flight performance behavior.

In addition to the analyses of the basic gaze-based metrics, Underwood et al. (2003) examined the gaze patterns of experienced and novice drivers based on transition probabilities and 3-element scanpaths. Differences between experienced car drivers and novices were most apparent in the fact that experienced drivers frequently fixated the road far ahead after fixating on off-road AOIs. The road far ahead delivers maximum preview, and thus, delivers a fast update of the traffic situation whenever drivers looked away from the road and fixate it again. The 3-element scanpath analysis showed that several patterns started and finished at an area where the car would arrive in one or two seconds. This visual behavior was explicitly denoted as a scanning strategy (Underwood et al., 2003), and the authors assumed an effect of *perceptual narrowing* for the novices in situations of high cognitive load (see van Leeuwen et al., 2015).

By the use of a Markovian analysis, Hayashi (2004) found that experienced pilots were able to handle more parallel flying tasks which also require visual scanning, compared to inexperienced pilots. She explained this finding by the higher workload experienced by the lesser-trained pilots (see Robinski and Stein, 2013), while other authors generally evaluated the relationship between gaze patterns and workload (Di Nocera, Camili and Terenzi, 2006; de Rivecourt et al., 2008; Schieber and Gilland, 2008), also highlighting the immense workload during a manual landing (Entzinger and Suzuki, 2012). Lesser-trained pilots had to drop tracking tasks because their capacity was insufficient for all possible tasks due to less expertise (Hayashi, 2004). Kang and Landry (2014) used scanpath analysis to teach experts' gaze patterns to novices in air traffic control. These novices showed better conflict detection performance after the intervention as compared to a control group.

Despite of all these evidences, one essential question remains whether behavioral deficits, which could be observed in novice drivers (Chapman et al., 2002; Underwood et al., 2003), can also be observed in long-haul pilots who are assumed to be experts (see Yang et al., 2013), but who do not have much practice due to flight time regulations, which is not well addressed in previous studies. We assume there will be shortcomings in their skills that are comparable to those found in novices, not only in their instrument sampling behavior (Haslbeck et al., 2012), but also in their instrument scanning behavior.

1.3. *Aim of this paper and research questions*

Based on the reviewed literature we identified a research gap regarding (1) detailed analysis of pilots' gaze patterns in a fine-motor flying task depending on the level of practice and (2) whether different visual strategies can be identified and evaluated. This paper focuses on the analysis of pilots' visual scanning when performing a manual approach (raw data) and landing task. The aim is to (1) examine the influence of practice on visual scanning, (2) identify frequently occurring gaze patterns, and (3) evaluate the effectiveness of several different gaze patterns. An important prerequisite for future cockpit design and training programs will be to know how pilots perform their visual scanning. Accordingly, the following research questions were developed:

RQ1: How does the level of practice affect visual scanning in airline pilots?

RQ2: Do pilots use recurring gaze patterns in manual flight?

RQ3: Is there a correlation between gaze patterns and flight performance?

With regard to RQ1 we hypothesize that highly practiced (short-haul) pilots show a better performance with respect to visual scanning (see Haslbeck et al., 2012) similar to fine-motor flying skills (Haslbeck and Hörmann, 2016). We also hypothesize that pilots repeat certain gaze patterns and recurring sequences can be found (RQ2) in the goal-driven task of manual flying (Schütz et al., 2011; Haslbeck and Bengler, 2016). Finally, we assume a correlation between pilots' visual scanning and their fine-motor flight performance (RQ3), because both belong to corresponding stages of human information processing (Wickens et al., 2013; Haslbeck and Bengler, 2016).

2. Method

2.1. *Independent variable and participants*

We report on the gaze pattern analysis derived from a manual flying experiment conducted in 2013 (Haslbeck and Hörmann, 2016). In the original study 120 randomly assigned professional commercial airline pilots (ATP licensed) participated. However, for this analysis only 51 out of 120 participants could be analyzed due to different flying tasks and few data of poor quality. The independent variable of this experiment was the level of practice. The number of landings within the past 30 days prior to the experiment was taken as a measure indicating more flight practice for the Airbus A320 short-haul crews (Table 1). Short-haul crews maintain a remarkably higher level of flight practice due to shorter but more frequent flights in line operation, while long-haul pilots only conduct few flights due to legal rest periods. All 51 pilots (Table 1) held a type rating for the Airbus aircraft family, differentiated into short-haul (A320) and long-haul (A340) pilots. All participants occupied the same seat as well as the same simulator type for which they were rated, e.g. a short-haul captain occupied the left seat in the A320 simulator. They assumed the role of the pilot flying, thus, actively performed all manual flying tasks, i.e. they scanned the primary flight display and operated the sidestick (SKYbrary, 2016a). All participants were supported by an accordingly licensed colleague as pilot monitoring on the other seat. To avoid small sample sizes, only two groups were assigned for the variation of the independent variable addressing different levels of practice: short-haul (more flight practice) and long-haul (less flight practice) crews.

2.2. *Apparatus*

We conducted this experiment in collaboration with a major European airline which provided two qualified Airbus-type full-

Table 1
Demographic data for participants.

Rank	N	Age		Flight hours		Landings in past 30 days		Years since flight school	
		mean	SD	mean	SD	mean	SD	mean	SD
Short-haul (Airbus A320) group									
FO	6	30.8	2.6	3676	520	11.3	4.9	6.3	2.0
CPT	20	43.4	4.4	11,636	1886	16.8	11.2	18.2	3.3
Long-haul (Airbus A340) group									
FO	20	36.0	3.2	6926	1787	2.4	1.6	11.9	2.7
CPT	5	49.3	4.0	14,000	1874	3.8	1.3	25.3	4.7

Note. Captains (CPT) have more experience according to flight hours compared to their first officer (FO) colleagues. Due to small sample sizes for A320 FOs and A340 CPTs and very small expected rank differences (see Haslbeck and Hörmann, 2016) all pilots of one aircraft type were merged to one group each.

flight simulators (JAR-FSTD A): an Airbus A320-200 simulator and an Airbus A340-600 simulator. All flight parameters, including control inputs and instrument landing system (ILS) tracking data, were recorded by the flight simulator data recorder at a sampling rate of 15 Hz. Head-mounted monocular DIKABLIS Essential eye-tracking devices were used throughout the experimental procedures, recording both pilots' gaze behavior synchronously at a sampling rate of 25 Hz.

2.3. Instruction and scenario

Prior to the experiment, all participants handled another scenario addressing operational tasks (35 min), also serving as a warm up for this experimental scenario. All participants had to fly a 10-minute manual flight and landing scenario approaching Munich Airport (26R EDDM) in normal law mode (see SKYbrary, 2016b). The scenario started at an altitude of 5000 ft, approximately 8 min before touching down. Data was collected during the instrument approach between 3300 ft and 270 ft above ground level (AGL). Weather parameters were set to a visibility of 1200 m, gusty wind of 220°/17–22 kts, a ceiling of 270 ft, and light rain, i.e. the runway was not visible during the measurement. Because of the difference between the runway orientation (260°) and the wind (220°), a certain wind correction angle had to be maintained during this straight approach. Shortly after the scenario started, a (simulated) malfunction of the autopilot (AP) and the flight director (FD) necessitated a manual approach. These conditions provide a medium to above average task load for pilots. However, pilots are clearly required to be able to manage such an approach and land safely.

Manual aircraft control had to be done by raw data, a very basic kind of fine-motor flying without AP or FD, but including Airbus-like *envelope protections* (see SKYbrary, 2016b). We included the exception of an available auto thrust, which was deactivated by all participants at an early stage. Thus lateral, vertical, and longitudinal manual aircraft control had to be done by referencing the primary flight instruments. These were primarily attitude, airspeed, altitude, heading, as well as the glideslope (GS) and localizer (LOC) indicators. At the same time, deviations from an ideal course on the ILS and airspeed had to be compensated by control inputs into the sidestick and the thrust levers. All participants were instructed to fly as accurately as possible according to licensing standards (European Union, 2011, p.117) and the company's standard operating procedures and to perform a *stabilized approach* (SKYbrary, 2016c).

2.4. Dependent measures

2.4.1. Flight performance measures

As a measure of manual control performance, deviations from an ideal glideslope (vertical guidance) and localizer (lateral guidance) measured in *dots*, were taken. One dot indicates a half scale deflection from glideslope or localizer (European Union, 2011, p.117) corresponding to a deviation of ±0.4° on the glideslope and ±0.8° on the localizer for Airbus aircraft. The root mean square error (RMSE), a recommended and widely applied measure to assess tracking performance (Scallen et al., 1995; Rantanen et al., 2004), was taken as the indicator to evaluate the objective manual flight performance.

2.4.2. Areas of interest (AOI)

The analysis of visual behavior concentrated on relevant AOIs on the primary flight display (18.4 × 18.4 cm), which are the most important for manual flying: attitude, speed, altitude, heading, vertical speed, and the indicators for glideslope and localizer tracking. However, due to the limited tracking accuracy of the eye-tracking device (average glance direction accuracy > 2° in this study), very close AOIs could not be analyzed as separate ones. Vertical speed and glideslope information were added to the altitude area while localizer information was added to heading, resulting in only four discriminable AOIs (see Fig. 1). The indicators within the joint displays offer similar information, e.g. both heading and localizer account for horizontal tracking, while vertical speed, glideslope, and altitude account for vertical tracking (see Hayashi, 2004, p. 17). Thus, we consider the combination of AOIs depicting very similar information as validated. Other AOIs for navigation, wind, or engines were not considered.

2.4.3. Gaze-based metrics

Three basic metrics according to ISO 15007-1 were utilized to gain an overview of pilots' information acquisition: the *percent time on AOI*, the *mean glance duration*, and the *glance rate*. Percent time on AOI provides general information about pilots' attention distribution on all AOIs and shows differences comparing the visual behavior across different practice levels. The mean glance duration indicates the effectiveness of information acquisition processes and the glance rate shows the activeness of eye movements with regard to certain AOI.

2.4.4. Gaze pattern analysis metrics

The visual behavior was analyzed for each participant performing the manual flying task for the manual approach between 3300 ft and 270 ft AGL (approx. 5 min). To characterize and compare the complex and long gaze patterns of each pilot, we considered the maximum transition-based MTAHC algorithm (Kang and Landry, 2015) as one suitable approach. The concept of this algorithm is to determine the set of AOIs with higher transition values compared with other AOIs to understand the cognitive process, which is named *Visual Grouping Set (VG)*. For this purpose an unordered transition matrix is first established then the two AOIs with the largest transition values in the matrix are clustered as one VG. By further clustering the other AOIs with lower transition values, multi-level VG sets can be formed.

In this study we modified this algorithm to focus more on the saccades rather than individual AOIs, since the cross-checking pattern between the instruments is of more interest. Analogous

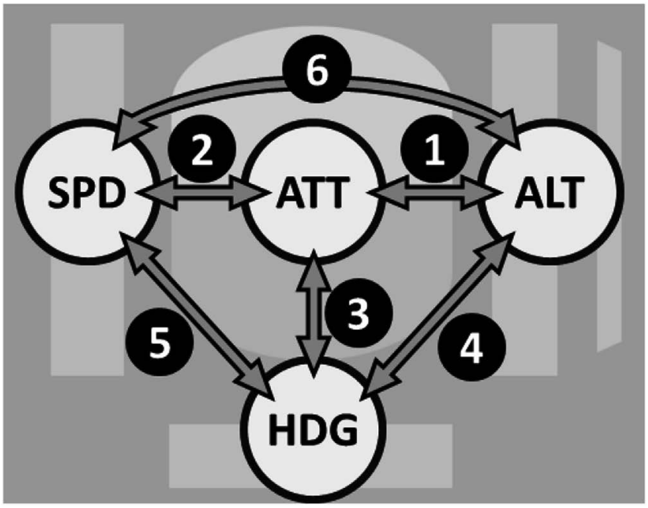


Fig. 2. Coding of the unordered saccades: 1 ATT-ALT, 2 ATT-SPD, 3 ATT-HDG, 4 ALT-HDG, 5 SPD-HDG, 6 SPD-ALT.

to the concept of VGs we first calculated the transition matrices for the six possible unordered saccades between the defined AOIs, coded with numbers from 1 to 6 (Fig. 2). The saccades with the highest occurring frequencies compared to other saccades form a saccade grouping set (SG), which represents the main scanning activity of a pilot. To facilitate the comparison, we only compare SGs including three saccades. The simpler 2-element SGs cannot sufficiently discriminate between most pilots to characterize a pilot's visual behavior, while a SG including 4 saccades would result in a large potential size for comparison, resulting in limited identification of common characteristics. For the nomenclature of SGs we selected the numbers shown in Fig. 2 in combination with braces; e.g. {123} represents the spokes-of-a-wheel pattern.

3. Results

All effects were reported as significant at $p < 0.05$ and r respectively η_p^2 was given as effect size. For statistical analyses the statistical package SPSS 22 was taken.

3.1. Manual flight performance

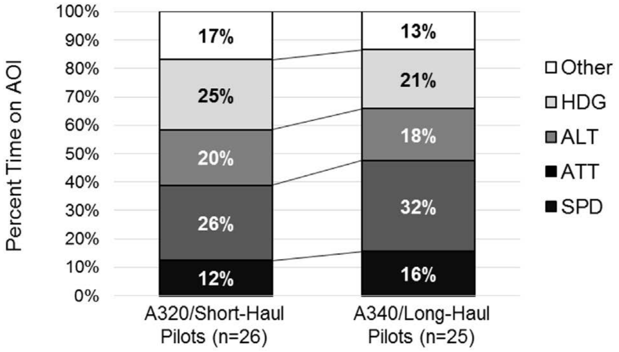
The pilots' RMSE of localizer and glideslope deviations are illustrated in Fig. 3. For flightpath deviation data the assumption of normality was violated due to a positive skew. Thus, log10-transformed data was taken for independent t-tests, while for visualizations and reported means non-transformed data was considered.

The mean RMSE on localizer of the short-haul pilots ($M = 0.11$ dot, $SD = 0.04$) was significantly smaller compared to the long-haul group ($M = 0.23$ dot, $SD = 0.11$), $t(49) = 6.22$, $p < 0.001$, $r = 0.66$ (one-tailed). Similarly, the RMSE for glideslope indicated significantly smaller tracking deviations of the short-haul pilots ($M = 0.14$ dot, $SD = 0.05$) compared to long-haul pilots ($M = 0.30$ dot, $SD = 0.12$), $t(49) = 7.13$, $p < 0.001$, $r = 0.71$ (one-tailed). The overall results showed that the short-haul pilots accomplished the flying task with smaller deviations regarding both dimensions. The two groups are thus a good representation for two levels of different practice and performance, respectively.

3.2. Gaze-based data analysis

A first and general measure was the percent time on AOI (normalized data depicted by Fig. 4) based on total glance times for all participants. Glances at the regions outside the designated AOIs were also presented, denoted as others to provide an overview of the pilots' attention allocation. While long-haul pilots focused more dominantly on the ATT with nearly one third of the time (32%), short-haul pilots showed a more balanced allocation with scanning ATT and HDG for about one fourth of the time each (26% and 25%, respectively).

To examine whether pilots of two fleet affiliations behaved differently in *mean glance durations* per AOI (Fig. 5), a 2 × 4 (Fleet × AOI) mixed ANOVA was conducted (Table 2, upper panel). Glances towards heading were found to last significantly longer compared to the other three AOIs, independent of the group



	mean [s]	SD [s]	n	mean [s]	SD [s]	n
Other	39.3	12.4	62	29.9	10.5	40
HDG	57.6	19.4	74	46.6	13.0	63
ALT	45.6	14.7	89	40.8	12.4	84
ATT	61.3	26.2	117	71.3	23.9	147
SPD	29.0	10.2	66	34.9	11.4	79

Fig. 4. Percent time on AOIs between groups shows normalized total glance times on different areas of interest on the primary flight display. The additional table depicts mean values, standard deviations (SD), and number of glances (n) of the related raw gaze data.

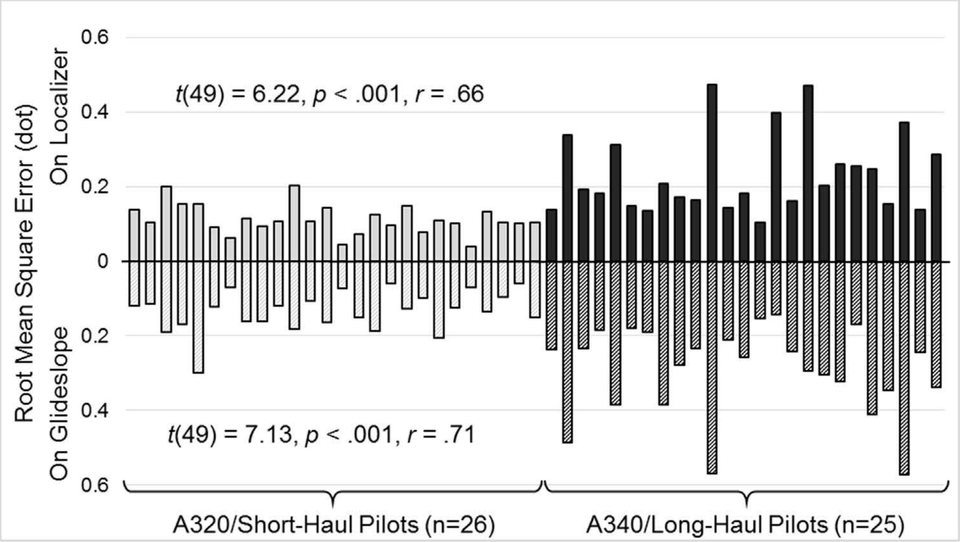


Fig. 3. Deviations (i.e. root mean square error in dot) from glideslope and localizer for both groups between 3300 ft and 270 ft above ground level based on non-transformed data. The reported independent t-tests are based on log10-transformed data; r indicates the effect size.

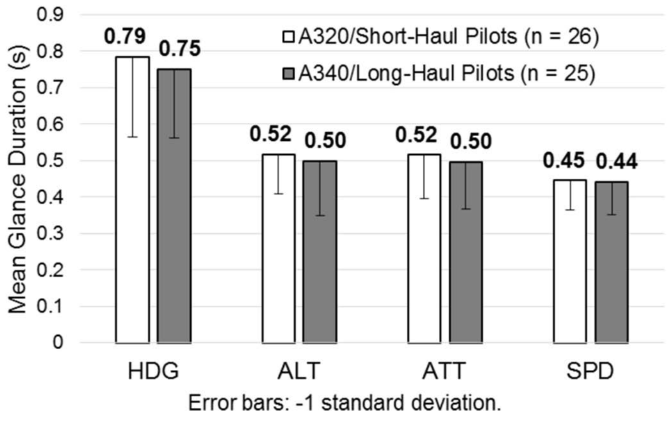


Fig. 5. Mean glance duration (in seconds) on AOIs compared between groups.

assignment.

Similarly, another 2 × 4 (fleet × AOI) mixed ANOVA was performed for the *glance rate* (Fig. 6, Table 2, lower panel). There was an interaction between both main effects, which seems plausible when analyzing glance rates. Attitude had the highest glance rates among the different AOIs and it was even significantly higher in long-haul pilots.

Altogether, seen from the results of the two analyses, the differences regarding *percent time on AOIs* were mainly due to the pilots' different sequential scanning strategies rather than glance duration.

3.3. Gaze pattern analysis

The dominant saccades were calculated for each of the 51 participants by transition matrices and were clustered into normal SGs of three components. Four main SGs were identified (Fig. 7): {123} for 28 pilots, {134} for 9 pilots, {126} for 6 pilots and {456} for 5 pilots, while three pilots showed other 3-element SGs.

The SG {123} – an attitude-centered pattern – indicates the dominant transitions between the attitude and other AOIs, which is consistent with the spokes-of-a-wheel scanning strategy. In contrast, the two triangular SGs, {134} and {456}, directly link the peripheral AOI partly without checking attitude. These patterns align with the cross-checking between altitude, speed, and heading, which were rarely found among long-haul pilots. As shown in Fig. 8, the spokes-type strategy was observed mainly among long-haul pilots, representing nearly 70%, while there was no single dominant pattern among short-haul pilots. Short-haul pilots showed a more balanced distribution between the spokes-type SG (42%) and triangular-type SGs (42%). For this exploratory approach the assumption of independence was met. Consequently, a two-sided chi-square test comparing the frequencies of spokes-type and triangular-type SGs between both groups of pilots (frequencies depicted by Table 3) indicated significant differences between both groups in the use of triangular patterns, $X^2(1) = 5.78$; $p = 0.016$. To evaluate spokes types against triangular types concerning flightpath deviations (Table 3) a comparisons based on short-haul pilots was performed. Pilots using the spokes-type strategy showed significant larger deviations on the localizer compared to pilots using triangular types. Long-haul pilots were not considered because of a bias due to only few cases for the triangular types. The same applied for the SG {126}.

4. Discussion

Fine-motor flying skills in terms of ILS flightpath deviations were taken as a factor to distinguish between pilots with a high and a low level of practice. This has already been analyzed in detail (Haslbeck and Hörmann, 2016), and thus will not be further discussed here. An earlier study found differences in certain checks (for wind, speed, and the flight mode annunciator) based on the same distinction between long-haul and short-haul pilots (Haslbeck et al., 2012).

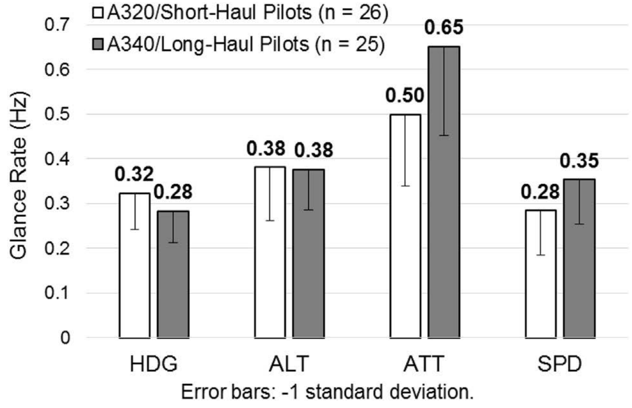


Fig. 6. Glance rate (in Hz) on AOIs compared between groups.

Table 2
Statistical analysis for mean glance duration and glance rate.

Source	Effects (Univariate Tests)	Bonferroni Post Hoc Comparisons
Mean Glance Duration		
Fleet	$F(1, 49) = 0.58, p = 0.449, \eta_p^2 = 0.01$, n.s.	
AOI	$\varepsilon = 0.70, F(2.09, 102.60) = 64.44, p < 0.001, \eta_p^2 = 0.57$	HDG-ALT ($p < 0.001$); HDG-ATT ($p < 0.001$); HDG-SPD ($p < 0.001$); ALT-SPD ($p < 0.001$); ATT-SPD ($p = 0.012$)
Fleet * AOI	$\varepsilon = 0.70, F(2.09, 102.60) = 0.12, p = 0.90, \eta_p^2 < 0.01$, n.s.	
Glance Rate		
Fleet	$F(1, 49) = 5.78, p = 0.020, \eta_p^2 = 0.11$	A320-A340 ($p = 0.020$)
AOI	$\varepsilon = 0.68, F(2.04, 99.96) = 58.10, p < 0.001, \eta_p^2 = 0.54$	HDG-ALT ($p < 0.001$); HDG-ATT ($p < 0.001$); ALT-ATT ($p < 0.001$); ALT-SPD ($p = 0.012$); ATT-SPD ($p < 0.001$)
Fleet * AOI	$\varepsilon = 0.68, F(2.04, 99.96) = 6.72, p = 0.002, \eta_p^2 = 0.12$	ATT: A320-A340 ($p = 0.004$)SPD: A320-A340 ($p = 0.014$)A320: HDG-ATT ($p < 0.001$); ALT-ATT ($p = 0.025$); ALT-SPD ($p = 0.003$); ATT-SPD ($p < 0.001$)A340: HDG-ALT ($p = 0.001$); HDG-ATT ($p < 0.001$); HDG-SPD ($p = 0.019$); ALT-ATT ($p < 0.001$); ATT-SPD ($p < 0.001$)

Note. Greenhouse-Geisser estimates of sphericity (ε) and corrected degrees of freedom are reported because the assumption of sphericity was violated for within-subjects tests.

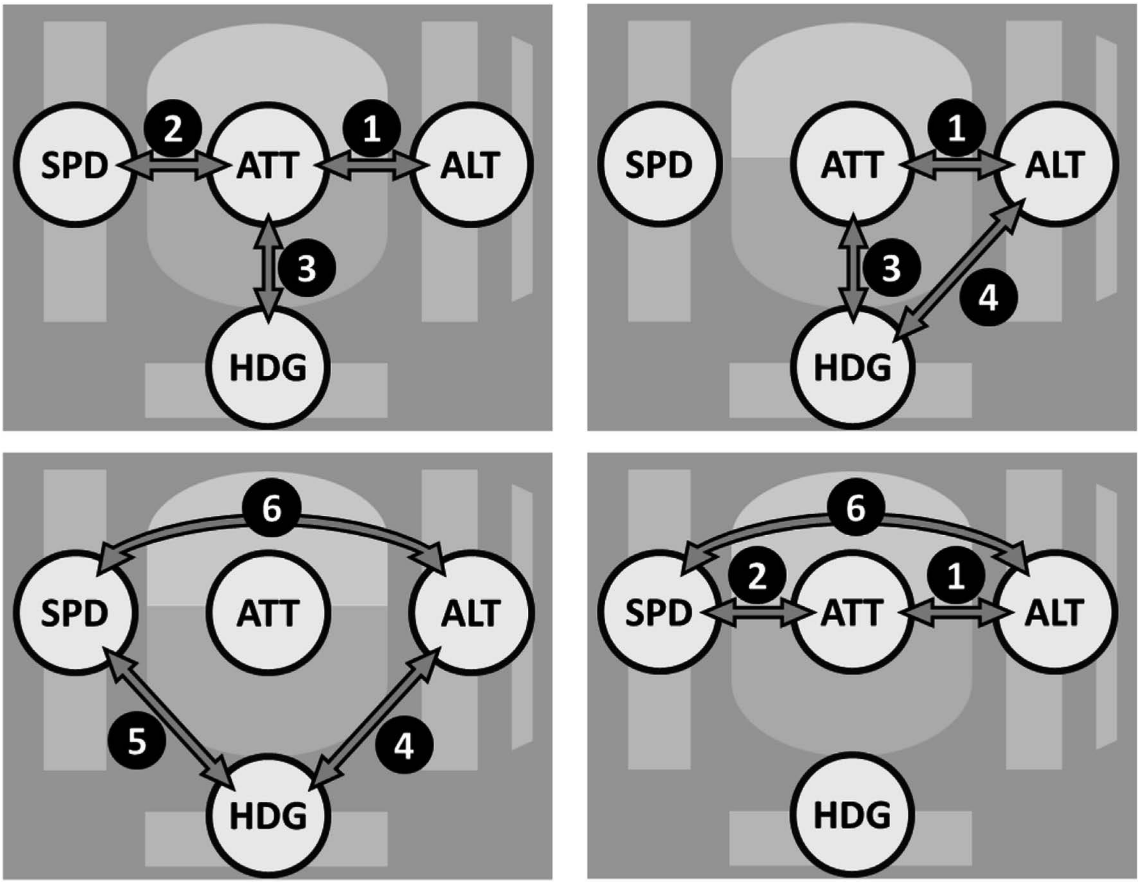


Fig. 7. Four most frequent saccade grouping sets. Top left 123 “spokes-of-a-wheel” showed by 12 pilots; top right 134 “small-right-triangle” showed by 8 pilots; bottom left 456 “big-triangle” showed by 5 pilots; bottom right 126 “long-and-short-horizontal” showed by 6 pilots.

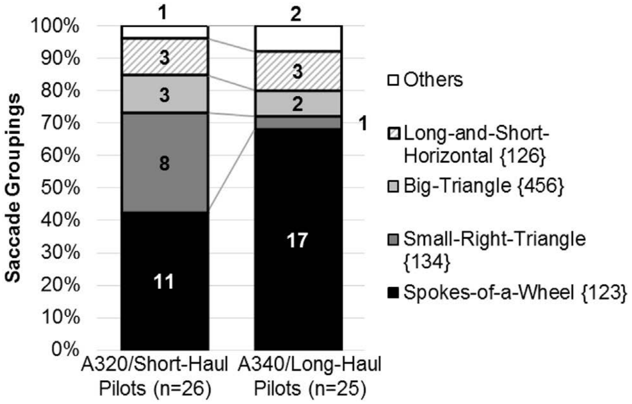


Fig. 8. Observed saccade grouping sets for both groups of participants.

4.1. Single glances

The most remarkable difference between groups was the higher concentration on attitude among long-haul pilots at the cost of most other AOIs. This is in line with the glance rates. The only meaningful prominent difference in this analysis was the high glance frequency in attitude. This display is the most important AOI on the primary flight display (Spady, 1978; Harris and Christhilf, 1980) for manual flying, providing a prediction of the near future (Colvin et al., 2005). This is similar to findings in road traffic studies (Underwood et al., 2003). Operators performing a dynamic tracking task seem to prefer a near future prediction. With a significantly higher glance rate on attitude long-haul pilots were more focused on the artificial horizon as compared to their short-haul colleagues.

The mean glance durations delivered one unexpected finding: the average glance duration on heading was about 800 ms, while a typical duration for glances on the primary flight display is 500 ms (Senders et al., 1967; Harris and Christhilf, 1980; Steelman et al., 2011; Chen and Milgram, 2011; Haslbeck and Bengler, 2016). A possible explanation for this might be the fact that a wind correction angle had to be maintained, complicating the monitoring of heading. Another reason may be attributed to the definition of AOIs. Some experts hold the opinion that beside localizer and glideslope, the vertical speed and heading indicators also provide important information for pilots when flying the ILS. That is to say, when pilots scanned the two integrated AOIs, altitude, and heading, they might scan multiple indicators at the same time. As reading the digital numbers on the heading indicator requires more time than examining whether the pointer of vertical speed is at the desired place, the averaged glance duration onto heading could also be longer as a result. Although gaze-based data showed some differences between groups, it neither satisfactorily describes pilots' visual behavior, nor sufficiently answers RQ1. The level of practice, operationalized by the fleet, has a significant effect on pilots’ visual behavior. However, this approach does not provide more details or an explanation.

4.2. Saccade grouping sets

Spokes-of-a-wheel was the most widely observed pattern and

Table 3
Evaluation of the spokes type SG against triangular type SGs concerning flightpath deviations.

Fleet	Dimension	Spokes Types			Triangular Types			SG {126}		
		n	mean	SD	n	mean	SD	n	mean	SD
A320/Short-Haul (n = 25)	Localizer$^{\circ}$	11	0.13°	0.04	11	0.10°	0.03	3	0.10	0.01
	Glideslope^{+}		0.16^{+}	0.06		0.12^{+}	0.04		0.12	0.05
A340/Long-Haul (n = 23)	Localizer	17	0.25	0.11	3	0.20	0.10	3	0.17	0.02
	Glideslope		0.32	0.12		0.31	0.14		0.22	0.03

Note. T-tests are based on log10-transformed data, while mean and standard deviation are based on non-transformed data. $^{\circ}$A one-tailed independent *t*-test indicated significant differences for short-haul pilots between spokes types and triangular types on localizer, $t(20) = 1.98$, $p = 0.031$, $r = 0.40$. $^{+}$A one-tailed *t*-test did not indicate significant differences for short-haul pilots between spokes types and triangular types on glideslope, $t(20) = 1.33$, $p = 0.100$, $r = 0.28$, n.s. For the sake of completeness data for the saccade grouping set {126} is also depicted.

was the dominant SG among the long-haul pilots. In contrast, short-haul pilots showed more balanced scanning, employing spokes-types and triangular-types at equal shares. Haslbeck et al. (2012) found similar results in an experiment comparing the percent time on AOI on the primary flight display of short-haul first officers and long-haul captains: the latter had a higher total glance time on attitude as well. They assumed that long-haul pilots fell into an accustomed visual behavior even when attitude did not provide any additional information. It can be inferred from the present study that such an accustomed gaze behavior is actually the spokes-type gaze pattern.

Some researchers suggested that spokes-of-a-wheel is a monitoring gaze pattern applied during flight phases with less control inputs in order to maintain the current course (Jones, 1985, p. 55; Hayashi, 2004, p. 47–48). This technique focuses a pilot's attention on attitude, a control instrument and preview display similar to the preview strategy found in novice drivers (Underwood et al., 2003) and maintains only reduced attention on the surrounding performance instruments: speed, altitude, and heading (see Federal Aviation Administration, 2012, p. 6–18 - 6–19). Since the main concept of this technique requires the pilots to use attitude as the primary preview on the flightpath and make control inputs primarily according to this display, it is assumed to be more suitable for flying tasks in which the information about how to maneuver the aircraft could be directly obtained from the attitude (see Colvin et al., 2005), e.g. when the FD provides an easy tracking task. Within this context the spokes-of-a-wheel strategy is plausible: first make inputs to change the attitude based on the indication of the FD, then check other performance instruments to examine the effect of the adjustment. However, when it comes to manual ILS flying with AP and FD deactivated – as presented in this experiment – the primary information to guide the pilot to the desired approach course and glideslope is provided by localizer and glideslope scales located within heading and altitude areas, respectively. Hence, these two AOIs change their meaning, shifting from performance instruments to control instruments. Consequently, the spokes-type does not sufficiently facilitate manually flying the ILS because frequent flightpath corrections require a shift of attention and, therefore, changes in visual scanning.

Frequent scanning of the localizer and glideslope indicators becomes important in order to notice the deviation trend as soon as possible and make the smallest correction necessary. Changing the attentional allocation from the attitude-centered pattern to address more heading and altitude areas is suggested as a suitable strategy in response to task demands. In accordance with previous studies (Jarodzka et al., 2010), better-skilled pilots could adjust their strategy to attend more to relevant information. Since long-haul pilots suffered from skill degradation, we observed that they maintained the accustomed scanning strategy suitable for the flying phase with active automated systems and failed to make adjustments with regard to the actual manual flying tasks. The pilots who maintained the spokes-type strategy were less sensitive to external stimuli about current aircraft status and more dependent on their accustomed scanning behavior. Since a considerable amount of fine-motor skill is required to maintain a complex control task, the long-haul pilots' significant lower level of practice (Table 1) could reflect their limited skills and resources (see Robinski and Stein, 2013) and consequently, their relative stable scanning behavior regardless of the context (e.g. dynamic changes in the aircraft status). Our interpretation is that long-haul pilots predominantly applied a mental model not suitable for the task. However, when an appropriate scanning strategy is accessible, flightpath tracking becomes a less demanding and more successful task, which is reflected by the short-haul pilots' superior ILS performance. Short-haul pilots frequently using triangular-type strategies showed a significant better flight performance on localizer, which confirms our hypotheses concerning the advantages of this strategy at least for one ILS dimension. The same tendency was found in long-haul pilots. However, the sample size was too small for statistical analysis. Finally, pilots using the SG {126} partly had the lowest ILS deviations, especially on the long haul. However, again only 3 pilots on every fleet do not allow for further conclusions.

4.3. Conclusions and application

In terms of RQ1, the level of practice was found to have an effect on visual scanning, with long-haul pilots fixating on attitude more frequently (Figs. 4 and 6) and predominantly showing attitude-based gaze patterns (Fig. 8). The influence of practice on manual aircraft control (Fig. 3) was also documented. With respect to RQ2 pilots showed recurring gaze patterns analyzed by the SG method, which addresses undirected saccades (Figs. 7 and 8). In terms of RQ3 there was a correlation between gaze patterns and flight performance (Table 3): pilots using triangular scanning strategies showed the best ILS performance, while their colleagues using spokes-typed strategies showed larger ILS deviations. Less-practiced long-haul pilots mostly used spokes-type gaze patterns, which seem appropriate for monitoring the correct flight path only. Their performance was poorer when manually flying an ILS. For a manual raw data precision approach, localizer, and glideslope indicators become the most important control instruments, especially when correcting flight path deflections. Attitude is too imprecise to show such deflections immediately after their appearance. Thus, a well-balanced mix between spokes-type and triangular types seems more appropriate for this tracking task. Long-haul pilots have exhibited behavioral deficits comparable to novice drivers (Chapman et al., 2002; Underwood et al., 2003; see also Yang et al., 2013).

There was partly no effect of practice (depicted by the fleet affiliation) on gaze-based behavior but a clear effect on gaze patterns. Gaze-based findings could explain *how* pilots performed their visual scanning based on simple metrics but could neither explain *why* and *where* exactly the observed performance differences occurred, nor could they identify sequences. Only the analysis of gaze patterns based on dominant saccades was able to identify meaningful behavioral differences and, therefore, deficits.

For application in the aviation industry this paper presented different gaze patterns in correlation with pilots exhibiting different levels of practice. Short-haul pilots who fly frequently showed a broader repertoire of various gaze patterns. However, long-haul pilots with few flights per month were not sufficiently proficient in this regard. This loss of skills needs to be addressed by selective training of visual scanning (see Shapiro and Raymond, 1989) in combination with manual aircraft control (see Wetzel et al., 1998; Chapman et al., 2002; Schütz et al., 2011; Yang et al., 2013; Kang and Landry, 2014). Long-haul pilots should enlarge their repertoire of gaze patterns again to shift their attention allocation from attitude to glideslope and localizer indicators in a more unburdened manner. However, the results of this study also indicate to be very careful with the idea of *learning from experts* (Duncan et al., 1991; Wetzel et al., 1998; Robinski and Stein, 2013; Kang and Landry, 2014; Kuebler et al., 2015) because the commonly used indicator for expertise in aviation, total flight hours, insufficiently correlates with practice, skill or proficiency (Ebbatson, 2009; Franks et al., 2014; Haslbeck and Hörmann, 2016).

4.4. Limitations

There is a potential confound of practice and different aircraft types. Airbus A320 and A340 have the same display layout. Visual scanning can be done the same way on both types but there are differences in the flight dynamics of the aircraft. Control of the two types is not identical, which has a residual influence on the whole closed-loop control information processing. The Airbus fly-by-wire flight control system was designed in alignment with the principle of commonality, providing the same flying and handling qualities (Brière and Traverse, 1993; Favre, 1994; Joint Aviation Authorities, 2004; Bissonnette and Culet, 2013). Thus both types have a similar look and feel and can be controlled with identical precision. Furthermore, we only studied pilots from one airline, all with very similar training experience and pilot career models. Pilots from completely different operations might show different visual scanning behavior. Finally, manual flying is not restricted to ILS approaches only. Other flight phases and maneuvers will prompt other gaze patterns potentially with other AOI.

Acknowledgements

Both authors have equally contributed to this paper. This work was funded by the German Federal Ministry of Economics and Technology via the Project Management Agency for Aeronautics Research (DLR) within the Federal Aeronautical Research Program (LuFo IV-2). The authors acknowledge the support of Patrick Gontar and Ekkehart Schubert in the flight simulator experiments, of David Schopf for literature review, of Hang Yin for his support in gaze data analysis and his creatorship of the saccade grouping approach, of Armin Eichinger for support in statistics, of Michael Dorr for literature recommendations, and of Joost de Winter in revising the structure of the paper.

References

Allsop, J., Gray, R., 2014. Flying under pressure: effects of anxiety on attention and gaze behavior in aviation. J. Appl. Res. Mem. Cognit. 3 (2), 63–71. http://dx.doi.org/10.1016/j.jarmac.2014.04.010.

Anderson, N.C., Bischof, W.F., Laidlaw, K., Risko, E.F., Kingstone, A., 2013. Recurrence quantification analysis of eye movements. Behav. Res. methods 45 (3), 842–856. http://dx.doi.org/10.3758/s13428-012-0299-5.

Bellenkes, A.H., Wickens, C.D., Kramer, A.F., 1997. Visual scanning and pilot expertise: the role of attentional flexibility and mental model development. Aviat. Space, Environ. Med. 68 (7), 569–579.

Bissonnette, N., Culet, J.P., 2013. *Airbus A318/A319/A320/A321/A330/A340* (Flight Standardization Board (FSB) Report). Renton. Retrieved from. http://fsims.faa.gov/wdocs/fsb/a320%20rev%204.htm.

Brière, D., Traverse, P., 1993. AIRBUS a320/a330/a340 electrical flight controls: a family of fault-tolerant systems. Proc. Twenty-Third Int. Symp. Fault-Toler. Comput. 616–623. http://dx.doi.org/10.1109/FTCS.1993.627364.

Chapman, P., Underwood, G., Roberts, K., 2002. Visual search patterns in trained and untrained novice drivers. Transp. Res. Part F Traffic Psychol. Behav. 5 (2), 157–167. http://dx.doi.org/10.1016/S1369-8478(02)00014-1.

Chen, H.-Y.W., Milgram, P., 2011. Determining fixed glance duration for visual occlusion research. Proc. Hum. Factors Ergon. Soc. Annu. Meet. 55 (1), 1904–1908. http://dx.doi.org/10.1177/1071181311551396.

Colvin, K., Dodhia, R., Dismukes, K., 2005. Is pilots' visual scanning adequate to avoid mid-air Collisions?. In: In Proceedings of the 13th International Symposium on Aviation Psychology, pp. 104–109.

Di Nocera, F., Camilli, M., Terenzi, M., 2006. Using the distribution of eye fixations to assess pilots' mental workload. Proc. Hum. Factors Ergon. Soc. Annu. Meet. 50 (1), 63–65. http://dx.doi.org/10.1177/154193120605000114.

Dick, A.O., 1980. Instrument Scanning and Controlling: Using Eye Movement Data to Understand Pilot Behavior and Strategies (NASA Contractor Report No. 3306). Rochester.

Dorr, M., Martinetz, T., Gegenfurtner, K.R., Barth, E., 2010. Variability of eye movements when viewing dynamic natural scenes. J. Vis. 10 (10), 28. http://dx.doi.org/10.1167/10.10.28.

Duncan, J., Williams, P., Brown, I., 1991. Components of driving skill: experience does not mean expertise. Ergonomics 34 (7), 919–937. http://dx.doi.org/10.1080/00140139108964835.

Ebbatson, M., 2009. The Loss of Manual Flying Skills in Pilots of Highly Automated Airliners (PhD Thesis). Cranfield University, Cranfield.

Einhäuser, W., Rutishauser, U., Koch, C., 2008a. Task-demands can immediately reverse the effects of sensory-driven saliency in complex visual stimuli, 2.1-19 J. Vis. 8 (2). http://dx.doi.org/10.1167/8.2.2.

Einhäuser, W., Spain, M., Perona, P., 2008b. Objects predict fixations better than early saliency, 18.1-26 J. Vis. 8 (14). http://dx.doi.org/10.1167/8.14.18.

Entzinger, J.O., Suzuki, S., 2012. Inside the Pilot's mind: visual information use in manual aircraft landings. In: ICAS 2012 Proceedings. 28th Congress of the International Council of the Aeronautical Sciences. Optimage Ltd, Edinburgh.

Commission Regulation (EU) No 1178/2011 of 3 November 2011, Volume 54Official Journal of the European Union (European Union 2011).

Favre, C., 1994. Fly-by-wire for commercial aircraft: the Airbus experience. Int. J. Control 59 (1), 139–157. http://dx.doi.org/10.1080/00207179408923072.

Federal Aviation Administration, 2012. *Instrument Flying Handbook* (No. FAA-h-8083–15B). Oklahoma City. Retrieved from. http://www.faa.gov/regulations_policies/handbooks_manuals/aviation/media/faa-h-8083-15b.pdf.

Field, E., Harris, D., 1998. A comparative survey of the utility of cross-cockpit linkages and autoflight systems' backfeed to the control inceptors of commercial aircraft. Ergonomics 41 (10), 1462–1477. http://dx.doi.org/10.1080/001401398186216.

Fitts, P.M., Jones, R.E., Milton, J.L., 1949. Eye Fixations of Aircraft Pilots (AF Technical Report No. 5967). Wright-Patterson Air Force Base.

Foulsham, T., Dewenhurst, R., Nyström, M., Jarodzka, H., Johansson, R., Underwood, G., Holmqvist, K., 2012. Comparing scanpaths during scene encoding and recognition: a multi-dimensional approach. J. Eye Mov. Res. 5 (4), 1–14. Retrieved from. www.jemr.org/online/5/4/3.

Franks, P., Hay, S., Mavin, T.J., 2014. Can competency-based training fly?: an overview of key issues for ab initio pilot training. Int. J. Train. Res. 12 (2), 132–147. http://dx.doi.org/10.5172/ijtr.2014.12.2.132.

Harris, R.L., Christhilf, D.M., 1980. What do pilots see in displays? Proc. Hum. Factors Ergon. Soc. Annu. Meet. 24 (1), 22–26. http://dx.doi.org/10.1177/107118138002400106.

Harris, R.L., Glover, B.J., Spady, A.A., 1986. Analytical Techniques of Pilot Scanning Behavior and Their Application (NASA Technical Paper No. 2525).

Haslbeck, A., Bengler, K., 2016. Pilots' gaze strategies and manual control performance using occlusion as a measurement technique during a simulated manual flight task. Cognition. Technol. Work 18 (3), 529–540. http://dx.doi.org/10.1007/s10111-016-0382-2.

Haslbeck, A., Hörmann, H.-J., 2016. Flying the needles: flight deck automation erodes fine-motor flying skills among airline pilots. Hum. Factors 58 (4), 533–545. http://dx.doi.org/10.1177/0018720816640394.

Haslbeck, A., Schubert, E., Gontar, P., Bengler, K., 2012. The relationship between pilots' manual flying skills and their visual behavior: a flight simulator study using eye tracking. In: Landry, S.J., Salvendy, G., Karwowski, W. (Eds.), Advances in Human Factors and Ergonomics, Advances in Human Aspects of Aviation. CRC Press, Boca Raton, pp. 561–568.

Hayashi, M., 2004. Hidden Markov Models for Analysis of Pilot Instrument Scanning and Attention Switching. Massachusetts Institute of Technology, Cambridge.

Holmqvist, K., Nyström, M., Andersson, R., Dewenhurst, R., Jarodzka, H., van de Weijer, J., 2011. Eye Tracking: A Comprehensive Guide to Methods and Measures. Oxford University Press, Oxford, New York.

International Air Transport Association, 2015. Safety Report 2014. Issued April 2015. Montréal.

ISO 15007-1:2015, 2015. Road Vehicles - Measurement of Driver Visual Behaviour with Respect to Transport Information and Control Systems - Part 1: Definitions and Parameters.

Jarodzka, H., Scheiter, K., Gerjets, P., van Gog, T., 2010. In the eyes of the beholder: how experts and novices interpret dynamic stimuli. Learn. Instr. 20 (2), 146–154. http://dx.doi.org/10.1016/j.learninstruc.2009.02.019.

Joint Aviation Authorities, 2004. Airbus A320-A330-A340 Cross Crew Qualification & Mixed Fleet Flying.

Jones, D.H., 1985. An Error-Dependent Model of Instrument-Scanning Behavior in Commercial Airline Pilots (NASA Contractor Report No. 3908).

Kang, Z., Landry, S.J., 2014. Using scanpaths as a learning method for a conflict detection task of multiple target tracking. Hum. Factors 56 (6), 1150–1162. http://dx.doi.org/10.1177/0018720814523066.

Kang, Z., Landry, S.J., 2015. An eye movement analysis algorithm for a multielement target tracking task: maximum transition-based agglomerative hierarchical clustering. IEEE Trans. Human-Mach. Syst. 45 (1), 13–24. http://dx.doi.org/10.1109/THMS.2014.2363121.

Kasarskis, P., Stehwien, J., Hickox, J. C., Aretz, A., Wickens, C. D., 2001. Comparison of expert and novice scan behaviors during VFR flight. 11th International Symposium on Aviation Psychology, Columbus, Ohio.

Kowler, E., 2011. Eye movements: the past 25 years. Vis. Res. 51 (13), 1457–1483. http://dx.doi.org/10.1016/j.visres.2010.12.014.

Kuebler, T.C., Eivazi, S., Kasneci, E., 2015. Automated Visual Scanpath Analysis Reveals the Expertise Level of Micro-neurosurgeons (MICCAI'15 Workshop on Interventional Microscopy).

Milton, J.L., Jones, R.E., Fitts, P.M., 1950. Eye Fixations of Aircraft Pilots (AF Technical Report Nr. 6018). United States Air Force, Dayton, Ohio.

Myers, C.W., 2007. Scan Pattern Adaptation during Repeated Visual Search (PhD Thesis). Rensselaer Polytechnic Institute, Troy.

Noton, D., Stark, L., 1971. Scanpaths in eye movements during pattern perception. Science 171 (3968), 308–311. http://dx.doi.org/10.1126/science.171.3968.308.

Ottati, W.L., Hickox, J.C., Richter, J., 1999. Eye scan patterns of experienced and novice pilots during visual flight rules (VFR) navigation. Proc. Hum. Factors Ergon. Soc. Annu. Meet. 43 (1), 66–70. http://dx.doi.org/10.1177/154193129904300114.

de Rivecourt, M., Kuperus, M., Post, W., Mulder, L., 2008. Cardiovascular and eye activity measures as indices for momentary changes in mental effort during simulated flight. Ergonomics 51 (9), 1295–1319.

Rantanen, E.M., Johnson, N.R., Talleur, D.A., 2004. The Effectiveness of Personal Computer Aviation Training Device, a Flight Training Device, and an Airplane in Conducting Instrument Proficiency Checks: Volume 2: Objective Pilot Performance Measures. Final Technical Report AHFD-04-16/FAA-04-6. Oklahoma City.

Robinski, M., Stein, M., 2013. Tracking visual scanning techniques in training simulation for helicopter landing. J. Eye Mov. Res. 6 (2), 1–17. Retrieved from. http://www.jemr.org/online/6/2/3.

Sarter, N.B., Woods, D.D., 1994. Pilot interaction with cockpit automation II: an experimental study of pilots' model and awareness of the flight management system. Int. J. Aviat. Psychol. 4 (1), 1–28. http://dx.doi.org/10.1207/s15327108ijap0401_1.

Scallen, S.F., Hancock, P.A., Duley, J.A., 1995. Pilot performance and preference for short cycles of automation in adaptive function allocation. Appl. Ergon. 26 (6), 397–403. http://dx.doi.org/10.1016/0003-6870(95)00054-2.

Schieber, F., Gilland, J., 2008. Visual entropy metric reveals differences in drivers' eye gaze complexity across variations in age and subsidiary task load. Proc. Hum. Factors Ergon. Soc. Annu. Meet. 52 (23), 1883–1887. http://dx.doi.org/10.1177/154193120805202311.

Schütz, A.C., Braun, D.I., Gegenfurtner, K.R., 2011. Eye movements and perception: a selective review. J. Vis. 11 (5) http://dx.doi.org/10.1167/11.5.9.

Senders, J. W., Kristofferson, A. B., Levison, W. H., Dietrich, C. W., Ward, J. L., 1967. The Attentional Demand of Automobile Driving. Paper sponsored by Committee on Highway Safety and presented at the 46th Annual Meeting (Highway Research Record No. 195).

Shapiro, K.L., Raymond, J.E., 1989. Training of efficient oculomotor strategies enhances skill acquisition. Acta Psychol. 71 (1–3), 217–242. http://dx.doi.org/10.1016/0001-6918(89)90010-3.

Simon, P., Rousseau, F., Angue, J.-C., 1993. Quantitative Analysis of Mental Workload Influence on Eye Scanning Movements (International Conference on Systems, Man and Cybernetics). Le Touquet.

SKYbrary, 2016a. Pilot Flying and Pilot Not Flying. Retrieved from. http://www.skybrary.aero/index.php/Pilot_Flying_and_Pilot_Not_Flying.

SKYbrary, 2016b. Flight Control Laws. Retrieved from. http://www.skybrary.aero/index.php/Flight_Control_Laws.

SKYbrary, 2016c. Stabilised approach. Retrieved from. http://www.skybrary.aero/index.php/Stabilised_Approach.

Spady, A.A., 1978. Airline Pilot Scan Patterns During Simulated ILS Approaches (NASA Technical Paper No. 1250). Hampton, Virginia. Retrieved from. http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19790011511.pdf.

Steelman, K.S., McCarley, J.S., Wickens, C.D., 2011. Modeling the control of attention in visual workspaces. Hum. Factors 53 (2), 142–153. http://dx.doi.org/10.1177/0018720811404026.

Tole, J.R., Stephens, A.T., Vivaudou, M., Ephrath, A., Young, L.R., 1983. Visual Scanning Behavior and Pilot Workload (NASA Contractor Report No. 3717).

Underwood, G., Chapman, P., Brocklehurst, N., Underwood, J., Crundall, D., 2003. Visual attention while driving: sequences of eye fixations made by experienced and novice drivers. Ergonomics 46 (6), 629–646. http://dx.doi.org/10.1080/0014013031000090116.

van Leeuwen, P.M., Happee, R., Winter, J. de, 2015. Changes of Driving Performance and Gaze Behavior of Novice drivers During a 30-min Simulator-based Training. Procedia Manuf. 3, 3325–3332.

Veillette, P.R., 1995. Differences in aircrew manual skills in automated and conventional flight decks. Transp. Res. Rec. 1480, 43–50.

Wetzel, P.A., Anderson, G.M., Barelka, B.A., 1998. Instructor use of eye position based feedback for pilot training. Proc. Hum. Factors Ergon. Soc. Annu. Meet. 42 (20), 1388–1392. http://dx.doi.org/10.1177/154193129804202005.

Wickens, C.D., 2002. Situation awareness and workload in aviation. Curr. Dir. Psychol. Sci. 11 (4), 128–133. http://dx.doi.org/10.1111/1467-8721.00184.

Wickens, C.D., Hollands, J.G., Banbury, S., Parasuraman, R., 2013. Engineering Psychology and Human Performance, fourth ed. Pearson, Boston.

Wiener, E.L., Curry, R.E., 1980. Flight-deck automation: promises and problems. Ergonomics 23 (10), 995–1011. http://dx.doi.org/10.1080/00140138008924809.

Yang, J.H., Kennedy, Q., Sullivan, J., Fricker, R.D., 2013. Pilot performance: assessing how scan patterns & navigational assessments vary by flight expertise. Aviat. Space, Environ. Med. 84 (2), 116–124. http://dx.doi.org/10.3357/ASEM.3372.2013.

Haslbeck, A., Kirchner, P., Schubert, E., & Bengler, K. (2014). A Flight Simulator Study to Evaluate Manual Flying Skills of Airline Pilots. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 58(1), 11–15. doi:10.1177/1541931214581003

A FLIGHT SIMULATOR STUDY TO EVALUATE MANUAL FLYING SKILLS OF AIRLINE PILOTS

Andreas Haslbeck[1], Paul Kirchner[1], Ekkehart Schubert[2], and Klaus Bengler[1]
[1]Institute of Ergonomics, Technische Universität München, [2]Institute of Aeronautics and Astronautics, Section Flight Guidance and Air Transportation, Technische Universität Berlin

This paper reports an experimental study with the objective to assess pilots' raw-data-based flight performance which is affected by long-term practice and structured training. Fifty-seven airline pilots with different levels of aviation experience scheduled on an Airbus fleet, representing contrary levels of practice and training, had to fly a simulated 45 minutes approach and landing scenario while flight performance data were objectively recorded. The level of practice and training was found to have a significant influence on manual flying skills. Pilots with low levels of practice and training showed a large variance in manual flight performance; pilots with high levels of practice and training demonstrated high and homogenous performance.

INTRODUCTION

For today's aviation experts, manual flying is a critical issue. The aviation industry, primarily aircraft manufacturers and air carriers, is trying to manage the trade-off between safe and economic flight operation. On the one hand, they put emphasis on the training of their pilots' manual skills, but on the other hand, obligatory manual training hours are to be reduced with every new type of aircraft introduced, in order to reduce the effort for necessary type rating training sessions. In preventing accidents, the pilot's manual flying skills are considered as the last line of defense: if all the automation breaks down and manual operation becomes necessary, a pilot is still in charge of conducting a safe landing on his own. Two very prominent disasters, indicating lacking manual flying skills are Air France flight 447 and Asiana Airlines flight 214 (BEA, 2012; NTSB, 2014)

Although much research has been done in the field of aviation, empirical studies on manual flight performance in specific comparable scenarios are still rare. Some of the more prominent works in this area are discussed below. In the present paper, the question is addressed whether airline pilots can maintain sufficient manual flying skills by recurrent training and daily flight practice over the course of a pilot's career. The experimental results are derived from a flight simulator study, which was performed in cooperation with a major European airline.

Acquisition of manual flight

Manual control of an aircraft is an active task relative to when pilots monitor the aircraft under automation (Flach, 1990; Sarter & Woods, 1994); also known as a closed-loop control problem (Field & Harris, 1998; Wickens, 2003). Manual flying is a psychomotor process requiring more than operating the control stick of an aircraft. Three main stages of information processing have to be considered in manual flying: perception, cognitive processing, and response execution (Childs & Spears, 1986). One model frequently referred to for these sequent stages was founded by Wickens (Wickens & Hollands, 1999). In flight school, pilots learn and intensively train these active processes before they are introduced to automation, which then switches their task as a pilot from handling an aircraft to managing it (Childs & Spears, 1986; JAA, 2006). From this point on, pilots are faced with automation induced skill degradation (Balfe, Wilson, Sharples, & Clarke, 2012), caused by the automation taking over the responsibility for tasks previously performed by the human operators (Parasuraman & Riley, 1997).

Automation-induced changes on the flight deck

As flying becomes more automated, pilot´s manual flying skills degrade. This inverse relationship is primarily caused by the automation altering the active flying task to a passive monitoring task (Sarter & Woods, 1995). The introduction of early glass cockpits (late 1970s), flight management systems, and fly-by-wire control (late 1980s) in commercial aviation were significant automation milestones. Billings (1991) also described these changes in terms of information, management, and control automation. Automation helps the human operator in difficult situations when incapacity, workload, fatigue or inaccuracy occur – just to name a few. Well known ironies of automation describe negative automation effects like skill degradation (Billings, 1991; Endsley & Kiris, 1996) or reduced operator vigilance (Endsley, 1999).

Evidence-based experimental studies to assess degrading manual flying skills

In the mid-1980s, empirical studies shed some light on this issue. While a large number of experiments explored exposure to automation and related situation awareness, only a few attempts were made at investigating and measuring the development and degradation of manual flying skills under automation.

An early effort to warn of diminishing flying skills was made by Childs and Spears (1986). They postulated a concern that ineffective perceptual processes lead to deteriorated motor responses. Sarter and Woods (1994) reported a study focusing on pilots' mental models of the flight management system. Their findings revealed that these mental models do not account for monitoring skills and, to some degree, manual skills, as they require active interventions by the pilot, like the loss of glideslope information. The transition from conventional flight decks towards automated ones was evaluated in a simulator study by Veillette (1995). In an experiment with non-voluntary airline pilots, he measured manual flight performance across varying degrees of automation. The results of this study showed significant differences in manual flight performance between the two groups: the pilots accustomed to automation had significantly larger deviations from the ideal flight paths – evidence of degrading manual skills due to automation. A more recent study analyzing manual flight performance was introduced by Gillen (2008). His comparison showed that pilots' self-assessment delivered higher ratings than pilots were able to perform in the simulator experiment – pilots' confidence in their own skills was subject to a bias. In addition, the pilots tested performed below certification standards, which means these subjects would have failed in a certification situation.

Ebbatson (2009) showed in a large-scale analysis of manual flight performance data a correlation between manual flying skills and practice rather than overall flight experience. Recent flight practice including manual flying occurring a few weeks prior to the experiment had more influence on the measured performance than flight hours accumulated over a pilot's entire career. Finally, Ebbatson suggests replicating his experiment with a group of long-haul pilots, where he assumes even stronger effects between manual flight performance and practice would be found.

These studies all had specific foci, but cannot deliver a comprehensive view on the performance of pilots with a low level of practice and only few opportunities for training, like pilots in long-haul operation. Attempting to fill this gap, the following study was conducted to focus on long-haul pilots. Considering different approaches already mentioned, the current study addresses the following aspects specifically (see also Haslbeck et al. (2012)):

- A randomized sample for manual flying experiments is necessary to avoid self-selection and volunteer biases, thus, participants should not be chosen on a voluntary basis. Otherwise pilots with fairly high skill levels tend to participate in such experiments.
- A highly realistic and valid standardized setting for experimental simulator studies is needed, using a certified full flight simulator with motion effects e.g. from light turbulence weather effects and having real air traffic control instructions, which would force pilots to handle a higher workload by distinguishing between remote messages and their own messages in radio communications.
- The difficulty of a scenario should deliver tasks that can be fulfilled but should also give participants a chance to fail due to their manual flying skills.
- The highly standardized progress sequence of the experimental simulation scenario should ensure that, in general, all participants face the same technical, environmental, and organizational conditions.

METHOD

Research Question and Aim of the Study

The main research question is: How do practice and training influence manual flying skills? The concept 'level of practice and training' (according to the German expression *Trainiertheit*) stands for the manual flying skill level of pilots and is affected by the following three aspects:

- passed time since initial flight school to account for long-time skill degradation;
- daily flight practice to consider aspects of on-the-job training of skill;
- the effect of flight simulator training lessons, when selected flying tasks and maneuvers are repeatedly practiced and tested under supervision.

Flight experience, for example in terms of flight hours, and the level of practice and training are inversely proportional: while their flight hours are continuously rising, for most pilots, sections including active handling are rare especially on the long-haul. In this context, experience is rather meant as declarative knowledge how to solve problems and tasks than implicit knowledge or skill how to fly an aircraft. Therefore, it is expected that long-haul captains (CPTs) would have a lower skill level than short-haul first officers (FOs), because it has been longer since they attended flight school including systematic initial flight training, and they have a significantly lower frequency of recent flights than their short-haul colleagues.

Participants

To investigate the influence of practice and training, younger FOs on short-haul schedules and elder CPTs on long-haul service were chosen at random (stratified random sample), to establish an extreme groups design, representing two typical populations on both evaluated fleets. All participants occupied the same seat as they do in line operation. Two Airbus-type qualified full-flight simulators (JAR-FSTD A) were used for this study because of the very comparable cockpit designs and the resulting ease of transferring between different types (communality). 27 male CPTs participated, representing a low level of daily practice and training but a high level of operational experience. Their simulator was operated in an Airbus A340-600 configuration. For the other group, representing a high level of daily practice and training but a low level of operational experience, 30 FOs (27 male, 3 female) took part in this experiment in an Airbus A320-200 simulator. They should have been in line operation for about five years. Two randomly selected CPTs reported sick and were replaced by two equally qualified but voluntary CPTs. All participating pilots experience four simulator events per annum. The CPTs had more operational tasks (executive decisions) in their last two simulator sessions prior to the experiment, while the FOs were said to have experienced manual flying tasks. This means that the kind of training for both groups of pilots ideally met the experiment's demands. Table 1 shows the demographical data, showing flight experience as overall flight hours and years since flight school, as well as the number of individually performed landings within the past 30 days. Participating pilots were scheduled for the experiment by their company's flight operations department, so participation was part of their service schedule and not on a voluntary basis. All pilots were airline pilots (ATPL) and in service of the cooperating airline. All CPTs held the type rating for A330/340 family aircraft, and the FOs held the type rating for A319/320/321 family types.

Table 1. Demographical data of participants (mean values).

	age	overall flight hours	indiv. landings in past 30 days	years since flight school
CPTs (n=27)	50.4	15,019.7	3.4	24.6
FOs (n=30)	30.4	3,373.9	16.6	4.5

Procedure

The participating pilots were prepared the same way as for a regular flight, wearing pilot uniforms and bringing their daily used computers for the electronic flight bag system. Subjects were always instructed to be the pilot flying (PF). A confederate pilot monitoring (PM) was instructed to have a passive but cooperative role, and not to cause errors. These confederate pilots (two alternating for each group of participants) were also scheduled by the partner airline on the correspondent fleet. The first subject began the experiment (three subjects per night) approximately two and a half hours after the starting time. The whole procedure resembles longer flights with landings during the early morning hours, representing long-haul flights from the east or mid-range flights operated with short-haul aircraft in the partner airline (Haslbeck et al., 2012).

Scenario

After an uneventful flight from the east toward Munich Airport, the PF returned from his last break to perform the approach and landing 25 minutes prior to scheduled touchdown. All flight crews had to perform a missed approach before intercepting the ILS (guide beam provided by the instrument landing system) for a second time. At this time the approach mode could not be armed and the autopilot was disabled by a scripted event. After this point, the pilots had to perform all flying activities manually without the flight director and autopilot assistance. When the localizer was manually intercepted – providing runway centerline guidance – the measurement of manual flight performance started. A hand-flown landing (raw data ILS) with touchdown ended the 45-min. scenario.

Dependent Measures

Pilots were instructed to act and fly according to standard operating procedures of their airline (including licensing standards) – the same as in a real flight. Flight performance data were objectively measured by the flight simulator's data recorder. Here, deviations from the ideal glide slope (vertical guidance), and localizer (lateral guidance) were measured. These metrics represent the resulting system performance according to a control loop including the pilot and the aircraft (Morris & Miller, 1996). Flight path deviations can be considered in two different ways: measurement of absolute values with averaging afterwards or comparing maximum deviations to licensing standards. Both approaches were pursued and are subsequently shown.

All data for the localizer and glide slope are standardized to aberrations in dots, a unit which can be monitored on the primary flight display in the cockpit and which gives pilots information about their actual attitude with respect to the ideal approach path. The individual glide slope variations for all participants are observed from 3,000 ft. AGL (above ground level) down to 200 ft. AGL, whereas localizer aberration is considered significant from 3,000 ft. AGL to the model height of the aircraft above the threshold of 50 ft. AGL.

According to partner airline manuals, guidelines and laws (JAA, 2006), a maximum variance of one dot deflection on each side of the primary flight display localizer and glide slope scale must be maintained on precision approaches at all times. All measures can be directly compared to pilots' licensing standards (JAA, 2006; Ebbatson, 2009). To complement the comparison of maximum deviation values to legal standards, the root mean square error (RMSE) was calculated for all pilots to give a combined measure of their accuracy, equally weighting mean error and standard deviation (Hubbard, 1987; Flach, 1990). Here this measure is taken to express the differences between both groups, rather than to distinguish between the directions of both the localizer and glide slope.

RESULTS

Pilots´ manual flying performance in terms of maximum localizer and glide slope deviations from the manually flown ILS approach is shown in figure 1 and 2. These two diagrams evaluate pilots' skill against licensing standards (± 1dot max.).

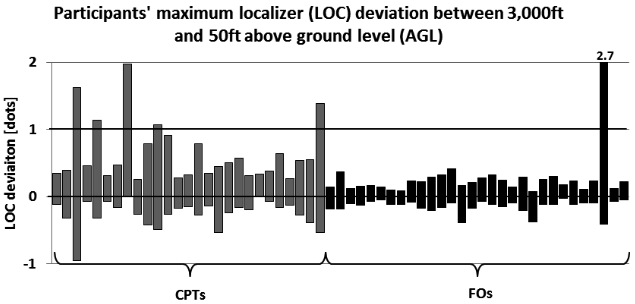


Figure 1: participants' maximum localizer deviations

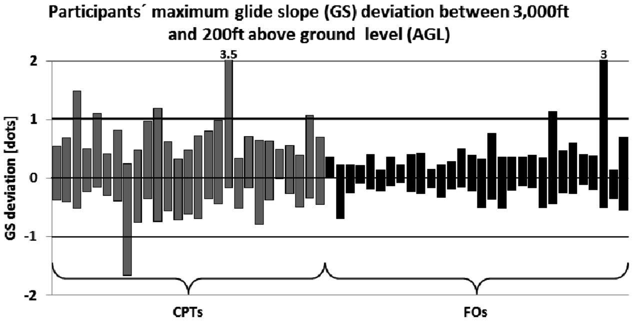


Figure 2: participant's maximum glide slope deviations

Each bar represents the maximum deviation for each pilot from the target value in figure 1 and 2. Positive localizer deviations imply a horizontal drift to the right of the runway centerline, with negative deviations correspondingly to the left. Vertical drift information is provided via the glide slope indicator. Positive deviances equal an aircraft position higher than the ideal glide path, while negative deviances in contrast represent a lower than ideal position of the airplane. Given the results depicted in Figure 1, six out of 57 subjects violated restrictions on the allowed localizer variance. For glide slope deviation, eight out of 57 participants could not perform within the acceptable limits. A total of nine different pilots (15.8 %) did not meet the mandatory skill test requirements in this scenario. Relative to test-person groups, seven (25.9 %) out of 27 CPTs did not fulfil at least one of the binding ILS deviation parameters, while two (6.7 %) out of 30 FOs did not. For localizer and glide slope deviations, the RMSE is shown in Figure 3.

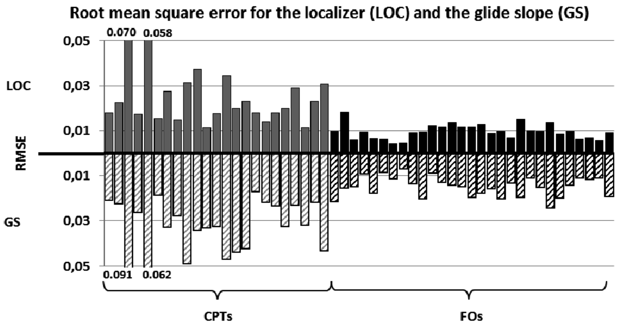


Figure 3: RSME for localizer and glide slope deviations

Table 2: statistical analysis of ILS deviations data

			mean	Independent t-Test	Mann-Whitney test
				t (df); p; r	U; z; p; r
max. localizer deviations	left	CPTs	.254	t(36.12)=2.552; p< .008; r= .39	U=249.5; z=2.489; p= .006; r= .33
		FOs	.143		
	right	CPTs	.646	t(55)=2.873; p= .003; r= .36	U=70; z=5.356; p< .001; r= .71
		FOs	.298		
RMSE localizer		CPTs	.025	t(23.86)=5.191; p< .001; r= .73	U=28; z=5.629; p< .001; r= .78
		FOs	.010		
max. glide slope deviation	high	CPTs	.796	t(55)=2.077; p= .021; r= .27	U=154; z=4.013; p< .001; r= .53
		FOs	.481		
	low	CPTs	.471	t(55)=3.08; p< .002; r= .38	U=211; z=3.102; p= .001; r= .41
		FOs	.271		
RMSE glide slope		CPTs	.035	t(24.48)=5.493; p< .001; r= .74	U=25; z=5.684; p< .001; r= .79
		FOs	.015		

The two groups were compared using parametric as well as non-parametric tests, as assumptions of normality could not be met for all deviation data. Both tests using an alpha-level of .05 indicate highly significant differences between the two groups' manual flying performances. In addition, the effect sizes expressed by a point-biserial r show moderate to very large effects.

DISCUSSION AND CONCLUSIONS

Evidence was found that participating CPTs with a lower level of practice and training (table 1) cause larger deviations from the ideal approach parameters than their more practiced FO colleagues; even if one limitation to this study is the fact that A320 and A340 differ in size, weight, and handling characteristics. From a technical or aeronautical perspective, these are completely different types of aircraft, while seen from a human factors view, both are successive milestones on a pilot's career and the human-machine-interfaces thinly differ because of the communality design principle of Airbus planes. When thinking about the differences in both types, the A340 is over four times heavier than its smaller counterpart and so it has larger flight path inertia. However both types are controlled by comparable roll and pitch rates, realized by larger control surfaces. In addition, the A340's higher flight path inertia might be an advantage in case of external perturbations like gusty wind. Based upon licensing standards on the one hand and flight operation realities on the other hand, all pilots have to perform within the same limits. Licensing standards neither differ between CPTs and FOs nor between long-haul or short-haul. Moreover both groups of pilots as well as both types of aircraft can use the same airports and runways. Thus requirements for manual aircraft handling are the same for all pilots. Another argument for choosing this comparison between A320 FOs and A340 CPTs is the assumed maximum range between the pilots' different skill levels (table 1) to utilize an extreme groups design.

That degrading manual flying skills have been observed in this rather small sample of pilots suggests that this is likely more prevalent than one could have suspected. As participants were active professional pilots, the results should be valid for other airline's personnel. In several cases, the deviations from ideal performance are large enough that pilots would have even failed a check situation – a dramatic finding that could reflect inadequate maintained skills. As training lessons normally cover equal contents for short-haul and long-haul pilots in longer sequences, one can assume that differences in manual flight performance are instead a consequence of everyday flight practice. This hypothesis is also supported by Ebbatson's (2009) study: accordingly, recent flight practice resulting from frequent flight operations, seem to be the most important factor in maintaining manual flying skills. Duncan, Williams and Brown (1991) have found some comparable insights in a real driving car experiment, "that adequate driving skills cannot be assumed, even for the `average´ experienced motorist, simply because they once were mastered […]". In comparison, simulator training can instead teach the right techniques for handling the aircraft (Buckley & Caple, 2009).

A further limitation to this study is that the level of practice and training is confounded with pilot's age and experience. Tsang (2003) describes and cites findings in her comprehensive review that "older, experienced individuals do not necessarily perform more poorly than their younger counterparts in tasks specific to their domain of expertise." Taylor, Kennedy, Noda, and Yesavage (2007) have reported a study investigating performance changes of pilots with different age and also under regard of different levels of expertise. Their results indicate no strong decline of landing skills by experienced (ATPL) pilots over time. In spite of these findings, in a field experiment with airline pilots, their level of practice and training will always be partially confounded with age and experience. For the measurement of manual flying skills, operational flight experience plays only a minor role and in an airline's daily operation, experience and the level of practice and training normally develop contrarily: CPTs have accumulated a vast amount of flight experience but experience only very few opportunities to practice flight skills – neither in simulator sessions nor in real operation. In spite of these limitations, the results of this study deliver a highly valuable picture of professional pilots' ability to manually control an aircraft.

Long-haul operation with its high degree of automation and pilots' long exposition to automated systems, was shown to have an eroding effect on manual flying skills; pilots with reduced flight duties and part-time schedules, like management pilots or ones who are on parental leave, should be kept in mind. Some examples to be supposed to airlines to implement strategies against deteriorating skills: additional simulator training sessions as well as type rating trainings concentrating on manual aircraft handling; combining short-haul and long-haul operation for long-haul pilots (mixed-fleet flying), especially for CPTs suffering from a lack of practice opportunities. For human-machine-interface designers, the approach of adaptive automation (Parasuraman, 2000) could also lead to a more flexible and dynamic task sharing between human and automation in the near future.

Future studies should further operationalize and analyze the influence of simulator training sessions. In addition, further groups of pilots with medium levels of practice and training, like short-haul CPTs and long-hauls FOs could complement insights in pilot's manual flying skills.

Balfe, N., Wilson, J. R., Sharples, S., & Clarke, T. (2012). Development of design principles for automated systems in transport control. Ergonomics, 55(1), 37–54.

BEA. (2012). Final Report on the accident on 1st June 2009 to the Airbus A330-203 registered F-GZCP operated by Air France flight AF 447 Rio de Janeiro - Paris. Le Bourget.

Billings, C. E. (1991). Human-Centered Aircraft Automation: A Concept and Guidelines: NASA-TM-103885. Moffet Field, California.

Buckley, R., & Caple, J. (2009). The theory & practice of training, 6th edition (6th ed.). London: Kogan Page.

Childs, J. M., & Spears, W. D. (1986). Flight-skill decay and recurrent training. Perceptual and Motor Skills, 62(1), 235–242.

Duncan, J., Williams, P., & Brown, I. (1991). Components of driving skill: experience does not mean expertise. Ergonomics, 34(7), 919–937.

Ebbatson, M. (2009). The Loss of Manual Flying Skills in Pilots of Highly Automated Airliners (PhD Thesis). Cranfield University, Cranfield.

Endsley, M. R. (1999). Level of automation effects on performance, situation awareness and workload in a dynamic control task. Ergonomics, 42(3), 462–492.

Endsley, M. R., & Kiris, E. O. (1995). The Out-of-the-Loop Performance Problem and Level of Control in Automation. Human Factors, 37(2), 381–394.

Field, E., & Harris, D. (1998). A comparative survey of the utility of cross-cockpit linkages and autoflight systems' backfeed to the control inceptors of commercial aircraft. Ergonomics, 41(10), 1462–1477.

Flach, J. M. (1990). Control With an Eye for Perception: Precursors to an Active Psychophysics. Ecological Psychology, 2(2), 83–111.

Gillen, M. (2008). Degradation of Piloting Skills (Master's Thesis). University of North Dakota, Grand Forks.

Haslbeck, A., Schubert, E., Onnasch, L., Hüttig, G., Bubb, H., & Bengler, K. (2012). Manual flying skills under the influence of performance shaping factors. Work: A Journal of Prevention, Assessment and Rehabilitation, 41(Supplement 1/2012), 178–183.

Hubbard, D. C. (1987). Inadequacy of root mean square error as a performance measure. In International Symposium on Aviation Psychology, 4th (pp. 698–704).

Joint Aviation Requirements - Flight Crew Licensing (Aeroplane), Joint Aviation Authorities 2006.

Morris, T. L., & Miller, J. C. (1996). Electrooculographic and performance indices of fatigue during simulated flight. Biological Psychology, 42(3), 343–360.

NTSB. (2014). Crash of Asiana Flight 214 Accident Report Summary: Descent Below Visual Glidepath and Impact With Seawall. Public Meeting of June 24, 2014. Washington, DC.

Parasuraman, R. (2000). Designing automation for human use: empirical studies and quantitative models. Ergonomics, 43(7), 931–951.

Parasuraman, R., & Riley, V. (1997). Humans and Automation: Use, Misuse, Disuse, Abuse. Human Factors, 39(2), 230–253.

Sarter, N. B., & Woods, D. D. (1994). Pilot Interaction With Cockpit Automation II: An Experimental Study of Pilots' Model and Awareness of the Flight Management System. The International Journal of Aviation Psychology, 4(1), 1–28.

Sarter, N. B., & Woods, D. D. (1995). How in the World Did We Get into That Mode? Mode Error and Awareness in Supervisory Control. Human Factors, 37(1), 5–19.

Taylor, J. L., Kennedy, Q., Noda, A., & Yesavage, J. A. (2007). Pilot age and expertise predict flight simulator performance: A 3-year longitudinal study. Neurology, 68(9), 648–654.

Tsang, P. S. (2003). Assessing Cognitive Aging in Piloting. In P. S. Tsang & M. A. Vidulich (Eds.), Principles and practice of aviation psychology (pp. 507–546). Mahwah, N.J: Lawrence Erlbaum.

Veillette, P. R. (1995). Differences in aircrew manual skills in automated and conventional flight decks. Transportation Research Record, (1480), 43–50.

Wickens, C. D. (2003). Pilot Actions and Tasks: Selections, Execution, and Control. In P. S. Tsang & M. A. Vidulich (Eds.), Principles and practice of aviation psychology (pp. 239–263). Mahwah, N.J: Lawrence Erlbaum.

Wickens, C. D., & Hollands, J. G. (1999). Engineering psychology and human performance (3rd). Upper Saddle River, NJ: Prentice Hall.

Haslbeck, A., & Hoermann, H.-J. (2016). Flying the Needles: Flight Deck Automation Erodes Fine-Motor Flying Skills among Airline Pilots. Human Factors. doi:10.1177/0018720816640394

Flying the Needles: Flight Deck Automation Erodes Fine-Motor Flying Skills Among Airline Pilots

Andreas Haslbeck, Technical University of Munich, Germany, and
Hans-Juergen Hoermann, German Aerospace Center, Hamburg, Germany

Objective: The aim of this study was to evaluate the influence of practice and training on fine-motor flying skills during a manual instrument landing system (ILS) approach.

Background: There is an ongoing debate that manual flying skills of long-haul crews suffer from a lack of flight practice due to conducting only a few flights per month and the intensive use of automation. However, objective evidence is rare.

Method: One hundred twenty-six randomly selected airline pilots had to perform a manual flight scenario with a raw data precision approach. Pilots were assigned to four equal groups according to their level of practice and training by fleet (short-haul, long-haul) and rank (first officer, captain).

Results: Average ILS deviation scores differed significantly in relation to the group assignments. The strongest predictor variable was fleet, indicating degraded performance among long-haul pilots.

Conclusion: Manual flying skills are subject to erosion due to a lack of practice on long-haul fleets: All results support the conclusion that recent flight practice is a significantly stronger predictor for fine-motor flying performance than the time period since flight school or even the total or type-specific flight experience.

Application: Long-haul crews have to be supported in a timely manner by adequate training tailored to address manual skills or by operational provisions like mixed-fleet flying or more frequent transitions between short-haul and long-haul operation.

Keywords: skilled performance, automation, perceptual-motor performance, manual controls, information processing

In his classical book about pilots' stick and rudder skills, Wolfgang Langewiesche (1944) explained that for learning the art of flying an aircraft, the pilot sometimes needs to withstand his or her natural responses. For example, in a stall situation at low altitude, the correct recovery requires to push the stick forward and point the aircraft's nose to the ground. A very strong skill is required to hold back powerful instinctive behaviors of pulling the stick backward in this situation. Hard and continuous drill is indispensable for pilots to acquire and maintain the adequate touch and feel essential to manually control the aircraft in any conceivable maneuver. However, in today's advanced technology, aircraft pilots are often lacking sufficient opportunities to practice when they are relieved too often from manual flying tasks by using automated systems (cf. SKYbrary, 2016a).

Manual control implies lateral (roll, heading), vertical (pitch, altitude, vertical speed), and longitudinal (airspeed) control of an aircraft (Puentes, 2011) mainly through adequate fine-motor inputs by the human pilot to a control yoke (Boeing types) or a sidestick (Airbus types) governing an aircraft's pitch and roll and the thrust levers. Yaw control by rudder pedals is a minor task performed in normal operation only momentarily during takeoff, the landing flare, and the deceleration after touchdown. In other words, manual control means hand flying by reference to raw data without highly automated systems like flight director, autopilot, autothrust, or other flight management systems (SKYbrary, 2016a). *Raw data* flying specifies the absence of the flight director (Casner, Geven, Recker, & Schooler, 2014). Under this basic but challenging condition, the pilot performs a compensatory tracking task and in parallel cognitively processes information about speed, altitude, and the flightpath. This task requires adequate knowledge and

---

Address correspondence to Andreas Haslbeck, Institute of Ergonomics, Technical University of Munich, Boltzmannstrasse 15, 85747 Garching, Germany; e-mail: haslbeck@tum.de.

HUMAN FACTORS
Vol. 58, No. 4, June 2016, pp. 533–545
DOI: 10.1177/0018720816640394

skills to control the dynamics of the aircraft to actively follow the intended trajectory. In addition, the localizer and glideslope indicators are among the most important information in the case of an instrument landing. For Airbus aircraft, manual flying is still supported by envelope protections in normal law mode.

Previous Work Investigating Pilots' Manual Flying Skills

Increased flight deck automation could reduce the opportunity for flight deck crews to practice their manual flying skills and therefore could degrade their levels of performance. Early warnings were raised by Wiener and Curry (1980) even prior to having broader data sets to examine this anticipated threat. Childs and Spears (1986) addressed perceptual and cognitive aspects of manual flying and its degradation, while Sarter and Woods (1994) reported few deficits in pilots' proficiency in standard tasks like aborting a takeoff or disengaging the approach mode. Veillette (1995) showed a significant influence of automation on manual flying skills. More recent experimental studies were performed by Gillen (2008), showing that pilots performed below licensing standards, and Ebbatson (2009) evaluated the effect of degradation on manual skills due to a lack of opportunities to practice among short-haul crews. In one of the latest studies, Casner et al. (2014) observed pilots having difficulties in cognitive tasks corresponding to manual flight. Quite recently though, aviation regulatory authorities have raised common concerns about the deterioration of basic manual flying skills among pilots flying highly automated aircraft and recommended some preventive actions (Civil Aviation Authority, 2014; European Aviation Safety Agency, 2013; Federal Aviation Administration, 2013a). On one side, there is a high level of agreement among pilot and training communities as well as manufacturers that manual and cognitive flying skills tend to decline because of a lack of practice due to increased use of automated systems. On the other side, as Civil Aviation Authority (2014) criticizes, scientific findings are still inconclusive as to which degree such decline occurs because evidence is often based on pilots' opinions and experiences or an analysis of narratives.

Evidence From Accident Statistics

With respect to long-haul crews, there is only anecdotal evidence that they suffer from an absence of practice opportunities, resulting in lower manual skills (Civil Aviation Authority, 2014; Drappier, 2008; Learmount, 2011). Aviation accidents with clear indications of a lack of manual flying skills, like the prominent Air France Flight 447 (Bureau d'Enquêtes et d'Analyses pour la sécurité de l'aviation civile, 2012) and Asiana Airlines Flight 214 (National Transportation Safety Board, 2014), are too rare to provide statistical evidence. Dismukes, Berman, and Loukopoulos (2007) analyzed 19 predominantly short-haul aviation accidents in the United States and found in 8 cases clear evidence for insufficient manual flying skills at least as a contributing factor. Lacking manual skills are also involved in many upset and loss-of-control accidents (Lambregts, Nesemeier, Wilborn, & Newman, 2008; Newman, 2012). The International Air Transport Association (IATA) published in 2015 an in-depth analysis of 415 accidents in commercial aviation, which occurred worldwide between 2010 and 2014. In 26% of these accidents (mostly landing accidents but also loss of control in flight), IATA found tangible evidence that manual handling flight crew errors were involved. As one of the recommendations to operators, IATA concluded that "Stable approaches are the first defense against runway excursions. The final, more important, defense is landing in the touchdown zone" (p. 77).

Degradation of Fine-Motor Flying Skills

While manual flying could be considered a rather simple tracking task in theory, pilots need regular practice and training to maintain this distinct set of fine-motor skills. Short-haul and long-haul operations support the maintenance of manual flying skills differently: For the former, 8 to 12 duty cycles per month with up to four legs each is typical, and for the latter, three to four long-range flights are performed monthly due to legal rest periods. Thus, both types of operation lead to different levels of practice in pilots' manual skills. None of the aforementioned studies directly addressed the influence of practice and training on fine-motor skills.

Based on the reviewed literature, we identified a research gap concerning a valid, holistic, and comparative evaluation of pilots' manual flight proficiency: (a) regarding different types of operations—long haul and short haul; (b) different levels of experience, responsibility, and tasks—captain (CPT) and first officer (FO); (c) under the recent amount of exposure to automation in today's advanced technology aircraft; and (d) in a realistic flying scenario familiar to the pilots. In our work, we are dealing with pilots who have different levels of practice, training, and experience while facing the same flying tasks within identical limits of licensing standards (European Union, 2011).

Research Questions and Hypotheses

The main research question of this paper is: What influence does the level of practice and training have on fine-motor flying skills? The *level of practice and training* is not a single and measurable metric or unit but rather a concept concerning flight proficiency that includes several influences: (a) elapsed time since initial flight training, addressing the long-term skill degradation (cf. Ebbatson, 2009; Franks, Hay, & Mavin, 2014); (b) daily flight practice, addressing on-the-job training (Fleishman, 1966; Savion-Lemieux & Penhune, 2005); and (c) the influence of flight simulator sessions, when periodically selected flying tasks and maneuvers are to be practiced (recurrent training) and tested (proficiency checks) under the supervision of trainers and examiners (Buckley & Caple, 2009). It was hypothesized that with a higher level of practice and training, pilots show better fine-motor flight performance (Haslbeck, Kirchner, Schubert, & Bengler, 2014). Referring to the *level of practice and training*, a secondary research question arises: whether the time dated back to flight school and expertise (determined by rank) or the daily flight practice (determined by fleet) has a stronger influence on manual flight proficiency. If the former aspect prevails, first officers would perform better because their elapsed time since flight school is shorter; in the latter case, short-haul crews would perform better because they have more daily flight practice. Ebbatson (2009) and Franks et al. (2014) have argued that (initial) training long ago cannot sufficiently support recent manual flying skills. Addressing these research questions, we expect to see a stronger effect from the daily flight practice and a slightly weaker effect from the time period since flight school, which consequently leads to the expected order of manual flight performance: FO short haul > CPT short haul > FO long haul > CPT long haul.

METHOD

This paper reports on the analysis of the fine-motor flight performance of airline pilots derived from two consecutive flight simulator studies. Both studies were funded by a German research program in cooperation with a major European airline. Experiment A took place in 2011, comparing manual flight performance of two groups of pilots: FOs scheduled on short-haul service, representing a high level of practice and training, as well as CPTs scheduled for long-haul operation, representing a low level of practice and training. To complement Experiment A with the two missing groups (i.e., CPTs on short haul and FOs on long haul), Experiment B was conducted in 2013 with CPTs scheduled for short-haul service and FOs scheduled for long-haul service.

Apparatus

Airbus types were selected for two reasons: First, the fly-by-wire technology designed under the commonality principle (Vadrot & Aubry, 1994) ensures very similar handling characteristics, and second, being equipped with second-generation electrical flight control systems, they expose pilots to high levels of automation (Brière & Traverse, 1993). Thus, we conducted both experiments in a southern German flight simulator training center equipped with two Airbus-type full-flight simulators (FFS Level D): one Airbus A320 device and another one in an Airbus A340-600 configuration.

Scenario and Instruction

Prior to both experiments, all participants completed another simulated flight scenario concerning operational problems (35 minutes) and simultaneously warmed up for the manual flying task. The flight scenario (10 minutes) for

TABLE 1: Demographic Data for Participants

			Age		Flight Hours: Overall/On Type		Landings in Past 30 Days		Years Since Flight School	
Rank	Fleet	*N*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
FO	A320	39	30.1	2.8	3,438/2,415	1,848/1,266	16.1	6.3	5.8	2.8
FO	A340	28	36.4	3.3	7,204/3,469	1,987/1,812	2.4	1.5	12.2	2.9
CPT	A320	30	43.0	4.3	11,276/3,847	1,931/2,355	16.6	10.2	18.1	3.3
CPT	A340	29	49.8	3.6	14,969/2,909	2,951/1,818	3.5	2.1	24.4	4.1

Note. ILS deviation data evaluation is based on these 126 participants; TD points were calculated for these and four more CPTs on A340. FO = first officer; CPT = captain; ILS = instrument landing system; TD = touching down.

our study was a manual approach to Munich Airport (26R EDDM) and was the same for all participants. It started shortly before a defect in the autopilot and the flight director occurred; thus, lateral and vertical control of the aircraft in an instrument landing system (ILS) approach had to be performed manually based on raw data. Pilots were, however, allowed to use autothrust for the longitudinal control, but very few did; auto-trim was engaged. The weather was set to the following parameters: visibility 1,200 meters, wind 220°/17–22 knots gusty, ceiling 270 feet, light rain. All pilots were instructed to perform a landing as accurately as possible according to company standard operating procedures and ATP licensing standards. Participants had the role of pilot flying (PF) and were supported by a pilot monitoring who was either a confederate pilot (Experiment A) or the second participating crewmember (Experiment B).

Participants

All participants were randomly selected by the crew scheduling department of the cooperating airline and occupied the same seat as well as the same aircraft type for which they were rated. Four groups (stratified random sample) of ATP licensed airline pilots, about 30 pilots per group, were scheduled as the PF in this experiment: FOs and CPTs on Airbus A320 as well as Airbus A340. All participating pilots experience routinely four annual 4-hour flight simulator training sessions: two recurrent training sessions and two legal licensing checks. Table 1 shows relevant demographical data for the 126 participants (see supplemental material available at http://hf.sagepub.com/supplemental). The overall flight hours are a general measure of flying experience. Landings within the last 30 days prior to the experiment account for short-term practice, and time period since flight school indicates long-term skill retention.

Dependent Measures

Deviations from ideal flight performance and landing parameters (cf. IATA, 2015, p. 77) were recorded and analyzed. ILS flightpath deflections were recorded in two different dimensions: For horizontal deviations from the localizer (LOC) and for vertical deviations from the 3°-glideslope (GS), maximum values were considered and the root mean square errors (RMSE) were calculated. The latter is a frequently used measure for fine-motor (flight) performance evaluations (McClernon, Miller, & Christensen, 2012; Rantanen, Johnson, & Talleur, 2004), even if it does not deliver the position information (Hubbard, 1987), which is of no interest for this study. Flight crew licensing standards require adhering deflections no larger than one dot for precision approaches (European Union, 2011). This unit corresponds to a deviation of ±0.8° on the LOC and ±0.4° on the GS for Airbus types and is indicated on two scales in the primary flight display. These flightpath deviations were measured for three different altitude segments. The upper segment, 3,000–1,000 feet above ground level (AGL), represents the initial instrument approach phase with medium difficulty, preparing the *stabilized approach* (SKYbrary, 2016b). The next segment, 1,000–270 feet AGL, stands for the increasingly demanding instrument approach phase within the limits for a stabilized approach, not exceeding deviations larger than one dot. The last segment, 270–50 feet AGL, represents the transition to the

TABLE 2: Results of Pilot Groups for ILS Flightpath and Touchdown Point Deviations

	RMSE ILS Flightpath Deviations (Dot)												TD Point Deviations (Meters)			
	3,000–1,000 Feet AGL				1,000–270 Feet AGL				270–50 Feet AGL							
	LOC		GS		LOC		GS		LOC		GS		LONG		LAT	
Group	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
FO A320	.10	.04	.14	.05	.12	.06	.19	.08	.14	.08	.74	.55	418	115	2.0	1.5
FO A340	.20	.08	.25	.10	.25	.14	.30	.13	.23	.18	.88	.67	501	133	2.1	2.0
CPT A320	.11	.04	.13	.06	.14	.06	.17	.07	.18	.12	.58	.28	428	104	1.2	1.1
CPT A340	.31	.22	.38	.20	.26	.10	.38	.18	.25	.11	1.2	.72	510	182	5.3	5.0

Note. RMSE = root mean square errors; ILS = instrument landing system; TD = touchdown; AGL = above ground level; LOC = localizer; GS = glideslope; LONG = longitudinal absolute distance to the threshold of the runway; LAT = lateral absolute distance to the centerline of the runway; FO = first officer; CPT = captain.

visual approach shortly before the landing flare. Apparently, GS data become somewhat unreliable in the last segment because some pilots seem to have commenced the flare above 50 feet AGL. The data sets were only analyzed if the participant had completely finished the approach without aborting it. A further measure for manual flight performance with high practical relevance is the first point of touching down (TD) upon landing. We measured these TD points in two dimensions: absolute longitudinal distances to the threshold of the runway and absolute lateral distances to the centerline. The ideal TD point is the boldly marked aiming point located about 400 meters (1,312 feet) behind the threshold. For this evaluation all data sets were included where the aircraft touched down, with or without a preceding go-around.

RESULTS

All effects will be reported as significant at $p < .05$, and η_p^2 is given as effect size. If the assumption of sphericity is violated, Greenhouse-Geisser estimates of sphericity are reported by ε. The mean values of all groups are presented in Table 2 for both measures: the flightpath deviations and the touchdown points.

Deviations From Localizer and Glideslope

The maximum deviations on localizer (Figure 1) and glideslope (Figure 2) in the segment between 1,000 and 270 feet AGL indicate that 10 (18%) out of 57 approaches (all on A340) exceeded with at least one flight parameter the limits of a stabilized approach or of the licensing standards of a precision approach.

All individual flightpath deviations averaged by the RMSE in the most relevant segment between 1,000 and 270 feet AGL are displayed in Figure 3. These flight performance data were analyzed with a 2 × 2 × 3 (between-subjects fleet [A320, A340] × between-subjects rank [FO, CPT] × within-subjects altitude [3,000–1,000, 1,000–270, 270–50]) multivariate analysis of variance with two dependent variables: deviations on localizer and glideslope. In most cases, RMSE and absolute deviation scores are not normally distributed but positively skewed. For that reason, all flightpath deviation data have been log-transformed for the further statistical analysis. The results of ILS deviations, which contain a number of significant between- and within-subjects effects, are depicted in Table 3. Average differences between groups are visualized in Figure 4 and Figure 5.

Analysis of Touchdown Points

A two-way MANOVA was conducted, including the aforementioned factors fleet and rank, and two dependent variables: longitudinal (LONG) and lateral (LAT) absolute distance to the threshold and to the centerline of the runway, respectively. Furthermore, Pillai's trace was

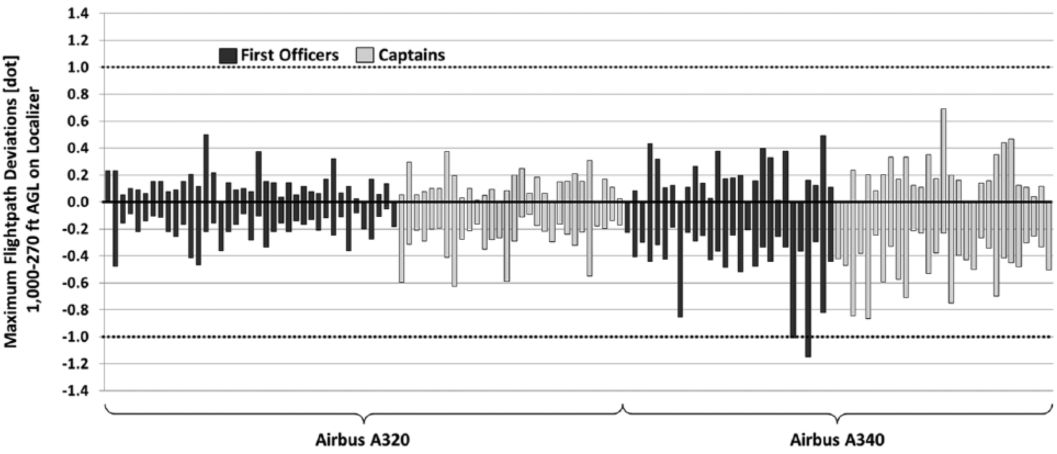


Figure 1. Individual maximum flightpath deviations on localizer. Negative values indicate an aircraft left of the extended centerline. Dotted lines indicate limits not to be exceeded for stabilized approach and license checks.

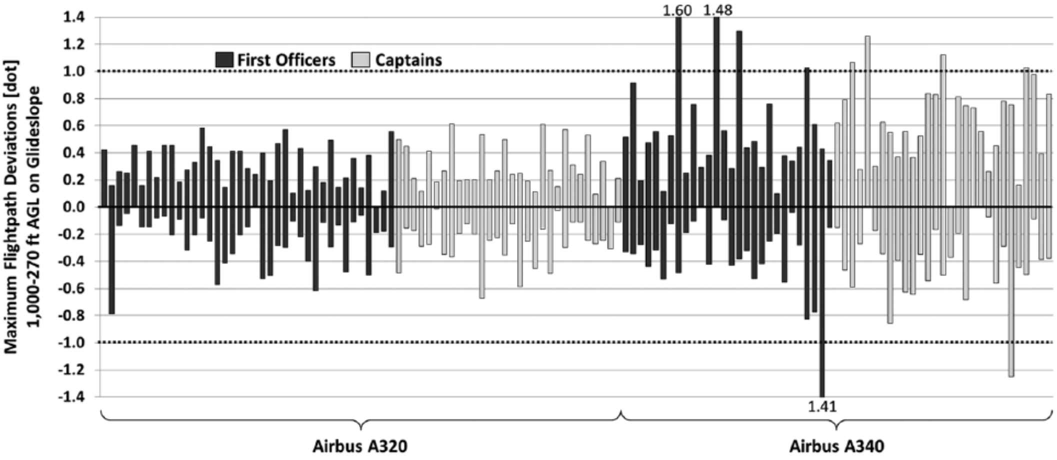


Figure 2. Individual maximum flightpath deviations on glideslope. Negative values indicate an aircraft below the glidepath. Dotted lines indicate limits not to be exceeded for a stabilized approach and license checks. Values larger than ±1.4 dot are depicted by numbers.

chosen as a rather robust multivariate test statistic showing significant differences between groups. Statistical analysis of TD points is given by Table 4, and differences between groups are illustrated in Figure 6.

Effect of Age and Time Since Initial Training

As shown in Table 1, the between-subjects effects of rank and fleet are correlated with the elapsed time since initial training and also with age. Based on the assumption that basic flight training is essential for building manual flying skills (cf. Langewiesche, 1944) and that these skills are prone to decay without regular practice in advanced-technology aircraft, the reported significant effects of fleet and rank could be confounded with differences in time since initial flight training or with age. In order to rule out the possibility that we simply found time-related effects, we included age and time since flight training as covariates in separate MANCOVAs, corresponding to the aforementioned MANOVAs in Table 3 and Table 4. In

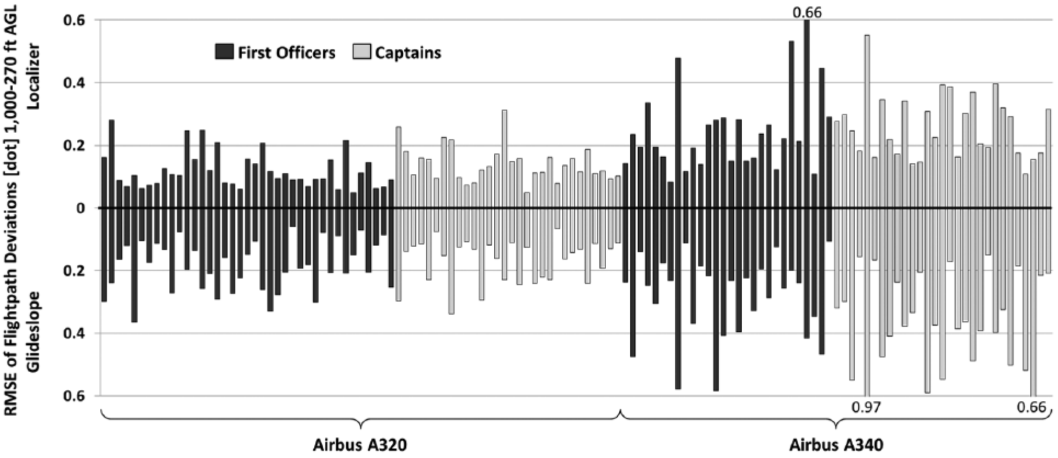


Figure 3. Root mean square error of all individual flightpath deviations on localizer (upper half) and glideslope (lower half). Values larger than ±0.6 dot are depicted by numbers.

fact, age and time since flight training are intercorrelated, $r_s = .95$, one-tailed $p < .001$. Consequently, the results are almost identical, and we only report the findings for the covariate time since initial training.

In the MANCOVA of the ILS deviations, with time since initial training as a covariate, only two between-subjects effects are significant. While the effect for rank disappears, fleet ($V = .31$, $F[2, 120] = 26.72$, $p < .001$, $\eta_p^2 = .31$) and Fleet × Rank ($V = .09$, $F[2, 120] = 5.98$, $p = .003$, $\eta_p^2 = .09$) are significant. The effect of time since initial training itself is not significant. The findings for the TD-point deviations are similar. Time since initial training has no significant direct effect on the absolute deviation scores, and rank is also not a significant between-subjects factor. Nevertheless, the differences between long-haul and short-haul pilots are still statistically significant ($V = .08$, $F[2, 123] = 5.13$, $p = .007$, $\eta_p^2 = .08$) as well as the Fleet × Rank ($V = .11$, $F[2, 123] = 7.71$, $p = .001$, $\eta_p^2 = .11$) interaction. Through these MANCOVAs, we can provide evidence that the differences in flying performance between long-haul and short-haul pilots cannot be interpreted as simple age or time effects. This confirms our second research question that daily flight practice has a stronger influence on manual flight proficiency than the time dated back to flight school.

DISCUSSION

General Findings According to Fine-Motor Flying Skills

With the comparison of flight performance of pilots on long-haul versus short-haul fleets, this study offers a quasi-experimental approach to the analysis of practice and training effects on the level of manual flying skills. The reported results clearly confirm that the level of practice and training as measured by daily flying practice and elapsed time since initial flight training does have significant influences on fine-motor flying skills of airline pilots. In summary, we found the following rank order for fine-motor flight performance: CPT A320 > FO A320 > FO A340 > CPT A340. According to Table 1, the A320 CPTs had at least two advantageous factors: (a) They performed the highest number of landings in the past month, and (b) they had more flight hours on type compared to all other groups. The A320 FOs had an equal amount of practice in the past month but less total flying experience and less time on type. When looking at the long-haul data, it seems that the total flight time and the time on type beyond 2,000 or 3,000 hours are less important factors for the level of manual skills than the daily practice and the time period since flight school. Therefore, the A340 FOs generally had more difficulties than the A320 FOs. Moreover, the A340 CPTs could not use their enormous flying

TABLE 3: Statistical Analysis of Log-Transformed ILS Flightpath Deviations

Source	Multivariate Tests	Univariate Tests
Between-subjects		
Fleet	$V = .45$, $F(2, 121) = 48.90$, $p < .001$, $\eta_p^2 = .45$	LOC: $F(1, 122) = 76.84$, $p < .001$, $\eta_p^2 = .39$ GS: $F(1, 122) = 75.77$, $p < .001$, $\eta_p^2 = .38$
Rank	$V = .08$, $F(2, 121) = 5.50$, $p = .005$, $\eta_p^2 = .08$	LOC: $F(1, 122) = 11.09$, $p = .001$, $\eta_p^2 = .08$ GS: $F(1, 122) = 3.24$, $p = .074$, $\eta_p^2 = .03$, *ns*
Fleet × Rank	$V = .09$, $F(2, 121) = 5.60$, $p = .005$, $\eta_p^2 = .08$	LOC: $F(1, 122) = .20$, $p = .655$, $\eta_p^2 = .00$, *ns* GS: $F(1, 122) = 9.20$, $p = .003$, $\eta_p^2 = .07$
Within-subjects		
Altitude	$V = .80$, $F(4, 119) = 118.53$, $p < .001$, $\eta_p^2 = .80$	LOC: $\varepsilon = .87$, $F(1.75, 213.18) = 1.78$, $p = .175$, $\eta_p^2 = .01$, *ns* GS: $\varepsilon = .88$, $F(1.76, 215.03) = 318.66$, $p < .001$, $\eta_p^2 = .72$
Altitude × Fleet	$V = .14$, $F(4, 119) = 4.73$, $p = .001$, $\eta_p^2 = .14$	LOC: $\varepsilon = .87$, $F(1.75, 213.18) = 8.00$, $p = .001$, $\eta_p^2 = .06$. GS: $\varepsilon = .88$, $F(1.76, 215.03) = 7.00$, $p = .002$, $\eta_p^2 = .05$
Altitude × Rank	$V = .02$, $F(4, 119) = .70$, $p = .595$, $\eta_p^2 = .02$, *ns*	
Altitude × Fleet × Rank	$V = .03$, $F(4, 119) = .81$, $p = .524$, $\eta_p^2 = .03$, *ns*	

Note. These analyses are based on log10-transformed flightpath deviation data. The assumption of normality was violated for only 1 out of 24 data sets: glideslope, 3,000–1,000 feet AGL, CPTs on A320. The assumption of equality of covariance matrices and error variances was violated; however, by having only two dependent variables, we assume this violation as minor (Tabachnick & Fidell, 2007). For multivariate tests, Pillai's statistic was reported. When the assumption of sphericity was violated (within-subjects tests), degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity (ε). The assumption of equality of error variances was violated for one out of six data sets: localizer, 270–50 feet AGL. ILS = instrument landing system; LOC = localizer; GS = glideslope; CPT = captain; AGL = above ground level.

experience as an advantage for the manual flying tasks. These senior long-haul pilots perform on average less than a quarter of the number of takeoffs and landings compared to short-haul pilots (Table 1). Hence, they have substantially less opportunity to practice their manual flying skills. If the level of skill is directly related to the amount of daily practice, long-haul pilots should show inferior performance in a manual flying task. Besides the type of operation (fleet), another factor is suspected as being responsible for reduced manual flying skills: the pilot generation (called *time since initial training* in this study). According to the opinion of the European Aviation Safety Agency (2013, p. 1), senior pilots may be less comfortable with automation, while younger pilots may lack basic flying skills because they normally have less flying time on non–glass cockpit aircraft types. But then, the time interval since basic flight training and hence the time for skill decay is shorter for younger pilots. Consequently, as expected, we found an interaction of the main effect for fleet with the factor rank.

A limiting factor of our previous research (Haslbeck et al., 2014) was that the different sources of variance (e.g., level of practice, flying experience, and type of aircraft) could not be separated because only A320 FOs had been compared to A340 CPTs. To reach a more conclusive comparison, we included two additional groups, A320 CPTs and A340 FOs, in this study. The findings concerning the importance of the amount of current practice for the level of

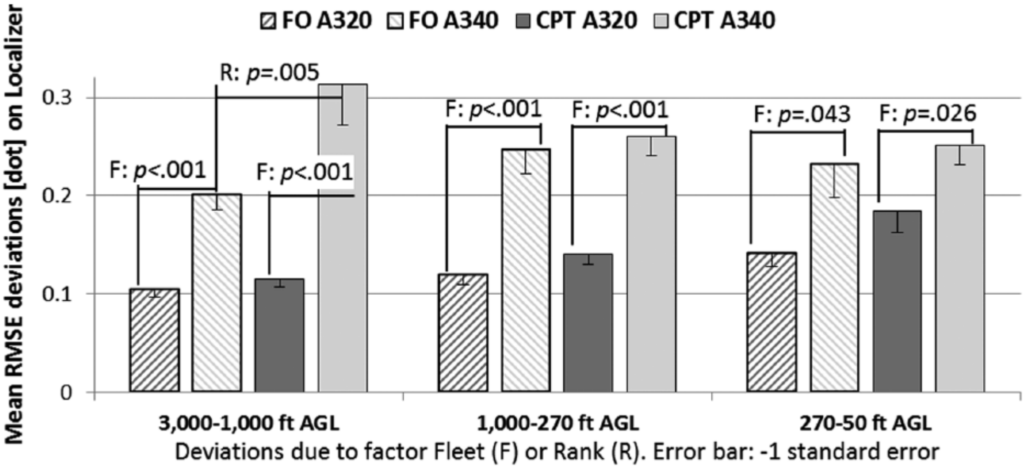


Figure 4. Visualization of localizer deviations based on non-transformed data indicating Bonferroni post hoc comparisons based on log-transformed data.

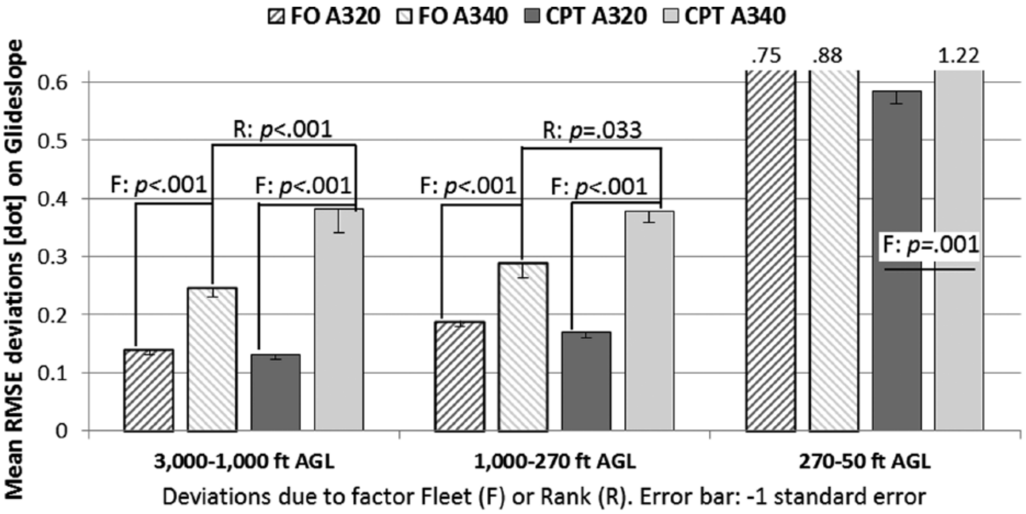


Figure 5. Visualization of glideslope deviations based on non-transformed data indicating Bonferroni post hoc comparisons based on log-transformed data. Values larger than 0.6 dot are depicted by numbers.

manual flying skills can generally be confirmed. After all, in this study we identified additional factors that are related to time since initial training, age, and experience.

All results in this study have clearly shown substantial influences of fleet on all manual flight performance scores. Many long-haul pilots have demonstrated consistently larger deviations from the ideal ILS flightpath, which can be explained by the lower level of practice. While the mean RMSE deviations from the localizer tend to remain constant across the three altitude segments, the deviations from the glideslope increase sharply for the final segment—the transition from instrument to visual flying. The first-order interaction effect (Altitude × Fleet) illustrates that the differences between the fleets become somewhat smaller when the aircraft approaches the ground, with the exception of the glideslope deviations during the visual segment.

TABLE 4: Statistical Analysis of Touchdown Point Deviations

Source	Multivariate Tests	Univariate Tests
Fleet	$V = .18$, $F(2, 125) = 13.43$, $p < .001$, $\eta_p^2 = .18$	LONG: $F(1, 126) = 11.67$, $p = .001$, $\eta_p^2 = .09$ LAT: $F(1, 122) = 17.25$, $p < .001$, $\eta_p^2 = .12$
Rank	$V = .04$, $F(2, 121) = 2.54$, $p = .083$, $\eta_p^2 = .04$, *ns*	LONG: $F(1, 126) = .15$, $p = .699$, $\eta_p^2 = .00$, *ns* LAT: $F(1, 126) = 5.06$, $p = .026$, $\eta_p^2 = .04$
Fleet × Rank	$V = .11$, $F(2, 125) = 7.90$, $p = .001$, $\eta_p^2 = .11$	LONG: $F(1, 126) = .00$, $p = .980$, $\eta_p^2 = .00$, *ns* LAT: $F(1, 126) = 15.84$, $p < .001$, $\eta_p^2 = .11$

Note. For multivariate tests, Pillai's statistic was reported. The assumption of equality of covariance matrices and error variances was violated. For rank, only a statistical trend was found in the multivariate test. Univariate results are reported to complete the picture (Tabachnick & Fidell, 2007). LONG = longitudinal absolute distance to the threshold of the runway; LAT = lateral absolute distance to the centerline of the runway.

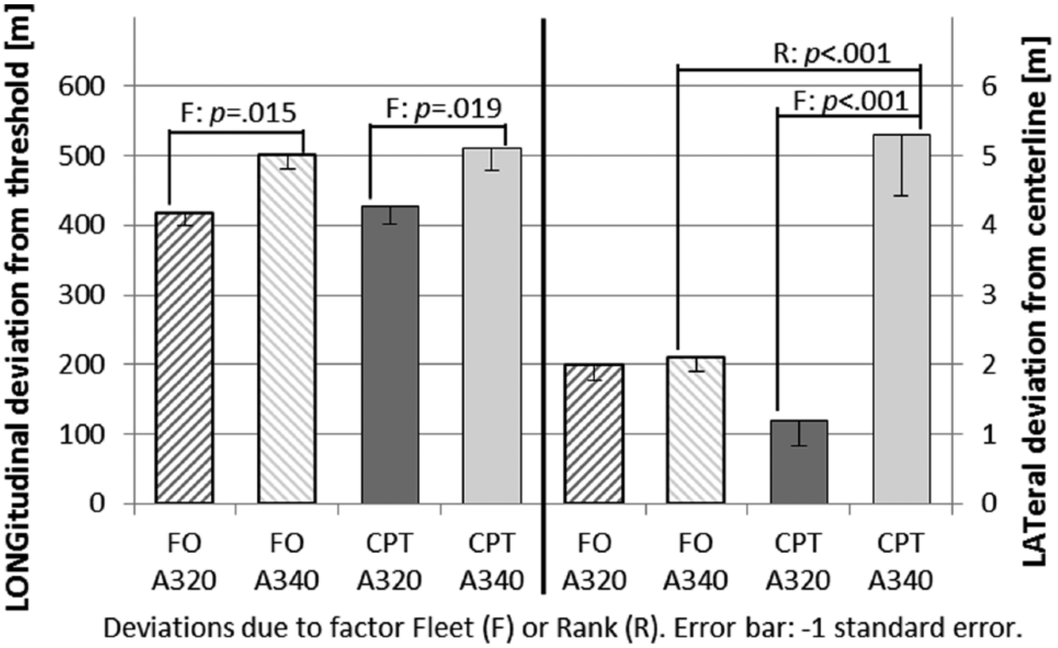


Figure 6. Touchdown point deviations indicating univariate test results.

Long-haul crews flew higher above the glideslope, obviously aiming at a TD point wider into the runway. For the sake of completeness, it must be said that Munich (EDDM) has 4,000 meter–long runways that significantly reduce the potential consequences of longer landings.

As a second performance parameter, we analyzed the absolute distances of the TD points from the threshold and from the centerline, respectively. Again, a strong effect for fleet was found. Short-haul pilots landed closer to the centerline and about 400 meters down the runway, while long-haul pilots performed longer landings (about 500 meters beyond the threshold) with larger deviations from the centerline. The interaction Fleet × Rank is due to the lower performance of long-haul CPTs.

The nature of effects for the between-subjects factor rank is more complex because CPTs were on average 13 to 14 years older than the FOs. Additionally, CPTs have accumulated about 8,000 hours more flight time, and the time since initial flight training was 12 to 13 years longer. Age, flight time, and time since flight school are highly correlated. In order to neutralize these confounding variables, we executed MANCOVAs with time since initial training as a covariate. In these analyses, all between-subject effects of rank were insignificant, while fleet still explained 31% of the variance for the ILS deviation measures and 8% of the variance for the TD points. Obviously, the level of practice measured by the number of executed landings per month contributed significantly to the decrement of manual flying skills in both CPTs and FOs of the long-haul fleet. This effect remained significant regardless of age or other time-related factors. The first-order interaction effects Fleet × Rank explained further variance in the analyses of ILS deviations and of the TD points. While CPTs showed better performance than FOs on the short-haul fleet, the long-haul FOs performed slightly better than the long-haul CPTs.

Besides the significance of current practice for fine-motor flying performance, our findings do not confirm recent concerns about a general lack of basic flying skills among the younger generation of pilots (Civil Aviation Authority, 2014; European Aviation Safety Agency, 2013). As the youngest group, the A320 FOs with an average age of 30 and a little less than six years of airline experience performed second best on the manual ILS and landing. At this stage of their career, they clearly had sufficient opportunity in practice and training to develop the necessary level of flying skills. In summary, pilots with little recurrent practice and extensive use of automation seem to be running the risk of losing Langenwiesche's (1944) touch and feel of how to *fly* an aircraft. Especially when considering Figure 3, the concern arises: What happens to pilots with even higher automated aircraft and longer working lifetime possibly spent on long-haul operation?

Limitations

Our analysis was carried out with pilots from one airline only. Findings could be different in other airlines with other training schemes and other pilot career models for their flight crews. We are also aware that *manual flying skills* in fact cover more tasks than a manual ILS approach with a precise landing within the touch-down zone. However, our aim was to complement existing research with objective performance data. By using scale deflections as the unit of accuracy, we assured that flightpath deviations are not weighed disproportionally against distance to touchdown.

One latent confound in this study deals with the different aircraft types. Both aircraft types are equipped with a fly-by-wire flight control system that has been designed to provide the same flying and handling qualities and maintain the highest applicable extent of commonality (Bissonnette & Culet, 2013; Brière & Traverse, 1993; Favre, 1994; Joint Aviation Authorities, 2004) providing a similar look and feel for the pilot. Nevertheless, differences in the dynamics between both types exist. However, the question is not whether both types can be controlled identically but whether both types can be controlled identically precisely when sufficient pilot training accounting for specific peculiarities of each type has been completed. This second question can be confirmed by the fact that runways, precision approaches, and certification standards for these types as well as licensing standards for the pilots are the same. According to the manufacturer's homepage, "a large majority of pilots prais[e] the handling qualities of Airbus aircraft and their commonality across the complete range of products" (Airbus SAS, 2016). Additional evidence for the commonality of the two aircraft types with respect to achievable precision in manual control comes from several A340 pilots in our sample who performed nearly on the same high level as the A320 crews did (Figures 1–3). However, none of the A340 pilots had the high level of daily flight practice as the A320 pilots did.

Recommendations

Until fail-proof automation outperforms human performance in all situations, the pilots remain the last line of defense in the cockpit. Under the described circumstances, these pilots need even stronger manual flying skills as proposed by Langewiesche (1944) in the earlier days of aviation. Based on our findings, we suggest a number of organizational and design recommendations for how to prevent a significant deterioration of manual flying skills. First of all, our findings indicate that the recent amount of regular simulator training is not sufficient to maintain manual flying skills of long-haul crews. Pilots with part-time schedules or reduced flight duties, like management pilots and pilots on parental leave, need special attention even when operating on short-haul service. Specific flight simulator training with a focus on manual flying tasks is one possible intervention. Mixed-fleet flying could be another powerful approach to increase a pilot's practice if negative transfer effects can be kept under control (Lyall & Wickens, 2005). In this case, pilots would perform short-haul and long-haul flights with type ratings for both types of aircraft in an alternating scheme. Another measure could be the operation of highly frequented short-haul connections with long-haul aircraft, like several flights within Japan. More manual flight practice could also be derived by changing companies' automation policies to encourage pilots to fly manually if the situation permits (European Aviation Safety Agency, 2013; Federal Aviation Administration, 2013b). Such interventions have to be applied in the earlier stages of a pilot's career before degradation can take place. Otherwise, avoidance behaviors and a feeling of discomfort according to manual flying could lead into a negative spiral of permanently less manual flight conduction. From a design perspective, intelligent (Geiselman, Johnson, & Buck, 2013) or adaptive automation (Parasuraman, 2000) could charge a pilot with several tasks to maintain his or her attention and situation awareness, thus keeping the pilot in the loop. Short-term effects can be avoiding automation surprises, which can lead to severe accidents, while a long-term effect can be the preservation of skills. However, all recommendations have to be considered for potential tradeoffs at the expense of safety by possible undesired side effects.

ACKNOWLEDGMENTS

This work was funded by the German Federal Ministry of Economics and Technology via the Project Management Agency for Aeronautics Research (DLR) within the Federal Aeronautical Research Program (LuFo IV-2). The authors acknowledge the support of Patrick Gontar and Ekkehart Schubert in both flight simulator experiments. The online supplemental material is available at http://hf.sagepub.com/supplemental.

KEY POINTS

- Commercial airline pilots showed different levels of practice and training according to their scheduled type of operation, short haul or long haul.
- Fleet (distinction between short haul and long haul) showed large significant effects on all analyzed manual flight performance indicators.
- Rank (distinction between captain and first officer) only showed little effects on manual flight performance.
- All results supported the conclusion that recent flight practice is a significantly stronger predictor for manual flying performance than the time period since flight school or even the total or type-specific flight experience.

REFERENCES

Airbus SAS. (2016). *Fly-by-wire*. Retrieved from http://www.airbus.com/innovation/proven-concepts/in-design/fly-by-wire/

Bissonnette, N., & Culet, J. P. (2013). *Airbus A318/A319/A320/A321/A330/A340* (Flight Standardization Board [FSB] Report). Renton, WA: FSB.

Brière, D., & Traverse, P. (1993). AIRBUS A320/A330/A340 electrical flight controls—A family of fault-tolerant systems. *Proceedings of the Twenty-Third International Symposium on Fault-Tolerant Computing*, 616–623. doi:10.1109/FTCS.1993.627364

Buckley, R., & Caple, J. (2009). *The theory & practice of training* (6th ed.). London: Kogan Page.

Bureau d'Enquêtes et d'Analyses pour la sécurité de l'aviation civile. (2012). *Final report on the accident on 1st June 2009 to the Airbus A330-203 registered F-GZCP operated by Air France flight AF 447 Rio de Janeiro - Paris*. Le Bourget: Author.

Casner, S. M., Geven, R. W., Recker, M. P., & Schooler, J. W. (2014). The retention of manual flying skills in the automated cockpit. *Human Factors*, *56*(8), 1506–1516. doi:10.1177/0018720814535628

Childs, J. M., & Spears, W. D. (1986). Flight-skill decay and recurrent training. *Perceptual and Motor Skills*, *62*(1), 235–242. doi:10.2466/pms.1986.62.1.235

Civil Aviation Authority. (2014). *Flight-crew human factors handbook: CAP 737*. Gatwick Airport South: Author.

Dismukes, K., Berman, B. A., & Loukopoulos, L. D. (2007). *The limits of expertise: Rethinking pilot error and the causes of airline accidents. Ashgate studies in human factors for flight operations*. Aldershot, UK: Ashgate.

Drappier, J. (2008). The erosion of manual flying skills in highly automated aircraft. *Flight Comment*, *2008*(2), 33–36.

Ebbatson, M. (2009). *The loss of manual flying skills in pilots of highly automated airliners* (Unpublished doctoral dissertation). Cranfield University, Cranfield.

European Aviation Safety Agency. (2013). *EASA automation policy: Bridging design and training principles*. Cologne: Author.

European Union. (2011). Commission regulation (EU) No 1178/2011 of 3 November 2011. *Official Journal of the European Union*, *54*.

Favre, C. (1994). Fly-by-wire for commercial aircraft: The Airbus experience. *International Journal of Control*, *59*(1), 139–157. doi:10.1080/00207179408923072

Federal Aviation Administration. (2013a). *Operational use of flight path management systems: Final report of the performance-based operations aviation rulemaking committee*. Washington, DC: Author.

Federal Aviation Administration. (2013b). *Safety alert for operators* (No. SAFO 13002). Washington, DC: Author.

Fleishman, E. A. (1966). Human abilities and the acquisition of skill: Comments on Professor Jones' paper. In E. A. Bilodeau (Ed.), *Acquisition of skill* (pp. 147–167). New York, NY: Academic Press.

Franks, P., Hay, S., & Mavin, T. J. (2014). Can competency-based training fly?: An overview of key issues for ab initio pilot training. *International Journal of Training Research*, *12*(2), 132–147. doi:10.5172/ijtr.2014.12.2.132

Geiselman, E. E., Johnson, C. M., & Buck, D. R. (2013). Flight deck automation: Invaluable collaborator or insidious enabler? *Ergonomics in Design: The Quarterly of Human Factors Applications*, *21*(3), 22–26. doi:10.1177/1064804613491268

Gillen, M. (2008). *Degradation of piloting skills* (Unpublished master's thesis). University of North Dakota, Grand Forks.

Haslbeck, A., Kirchner, P., Schubert, E., & Bengler, K. (2014). A flight simulator study to evaluate manual flying skills of airline pilots. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 58, pp. 11–15). Thousand Oaks, CA: SAGE.

Hubbard, D. C. (1987). Inadequacy of root mean square error as a performance measure. In R. S. Jensen (Ed.), *Proceedings of the Fourth International Symposium on Aviation Psychology* (pp. 698–704). Columbus, OH: Ohio State University.

International Air Transport Association. (2015). *Safety report 2014*. Montréal: Author.

Joint Aviation Authorities. (2004). *Airbus A320-A330-A340 cross crew qualification & mixed fleet flying*. Cologne: Author.

Lambregts, A. A., Nesemeier, G., Wilborn, J. E., & Newman, R. L. (2008). *Airplane upsets: Old problem, new issues*. AIAA Modeling and Simulation Technologies Conference and Exhibit No. AIAA 2008-6867, Honolulu, HI.

Langewiesche, W. (1944). *Stick and rudder: An explanation of the art of flying*. New York, NY: McGraw-Hill.

Learmount, D. (2011). *Industry sounds warnings on airline pilot skills*. Retrieved from https://www.flightglobal.com/news/articles/industry-sounds-warnings-on-airline-pilot-skills-352727/

Lyall, B., & Wickens, C. D. (2005). Mixed fleet flying between two commercial aircraft types: An empirical evaluation of the role of negative transfer. In *Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting* (pp. 45–48). Santa Monica, CA: Human Factors and Ergonomics Society.

McClernon, C. K., Miller, J. C., & Christensen, J. C. (2012). Variance as a method for objectively assessing pilot performance. In *Proceedings of the Human Factors and Ergonomics Society 56th Annual Meeting* (pp. 85–89). Santa Monica, CA: Human Factors and Ergonomics Society.

National Transportation Safety Board. (2014). *Crash of Asiana Flight 214 accident report summary: Descent below visual glidepath and impact with seawall*. Retrieved from http://www.ntsb.gov/news/events/2014/asiana214/abstract.html

Newman, R. L. (2012). *Thirty years of airline loss-of-control mishaps*. AIAA Modeling and Simulation Technologies Conference and Exhibit No. AIAA 2012-4495, Minneapolis, MN.

Parasuraman, R. (2000). Designing automation for human use: Empirical studies and quantitative models. *Ergonomics*, *43*(7), 931–951. doi:10.1080/001401300409125

Puentes, A. (2011). *The manual flight skill of airline pilots* (Unpublished master's thesis). San Jose State University, San Jose. Retrieved from http://scholarworks.sjsu.edu/etd_theses/4109

Rantanen, E. M., Johnson, N. R., & Talleur, D. A. (2004). The effectiveness of personal computer aviation training device, a flight training device, and an airplane in conducting instrument proficiency checks: Volume 2: Objective pilot performance measures (Final Technical Report AHFD-04-16/FAA-04-6). Oklahoma City, OK: FAA.

Sarter, N. B., & Woods, D. D. (1994). Pilot interaction with cockpit automation II: An experimental study of pilots' model and awareness of the flight management system. *The International Journal of Aviation Psychology*, *4*(1), 1–28. doi:10.1207/s15327108ijap0401_1

Savion-Lemieux, T., & Penhune, V. B. (2005). The effects of practice and delay on motor skill learning and retention. *Experimental Brain Research*, *161*(4), 423–431. doi:10.1007/s00221-004-2085-9

SKYbrary. (2016a). *Pilot handling skills*. Retrieved from http://www.skybrary.aero/index.php/Pilot_Handling_Skills

SKYbrary. (2016b). *Stabilised approach*. Retrieved from http://www.skybrary.aero/index.php/Stabilised_Approach

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston, MA: Pearson/Allyn & Bacon.

Vadrot, R., & Aubry, C. (1994). Cross crew qualification and mixed fleet flying: The Airbus family concept. *FAST*, *17*, 21–26.

Veillette, P. R. (1995). Differences in aircrew manual skills in automated and conventional flight decks. *Transportation Research Record*, *1480*, 43–50.

Wiener, E. L., & Curry, R. E. (1980). Flight-deck automation: Promises and problems. *Ergonomics*, *23*(10), 995–1011. doi:10.1080/00140138008924809

Andreas Haslbeck received a diploma degree in mechanical engineering from Technical University of Munich (TUM) in 2007. He is currently pursuing a doctorate in ergonomics. At the Institute of Ergonomics (TUM), where he was involved in experimental research using eye tracking and simulation, he conducted a large flight simulator study evaluating manual flying skills under the influence of practice and training. Besides eye tracking data, this experiment also recorded and analyzed workload, fatigue, and complacency effects.

Hans-Juergen Hoermann is an accredited aviation psychologist at the German Aerospace Center (DLR) since almost 30 years. He holds a PhD in applied psychology (1987) from the Free University in Berlin. His research focuses on cockpit human factors and crew resource management training. As a ground instructor, he is involved in pilot training.

Date received: October 1, 2015
Date accepted: February 29, 2016

Haslbeck, A., Hoermann, H.-J., & Gontar, P. (2017). Stirring the Pot: Comparing Outer and Inner Loop Control Behavior in Airline Pilots. (Manuscript under review)

This paper is currently under review and, therefore, only included in the examiners' copies of this dissertation.