**ORIGINAL ARTICLE**

# The influence of training level on manual flight in connection to performance, scan pattern, and task load

Maik Friedrich[1] · Seung Yong Lee[2] · Paul Bates[2] · Wayne Martin[3] · Anja K. Faulhaber[4]

## Abstract

This work focuses on the analysis of pilots' performance during manual flight operations in different stages of training and their influence on gaze strategy. The secure and safe operation of air traffic is highly dependent on the individual performances of the pilots. Before becoming a pilot, he/she has to acquire a broad set of skills by training to pass all the necessary qualification and licensing standards. A basic skill for every pilot is manual control operations, which is a closed-loop control process with several cross-coupled variables. Even with increased automation in the cockpit, the manual control operations are essential for every pilot as a last resort in the event of automation failure. A key element in the analysis of manual flight operations is the development over time in relation to performance and visual perception. An experiment with 28 participants (including 11 certified pilots) was conducted in a Boeing 737 simulator. For defined flight phases, the dynamic time warping method was applied to evaluate the performance for selected criteria, and eye-tracking methodology was utilized to analyze the gaze-pattern development. The manipulation of workload and individual experience influences the performance and the gaze pattern at the same time. Findings suggest that the increase of workload has an increased influence on pilots depending on the flight phase. Gaze patterns from experienced pilots provide insights into the training requirements of both novices and experts. The connection between workload, performance and gaze pattern is complex and needs to be analyzed under as many differing conditions. The results imply the necessity to evaluate manual flight operations with respect to more flight phases and a detailed selection of performance indications.

**Keywords** Task load · Pilot performance · Gaze pattern · Eye tracking · Pilot versus novices

## 1 Introduction

Becoming a pilot is a dream that many people pursue within their life time. The reasons for this are endless and include career opportunities, curiosity towards technical systems and flying as a recreational activity. Training, qualification, and licensing standards for pilots flying aircraft under the full range of operations require the demonstration of a broad set of skills. Currently, automation plays an important role in the operation of an aircraft and the importance of manual flying skills seems to be decreasing. However, in the case of malfunctions or total failure of the automation, the pilot is expected to perform manual control operations as a last resort (Landry 2014). Field and Harris (1998) describe manual control operations as a closed-loop control process with cross-coupled variables: pitch, roll, and yaw of the aircraft and altitude, lateral, and longitudinal change of the flight path. These variables have to be monitored and adjusted continuously while flying the aircraft.

Studies focusing on manual control operations with different stages of skill acquisition during normal and abnormal scenarios are limited. Research in this area is necessary to understand the development of manual control operations and to provide input for today's training of aircraft pilots. To support this research, we specify and elaborate the following research questions (RQ). Firstly, how does the level of training affect performance during manual flight operations in different task-load situations? Secondly, how does the level of training affect gaze strategies for manual flight

✉ Maik Friedrich
  maik.friedrich@dlr.de

1 German Aerospace Center, Lilienthalplatz 7, 38108 Braunschweig, NI, Germany

2 University of Southern Queensland, Toowoomba, Australia

3 Central Queensland University, Rockhampton, Australia

4 Institute of Flight Guidance, Braunschweig, TU, Germany

operations in high and low task-load situations? And thirdly, is there a connection between workload, performance, and gaze strategies?

## 1.1 Influences on manual flight and performance

Manual control of aircraft is defined by the Federal Aviation Administration (2017) as "managing the flight path through manual control of pitch, bank, yaw and/or thrust. Manual flight operations may be conducted with or without a flight director and require foundational knowledge and skill proficiency". Childs and Spears (1986) state that response execution, perception, and cognitive workload are considered as an information basis when analyzing the manual control of an aircraft. Cognitive workload is subject to the impact of stressors, e.g., complexity or time pressure, on the person, depending on their individual constitution, resources, ability, etc. (Edwards 2013). As external equivalent, task load is the demand by the task or the system. Childs and Spears (1986) also state that the training for physically manipulating the primary flight controls requires complex psychomotor and cognitive skills. Pilots need to acquire these skills before being introduced to automation (Billings 1997). Evidently, the overuse of automation results in manual flying skill degeneration (Landry 2014) and a negative complacency effect (Parasuraman and Manzey 2010).

With the beginning of increased cockpit automation in the early 1980s, the demand on the crew was altered and several studies have focused on the exposure of pilots to this changed environment (Parasuraman and Riley 1997; Rigner and Dekker 2000). Other research has focused on the skill sets required to operate the new technologies, and issues accompanying the increase in automation (e.g., Billings 1997; Wiener et al. 1999). Only a few studies have looked at different stages of manual operation of aircraft in relation to performance.

Veillette (1995) used a one-factor experiment with conventional flight deck pilots and automated flight deck pilots. The participants were classified based on the degree of automation that their typical aircraft possessed. Performance was measured as deviation of the instantaneous pitch-and-bank angle during a time interval. The results revealed that pilots flying more-automated aircraft showed statistically significant differences in manual control in comparison to pilots flying less-automated aircraft, especially in abnormal operations. Veillette (1995) also raised the issue that pilots who start their training in a highly automated flight deck might not even have the chance to develop extensive manual flying skills (see also Billings 1997; Ebbatson 2009).

Furthermore, the following studies concentrated only on the performance as a main indicator for pilot behavior. Ebbatson et al. (2008) analyzed performance increase in manual flight with training. Seventeen male cadet pilots repeated a scenario in a fixed-base JAA-approved flight training device simulating a Boeing 737NG series aircraft. Performance was measured by analyzing the deviation from the optimum flight path while flying an instrument landing system (ILS) approach. Even though no significant performance differences between normal and single-engine approaches could be identified, control input frequency measures revealed different control strategies in pitch and yaw axis. The results demonstrated the need to fully evaluate pilots' behavior during each flight phase.

Haslbeck et al. (2014) assessed raw-data-based flight performance of airline pilots with different levels of experience in a 45-min ILS approach scenario where the participants' performance was measured as deviation from the ideal glide slope and localizer. The results showed that in accordance with other previous research (e.g.,Taylor et al. 2007; Tsang 2003), a significant influence of practice and training levels confounded with the pilots' age and experience on manual flight operations was observed, necessitating the need for further studies with the inclusion of medium-level groups (Haslbeck et al. 2014). The studies above focused on manual aircraft control by analyzing the performance in terms of how the pilots responded in execution. Following Childs and Spears (1986), this should be supported by the analysis of the cognitive system to gain a more comprehensive understanding of the manual task performance.

## 1.2 Cognitive systems and gaze pattern

"A complex socio-technical system such as an airliner cockpit, accordingly, is a joint cognitive system that consists of several component cognitive systems (pilot, copilot, autopilot)" (Blomberg 2011). A common way to study cognitive systems is via eye movements, e.g., by means of so-called scanpaths. A scanpath is the description of glances with geometric vectors (Kang and Landry 2015) in relation to the task environment. By assigning each possible source of information a single Area of Interest (AoI), the order in which the information is gathered is described as a scanpath with symbol representation (Holmqvist et al. 2011) or gaze pattern (Dorr et al. 2010). When dealing with their analysis, the issue of connecting gaze patterns to specific cognitive processes is still unresolved (Holmqvist et al. 2011). Several studies have failed to identify a cognitive process and its direct mapping to a scanpath or gaze pattern (Goldberg and Kotval 1999; Salvucci 1999). Even though defining a direct relationship between a cognitive process and a gaze pattern seems unlikely, using an experimental design and a well-defined task environment can reduce the error in interpreting the scanpath.

In the task environment of the cockpit, an efficient gaze pattern is essential to keep the pilot updated about the system's states, which is the basis for safe and efficient control

of the aircraft. The main source of information for manual control operations is the primary flight display (PFD) that contains the attitude indicator (PFD_ATT), altitude tape (PFD-ALT), heading indicator (PFD_HDG), airspeed tape (PFD_AIS), and vertical speed indicator (PFD_VES). The Federal Aviation Administration (2012) describes the radial cross-checking technique as a visual scanning strategy for covering the PFD adequately. The radial cross-check starts at the PFD_ATT and returns to it after checking one other source of information. In parallel, the pilot performs control inputs and uses the checks to monitor their effect (Federal Aviation Administration 2012). Continuous monitoring provides input in a top-down process (Holmqvist et al. 2011; Papenfuss and Friedrich 2016; Schütz et al. 2011) in contrast to bottom-up or event-driven monitoring (Schütz et al. 2011). Casner e al. (2014) found that participants' instrument scanning and manual control skills were mostly intact even when the participants reported that they were infrequently practiced with increased automation within the cockpit. However, when the participants were asked to manually perform the cognitive tasks needed for manual flight with decreased automation (e.g., tracking the aircraft's position without the use of a map display), more frequent and significant problems were observed.

Haslbeck and Zhang (2017) argue that gaze-based metrics (e.g., percentage time on AoI or frequency on AoI) are inadequate for understanding pilots' scanning behavior and how the information-gathering process is performed under different circumstances. They prefer a comparison of gaze-pattern sequences, even though the process is challenging (Foulsham et al. 2012), especially with increasing length of sequences (Kang and Landry 2015). Simon et al. (1993) propose a reduction of complexity by breaking the length of the gaze pattern before analysis. Methods for comparing different scanpaths can also be used to compare different gaze patterns, e.g., generating an average scanpath (Hembrooke et al. 2006; Holšánová 2008) or analyzing the most frequent gaze patterns for each participant (Haslbeck and Zhang 2017).

However, an underrepresented issue in the literature is the comparison between different groups (Feusner and Lukoff 2008) and the interpretation of the results without any additional information. The frequency of gaze patterns in which the same information is acquired can differ between single individuals and also between groups. Therefore, the selection of a gaze pattern that is worthwhile for the task could provide utility as a baseline comparator. In the driving domain, Underwood et al. (2003) defined a gaze pattern with the length of three information sources as the shortest length of a scanning strategy and compared experts against novices for the driving task. Within the expert group, better scanning strategies that focused more on the traffic far view were identified and therefore allowed for better planning.

Haslbeck and Zhang (2017) analyzed the scanning strategies for manual flight operations of commercial airline pilots by accounting for their level of practice in relation to their normal flight operations (short-haul or long-haul). Each participant flew an approximately 8-minute manual approach that can be categorized as a high workload situation (see e.g., Lee and Liu 2003). The autopilot, flight director and auto throttle were disabled. The short-haul pilots accomplished the manual flight with less deviation from the expected glide slope. An analysis of the gaze patterns identified four main gaze strategies that were used by both groups but with unequal distribution. Long-haul participants tended to adhere longer to the PFD_ATT, which is the main source of information if the flight director is enabled. The results indicate that the total number of flight hours as an indication of expertise correlates insufficiently with practice or skill.

The above cited literature shows that research on manual flight operations has a long history with a relatively small number of new publications per year. This is due to fact that are almost no updates on the concept of manual control and that there are high costs for a realistic simulator and with licensed airline pilots. This will probably continue beyond the point of full automation, and makes this publication even more necessary.

The existing research on experts performing manual flight operations and their skill degeneration with increased automation is a well-established part of the literature, as are the consequences for safety due to degraded performance (Haslbeck and Hoermann 2016; Landry 2014). However, the question remains as to whether this behavior can be observed in comparison to novices as shown in other domains, e.g., driving (Underwood et al. 2003) and in relation to perception, task load, and response execution. An answer to this question would allow insights into the general skill acquisition and transferable knowledge that could be provided early in the teaching process.

## 2 Objectives of this paper

Current research on manual flight operations indicates a strong connection between performance, perception (Haslbeck and Zhang 2017), and task load. To help explain the connections, Fig. 1 presents the model for manual flight operation by Wickens (1999) extended with perception and the influence of workload (adapted from McRuer 1982). As mentioned above, PFD_ATT, PFD-ALT, PFD_HDG, PFD_AIS, and PFD_VES are the essential sources of information for the task of manual flight operations. Pilot perception is connected to gaze strategy that assesses the order in which those areas are attended to. The gathered information influences the pilot's actions in keeping the aircraft on the flight path and is then transformed into attitude commands. The
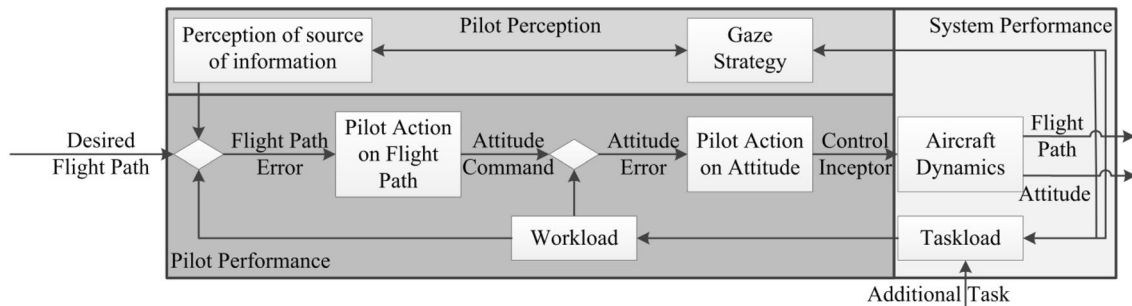
**Fig. 1** The model of manual flight operation in relation to perception and workload (adapted from McRuer 1982; Wickens 1999)

task load is derived from the system performance and additional tasks. This influences the workload of the pilot who has to maintain the path and input commands.

Using Fig. 1 as an initial step for analyzing manual flight operations with a perspective on performance, perception, and task load allows for a first approach to a comprehensive view on the interactions between these factors. The connections in Fig. 1 can help to answer important research questions addressed in this paper. In the previous sections, we identified two gaps in the existing research on manual flight operations. The first gap concerns different stages of training in combination with performance, perception, and the interconnectivity between these factors in dependence of high workload phases such as landing or takeoff. The second gap also relates to different stages of training but with a focus on gaze strategies for the task itself and their development throughout training. Therefore, we focus on the analysis of (1) pilots' performance during manual flight operations in different stages of training and situations and (2) examine the influence of training on gaze strategy.

The development of future cockpits and training strategies depends on validated models describing the connection between workload, performance, and perception. To support the model in Fig. 1, we address the cognitive system of the airline pilot. As Hollnagel (1993) and other cognitive scientists repeatedly point out, cognition can only be studied in a realistic scenario; this study was therefore performed using a real-time simulation as experimental technique. This allows for a realistic manual control of an aircraft by airline pilots with different levels of training in normal and abnormal scenarios.

# 3 Methods

## 3.1 Sample

The sample consists of a total of 28 participants. These were recruited in three groups based on their level of training—novices, experts, and pilots. The novices (N) were first-semester Bachelor of Aviation students from the University of Southern Queensland with a minimum flight experience of 10 h, either in a simulator or on aircraft. The experts (E) were third-semester students with a minimum flight experience of 160 h, either in a simulator or on aircraft, but who did not yet have a commercial pilot's license. The requirements for pilots (P) were a valid pilot's license, a type rating for the Boeing aircraft family, and active flying for an airline either as first officer or captain.

Table 1 presents the demographic data for each group separately. Experience of the participants was evaluated by means of the self-rated question "How would you rate your experience in flying an aircraft on a scale from 1 (Novices) to 10 (Expert)?" The participants obtained no monetary compensation or travel expenses. This research complied with the National Statement on Ethical Conduct in Human Research (2007) and was approved by the Institutional Review Board at the University of Southern Queensland. Informed consent was obtained from each participant.

## 3.2 Design

For this study, the manual flight operation task for a Boeing 737 was simulated in a high-fidelity setting using a

| Table 1 Demographic data for participants | Group | Number | Age (M/SD/Min/Max) | Gender (F/M) | Self-rated experience 1—novice to 10—expert (M/SD/Min/Max) |
|---|---|---|---|---|---|
| | Novices (N) | 10 | 24.3/4.8/19/33 | 3/7 | 2.4/1.5/1/5 |
| | Experts (E) | 7 | 24.4/3.4/18/29 | 1/6 | 6.9/1.1/5/8 |
| | Pilots (P) | 11 | 33.7/9.6/23/56 | 3/9 | 8.1/1.2/6/10 |

**Table 2** Steps within each scenario for the participant, with changes to altitude and heading

| Step ID | Step description | Altitude (feet) | Heading (degree) | Flight phases |
|---|---|---|---|---|
| 0 | Start of the experiment at initial position and altitude | 1500 | 017 | |
| 1 | Climb to 2500 ft at 500 feet per minute | Climbing | 017 | Climb |
| 2 | Once established in the climb, right turn heading 090 | Climbing | 017 | Climb; Change-Course-1 |
| 3 | Once levelled off at 2500 ft, left turn heading 020 | 2500 | 090 | Change-Course-2 |
| 4 | Fly straight ahead for 90 s | 2500 | 020 | Straight ahead |
| 5 | After 90 s, turn left heading 290 | 2500 | 020 | Change-Course-3 |
| 6 | Three miles before centerline, turn left heading 230 for localizer interception | 2500 | 290 | Change-Course-4 |
| 7 | Localizer alive, turn left heading 196 | 2500 | 230 | Change-Course-5 |
| 8 | Glideslope alive | Descending | 196 | Landing |
| 9 | Leave auto throttle on for landing (non-standard) and select reverse thrust after landing | Descending | 196 | Landing |

between-subject design $3 \times 2$ with repeated measures with the first factor investigating the dependent variables performance, gaze strategy, and workload and the second factor regarding (a) varying manual flight operations (straight ahead, change altitude, change heading) and (b) different task-load conditions (high vs. low task load). The scenario was divided into abnormal and normal scenarios. The varying of manual flight operations was equal for both scenarios. In the abnormal scenario, the alert for an incorrect flaps position was played throughout the scenario in an infinite loop to increase the task load.

The scenarios used in this study represent an approach to Brisbane airport and end with touchdown on the runway. Both scenarios are equal in the order of steps each participant had to complete (Table 2). They start at the same initial position (Lat S27.39.01; Lon E153.16.8) and altitude (1500ft) with heading 017. The planned duration for each scenario was approximately 13 min. As the focus of this study was on manual flight operations, auto throttle and flight director were engaged throughout the scenarios. Additionally, inputs to speed, flaps, landing gear, and flight management system were performed by the experimenter.

The flight phases (Table 2) were designed to cover a wide range of possible situations for manual flight operations and to influence each other as little as possible within a realistic setting. Figure 2 presents the vertical profile of the simulator flying the scenario with the autopilot engaged. Except for Climb and Change-Course-1, all flight phases do not overlap in time. The scenarios were previously tested for feasibility and realism by a domain expert.

### 3.3 Apparatus

Figure 3 shows the experimental setup, a fixed-base, fully enclosed Flight Training Device at the University of Southern Queensland Springfield campus. The Flight Training
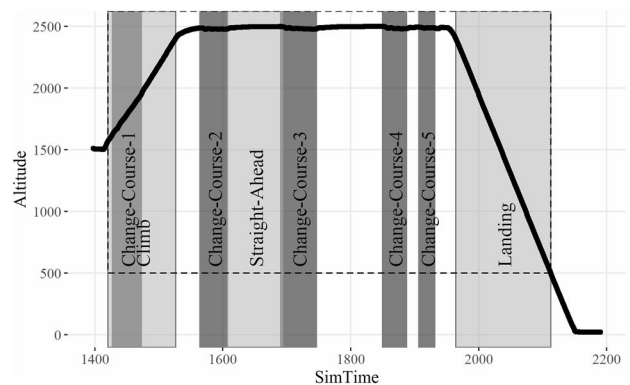


**Fig. 2** The vertical profile (black line) of the autopilot flying the scenario in simulation time (SimTime in seconds). The evaluated flight phases are marked by the boxes. The data used for analysis are marked with a dashed line
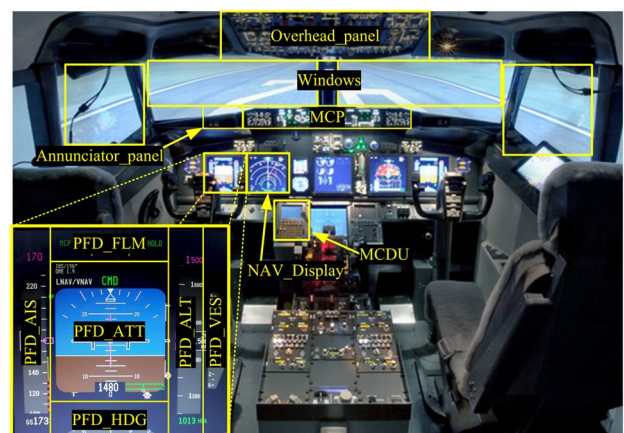


**Fig. 3** Cockpit of the 737, PS4.5 flight simulator with the defined Areas of Interest

Device PS4.5 was produced by Pacific Simulators. The participants were seated in the left cockpit seat, wearing eye-tracking glasses. The simulator has realistic flight controls, rudder pedals, steering tillers, and a flight management system. Manual control operations were performed with the yokes with stick shakers. Radio communication was not necessary.

### 3.3.1 Workload

In this study, the NASA-TLX (Hart and Staveland 1988) was used as subjective offline workload measurement. The NASA-TLX is a well-established tool with a high reliability (retest reliability 0.52–0.75) and validity, widely accepted in the research community and often applied by Eurocontrol (EUROCONTROL 2012). The NASA-TLX consists of six dimensions: mental demand, physical demand, temporal demand, performance, effort, and frustration level. The relation between the dimensions is evaluated in a pre-run evaluation and used in all runs for the calculation of the NASA-TLX score.

### 3.3.2 Performance measurements.

The performance of the participants was evaluated in connection to the recorded aircraft flight condition. According to Fig. 1, the deviation between planned and flown flight path within each flight phase was interpreted as performance measurement. The planned flight path or baseline was the path of the autopilot (Fig. 2) flying the scenario defined in Table 2. For a realistic baseline of the optimal flight conditions within the flight phases, the autopilot was used to control the aircraft instead of a linear description of the expected values. The aircraft flight condition recorded during the autopilot run was used to determine the performance of each participant.

Table 3 presents the aircraft flight parameters selected as a measurement to describe the flight path. The measurements were recorded with 1 Hz by the simulator. Based on the assumption that the baseline represents the optimal flight path, the deviations from each of the measurements in Table 3 are interpreted as an indication of performance.

**Table 3** Measurements of aircraft flight parameters recorded by the simulator

| Measurement | Description |
| --- | --- |
| Airspeed_indicated | The indicated airspeed of the aircraft |
| Altitude | The altitude of the aircraft |
| Vertical_speed | The vertical speed of the aircraft |

### 3.3.3 Eye tracking

The SMI Eye-Tracking Glasses 2 Wireless (SensoMotoric Instruments, Germany) were used to collect eye-tracking data. This is a head-mounted system with a sampling rate of 60 Hz that collects eye direction, gaze, and head position and calculates a position on an AoI (Fig. 3) in a pre-defined environment. The accuracy within the gaze-tracking range of 60° vertical and 80° horizontal is 0.5°. The iView System uses a one-point calibration to ensure the correctness of the eye-tracking process. The eye data and the simulator data were recorded with a synchronized time stamp and analyzed with the DLR software EyeTrackingAnalyser (Friedrich et al. 2016). The gaze on a source of information was defined as raw eye data within a defined AoI. The AoIs are defined for the complete cockpit, to measure as much gaze as possible. The AoIs relevant for the task were all on the PFD. The order in which the sources of information were attended to is interpreted as gaze strategy and set in relation to performance and workload.

## 3.4 Procedure

The experiment had a total runtime of approximately 1 h. Upon arrival, the participants were briefed on the procedure of the experiment and it was emphasized that they could leave at any given moment. All participants signed written informed consent for the use of their data. The participants were advised to perform the task of the pilot flying and use only manual flight operations. They were instructed that all clearances had already been given by ATC. The participants completed a demographic questionnaire and the width subscale questionnaire of the NASA-TLX. The participants signed a consent form in which the anonymity of data collection and analysis was guaranteed. The participants then had a training session of 15 min to familiarize themselves with the simulator. Each participant was allowed to finish the training early and only one used the full amount of time.

After training, the eye-tracking glasses were calibrated with a one-point method and the first scenario run was started. The order of scenarios, first normal and second abnormal or the other way around, was randomly assigned to each participant. After the first run, the participants had to complete the NASA-TLX followed by a 5-min break. After the break, the second scenario run was started and finished with the second completion of the NASA-TLX. At the end, the participants were debriefed.

## 4 Results

To increase comparability in total duration between the manual flight phases, the data used for analysis starts with the first climb maneuver and ends when the altitude drops

below 500ft (see dashed line in Fig. 2). Therefore, the touch-down is not included in the data analysis. The duration of the selected maneuvers is comparable to each other; this means that the scenario was comparable in total for each participant.

Due to the fixed sequence of flight phases throughout the scenarios, the initial analysis focused on their comparability. All participants performed manual flight operations following the flight director's instruction during the complete scenario and therefore also throughout the defined flight phases. Figure 4 shows boxplots per N, E, and P groups and flight phase, separated into normal (left) and abnormal (right) scenario. Except for minor outliers, none of the flight phases in each scenario showed a significant difference in average duration between the groups. The sequence effect of the scenarios was accounted for in each analysis in the following sections.

## 4.1 Effects on workload

The average NASA-TLX score increases between the normal ($M = 7.6$; $SD = 2.77$) and the abnormal condition ($M = 9.48$; $SD = 3.1$). Figure 5 presents the results for the NASA-TLX scores for each individual group. The inference analysis of the workload for each group showed a significant increase for the N group (F $(1.26) = 14.612$, $p < 0.001$). The post hoc analysis was performed with the Tukey range test. Therefore, besides a general increase in workload, the results

show a significant increase in workload for the lower level of training.

## 4.2 Performance

The performance analysis concentrated on the deviation between the recorded aircraft flight parameters (Table 3) of the baseline (autopilot) and the participant. Each participant completed each run with a landing on the runway. For each flight phase, the selected measurements were compared against the baseline of that flight phase. Because of the duration difference between the baseline and the participants, the dynamic time warping (DTW) (Salvador and Chan 2007) method was used to calculate the minimal distance between two time series independent from their duration.
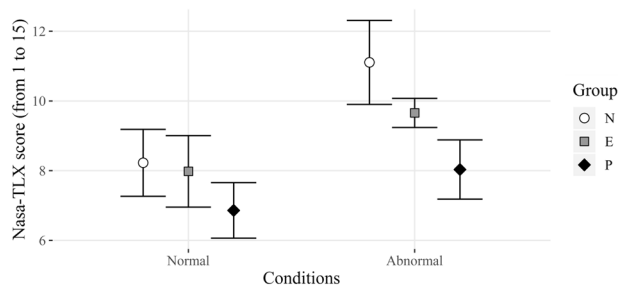


**Fig. 5** NASA-TLX scores separated per group and condition. The error bars represent the standard error
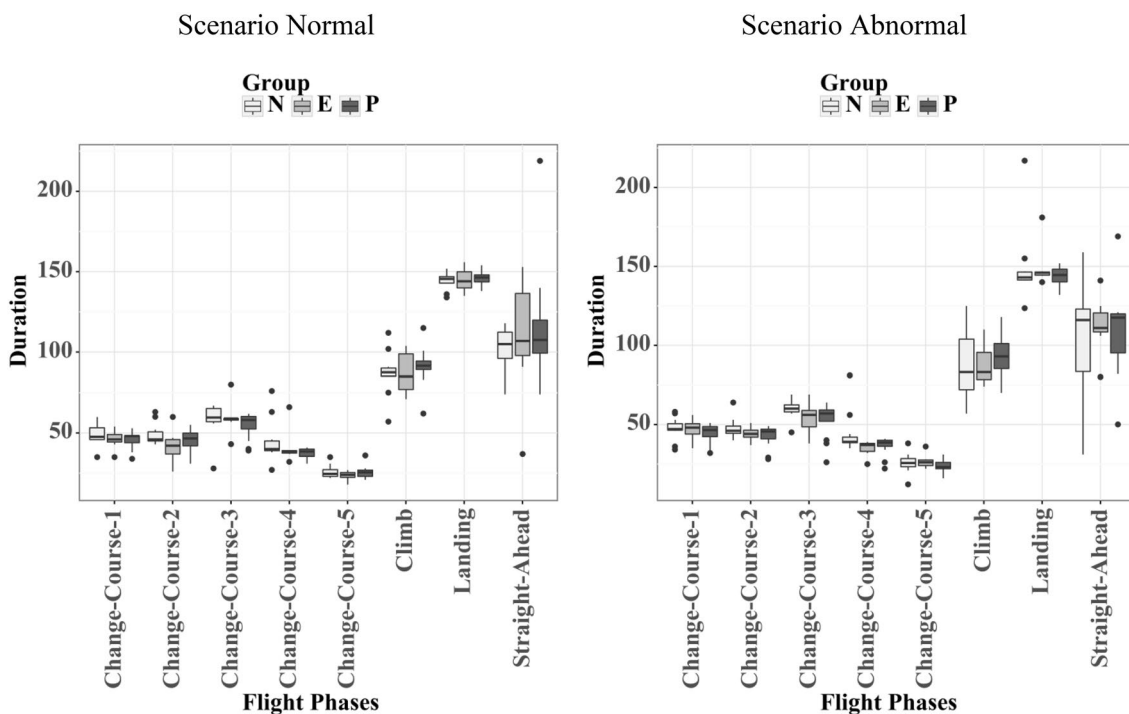


**Fig. 4** Boxplots of the duration of each flight phase for the normal (left) and the abnormal (right) scenario

The minimal distance represents a sum of absolute differences between the analyzed time series and was used as an indicator of how equal both time series are.

Figure 6 presents the results of the DTW for the parameter altitude during the Change-Course-2 flight phase. The path between them represents the minimal mapping of a time series with unequal duration. During Change-Course-2, the altitude is set to 2500ft and the task is to change heading from 017 to 090. The triangles represent the baseline performance of the autopilot which is a realistic estimation of the aircraft flight condition within this flight phase. The
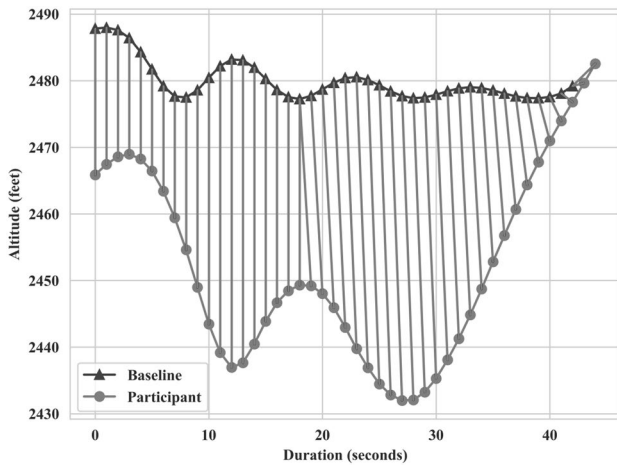


**Fig. 6** Example of the DTW analysis for the parameter altitude during a Change-Course-2 performed by participant 13, with the path between them

circles represent the pilot performance (participant 13, normal condition, novice group) in the same flight phase. For the example in Fig. 6, the deviation from the baseline (DFB) is 1262.22. The DFB represents an indicator of performance differences between participants, where a greater DFB can be considered as lower performance. The DTW method was used to calculate the DFB for all measurements in Table 3 separately for each flight phase and participant.

For the following results, the DFB depending on the defined parameters for the N, E, and P groups were analyzed separately for flight phases and scenarios. Due to their similarity in altitude, the DFB results for Change-Course-2, Change-Course-3, Change-Course-4, and Change-Course-5 were summarized to Change-Course-2–5 for the following analysis. Table 4 presents the average DFB for each group in relation to the scenario for each flight phase and parameter. The DFB for each flight phase is represented by the parameters Airspeed, Altitude, Vertical_speed.

Table 5 presents the results of a multiple-factor ANOVA. The results are separated by main effects for group (N, E, and P) and condition (Normal, Abnormal). The column post hoc indicates which DFB is significantly high between the groups or conditions. All $p$ values marked with a star are statistically significant at the $p < 0.05$ significance level. No interaction effects between group and condition were statistically significant. The post hoc analysis was performed with the Tukey range test.

An analysis of the results in Table 4 shows that independent from the scenario, the pilot group always has the lowest DFB, except for the parameter Altitude during

**Table 4** The average mean DFB for each group per flight phase, separated by scenario

| Flight Phase | Parameter | Condition | | | | | |
|---|---|---|---|---|---|---|---|
| | | Normal | | | Abnormal | | |
| | | novices (M/SD) | Experts (M/SD) | Pilots (M/SD) | Novices (M/SD) | Experts (M/SD) | Pilots (M/SD) |
| Change-Course-1 | Airspeed | 34.1/14.6 | 32.7/15.4 | 26.4/20.4 | 39.1/20.9 | 35/14.9 | 17.8/12.2 |
| | Altitude | 2646.3/2176.0 | 5037.1/3204.8 | 1599.4/1362.1 | 6306.6/6822.0 | 1339.1/723.9 | 3847.7/5095.8 |
| | Vertical_speed | 6927/3699.7 | 9197.7/7524.3 | 6288.2/3713.1 | 10,796.3/7494.7 | 8862.7/3341.8 | 4104.5/2212.6 |
| Change-Course-2–5 | Airspeed | 64.77/34.61 | 55.47/33 2 | 34.12/16.5 | 64 5/33.77 | 42.71/11.6 | 33.11/11.54 |
| | Altitude | 1575.6/1265.29 | 2106.1/3581.36 | 879.29/643.18 | 1889.7/1589.89 | 1038.94/431 7 | 701.88/470.62 |
| | Vertical_speed | 7448.4/2393 | 6106.3/4335.6 | 3686.6/2071.3 | 7400.4/2100.1 | 6095.9/1598.3 | 3437.4/1522 |
| Climb | Airspeed | 70.6/32 | 74.4/28.8 | 56.7/18.7 | 74.6/64.2 | 63.1/12.6 | 45.2/13.8 |
| | Altitude | 354.7/105.3 | 455.2/159 | 329.2/51.4 | 444/226.7 | 347.8/42.1 | 314.2/63.5 |
| | Vertical_speed | 15,508.6/7018 | 20,289/11,507.6 | 12,945.7/5344.9 | 20,086/12,157.7 | 16,351.9/4731 | 10,848.8/5422 |
| Landing | Airspeed | 164.9/64.8 | 233.6/183.6 | 153.3/50.6 | 207/172.4 | 513.4/802.7 | 144.2/68.3 |
| | Altitude | 576.8/54.2 | 645.6/171.7 | 562.1/52.0 | 585.0/123.5 | 587.8/85.5 | 577.5/51.3 |
| | Vertical_speed | 19,080.1/8809.6 | 30,248.9/23,333.7 | 15,588.5/5664.1 | 22,550/16,287.6 | 19,282./10,672.6 | 16,042/8430.4 |
| Straight-Ahead | Airspeed | 95.2/ 60.5 | 131.2/ 68.8 | 86.1/38.9 | 104.6/59.4 | 87.9/42.6 | 83.0/54.7 |
| | Altitude | 3920.2/3729.3 | 3306.1/1911.5 | 2268.5/1272.4 | 3892.0/3051.9 | 2896.9/2573.6 | 1860.4/1076.3 |
| | Vertical | 12,606.3/7541.3 | 14,629.3/5805.0 | 11,907.4/5265.9 | 12,781.3/7781.5 | 11,586.9/7030.7 | 9541.8/5163.4 |

**Table 5** The $F$ Test results for the deviation between participant and baseline separated by Group (N, E, and P) and Condition (Normal, Abnormal) extended with the post hoc analysis

| Flight phase | Parameter | Group | | |
|---|---|---|---|---|
| | | $F$ tests with $p$ value | Condition separated Post hoc | |
| | | | Normal | Abnormal |
| Change-course-1 | Airspeed | $F(2,52) = 5.34, p = 0.007*$ | | E > P<br>N > P |
| | Altitude | $F(2,52) = 1.38, p = 0.25$ | | |
| | Vertical _speed | $F(2, 52) = 6.50, p = 0.003*$ | | N > P |
| Change-course-2–5 | Airspeed | $F(2,222) = 12.03, p < 0.001*$ | N > P | N > P |
| | Altitude | $F(2, 222) = 10.32, p < 0.001*$ | | N > P |
| | Vertical _speed | $F(2, 222) = 35,19, p < 0.001*$ | E > P<br>N > P | E > P<br>N > P |
| Climb | Airspeed | $F(2,52) = 5.94, p = 0.004*$ | [1] | [1] |
| | Altitude | $F(2,52) = 2.60, p = 0.08$ | E > P | |
| | Vertical _speed | $F(2, 52) = 4.47, p = 0.017*$ | | (N > P)[2] |
| Landing | Airspeed | $F(2, 52) = 1.24, p = 0.29$ | | |
| | Altitude | $F(2, 52) = 0.48, p = 0.61$ | | |
| | Vertical _speed | $F(2, 52) = 1.64, p = 0.20$ | | |
| Straight-ahead | Airspeed | $F(2,52) = 0.791, p = 0.458$ | | |
| | Altitude | $F(2,52) = 3.52, p = 0.036*$ | [1] | [1] |
| | Vertical _speed | $F(2, 52) = 0.983, p = 0.38$ | | |

[1]None of the post hoc tests revealed significant differences; [2]The p value was above 0.05

Change-Course-1. The expert group is, on average, better than the novice group in 4 cases for the normal scenario and in 13 cases for the abnormal scenario. The results in Table 5 show that the scenarios have no effect on the DFB between groups. The post hoc test that was separated for each scenario revealed a significant difference between novice and pilot groups in two cases for the normal and in six cases for the abnormal scenario. All significant effects showed that the pilot group performed better than the novice group that the pilot group performed better than the expert group, or both. No significant difference between expert and novice groups was identified.
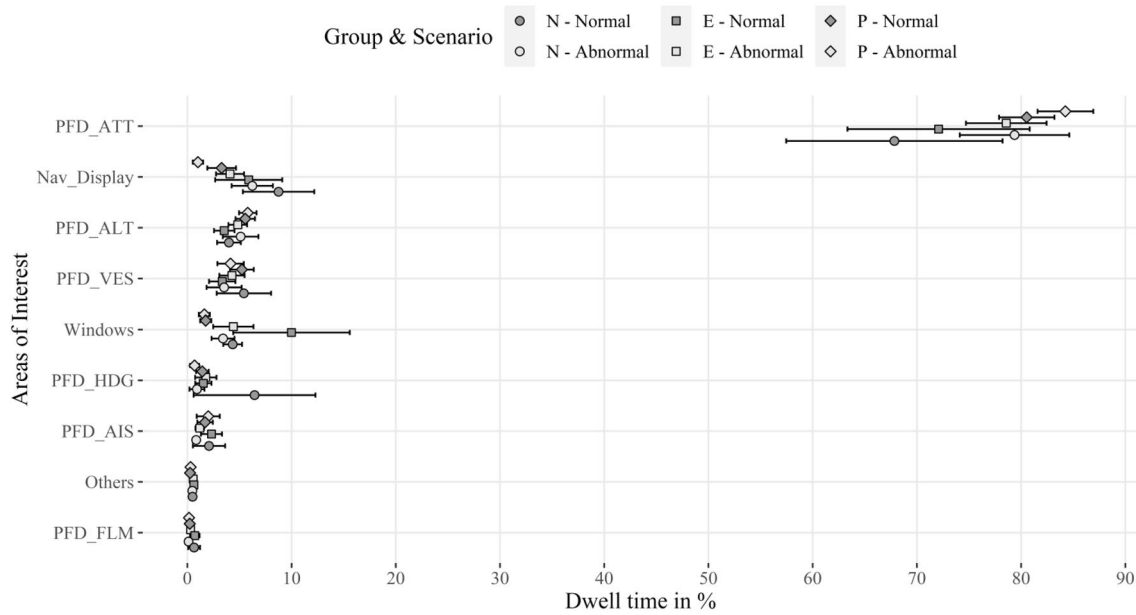
## 4.3 Effects on gaze strategy

The identification of the gaze strategy was based on the order in which the sources of information were attended to by the participant's eye movement. A valid eye movement is defined as a gaze vector that is inside of a defined AoI (Fig. 3). The percentage of a valid eye movement is calculated by the amount of valid eye movements divided by the total number of recorded eye movements that included, e.g., blinks, measurement problems, and gaze vectors outside of an AoI. The average of valid eye movements reached $M = 92.2\%$ (SD = 5.9), after removing one novice and three

runs from two pilots due to their percentage of valid eye movement below 75%. These three participants wore glasses together with the eye-tracking device, which resulted in a decrease in the quality of the recorded data. As for the performance analysis, the results for the flight phase Change-Course-2 to Change-Course-5 were combined.

As a first indicator for gaze strategies, an analysis of the average dwell times was conducted to identify AoIs with increased attention. Figure 7 shows all dwell times in percentage separated by group and condition averaged for the complete run. The summarized percentage for the AoIs MCP, Annunciator_Panel, MCDU, and Overhead_panel was below 0.5 percent and therefore combined to others. Independent from group and scenario and in accordance with the expected behavior of the manual flight operations, participants never spent less than 67 percent on average of their dwell times on the PFD_ATT. The pilot group in the abnormal condition even reached an average of $M = 84.26$ (SD = 8.45).

The second part of the gaze-strategy analysis concentrated on the differences in dwell-time percentages for each flight phase. Table 6 presents the results of the $F$ Test regarding the deviations in dwell times per AoI and flight phase. The results show that no significant differences were identified for Change-Course-1 and only $p$ values between 0.1

gaze-pattern behavior independent from the task load for the pilot group. The AoI PDF_ATT is represented in each of the highest gaze patterns for the pilot group. Gaze patterns involving only PFD-ATT and PFD-VES always reached the highest position in the P group, with the exception of the abnormal condition during Straight-Ahead and both conditions during Landing. When comparing gaze patterns involving only PFD-ATT and PFD-ALT versus gaze patterns involving only PFD-ATT and PFD-VES, it can be shown that the N group is using the first gaze pattern as often as E or P in contrast to the second gaze pattern that is never or rarely used by the N group. The N group used gaze patterns involving PFD-ATT and Nav Display with a higher dwell-time percentage than the E and P groups in all cases.

In comparison to the highest-ranked gaze pattern for the P group (Fig. 8), Table 7 presents the highest-ranked gaze patterns for the N group. By ignoring the order, the first two gaze patterns for all five flight phases are equal for normal and abnormal conditions. Four flight phases have the same gaze pattern in the first position for normal and abnormal conditions. A difference between normal and abnormal conditions is the increase in percentage of the first gaze pattern.

## 5 Discussion

The study presents interesting indications with respect to the influence of task load on manual flight operations with different stages of pilot training. Even though the sample size of each group is relatively low, the insights from the group comparison present a valuable method for identifying different stages of training. The increased task load influenced performance and gaze strategy in different flight phases and therefore presents feedback for their interconnection. Performance (DTW) and perception (eye tracking) were measured with objective methods. Within this study only workload (NASA-TLX) was measured with a subjective method. In comparison to state-of-the-art indicators for task load, as e.g. heart rate (Vanderhaegen et al. 2020), the NASA-TLX was selected as a simple and well established method in the aviation community to measure workload in different scenarios.

Inducing different levels of task-loads is complex because manual flight operations require visual and manual attention. Due to the eye-tracking measurement and the realistic task environment, we decided to use noise as an additional audio input (Van Gemmert and Van Galen 1997) instead of a secondary task (Knowles 1963) to increase the task load. On the other hand, the complexity of the task had to be reduced by engaging auto throttle and flight director throughout the experimental runs to generate a task all groups could perform and therefore allow for comparability between them. However, all participants were able to land the aircraft even

in the abnormal condition; the increase in task load did not overload any participant and did not impair the results.

As presented in the model by McRuer (1982); (extended with Wickens 1999 in Fig. 1), the workload influences manual flight operations on the flight path and on the attitude actions performed by the pilot. The NASA-TLX results indicate that the effect of increased workload was higher on the groups with less experience. This is also supported by the fact that the performance difference between the groups is increased in the abnormal scenario. A reason could be that group N simply does not look at task specific AoIs. However, the results on dwell time percentage per scenario indicate a tendency of group N to focus within the abnormal scenario even more on AoIs, similar to the ones group P is looking at. The analysis of the gaze pattern then reveals (for details see RQ three) that even with similar dwell time percentage, the N group uses different gaze patterns, and especially incorporates the Nav_Display. Considering that the flight path did have a minor role in the experimental condition, the influence of the gaze pattern on performance indicates a direct connection between gaze strategy and attitude action in the model.

The experimental focus was set on the three RQ concerning level of training, effects on gaze strategy, and their relation to task load. The results concerning the first RQ on the influence of the level of training on performance in high and low task-load situations differ depending on the flight phase. There is almost no significant influence of task load on performance during the flight phases climb, straight-ahead, and landing. These results stand in contrast to the flight phases Change-Course 1 to 5 that show a significant decrease in performance for the abnormal condition. Therefore, in accordance with previous research (e.g., Haslbeck et al. 2014; Taylor et al. 2007), we could show that the level of training affects performance in different ways and is dependent on the flight phase. Age differences between the N and P groups cannot explain the decrease in performance between the groups, because there was no age difference between the N and E groups in our sample.

The second RQ concerns the level of training in connection to the gaze strategy during manual flight operations. The general gaze pattern in the pilot group indicates that the radial cross-checking technique (Federal Aviation Administration 2012) was selected, with PFD_ATT as center due to its high importance for the task (Harris and Christhilf 1980). As no event-driven (Schütz et al. 2011) monitoring was induced, task-related gazes away from the PFD were only necessary in the landing phase. In extension to the proposed analysis (Foulsham et al. 2012; Haslbeck and Zhang 2017), we compared the most frequent gaze patterns within the P group to the other groups. Similarly to the driving task (Underwood et al. 2003), the results show that the P group is focusing more on AoIs related to the task and the expected

**Fig. 8** The five highest gaze-strategy patterns for the P group dwell times in percentage, separated by condition and flight phase. The error bars represent the standard error
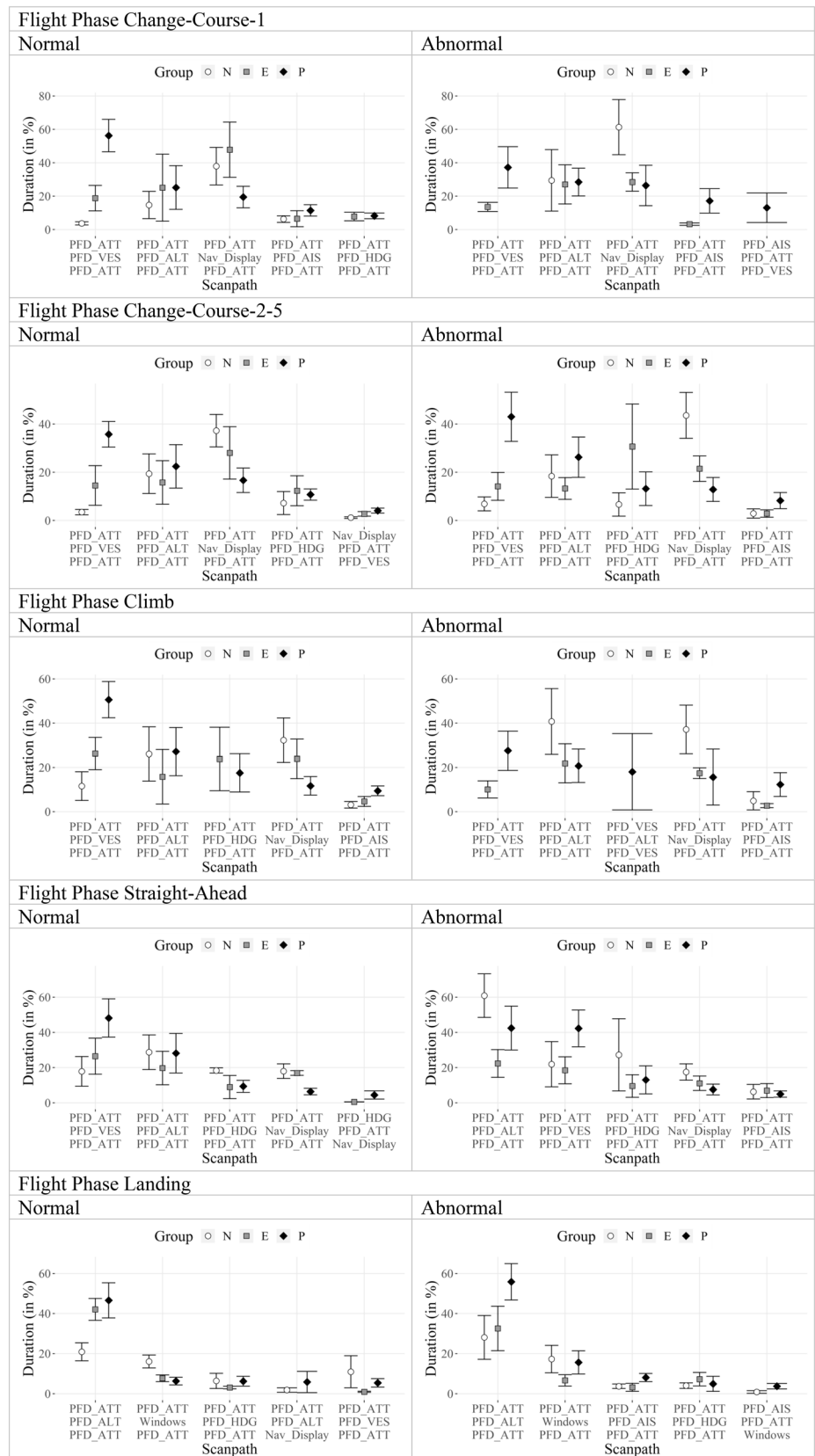
**Table 7** The five highest gaze patterns for novices (N), dwell times in percentage, separated by condition and flight phase

| Flight phase | Condition | | | |
| --- | --- | --- | --- | --- |
| | Normal | | Abnormal | |
| | Gaze pattern | Mean (SD) | Gaze pattern | Mean (SD) |
| Change-Course-1 | PFD_ATT-Nav_Display-PFD_ATT | 37.96 (31.84) | PFD_ATT-Nav_Display-PFD_ATT | 61.34 (36.83) |
| | PFD_ATT-PFD_ALT-PFD_ATT | 14.72 (14.13) | PFD_ATT-PFD_ALT-PFD_ATT | 29.45 (41.16) |
| | PFD_ALT-PFD_ATT-Nav_Display | 7.08 (5.61) | Nav_Display-PFD_ATT-PFD_ALT | 4.49 (3.47) |
| | PFD_VES-PFD_ALT-PFD_ATT | 6.46 (6.56) | PFD_ALT-PFD_ATT-Nav_Display | 4.2 (3.17) |
| | PFD_ATT-PFD_AIS-PFD_ATT | 6.28 (2.79) | PFD_ATT-Nav_Display-PFD_ALT | 2.88 (1.56) |
| Change-Course-2–5 | PFD_ATT-Nav_Display-PFD_ATT | 37.22 (19.05) | PFD_ATT-Nav_Display-PFD_ATT | 43.58 (28.51) |
| | PFD_ATT-PFD_ALT-PFD_ATT | 19.39 (20.16) | PFD_ATT-PFD_ALT-PFD_ATT | 18.39 (24.86) |
| | PFD_ATT-PFD_HDG-PFD_ATT | 7.22 (8.21) | PFD_ATT-Windows-PFD_ATT | 12.88 (16.59) |
| | PFD_ATT-PFD_ALT-PFD_VES | 6.62 (6.13) | PFD_ATT-PFD_VES-PFD_ATT | 6.86 (6.47) |
| | PFD_ATT-PFD_AIS-PFD_ATT | 6.27 (8.86) | PFD_ATT-PFD_HDG-PFD_ATT | 6.65 (8.36) |
| Climb | PFD_ATT-Nav_Display-PFD_ATT | 32.3 (28.41) | PFD_ATT-PFD_ALT-PFD_ATT | 40.75 (39.43) |
| | PFD_ATT-PFD_ALT-PFD_ATT | 26.07 (27.37) | PFD_ATT-Nav_Display-PFD_ATT | 37.12 (24.61) |
| | PFD_ALT-PFD_ATT-PFD_ALT | 21.76 (30.15) | PFD_ALT-PFD_VES-PFD_ALT | 8.09 (8.01) |
| | PFD_ATT-PFD_VES-PFD_ATT | 11.58 (11.24) | PFD_ALT-PFD_ATT-PFD_VES | 7.81 (9.4) |
| | PFD_ALT-PFD_VES-PFD_ALT | 6.08 (8.2) | PFD_ALT-PFD_ATT-Nav_Display | 5.45 (5.68) |
| Straight-Ahead | PFD_ATT-PFD_ALT-PFD_ATT | 28.72 (26.03) | PFD_ATT-PFD_ALT-PFD_ATT | 60.89 (27.68) |
| | PFD_ATT-PFD_HDG-PFD_ATT | 18.37 (2.06) | PFD_ATT-PFD_HDG-PFD_ATT | 27.25 (29.02) |
| | PFD_ATT-Nav_Display-PFD_ATT | 17.98 (11.7) | PFD_ATT-PFD_VES-PFD_ATT | 21.9 (22.26) |
| | PFD_ATT-PFD_VES-PFD_ATT | 17.83 (20.68) | PFD_ATT-Nav_Display-PFD_ATT | 17.51 (13.18) |
| | PFD_ATT-PFD_AIS-PFD_ATT | 10.77 (9.82) | PFD_ALT-PFD_ATT-PFD_ALT | 13.3 (17.01) |
| Landing | PFD_ATT-PFD_ALT-PFD_ATT | 20.9 (12.66) | PFD_ATT-PFD_ALT-PFD_ATT | 28.08 (30.92) |
| | PFD_ATT-Windows-PFD_ATT | 16.06 (9.1) | PFD_ATT-Windows-PFD_ATT | 17.28 (18.13) |
| | PFD_ATT-PFD_VES-PFD_ATT | 10.96 (15.97) | PFD_ATT-Nav_Display-PFD_ATT | 12.61 (7.88) |
| | PFD_ATT-PFD_HDG-PFD_ATT | 6.39 (9.31) | PFD_ALT-PFD_ATT-Windows | 4.34 (4.22) |
| | PFD_ATT-Nav_Display-PFD_ATT | 4.7 (4.44) | PFD_ATT-PFD_HDG-PFD_ATT | 4.03 (3.4) |

aircraft behavior, e.g., interest in PFD_VES increased when flying a Change-Course due to the expectation of a decline at the end of this flight phase. The task load-increasing noise did not have any effect on the gaze pattern of group P.

The third RQ tackles the relationship between task load, performance, and gaze pattern. The connection between workload and performance shows that the influence of workload on performance is stronger if the level of training or task-related experience is low. This is in line with results from other domains (e.g., De Waard 1996; Volpe et al. 1996). In connection to that, gaze pattern analysis shows that all groups used the radial cross-check technique (Federal Aviation Administration 2017) with PFD_ATT as the center. The most important cross check for the P group was PFD_ATT PFD_VES PFD_ATT, except for the landing phase where PFD_ATT PFD_ALT PFD_ATT becomes more important. The P group seems to use the vertical speed and incorporates it into their attitude actions, to hold the requested altitude. This shows a shift in the cognition of the P group when switching the input information.

In contrast to the P group, the N group paid additional attention to the Nav_Display, especially during Change-Course flight phases, and almost non to the PFD_VES. In the abnormal condition, the N group's dwell duration on the PFD_ATT increased and significant increases for the gaze pattern PFD_ATT PFD_ALT PFD_ATT were found. The increase could indicate a focus on the task relevant information within the abnormal condition that should also improve performance (Cox-Fuenzalida 2007; Drory 1985; Naveh-Benjamin et al. 2005). This was, however, not the case. The distribution of gaze patterns shows that the N group did not look on PFD_VES with the same frequency as the other groups and therefore could not incorporate the information into their attitude actions. This leads to the assumption that with an increased task load, the N group did not have the adequate task knowledge, which led to a decreased performance compared to the P group. We believe that the change in information gathering is connected to the level of training and can be used as one marker for the early improvement of pilot training.

After the discussion on the RQ, we look at the evaluation of the DTW against the autopilot in the context of manual flight operation performance assessment. Different performance measurement techniques have been proposed and applied throughout the years (e.g., Baron and Levison 1975; Ebbatson et al. 2010; Hubbard 1987). The literature reveals two main methods to measure performance: comparisons with an optimal track (e.g., Ebbatson et al. 2008) and the selection of behavioral or environmental markers (e.g., Faulhaber 2019). The DTW of participant performance against autopilot is associated with the comparison of tracks. We propose two adaptations to the measurement based on the results by Ebbatson et al. (2008) claiming highly different control strategies by pilots in the pitch and yaw axes. Firstly, we propose the selection of aircraft parameters (e.g., airspeed, altitude, vertical speed) rather than pilot inputs, due to a variety of pilot inputs that could lead to the same flight behavior. Secondly, we suggest the selection of the autopilot behavior as the optimal track to evaluate pilot performance. Even though the autopilot might not be considered as optimal behavior for each flight phase, it represents a realistic and standardized behavior of all aircraft states throughout the scenario. We therefore believe that the performance results are reliable and valid, which is more important than the ongoing discussion about the optimal flight track.

## 6 Conclusion

Even with the small group sizes, some important indications were found, confirming the connection between workload, performance, and gaze pattern. During the experiment, workload was manipulated via noise for three groups with different experience. Regarding performance and workload between the runs, the relevant factor had to be the gaze pattern and the amount of attention spent on selected AoIs. The task of manual flight operations is relatively straightforward in keeping the aircraft in the respective corridor of each scenario by gathering information only from the AoIs on the PFD. The differences between the groups show a different focus on some AoIs in connection to the flight phases, which can be of special importance for optimizing pilot training.

The results show a necessity to evaluate manual flight operations with respect to more flight phases. Takeoff, approach, and landing are, of course, the most important phases of flight, but considering the unpredictable situations in air traffic, we need to take more flight phases into account. The results also showed that DTW can be used to determine the deviation for each aircraft from the baseline and is therefore a promising measurement for performance.

Further research building upon these results should focus on the performance-shaping factors and their interaction. A next step would be to validate the model presented in this paper with more experimental research and to review the results dependent on the flight phases. The different flight phases should be evaluated in relation to their general influence on workload and how this influence can be considered for analyzing the results. In terms of external validity, the higher workload should be induced by tasks the participants have to perform, such as communication with air traffic control. For the DTW method of performance analysis a broader basis for determining a suiting optimum should be evaluated. In relation to the ceiling effect, the workload should also be systematically increased until the performance of the manual flight operations shows a strong decrease for the novice group, assuming that the experts and pilots can still perform at this level.

**Author contributions** All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by MF and SYL. The first draft of the manuscript was written by MF and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

**Data availability** The data that support the findings of this study are available from the corresponding author, Maik Friedrich, upon reasonable request.

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

**Ethics approval** This research complied with the National Statement on Ethical Conduct in Human Research (2007) and was approved by the Institutional Review Board at the University of Southern Queensland.

**Consent to participate** Informed consent was obtained from each participant.

**Consent for publication** All authors have seen and approved the manuscript being submitted. We warrant that the article is the authors' original work.

## References

Baron S, Levison WH (1975) An optimal control methodology for analyzing the effects of display parameters on performance and workload in manual flight control. IEEE Transactions Syst Man Cyber SMC 5(4):423–430. https://doi.org/10.1109/TSMC.1975.5408433

Billings CE (1997) Aviation automation: the search for a human-centered approach. Lawrence Erlbaum, Mahwah, NJ

Blomberg O (2011) Conceptions of cognition for cognitive engineering. Int J Aviat Psychol 21(1):85–104. https://doi.org/10.1080/10508414.2011.537561

Casner SM, Geven RW, Recker MP, Schooler JW (2014) The retention of manual flying skills in the automated cockpit. Hum Factors 56(8):1506–1516. https://doi.org/10.1177/0018720814535628

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

Cox-Fuenzalida L-E (2007) Effect of workload history on task performance. Hum Factors 49:277–291

De Waard D (1996) The measurement of drivers' mental workload. Groningen University, Traffic Research Center Netherlands

Dorr M, Martinetz T, Gegenfurtner KR, Barth E (2010) Variability of eye movements when viewing dynamic natural scenes. J Vis 10(10):28–28

Drory A (1985) Effects of rest and secondary task on simulated truck-driving task performance. Hum Factors 27(2):201–207. https://doi.org/10.1177/001872088502700207

Ebbatson, M. (2009). The loss of manual flying skills in pilots of highly automated airliners. (PhD Thesis), Cranfield University, Retrieved from http://hdl.handle.net/1826/3484

Ebbatson M, Harris D, Huddlestone J, Sears R (2008) Combining control input with flight path data to evaluate pilot performance in transport aircraft. Aviat Space Environ Med 79(11):1061–1064. https://doi.org/10.3357/ASEM.2304.2008

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. https://doi.org/10.1080/00140130903342349

Edwards, T. (2013). Human performance in air traffic control. (PhD), Dissertation, University of Nottingham

EUROCONTROL. (2012). NASA task load index (NASA-TLX). HP repository. Retrieved from https://ext.eurocontrol.int/ehp/?q=node/1583. Accessed 4 July 2019

Faulhaber A (2019) From crewed to single-pilot operations: Pilot performance and workload management. Paper presented at the 20th international symposium on aviation psychology (ISAP 2019 Dayton Ohio United States

Federal Aviation Administration. (2012). Instrument flying handbook (FAA-H-8083–15B). Retrieved from https://www.faa.gov/regulations_policies/handbooks_manuals/aviation/media/FAA-H-8083-15B.pdf. Accessed 12 Aug 2018

Federal Aviation Administration. (2013). Operational use of flight path management systems. Retrieved from https://www.skybrary.aero/bookshelf/books/2501.pdf. Accessed 1 May 2019

Federal Aviation Administration. (2017). Manual flight operations proficiency. Retrieved from http://www.faa.gov/other_visit/aviation_industry/airline_operators/airline_safety/safo. Accessed 12 June 2018

Feusner M, and Lukoff B (2008) Testing for statistically significant differences between groups of scan patterns. Paper presented at the Proceedings of the 2008 symposium on Eye tracking research and applications

Field E, Harris DON (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. https://doi.org/10.1080/001401398186216

Foulsham T, Dewhurst 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 Move Res 5(3):1–14

Friedrich M, Rußwinkel N, Möhlenbrink C (2016) A guideline for integrating dynamic areas of interests in existing set-up for capturing eye movement: looking at moving aircraft. Behav Res Methods 49:3. https://doi.org/10.3758/s13428-016-0745-x

Goldberg JH, Kotval XP (1999) Computer interface evaluation using eye movements: methods and constructs. Int J Ind Ergon 24:631–645. https://doi.org/10.1016/S0169-8141(98)00068-7

Harris RL, Christhilf DM (1980) What do pilots see in displays? Proceed Human Fact Soc Ann Meet 24(1):22–26. https://doi.org/10.1177/107118138002400106

Hart SG, Staveland LE (1988) Development of NASA-TLX (task load index): results of empirical and theoretical research. Human mental workload 1:139–183

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

Haslbeck A, Zhang B (2017) I spy with my little eye: analysis of airline pilots' gaze patterns in a manual instrument flight scenario. Appl Ergonom 63:62–71. https://doi.org/10.1016/j.apergo.2017.03.015

Haslbeck A, Kirchner P, Schubert E, and Bengler K (2014) A flight simulator study to evaluate manual flying skills of airline pilots. Paper presented at the Human factors and ergonomics society annual meeting.

Hembrooke H, Feusner M, Gay G (2006) Averaging scan patterns and what they can tell us. Paper presented at the Symposium on Eye tracking research and applications, San Diego, California

Hollnagel E (1993) Human reliability analysis: context and control. Academic Press Inc, London

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). Eye tracking: A comprehensive guide to methods and measures: OUP Oxford.

Holšánová J (2008) Discourse, vision, and cognition. John Benjamins Publishing Company, Amsterdam

Hubbard D (1987). Inadequacy of root mean square error as a performance measure. Paper presented at the International Symposium on Aviation Psychology, 4 th, Columbus, OH.

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

Knowles WB (1963) Operator loading tasks. Hum Factors 5(2):155–161. https://doi.org/10.1177/001872086300500206

Landry D (2014) Keeping manual flying skills sharp. Air Line Pilot 83:6

Lee Y-H, Liu B-S (2003) Inflight workload assessment: comparison of subjective and physiological measurements. Aviat Space Environ Med 74(10):1078–1084

McRuer D (1982) Pitfalls and progress in advanced flight control systems (AGARD CP-321). AGARD/NATO, Neullity-sur-Seine

Naveh-Benjamin M, Craik FIM, Guez J, Kreuger S (2005) Divided attention in younger and older adults: effects of strategy and relatedness on memory performance and secondary task costs. J Exp Psychol Learn Mem Cogn 31(3):520–537. https://doi.org/10.1037/0278-7393.31.3.520

Papenfuss, and Friedrich M (2016) Head up only: a design concept to enable multiple remote tower operations. Paper presented at the 35th Digital Avionics Systems Conference, Sacramento

Parasuraman R, Manzey DH (2010) Complacency and bias in human use of automation. Attent Integrat 52(3):381–410. https://doi.org/10.1177/0018720810376055

Parasuraman R, Riley V (1997) Humans and automation: use, misuse, disuse, abuse. Hum Factors 39:230–253

Rigner J, Dekker S (2000) Sharing the burden of flight deck automation training. Int J Aviat Psychol 10(4):317–326. https://doi.org/10.1207/S15327108IJAP1004_1

Salvador S, Chan P (2007) Toward accurate dynamic time warping in linear time and space. Intell Data Anal 11(5):561–580

Salvucci DD (1999). Inferring intent in eye-based interfaces: tracing eye movements with process models. Paper presented at the Conference on Human Factors in Computer Systems, New York, NY

Schütz AC, Braun DI, Gegenfurtner KR (2011) Eye movements and perception: A selective review. Journal of vision 11(5):9–9

Simon P, Rousseau F, and Angue J (1993, 17–20 Oct. 1993). Quantitative analysis of mental workload influence on eye scanning movements. Paper presented at the Proceedings of IEEE Systems Man and Cybernetics Conference: SMC

Taylor JL, Kennedy Q, Noda A, Yesavage JA (2007) Pilot age and expertise predict flight simulator performance. Neurology 68:648–654

Tsang PS (2003) Assessing cognitive aging in piloting. In: Tsang PS, Vidulich MA (eds) Principles and practice of aviation psychology. Lawrence Erlbaum, Mahwah, N.J, pp 507–546

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. https://doi.org/10.1080/0014013031000090116

Van Gemmert AWA, Van Galen GP (1997) Stress, neuromotor noise, and human performance: a theoretical perspective. J Exp Psychol Hum Percept Perform 23(5):1299–1313. https://doi.org/10.1037/0096-1523.23.5.1299

Vanderhaegen F, Wolff M, Mollard R (2020) Non-conscious errors in the control of dynamic events synchronized with heartbeats: a new challenge for human reliability study. Saf Sci 129:104814. https://doi.org/10.1016/j.ssci.2020.104814

Veillette PR (1995) Differences in aircrew manual skills in automated and conventional flight decks. Transp Res Rec (1480)

Volpe CE, Cannon-Bowers JA, Salas E, Spector PE (1996) The impact of cross-training on team functioning: an empirical investigation. Hum Factors 38(1):87–100

Wickens CD (1999) Automation in air traffic control: The human performance issues. Current Research and Trends, Automation Technology and Human Performance, pp 2–10

Wiener EL, Chute RD, and Moses JH (1999). Transition to glass: Pilot training for high-technology transport aircraft (20000032979). Retrieved from https://ntrs.nasa.gov/search.jsp?R=20000032979. Accessed 1 July 2019