Fatigue in pilots who commute by driving: identification of the problem and investigation with driving and flying simulation

by

Alexandros Kontotasios

Submitted in accordance with the requirements for the degree of Doctor of Philosophy

The University of Leeds

Institute for Transport Studies

August 2020

Intellectual Property and Publication Statements

Acknowledgments

The completion of this thesis would not be possible without the support of a number of people. First of all, I am deeply grateful to my supervisors Professor Samantha Jamson, Edward C. Killan, and Dr Stephen J. Wright for their constructive feedback and warm encouragement. They were always there to support me during this marathon.

A big thank you should also go to the University of Leeds Driving Simulator team and especially to Hrvoje Jukic and Michael Daly, who helped me with the design of the simulated drives, and to Richard Romano who provided the eye-tracking device I used in my experiments. Special thanks also to Hamish Jamson, who helped me with the design of the X-Plane flight scenario as an airline pilot and pilot instructor. I am also indebted to Dr Charisma Choudhury for providing the wrist-worn device and all the participants in my studies.

I am grateful to Andriana, who stood by me throughout these years, was patient when I did not have time to spend with her and reminded me that I was not alone when I had a problem. A big thank you should also go to my family for supporting me in every way possible throughout the years and motivating me to achieve my goals. Finally, I am proud of myself because doing this PhD required leaving my country, studying extra-long hours, and writing up the thesis while working full-time amid the Covid-19 lockdown.

Abstract

Aviation authorities suggest that commuting to the airport can impair pilots' flying performance by inducing fatigue. That risk might be particularly high after long drives of high task load because they deplete individuals' attentional resources. Flying may then increase commuting pilots' levels of fatigue further leading to performance decrements while driving after duty. This thesis explored the extent of this issue among professional pilots worldwide and investigated if the type of fatigue induced by driving (i.e. active or passive) subsequently affects the levels of fatigue experienced in flying and driving. The effectiveness of a self-initiated intervention strategy for fatigue experienced in the commuting cycle after a long and demanding drive was explored too.

The extent of the issue was investigated with an online survey of 419 professional pilots flying from 52 countries. The pilots were asked about the usual duration and levels of mental demand of their drives to the airport and their usual levels of fatigue during the commuting cycle. Moreover, they answered questions about their difficulty in staying awake and performance in flight and while driving after duty. The survey suggested that many professional pilots worldwide are at an increased risk of fatigue-related aeroplane accidents and car crashes after duty when driving to the airport is long and mentally demanding. Therefore, the role of active fatigue was indicated.

Exploring the role of the type of fatigue induced by driving in fatigue experienced in subsequent flying and driving is essential in order to develop effective intervention strategies. This considered, an experiment with 60 non-pilots was conducted. Three groups (i.e. control, manual, and automated) completed for 32 minutes simulated flying tasks that did not require having pilot training. These tasks were followed by a 14-minute, simulated drive (level 0 of automation). The manual and the automated group also completed a 60-minute simulated drive (level 0 of automation for the manual and level 2 for the automated group) before the flying tasks. This study suggested that completing a long drive can increase individuals' levels of fatigue in flying and driving tasks that follow, especially when that long drive induces active fatigue.

Active fatigue can reduce with short bursts of physical activity. Hence, the second experiment investigated if such breaks can help individuals to counteract fatigue in flying and driving tasks that are completed after a long and demanding drive. Fifteen individuals with a Private Pilot's Licence completed a 60-minute simulated drive (level 0 of automation) followed by a 52-minute simulated flight of higher fidelity compared to the first experiment and a 20-minute, simulated drive (level 0 of automation). In one of their two sessions, the participants took two 6-minute breaks; one before and one after the flying tasks. It was found that the participants were less

fatigued and performed better in flight and during the short drive when they took the breaks. Hence, short bursts of physical activity could be one of the intervention strategies that pilots could use to reduce their levels of fatigue after long and demanding drives to the airport.

The outcomes of this thesis could be used by regulators, airlines, pilots, and researchers to work towards understanding in more depth fatigue in pilots who commute to the airport by completing long and demanding drives. Then, appropriate self-initiated intervention strategies, training, policies, work schedules, and fatigue regulations could be developed.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

1. **ANOVA** – analysis of variance
2. **ATC** – air traffic control
3. **CAA** – civil aviation authority
4. **CSS** – crew status survey
5. **EASA** – European union aviation safety agency
6. **ECG** – electrocardiogram
7. **EDA** – electrodermal activity
8. **EEG** – electroencephalography
9. **EOG** – electrooculogram
10. **ESS** – Epworth sleepiness scale
11. **FAA** – federal aviation administration
12. **fMRI –** functional magnetic resonance imaging
13. **FRMS** – fatigue risk management system
14. **HR** – heart rate
15. **HRV** – heart rate variability
16. **ICAO** – international civil aviation organization
17. **KSS** – Karolinska sleepiness scale
18. **MATB-II –** the second version of the multi-attribute task battery
19. **NASA-TLX** – national aeronautics space administration task load index
20. **NHTSA -** National Highway Traffic Safety Administration
21. **NTSB** – national transportation safety board
22. **OSA** – obstructive sleep apnoea
23. **PERCLOS** – percentage of eye closure
24. **PPL** – private pilot license
25. **PSG** – polysomnography
26. **PVT** – psychomotor vigilance task

27. **RMSD** – root mean square deviation

28. **RSME** – rating scale mental effort

29. **SD** – standard deviation

30. **STSS** – short-term sensory states

31. **TTC –** time to collision

32. **VAS** – visual analogue scale

33. **VAS-F** – visual analogue scale-fatigue

Glossary

1. **Active fatigue -** the state change resulting from continuous and long, task-related psychomotor adjustment.
2. **Alertness -** a state or degree of readiness to detect stimuli and maintain attention.
3. **Arousal** – the state (psychological or physiological) of being awaken.
4. **Attentional resources** – individuals' attentional capacity. This is limited so after some time of paying attention these resources are depleted.
5. **Baseline measurements** – measurements taken either when the participants arrived for the trials or just before/at the start of the driving/flying task sections.
6. **Boredom** – a state of decreased interest for the task at hand and low cognitive activation. In this thesis, it is considered part of the mechanism that causes passive fatigue by contributing to disengagement from tasks.
7. **Commuting -** the time and activity required for pilots to travel from their homes to the airport (domicile), where their duty assignment begins and again back home.
8. **Drowsiness -** the transitional state between wakefulness and sleep associated with some subjective feelings and symptoms. In this thesis, it is considered an expression of fatigue.
9. **Fatigue** – the definition used in this thesis is: 'the state of reduced performance capability (mental and/or physical) resulting from the interaction of fatigue induced by driving to the airport and flying due to time on task and task load, sleep-related factors, inter-individual differences, and environmental factors that can impair a commuting pilot's performance in flight and while driving after duty'.
10. **Fatigue-related data** – the data collected when measuring the subjective, physiological, and performance expressions of fatigue.
11. **Home base** – the airport, where a pilot's duties normally begin.
12. **Lapse** – In the Psychomotor Vigilance Task, the reaction times that are longer than 500 milliseconds (Matthews et al., 2017).
13. **Level 0 of automation** – only the human drives. No car features that steer, brake, or accelerate.
14. **Level 2 of automation** – these features provide steering and brake/acceleration support to the driver. The human drives whenever these driver support features are engaged and must constantly supervise them (SAE, 2019).

15. **Mental disengagement** – in this thesis, the sense of being away from the driving or flying tasks.
16. **Mental fatigue** - a state that arises from situations requiring long-term, continuous, repetitive performance on some mental task.
17. **Passive fatigue** - a mental state that develops in tasks that require system monitoring with either rare or even no overt perceptual-motor requirements.
18. **Physical fatigue** - a state caused by physical activity and is experienced as reduced muscular power and movement.
19. **Professional pilots** - for the purposes of this thesis, this refers to airline pilots and pilots of charter, cargo, and business aviation flights.
20. **Sleepiness** - the tendency to fall sleep. In this thesis, it is considered an expression of fatigue.
21. **Task load** – the perceived difficulty of completing a task.
22. **Tiredness** - an expression of fatigue identified by the feeling of having low levels of energy.
23. **Vigilance** – the ability of organisms to maintain their focus of attention and to remain alert to stimuli over long periods. In this thesis, it is considered an expression of fatigue.

Chapter 1. Introduction

Fatigue is a worrying issue in today's fast-paced society. The absence of a balance between excessive work demands and rest opportunities contributes to an increased risk of fatigue-related work accidents, which cost the United Kingdom between 115 and 240 million pounds per year (Health and Safety Executive, 2019). Accidents are more likely to occur when individuals are fatigued because of slower reaction times (Dinges et al., 1997; Lorist et al., 2000; Sanders, 1998), difficulties in storing information in working memory (Hockey, 1997), and planning errors (Lorist et al., 2000; Van der Linden et al., 2003). Fatigue has also been associated with deterioration of attention (Sanders, 1998), mental confusion (Dinges et al., 1997), distress (Saxby et al., 2008), and impairment of the ability to detect deviations in system parameters and retain new information (Falleti et al., 2003). Furthermore, fatigue may affect adversely the ability to focus on stimuli of interest and avoid the salient but irrelevant ones (Boksem et al., 2005).

The reduction of the fatigue-related risk of accidents in driving and flying is on the list of the most wanted safety improvements of the National Transportation Safety Board-NTSB (2019[a]). The number of fatalities in car crashes has reduced in the last twenty years (Figure 1.1) (Eurostat, 2017). However, fatigue remains a risk for drivers' safety since it has been identified as a contributory factor in 10-20% of the road accidents in the European Union (European Commission, 2020) and 2.5% of the fatal crashes in the US (National Highway Traffic Safety Administration - NHTSA, 2015). The higher percentage of fatal accidents in which fatigue was identified as a contributory factor in the study conducted by the European Commission compared to that by NHTSA might be attributed to a difference in the data collection approach followed. That is, the percentage was calculated based on drivers' self-reports and naturalistic driving data in the former and police records in the latter. Irrespective of how high these percentages are, they indicate an existent risk that needs managing.

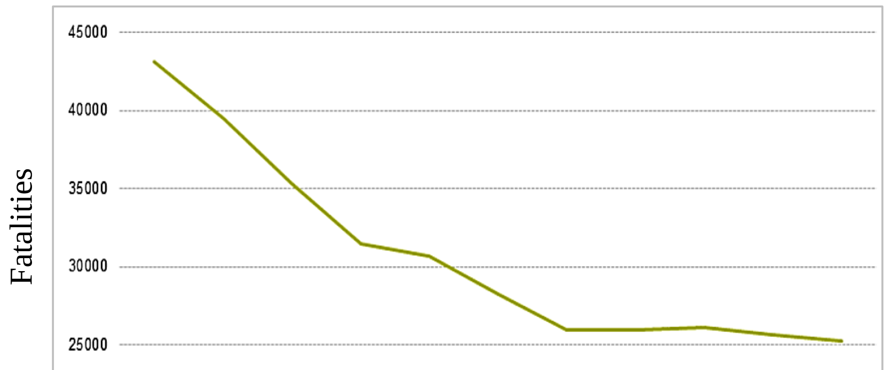


Figure 1.1 Fatalities in car crashes in the European Union from 2007 to 2017 (Eurostat, 2017).

Although air travel is safe, fatal aeroplane accidents occur too. According to the International Civil Aviation Organization (ICAO) (2019), more than 500 fatalities occurred in commercial aviation (i.e. aircraft operation for passenger and cargo transportation) accidents in 2018 without any apparent reduction noticed over the last decade (Figure 1.2). Therefore, air safety still needs improvement. In a review of the contributory factors to the air accidents investigated from 2001 to 2012 by the National Transportation Safety Board (NTSB), Marcus and Rosekind (2017) identified fatigue as one of them in almost 20% of the cases. This percentage could be even higher because accident investigations may not capture pilots' high levels of fatigue unless they are clearly stated and flying fatigued does not necessarily contribute to accidents. Regardless of the exact percentage, these results suggest the importance of managing the risk of being fatigued in flight.

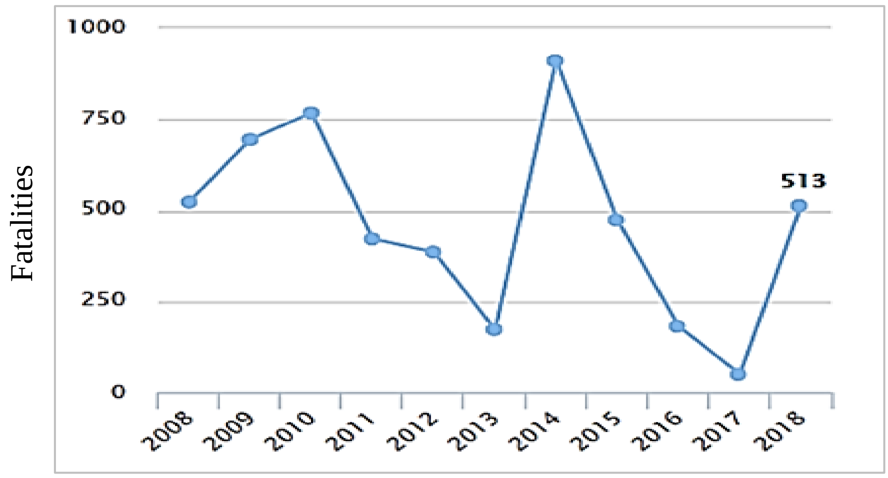


Figure 1.2 Fatalities in commercial flight accidents from 2008 to 2018 worldwide (aeroplanes above 5.7 tones) (International Civil Aviation Organization, 2019).

Several interventions have been developed to mitigate the risk of fatigue-related aeroplane accidents. That is, airlines provide fatigue training to educate pilots about that risk and fatigue regulations require that pilots report fit-for-duty (Commission Regulation, 2014; Federal Aviation Administration-FAA, 2012a, 2011; International Civil Aviation Organization, 2016). Moreover, airlines use biomathematical models to predict pilot fatigue depending on factors, such as work schedules, sleep duration, and duration of travelling (i.e. commuting) to the airport. There are also some airline policies that state the maximum distance between pilots' homes and their home base (i.e. the airport that a pilot usually starts the duties from) to make sure that they can quickly arrive at the airport when on stand-by duty. These policies may also result in avoiding fatigue caused by long travels to the airport.

The risk of flying fatigued due to travelling to the airport has been recently identified by aviation organisations (Federal Aviation Administration, 2012a; National Transportation Safety Board, 2010). However, this has not been clearly stated in the fatigue regulations and pilots do not receive detailed information about avoiding and managing it through training and policies. Additionally, the biomathematical models of fatigue do not consider the characteristics of that travel (e.g. the amount of activity required by the pilot); thus, they may miss information about how fatiguing journeys to the airport are. These gaps are reasonable since the effects of travelling to the airport on pilots' fatigue in flight have not been adequately explored.

Drawing from studies solely on drivers (see section 2.1.3), pilots who commute by driving may arrive fatigued at the airport, especially after long drives of high task load (i.e. the perceived difficulty of a task). Unless they reduce their levels of fatigue while at the airport or report being fatigued and do not fly, fatigue-related performance decrements may occur in flight. Although the link between driving and flying can be inferred from the literature on fatigue in drivers and pilots, that would be a logical leap because it has not been directly researched in commuting pilots. As a result, the extent of this issue among pilots remains unknown, the mechanisms behind any effects of driving on fatigue in flight are not understood, and appropriate intervention strategies cannot be developed.

Besides commuting to the airport, it is expected that many pilots drive after duty, for example, to return home. Studies suggest that flights can increase pilots' levels of fatigue (Petrilli et al., 2006) and car crashes are more likely when drivers are fatigued (Ting et al., 2008). Drawing from that, pilots may crash when driving after a fatiguing flight. This risk may be even higher when they have completed a long and demanding (thus fatiguing) drive to the airport. The risk of fatigue-related car accidents after duty has neither been mentioned in any documents published by aviation authorities nor is included in the fatigue regulations for pilots. Probably due to the lack of relevant literature, fatigue training and the biomathematical models of fatigue lack details about this risk. Moreover, most airlines do not have policies that aim to reduce the risk of car crashes after duty. Covering these gaps would, first, require exploring the extent of this issue among professional pilots. Second, it would be helpful to understand the human factors mechanisms behind any effects of driving and flying on fatigue in subsequent driving tasks in order to use effective intervention strategies. To this end, experiments that would include both driving and flying would be needed.

To sum up, pilots may be at an increased risk of aeroplane accidents and car crashes after duty due to fatigue induced by long and demanding drives to the airport and flying. This thesis will explore the extent of this issue among professional pilots worldwide. For the purposes of this thesis, the term 'professional pilots' will refer to airline pilots and pilots who fly charter, cargo, and business aviation flights. Although the findings of this thesis might apply to military pilots and ground instructors too, fatigue in their commuting cycles will not be explored. The reason is that their commutes to the airport are often not followed directly by pre-flight tasks and they may also spend several hours at work after their flights. Therefore, factors not related to a specific flight may often mediate commuting-related fatigue. Besides the extent of the issue, this thesis will investigate if the type of fatigue induced by driving (i.e. active or passive, see section 1.1.1) subsequently affects the levels of fatigue experienced in flying and driving. Finally, it will be explored if short bursts of physical activity can reduce the levels of fatigue experienced in the commuting cycle after a long and demanding drive. The findings of this thesis could improve flight safety and help pilots to avoid car accidents after duty.

This chapter defines the main concepts of the thesis, describes the gaps in the existing knowledge, and outlines the aims and the objectives of the thesis. Then, it continues with the practical implications of any findings of the research described in the thesis, a description of its novelty and original contribution, and an outline of the content of the next chapters.

1.1 Main concepts of the thesis

1.1.1 Fatigue

Fatigue is a multidimensional construct difficult to define (Desmond and Hancock, 2001) with some researchers using broad definitions. For example, Williamson et al. (2011, p.499) referred to fatigue as '*a biological drive for recuperative rest*', which is a definition that does not provide any information about the causes and effects of fatigue. This information is essential in this thesis because the focus is on exploring the contributors to fatigue and the effectiveness of an intervention strategy. Other definitions of fatigue are field-related. For example, Thiffault and Bergeron (2003, p.381) defined driver fatigue as '*a general psychophysiological state, which diminishes the ability of the individual to perform the driving task by altering alertness and vigilance*'. Despite mentioning the adverse impact of fatigue on driving performance, definitions like this do not refer to any causal factors.

A definition of fatigue in pilots that matches better the criteria set in this thesis was proposed by the International Civil Aviation Organization (ICAO) (2016, p.8). According to that, '*fatigue is a physiological state of reduced mental or physical performance capability resulting from sleep loss or extended wakefulness, circadian stage, or workload (mental and/or physical activity) that can impair a crewmember's alertness and ability to safely operate an aircraft or perform safety-related duties*'. Although some of the contributory factors of fatigue are included in this definition, the role of the time on task, inter-individual differences, and environmental factors is missing. Moreover, this definition mentions only the physiological aspects of fatigue missing the subjective state associated with that and does not highlight the potential interaction of causal factors. Drawing from the ICAO definition, a definition of fatigue in commuting pilots was developed for this thesis to describe '*the state of reduced performance capability (mental and/or* *physical) resulting from the interaction of fatigue induced by driving to the airport and flying due to time on task and task load, sleep-related factors, inter-individual differences, and environmental factors that can impair a commuting pilot's performance in flight and while driving after duty*'.

Due to difficulties in its definition, fatigue is often used interchangeably with other terms, such as sleepiness, drowsiness, and vigilance (Johns, 1998). In this thesis, these constructs are not considered synonymous to fatigue. However, they are closely related to it since fatigue can be often expressed through them (Netherlands Aerospace Centre and Netherlands Institute of Neurosciences, 2019). One of the terms that have been widely confused in the literature with fatigue is sleepiness, which Sagberg et al. (2004) defined as the tendency to fall asleep. Most studies have not distinguished between the two constructs and the literature remains inconclusive regarding their relationship. Even when a distinction was attempted, this was usually not based on research evidence (e.g. Watling, 2016). Balkin and Wensesten (2011) identified the difficulty in distinguishing between fatigue and sleepiness because they usually coexist in operational settings, they are difficult to isolate even in controlled environmental conditions, and both increase the objective and subjective difficulty in performing a task. However, they believed that differentiating between them is essential to identify the contributors to performance decrements and recommend effective intervention strategies.

Sleepiness is usually considered a symptom of sleep-related fatigue, but, in this thesis, it is perceived as an expression of both sleep- and task-related fatigue. This notion is based on studies that showed that the levels of sleepiness increased differently in non-sleep-deprived drivers depending on the levels of task load of the drives completed (Schömig et al., 2015; Vogelpohl et al., 2019). For example, Vogelpohl et al. (2019) manipulated the levels of task load during driving by changing the level of automation of the car (see Figure 2.2 for a description of the levels of automation). They found that the levels of sleepiness inferred from yawning and nodding (see section 3.2.2.5) increased more quickly in the group that drove the highly automated driving condition. If sleepiness was not an expression of task-related fatigue but only the effect of sleep-related factors like sleep deprivation, then task load should not have made a difference in sleepiness in that case.

As mentioned, sleepiness is accepted in this thesis as an expression of fatigue caused by either sleep- or task-related factors. Therefore, the methods used to measure sleepiness may not reveal which of the two types of contributors resulted in fatigue. Similar, in the case of commuting pilots, as a drive to the airport progresses (the time on task is a task-related factor – see section 2.1.3), the time awake since last sleep (i.e. sleep-related factor – see section 2.1.2) increases too. Therefore, finding increased levels of sleepiness at the end of a drive may indicate task-related fatigue, sleep-related, or a combination of the two. Although task- and sleep-related fatigue often co-exist, the experiments described in chapters 5 and 6 will be designed so that the effect of the latter can be controlled.

Fatigue differs from vigilance, which is '*the ability of organisms to maintain their focus of attention and to remain alert to stimuli over long periods*' (Warm et al., 2008, p.433). However, identifying lower levels of vigilance can be used to deduce increased levels of fatigue because fatigued individuals are expected to experience difficulties in detecting small environmental changes and responding to them (Thiffault and Bergeron, 2003). Fatigue induced by completing a task of high or low task load can reduce vigilance due to affecting the way individuals invest their attentional resources (see section 2.2.1). That is, vigilance decrements can occur when people become fatigued by completing a demanding task because the task depletes their attentional resources and, thus, investing more of them is difficult (Grier et al., 2003; Helton et al., 2005). In contrast, fatigue induced by performing a low task load activity relates to withdrawal from that task. In that case, vigilance reduces because of the disengagement of the attentional system. Due to the link between fatigue and vigilance decrements, fatigue will be inferred in the experiments of this thesis by measuring vigilance with the PVT.

Fatigue should be distinguished from arousal, which is the state (psychological or physiological) of being awake. Moreover, fatigue is not synonymous to tiredness, which is the feeling of having low levels of energy that usually accompanies fatigue induced by completing a task. However, task-related fatigue is a broader construct than tiredness and can also manifest itself in other ways, such as sleepiness. Similar, fatigue caused by a task may be expressed as drowsiness. Drowsiness is '*the transitional state between wakefulness and sleep associated with some subjective feelings and symptoms*' (Shen et al., 2006, p.64). In other words, drowsiness is the stage just before sleep, so it entails high levels of sleepiness. However, as explained, fatigue induced by completing a task can be expressed as sleepiness, and, therefore, drowsiness too. Similar to sleepiness, although fatigued individuals may experience drowsiness, drowsiness disappears with sleep but not with rest.

Due to the multidimensional nature of fatigue, several types of fatigue have been suggested in the literature. People often experience more than one types of fatigue at the same time, so it can be difficult to clearly distinguish between them. Figure 1.3 depicts a model of fatigue in drivers with the factors that contribute to task-related (active and passive) and sleep-related fatigue. Although different tasks are completed in flight, the same categorisation can be also used for pilots. Task-related fatigue is caused by completing a task, whereas the sleep-related one by factors, such as sleep deprivation and the time of the day. Commuting pilots may experience a combination of the two types of fatigue; for example, a long drive to the airport can cause task-related fatigue because of investing effort for a long period but sleep-related fatigue may be experienced too in that case due to sleep deprivation before the commute. The focus of this thesis is on task-related fatigue caused by driving to the airport because sleep-related fatigue is much better researched in pilots and relevant intervention strategies are in place (e.g. training).

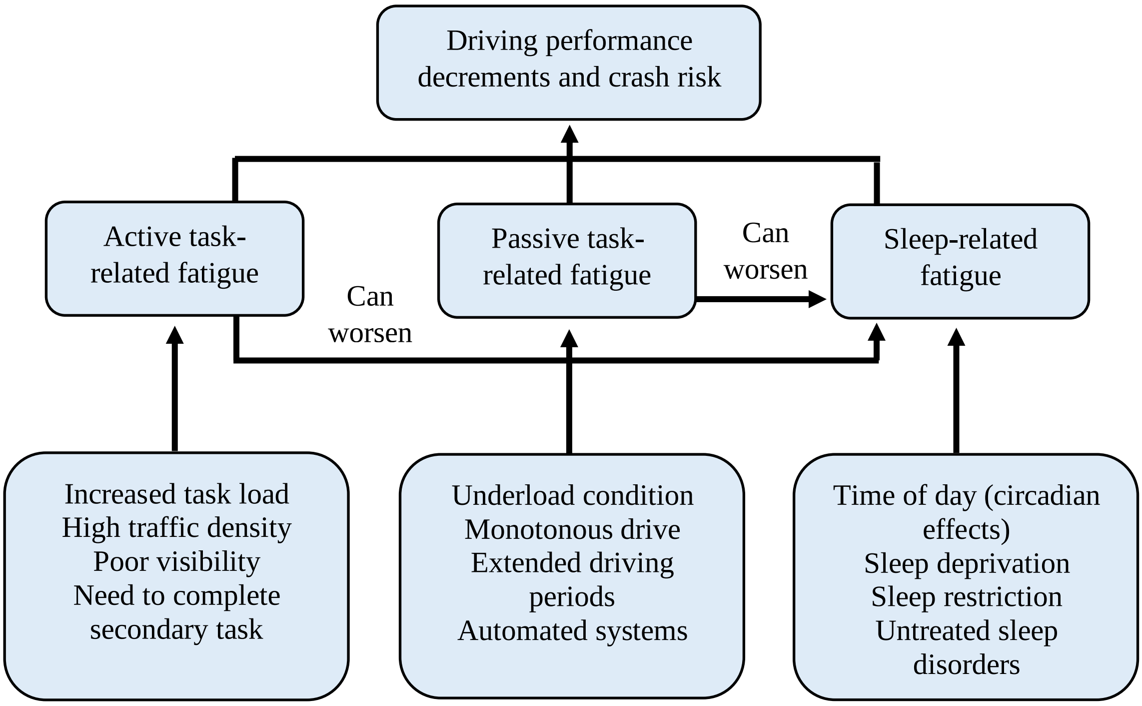


Figure 1.3 A model of fatigue (May and Baldwin, 2009).

Task-induced fatigue can be further broken down into active and passive. This distinction has been mainly made in the literature on driving. Nonetheless, it could apply to flying too since the mechanism of fatigue causation is not expected to differ between drivers and pilots. Active fatigue is '*the state change resulting from continuous and long, task-related psychomotor adjustment*' (Desmond and Hancock, 2001, p.601). Desmond and Hancock considered active fatigue as the type of fatigue typically experienced in manual driving because that imposes high motor-perceptual demands by requiring regular speed and steering adjustments. Similar, active fatigue may be induced in high task load phases of flights, such as the climb and descent. Since motor-perceptual demands require both physical and mental activity, active fatigue is experienced as a combination of mental and physical fatigue (more information about these types of fatigue later in this section).

In contrast to active fatigue, passive fatigue is a mental state that develops in tasks that require '*system monitoring with either rare or even no overt perceptual-motor requirements*' (Desmond and Hancock, 2001, p.601). Examples of such tasks are the highly automated driving and monitoring of systems when the autopilot of an aeroplane is engaged. In those cases, fatigue is expected to be mainly mental because of the limited motor requirements to complete tasks. Fatigue caused by low task load is often used interchangeably with boredom (Lal and Craig, 2001). Both passive fatigue and boredom are associated with a decreased interest for the task at hand, low cognitive activation and negative feelings, and are expected to reduce when engaging with interesting tasks. Nevertheless, in this thesis, the two constructs are perceived as distinct with boredom being considered part of the mechanism that causes passive fatigue by contributing to task disengagement. Passive fatigue is also distinct from habituation, which is caused by the repeated exposure to stimuli and persists after rest (in contrast to passive fatigue).

Another difference between active and passive fatigue is in how quickly they develop with research suggesting that individuals' levels of fatigue can increase sooner when driving is highly automated (i.e. passive fatigue) than manual (i.e. active fatigue) (see section 2.1.3). Active and passive fatigue also differ in the subjective experience and behavioural response to task load. That is, although both types of fatigue can be experienced as tiredness (i.e. the feeling of being tired), studies suggest that active fatigue induced by manual driving relates to increased distress (e.g. negative mood) and high levels of effort and mental workload (Saxby et al., 2013). Mental workload is the effect of task load on a person and depends on their available cognitive capacity (Zijlstra, 1993). In contrast, passive fatigue relates to low levels of effort and mental workload, a quick reduction in task engagement (Saxby et al., 2008), and mind wandering (Korber et al., 2015). The disengagement from the task in low task load driving can be explained by the 'effort/reward imbalance' theory (Tops et al., 2004). Drawing from that theory, drivers stop investing effort when they feel that this will not result in sufficient rewards. In the case of highly automated driving, drivers might believe that they could stop monitoring the driving scene since the car drives itself and their performance does not make a difference in safety.

Fatigue is not only distinguished based on its causes but also on its nature; that is, fatigue can be mental or physical. Physical fatigue is a state caused by physical activity, static muscle load, and/or spinal loading. This is experienced as reduced muscular power (Lal and Craig, 2001) and discomfort (Van Veen, 2016). In driving and flying, physical fatigue is caused, for example, by moving the steering wheel and the joystick, respectively. In contrast, mental fatigue is a state that arises from situations requiring '*long-term, continuous, repetitive performance on some mental task*' (Zhao et al., 2012, p.83) (e.g. system monitoring in flying) and is experienced as a general feeling of weariness and impaired mental performance (Lal and Craig, 2001). In real life, physical and mental fatigue often coexist.

To sum up, this thesis focuses on task-related, active fatigue, which includes both mental and physical fatigue. Passive fatigue will also be explored in the study described in chapter 5 to explore the fundamental aspects of the effects of active fatigue induced by driving on fatigue in flying and subsequent driving tasks by focusing on the role of task load. Some low task load sections of the flights used in the experiments of this thesis may also induce passive fatigue. In real life, fatigue is usually a combination of task- and sleep-induced fatigue, so data about sleep-related fatigue will be collected in the studies of this thesis too and the experiments will be designed so that their effect on fatigue is controlled. Finally, despite the differences of fatigue with sleepiness and vigilance, these are considered expressions of fatigue in this thesis and relevant data will be collected to infer it.

1.1.2 Commuting

Commuting was defined by the National Research Council (2011) as the time and activity required for pilots to travel from their homes to the airport (domicile), where their duty assignment begins and again back home. According to the National Research Council, all pilots commute irrespective of the duration of their journeys. In contrast, the Federal Aviation Administration (2010) defined commuting as the travels to the airport of more than two hours. Setting a time threshold like this is simplistic because the literature is inconclusive regarding how long a journey should be to induce fatigue. Moreover, travels of the same duration may not induce the same amount of fatigue due to differences in their task load (see section 2.1.3), the means of transport used (e.g. driving or being a passenger on a train), and the additional effect on fatigue of sleep-related factors (section 2.1.2). This considered, the definition of commuting provided by the National Research Council will be used in this thesis.

Commuting habits may vary amongst pilots because of differences in factors, such as the means of transport used and the duration of commutes. Commutes may also differ between days because, for example, a pilot who usually drives to the airport may on some days decide to travel by train instead. Moreover, some pilots may choose to commute from their home to a rest facility (e.g. hotel) near the airport the night prior an early-starting duty to avoid getting up too early rather than travelling just before duty. Pilots may also sleep at a rest facility near the airport after duty to avoid driving fatigued back home (or to any other rest facility), whereas others may commute straight after duty. Furthermore, commutes to and from the airport may be interrupted by short breaks (e.g. while driving) and pilots may increase their alertness with certain substances (e.g. caffeine).

Indirect commutes, commutes interrupted by breaks, and commutes under the effect of substances are not investigated in this thesis because these strategies might reduce the levels of fatigue induced by commuting. Thus, the safety risk potentially associated with commuting might be masked, resulting in a difficulty in identifying the role of commuting on pilots’ fatigue. In turn, not understanding if and how commuting to the airport relates to fatigue in flight and while driving after duty hinders the development of appropriate intervention strategies. The commutes of interest in this thesis are shown in green in Figure 1.4. Pilots who drive directly to the airport are expected to park their cars and walk to the location of the airport, where their duty assignment begins. If a duty does not finish at the airport it started at or commuting by driving after flight is too long, pilots may choose to commute to a rest facility (e.g. hotel) before driving home. In contrast, if a duty finishes at the airport it started at, pilots may decide to drive directly after duty back home.

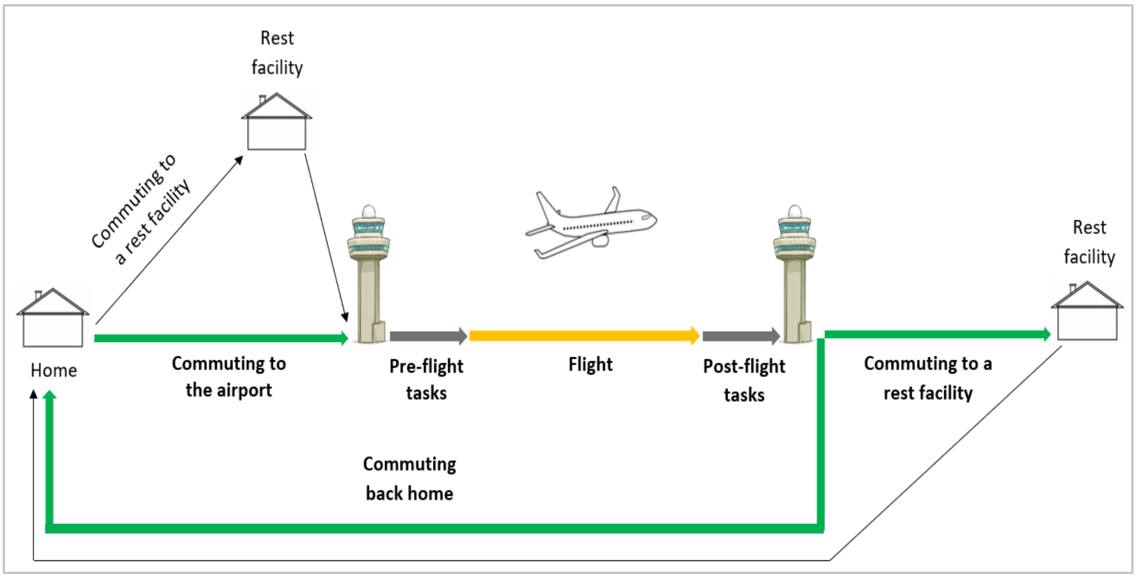


Figure 1.4 Parts of the commuting cycle of interest in this thesis (shown in green).

Pilots may commute by any means (or any combination thereof) of transport, but only commuting by driving is investigated in this thesis. The first reason for that decision is that the thesis focuses on active fatigue induced by commuting to the airport. In contrast, commuting as a passenger on public transport or by car can induce passive fatigue since passengers do not control the vehicle. As will be explained, it is expected that the risk for accidents in flight and while driving after duty is higher when driving to the airport induces active than passive fatigue. The second reason is that many pilots may drive to work, as is the case with other professionals (Department for Transport, 2015; Rapino and Fields, 2013; Statistics Canada, 2011). If that was the case, the risk of being fatigued in flight and while driving after duty could be widespread in the population of pilots and any findings of this thesis could be greatly applied.

1.2 Commuting-related fatigue in pilots

The potentially adverse effects of commuting-induced fatigue on fatigue and performance in flight have only recently drawn the attention of aviation authorities and researchers. This link was first mentioned in the investigation report of the Shuttle America flight 6448 accident at Cleveland in 2007, and, then, in the report of the Colgan Air Flight 3407 crash in New York in 2009 (Federal Aviation Administration, 2012[a]; Federal Aviation Administration, 2012[b]; National Transportation

Safety Board, 2010). The NTSB (2010) concluded that commuting to the airport by aeroplane had likely increased pilots' levels of fatigue in both accidents by contributing to sleep loss. However, these reports did not refer to commuting specifically by driving, task-related fatigue was not mentioned, and fatigue was not measured.

The necessity to research the link between commuting to the airport and fatigue in flight also emerges from three documents that followed the reports published by the NTSB. In the first, the FAA (2012[a]) recommended the inclusion of an analysis of the risks related to commuting to the airport in the Fatigue Education and Awareness Training Programmes. These programmes refer to the training provided by airlines to inform pilots about the factors that contribute to fatigue and its adverse effects on health and performance. They also provide information about the strategies that pilots can use to mitigate these negative effects of fatigue and the legal requirements to avoid flying fatigued. In the second study (2012[b]), the FAA stated that commuting to the airport is a concerning issue because it can contribute to higher levels of fatigue in flight by increasing the length of a pilot's day. They also recommended more research on commuting to the airport and suggested that pilots evaluate their commuting habits. Similar, the ICAO (2016) mentioned that longer commutes to the airport can contribute to increased levels of fatigue in flight.

The documents published by the FAA and the ICAO highlighted the necessity of knowing more about the role of commuting to the airport on pilots' fatigue and performance in flight. Nevertheless, they lacked details about the role of certain characteristics of commutes (i.e. their duration and task load and the means of transport used) on fatigue experienced in flight. Moreover, the potential adverse effects of fatigue induced by commuting to the airport and flying on fatigue while commuting after duty were not mentioned and no information was provided regarding relevant intervention strategies.

Despite aviation organisations highlighting the increased safety risk associated with commuting to the airport, only two studies have explored the link between fatigue induced by commuting to the airport and fatigue in flight. In the first, the National Research Council (2011) used databases with information about where 17,519 mainline pilots (i.e. flying aircrafts with more than 90 seats, often intercontinentally) and 7,553 regional pilots (i.e. flying aircrafts with 90 or fewer seats) in the US lived. They speculated that those who commuted longer distances to the airport might report more fatigued for duty. In the second relevant study, Zakariassen et al. (2019) sent questionnaires to 38 air ambulance pilots in Norway and Austria. They found that 18.4% of them believed that the time spent commuting to work can cause higher levels of fatigue in flight. Although both studies suggested that the duration of commuting to the airport can relate to in-flight fatigue, fatigue was not measured. Moreover, data about the means of transport used and the duration and task load of the commutes were not collected. Finally, the potential link between fatigue induced by driving to the airport and fatigue while driving after duty was not investigated and no intervention strategies for fatigue in the commuting cycle were tested.

The link between commuting to the airport and flying performance has not been adequately researched either. Brown and Whitehurst (2011) asked pilots whether the time spent on commuting related to a deterioration of their performance in flight, but a preliminary analysis showed that none of the participants reported this link. It should be noted that no details were located regarding the sample and the methods of that study. In contrast, the link between commuting to the airport and flying performance was supported in a survey, where Friesacher and Greaves (2016) explored stress induced by commuting to the airport. They found that more than half of their participants (i.e. 528 airline pilots with domiciles in Europe) were aware of the safety issues related to their commutes. Moreover, more than 40 per cent agreed that commuting affected the quality of their colleagues' work, whereas longer commutes to the airport related to worse reported performance in flight for themselves. Friesacher and Greaves' results might indicate adverse effects of fatigue induced by commuting to the airport on flying performance. Nevertheless, fatigue was not mentioned in the questionnaire used with the authors attributing their findings to commuting-induced stress. No studies have explored flying performance after commutes to the airport specifically by driving.

Some studies in professionals other than pilots have explored the effects of fatigue induced by commuting to work on fatigue and performance during the shift. For example, Azwar et al. (2018) asked 105 production workers to rate their levels of fatigue at their workplace and found that longer commutes to work related to higher levels of fatigue during the shift. The effects of commuting to work specifically by driving were not explored in that study, fatigue and performance were not measured with objective measures (see sections 3.2.2 and 3.2.3), and the impact of other factors that can contribute to fatigue (e.g. sleep deprivation) were not controlled. In another study, Roma et al. (2012) investigated whether longer commutes to the airport related to slower reaction times of cabin crew measured with the Psychomotor Vigilance Task (PVT) (see section 3.2.3.1) before flight. The PVT is a simple reaction time test during which individuals push a button as soon as a stimulus appears on a screen (Dorrian et al., 2008). Higher levels of fatigue are expected to relate to lower vigilance, which, in turn, relates to slower reaction times in the PVT. This study did not support the link between the duration of commutes and performance with the authors concluding that commuting duration per se may not predict the performance of cabin crew at the beginning of flights. Nevertheless, they stated that their results might have been different if all their participants had commuted to the airport by driving because then they would not have been able to sleep or rest during commuting.

In contrast to the risk of flying fatigued due to driving to the airport, no aviation authorities have mentioned the risk of car crashes when pilots drive fatigued after duty. Moreover, there are no studies about the extent of this issue and the contributors to it. Some links might be drawn from the literature in other professionals. That is, there is some evidence of an increased risk of crashes after work from studies with nurses (Dorrian et al., 2008; Scott et al., 2008), medical residents (Barger et al., 2005; Marcus and Loughlin, 1996), and air traffic controllers (Nesthus et al., 2006). That risk was associated there to extended and night work shifts, sleep deprivation, consecutive shifts, and drives after work longer than 20 miles. Nevertheless, the link between fatigue induced by driving to work, fatigue during a shift, and fatigue experienced after it has not been explored in any profession.

Before exploring pilots' fatigue during the commuting cycle and the contributors to it, it is essential to collect information about the characteristics of their commutes. That information will help to identify the extent of the issue among the population of professional pilots and, thus, the likelihood of safety events associated with them. Many pilots are expected to commute to and from work by driving a car as is the case with other professionals in several countries (e.g. Department for Transport, 2015). Nevertheless, not much is known about the means of transport pilots usually use to commute.

In contrast to the literature in other professionals, two studies have suggested that most pilots do not commute by driving. In the first one, the National Research Council (2011) used databases of pilots' addresses in the United States (US) and speculated that more than 50% of them might commute by aeroplane because of the long distances between their home addresses and their home bases. More specifically, they suggested that approximately half of the pilots commuted more than 150 miles one-way per duty. However, the addresses in that database might not have been updated and the home addresses might not have been the starting points of the commutes. Moreover, drawing straight lines between two points on a map shows shorter distances than the ones travelled. If this was the case, then shorter distances could have been covered by other means of transport, such as by driving a car.

In the second relevant study, Kleinfehn (2016) surveyed regional airline pilots in the US and found that 62.8% of them commuted by aeroplane and only 26.1% by traditional (not explained further) means of transport. They also found long commuted distances with 28% of the pilots travelling between 151-750 miles one-way and 32.8% more than 750 miles. Similar to the study conducted by the National Research Council, the data in the study by Kleinfehn were collected from pilots with a home base in the US, which has a big geographical spread. Therefore, travelling by aeroplane may be a more practical way of covering such long distances.

Although the National Research Council and Kleinfehn provided an indication of the distances covered by pilots in the US to commute, they did not collect any data about the actual duration of the commutes. That is, a very long distance can be travelled quickly by aeroplane, whereas pilots may need a few hours to cover a relatively small distance by driving in high traffic density. As will be explained in section 2.1.3, the time spent on driving plays an important role in fatigue induction. Therefore, this information can be useful in studies that explore fatigue in pilots that commute by driving. In the only study that collected data about the duration of pilots' commutes, Friesacher (2015) found in a survey of 528 airline pilots with a home base in Europe (data from several countries were collected) that 56.8% of them reported commuting for more than 45 minutes (one-way). No findings from that study were sourced regarding the means of transport used to commute.

The focus of the thesis is on the effects of fatigue induced by long and demanding drives on fatigue experienced in subsequent flying and driving tasks. Thus, chapter 2 will also describe the literature on the role of high and low task load in fatigue in driving and flying. Despite the literature on the link between task load and fatigue in drivers and pilots, no study was sourced on how task load in driving relates to fatigue in flying and driving tasks that follow. No relevant study has been conducted in other professionals either. In addition, there is no literature on the intervention strategies that commuting pilots can use before and after their flights to reduce their levels of fatigue. Similar, relevant research has not been conducted in other professions. Testing the effectiveness of such strategies would require first exploring the extent of the issue and understanding its fundamental aspects; that is, the mechanisms behind fatigue induction and its effects on subsequent flying and driving tasks.

To sum up, the limited literature on commuting pilots and other professionals and the studies solely on pilots and drivers suggest that pilots may be at an increased risk of aeroplane accidents and car crashes after duty due to long and demanding drives to the airport and flying. However, there is no literature on the extent of this issue among professional pilots worldwide. Exploring that is essential in order to assess the risk and decide if and what intervention strategies are needed. That is, it is impractical to manage all the risks of the everyday life in high hazard industries like aviation, so only those that are considered very probable to happen and would have severe consequences are usually controlled. This considered, finding evidence that many pilots worldwide can be adversely affected by driving to the airport and that fatigue induced by commuting can contribute to accidents would indicate that the risk of commuting-related fatigue should be better understood and managed. The next sections of this chapter describe what the research included in this thesis aims to achieve, the practical implications of any findings, the novelty and original contribution of the thesis, and the next chapters.

1.3 Thesis aim and objectives

This thesis aims to explore the relationship between fatigue induced by long and demanding drives to the airport and pilots' levels of fatigue in flight and while driving after duty. The following three objectives were formed;

Objective one: to explore the extent to which fatigue induced by long and demanding drives to the airport affects negatively professional pilots' fatigue in flight and while driving after duty. This objective will be achieved by conducting an online survey with professional pilots. The survey will be designed based on the relevant literature and the findings of an interview with a pilot that aims to identify key aspects of the issue.

Objective two: to explore if the type of fatigue induced by driving (i.e. active or passive) subsequently affects the levels of fatigue experienced in flying and driving. Improving understanding of this effect is needed in order to develop effective intervention strategies. This objective will be achieved by conducting an experiment, where three groups will complete simulated flying followed by simulated driving tasks. Two of the groups will also complete a long, simulated drive before the flight but the cars driven will differ in the level of automation in order to cause either active or passive fatigue. Subjective and performance data will be collected to infer fatigue.

Objective three: to investigate if short bursts of physical activity can reduce the levels of fatigue experienced in the commuting cycle after a long and demanding drive. This objective will be achieved by comparing fatigue during simulated flying and driving tasks in individuals with a Private Pilot's License (PPL) between a session with and another without such breaks. Fatigue in these tasks will be measured by collecting subjective and performance data.

1.4 Practical implications

The findings of the research described in this thesis could be used to:

- Change the flight time limitation regulations, which specify the maximum flight time allowed in a duty period. Currently, the time spent on commuting to the airport is not considered when calculating these flight times.
- Influence future airline policies about commuting. At the moment, only some airlines use relevant policies.
- Help airlines to provide fatigue training through Fatigue Risk Management Systems (FRMSs) so that pilots can avoid, detect, and manage commuting-related fatigue.
- Improve the biomathematical models used to predict pilots' fatigue in flight by considering the task load of their commutes to the airport. These models are used by airline schedulers to provide quantitative and qualitative predictions of fatigue and performance in flight using mathematical equations.
- Help other professionals in high-hazard industries to manage the risk associated with commuting-related fatigue.

1.5 Novelty and original contribution

This thesis is novel in collecting data about pilots' commuting habits worldwide and exploring the extent to which fatigue induced specifically by long and demanding driving to the airport relates to pilots' fatigue in flight and while driving after duty. Moreover, it is the first to investigate the role in this issue of the type of fatigue induced by driving. In contrast to previous studies, the use of driving simulation will also allow exploring the effects of fatigue induced by travelling by one means of transport on fatigue while travelling by another means of transport. Finally, this thesis is the first to test the effectiveness of an intervention strategy for fatigue experienced in flying and driving tasks after a long and demanding drive.

Chapter 2 includes a literature review of the sleep- and task-related factors that contribute to fatigue in drivers and pilots and provides a description of the theoretical frameworks that explain fatigue causation and fatigue-related performance decrements. It also outlines the self-initiated intervention strategies for fatigue that might help commuting pilots.

Chapter 3 provides a literature review of the methods that can be used to collect data about the contributors to fatigue and measure its expressions.

Chapter 4 describes the interview with an airline pilot that aimed to identify key aspects of the issue of fatigue in pilots who commute by driving and focuses on the online survey that investigated the extent of that issue (objective 1).

Chapter 5 describes the first experiment of the thesis, which explored objective 2. This study built on the findings of the interview and the survey and extended them with experimental work.

Chapter 6 describes the second experiment of the thesis, which investigated objective 3. The methodology used and the intervention strategy tested were selected based on the relevant literature and the findings of the study of chapter 5.

Chapter 7 discusses and summarises the findings of the interview and the studies of the thesis. The limitations are outlined, recommendations for further research are made, the contributions to existing knowledge are discussed, and the practical implications of the findings are described.

Chapter 2. Contributors to fatigue, theoretical frameworks, and self-initiated intervention strategies

Fatigue is usually the output of the interaction of several contributory factors, so this chapter starts with a literature review of their role in fatigue. A description of the theoretical models that explain fatigue induction due to the time on task and task load and the theories about how fatigue impairs performance follows. Reviewing these theories is important because the focus of this thesis is on active fatigue induced by driving and on how that affects fatigue and performance in subsequent flying and driving tasks. Finally, this chapter outlines the self-initiated intervention strategies for fatigue that pilots and drivers can use. The contributors to fatigue and the self-initiated intervention strategies that will be described in this chapter are shown in Figure 2.1.

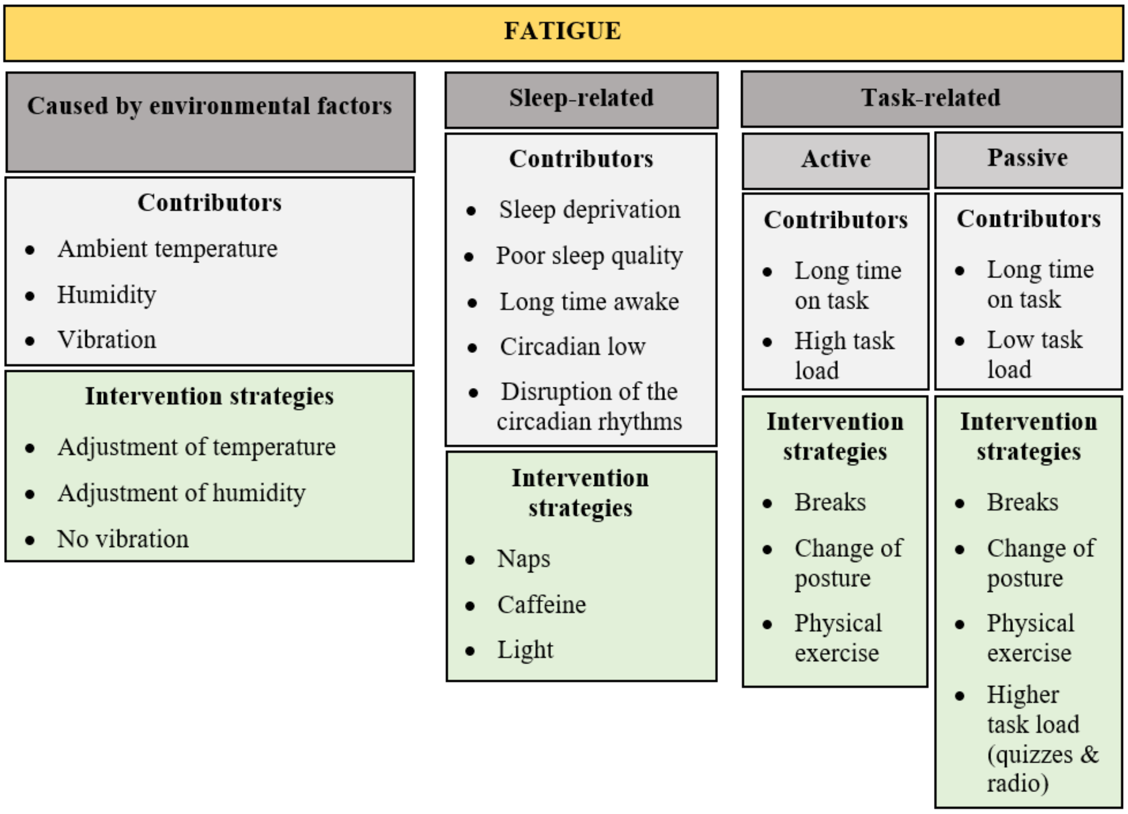


Figure 2.1 The contributors to fatigue and the self-initiated intervention strategies for it.

2.1 Factors that contribute to fatigue

Fatigue in pilots who commute by driving is expected to be the outcome of the interaction of various factors. For example, pilots may fly fatigued because of both task-related fatigue caused by a long drive of high traffic density to the airport and sleep-related fatigue due to inadequate sleep the night prior to their flight. At the same time, vibration in flight can increase the levels of fatigue experienced even further. When researching commuting-induced fatigue in pilots, the concurrent effects of various factors may inhibit drawing conclusions about which factor or combination thereof contributed to fatigue. Reviewing the environmental, sleep-related, and task-related factors that have been found to contribute to fatigue will help to understand how these can affect commuting pilots, manipulate the factors of interest (i.e. time on task and task load), and control for the environmental and the sleep-related ones in the experiments of this thesis.

2.1.1 Environmental factors

High ambient temperature is one of the environmental factors that can contribute to fatigue. There is evidence that the optimal ambient temperature for sedentary work is 22-25°C (Landstrom et al., 2000). Pilots have reported that high ambient temperature induces sleepiness (Co et al., 1999), whereas drivers often open the window to reduce their levels of sleepiness (Anund et al., 2008). The adverse impact of heat on drivers' alertness was suggested in a naturalistic driving study, where McKie and O'Hanlon (1978) found that driving at 32 degrees Celsius reduced participants' self-reported levels of alertness more than driving at 20 degrees Celsius. In another study (Zhu et al., 2019), the effects of both high ambient temperatures and high levels of humidity on sleepiness were explored in a climate chamber. Zhu et al. found that activity in the delta band measured with EEG (see section 3.2.2.4) increased significantly with ambient temperature (37°C compared to 26°C and 30°C) and relative humidity (70% compared to 50%). Increases in the activity in the delta band indicate higher levels of sleepiness (Ferreira et al., 2006).

As far as vibration is concerned, studies suggest that being exposed to whole-body vibration can also contribute to increased levels of fatigue as inferred from its expressions (Azizan et al., 2016; Satou et al., 2007; Satou et al., 2009). For example, in a laboratory study of eighteen university students that sat on a chair mounted on a vibration plate, Azizan et al. found that 20 minutes of exposure to vibration resulted in an increase in the reaction times and lapses in a PVT (see section 3.2.3.1) and higher ratings on the KSS (see section 3.2.1). These findings indicate a reduction in vigilance and an increase in the levels of sleepiness respectively.

2.1.2 Sleep-related factors

Commuting pilots are expected to experience a mix of task- and sleep-related fatigue. Although the focus of this thesis is on the former, understanding the latter too will help to understand fatigue in commuting pilots better and control sleep-related fatigue in the experiments of chapters 5 and 6. The literature on the following sleep-related factors will be described in this section: driving and flying during the circadian low, the disruption of the circadian rhythms due to time-zone crossing in long flights, being awake for too long, sleep deprivation, and poor sleep quality.

Circadian rhythms are the 24-hour rhythm driven by the circadian body clock, which is a circadian pacemaker located at the hypothalamus (Klein et al., 1991). This pacemaker is synchronised to the solar day when exposing to daylight to regulate nocturnal sleep and the levels of sleepiness in the day. Besides light, the circadian body clock can also be affected by the timing of meals (Wehrens et al., 2017). Because of the circadian rhythms, individuals feel the urge to sleep at night and experience two distinct time periods in a 24-hour day with maximal levels of sleepiness: one between 4 and 5 am and another between 3 and 5 pm (Carskadon, 1989).

The literature suggests higher levels of sleepiness when driving and flying during the circadian low. For example, Lenne et al. (1997) found higher speed deviations and longer reaction times (impaired performance has been related to higher levels of sleepiness) between 2 and 6 am. Furthermore, higher rates of car accidents have been found when driving between 2 and 5 am (Connor et al., 2001) and during the mid-afternoon (i.e. 3 pm) (Pack et al., 1995). As far as pilots are concerned, Wright and McGown (2001) explored sleepiness in long-haul flights and found that sleep episodes occurred more often around 4 pm and 4 am. In contrast to driving, flying can also cause sleepiness by desynchronising the circadian rhythms when multiple time zones are crossed (i.e. jet lag) (Rosekind et al., 2000). Changing from day to night shifts and vice versa may desynchronise the circadian rhythms as well.

Besides the time of the day, commuting pilots' levels of fatigue are also expected to increase with the time they have been awake since their last sleep because that increases the pressure to fall asleep (Walker, 2017). Caldwell et al. (2004) provided evidence of that effect in a study with military pilots who completed five sessions of simulated flights over a period of 37 hours without sleep. That study found a lower accuracy in completing turns, straight and level flying, and changes of altitude in a full flight simulator and keeping a continuously moving ball at the centre of a cross in the Multi-Attribute Task Battery (MATB). The MATB is a software that simulates activities completed in flight (see section 5.3.4.1).

Not sleeping enough may contribute to being fatigued too. Although the optimal duration of sleep may differ between individuals, Hirshkowitz et al. (2015) recommended that adults should sleep at least seven hours every night. In contrast, the Federal Aviation Administration (2009) defined sleep deprivation in pilots as sleeping less than eight hours per night. Besides any personal and operational reasons for sleep deprivation (e.g. early-morning starts) (Bourgeois-Bougrine et al., 2003), the need to commute to the airport by completing long drives may also contribute to sleep deprivation in pilots. Sleep deprivation caused by driving to the airport has not been explored. However, Walsleben et al. (1999) found that other professionals who commuted by train to work by completing journeys of more than 75 minutes one-way reported a lower mean weeknight sleep of approximately 40 minutes compared to those with commutes of less than 45 minutes.

The higher likelihood of driving fatigued due to the lack of sleep has been supported in studies on drivers that found higher road accident rates when individuals were sleep deprived (Connor et al., 2001; Cummings et al., 2001; De Valck and Cluydts, 2001; Stutts et al., 2003). Irrespective of whether sleep deprivation finally contributes to accidents, there is evidence that driving performance is more likely to deteriorate when drivers have not slept enough. For example, Philip et al. (2005[b]) found in a naturalistic driving study that higher levels of sleepiness were reported at the end of the drive and the drivers crossed the road lines more often when they had slept two compared to eight hours. Sleepiness caused by sleep deprivation has also been supported in studies with pilots. For example, Le Duc et al. (1999) found that sleep deprivation in pilots related to spatial disorientation in flight, whereas Wilson et al. (2006) found longer response times to lights that changed colour in the MATB software (see section 5.3.4.1).

Besides sleep deprivation, sleep-related fatigue in driving and flying can also be caused by poor sleep quality. The role of sleep quality in fatigue in drivers was suggested by Stutts et al. (2003) who found that individuals that had been involved in car accidents were more likely to report poor quality of sleep before driving. Similar, Reis et al. (2016) found that bad sleep quality measured on a scale in a survey with pilots related to higher reported levels of fatigue the week before the study.

One of the challenges in exploring sleep-related fatigue is that inter-individual differences can affect it. To begin with, it appears that individuals differ in their response to sleep deprivation. For example, Van Dongen et al. (2004) found high inter-individual variability in the self-reported sleepiness, performance in cognitive tasks, and PVT reaction times (PVT description in section 3.2.3.1) in a laboratory study with individuals that were sleep deprived for 36 hours. They concluded that the vulnerability to performance decrements due to sleep deprivation is a trait. Rupp et al. (2012) also suggested that resilience to sleep loss is a trait characteristic based on a laboratory study, where cognitive performance was measured in total and partial sleep restriction conditions. That study found that the participants who expressed a higher vulnerability to total sleep deprivation were more vulnerable to partial sleep restriction too. People also differ in their preference for a time of the day with some being morning types (i.e. they like sleeping and waking up early) and others evening types (i.e. they like to stay awake late at night and wake up late) (Vink et al., 2001). Depending on morningness and eveningness, the times of the circadian low (and with that sleepiness too) differ between individuals.

Besides the trait-like response to sleep-related factors of fatigue, sleep-related fatigue may also be mediated by age. That is, older individuals may find it more difficult to adjust their circadian rhythms to changes in the wake/sleep cycle (Harma et al., 1994). Their sleep may also be shorter and of lower quality than in younger individuals (Ohayon et al., 2004). In addition, older drivers are more likely to fall asleep that younger ones when they receive the same amount of sleep (National Highway Traffic Safety Association, 2002). In contrast to drivers, the literature on pilots has been inconclusive regarding the link between age and fatigue. For example, Reis et al. (2016) did not find a relationship between age and the reported levels of fatigue, whereas Van Drongelen et al. (2017) found in their survey that older pilots reported higher levels of fatigue. Maybe finding a link between age and fatigue is more difficult in professional pilots than in drivers because, in contrast to drivers, pilots are required to stop flying at the age of sixty-five (European Union Aviation Safety Agency, 2017).

To sum up, operating during the circadian low, the disruption of the circadian rhythms, the extended wakefulness, sleep deprivation, and poor quality of sleep can induce sleepiness in drivers and pilots. Inter-individual differences may mediate the effect of sleep-related factors on fatigue. The focus of this thesis is on task-related fatigue, but data about sleep-related factors will be collected in the studies of the thesis to understand fatigue better. Moreover, the experiments of chapters 5 and 6 will be designed so that the effect of sleep-related factors is controlled.

2.1.3 Task-related factors

The literature on the effects of time on task and task load on fatigue is described in this section. To begin with the time on task, the literature is inconclusive regarding how long a drive should be to induce fatigue. The main reason for that is the inconsistency in the methodologies used to measure fatigue (e.g. no fatigue baselines and different metrics) and the levels of task load of the drives. In one of the relevant studies, Kecklund and Akerstedt (1993) measured fatigue in a naturalistic driving study, where the participants drove a truck of level 0 automation. SAE International (2019) described six levels of car automation (Figure 2.2) with level 0 referring to driving without any systems that provide steering or acceleration/brake support. Kecklund and Akerstedt found that alpha wave activity measured with Electroencephalography (EEG) (described in chapter 3.2.2.4) increased in the last 2-3 hours of the 10-hour drives (i.e. an indication of sleepiness) as did the subjective ratings of sleepiness (measured on the Karolinska Sleepiness Scale, see section 3.2.1.2). The effects of task load on fatigue were not explored there.

In contrast to Kecklund and Akerstedt, other researchers investigated fatigue in driving simulator studies with motorway driving of low task load. For example, Thiffault and Bergeron (2003) found that driving a car of level 0 automation as short as 40 minutes induced fatigue, which was inferred from fewer small steering wheel movements (since straight-line driving is a vigilance task, Brookhuis et al., 1994). In another study with motorway driving (Ting et al., 2008), fatigue was induced after 80 minutes as evidenced by higher ratings on the Stanford Sleepiness Scale (a Likert item ranging from ‘*very alert*’ to ‘*very sleepy*’), a higher standard deviation of speed and lateral position, more edge-line crossings, and higher means of headway from the lead car. Subjective fatigue in Ting et al.’s study was measured only at the beginning and end of the drives. Therefore, sooner increases of the levels of fatigue could have been missed since driving performance may not deteriorate until drivers are fatigued enough (Ingre et al., 2006[a,b]; Williamson et al., 2014).

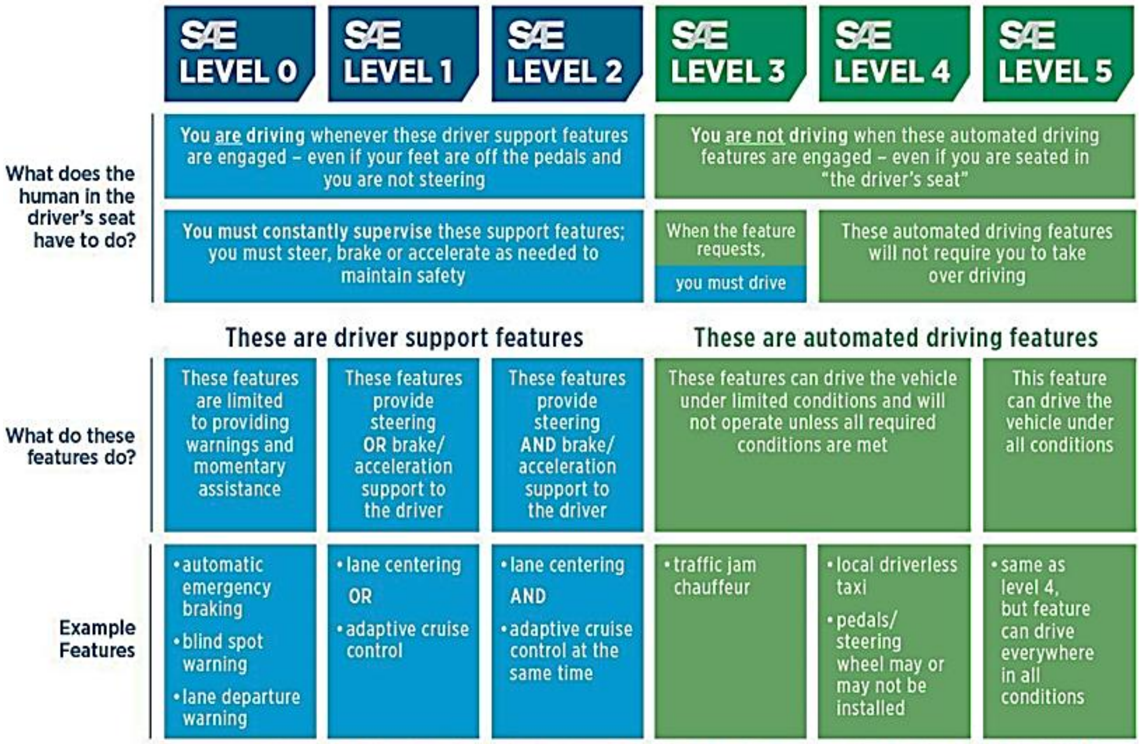


Figure 2.2 Levels of car automation (SAE International, 2019).

Vogelpohl et al. (2019) found a quicker induction of fatigue in a driving simulator study with a car of level 0 automation being driven in a low task load motorway scenario. That is, trained raters detected trends for increased levels of sleepiness in non-sleep-deprived individuals after 30 minutes. In Thiffault and Bergeron's, Ting et al.'s, and Vogelpohl et al.'s studies, low task load should have induced passive fatigue. In contrast, driving should have caused a mix of active and passive fatigue in Oron-Gilad and Ronen's study (2007), where the participants drove a combination of simulated drives of low, moderate, and high task load. In that study, the subjective ratings of fatigue increased after 45 minutes of driving. The studies above suggest that the time needed for driving to induce fatigue can differ depending on the task load of the drives.

Some researchers focused specifically on the role of task load in fatigue induction in driving. For example, Gimeno et al. (2006) suggested that driving causes fatigue when mental effort is applied in high task load driving for a long period. Driving conditions that relate to higher levels of mental workload are the higher variability of the speed of the lead car in car following (Young and Stanton, 2004), the close distances (i.e. headways) to the car in front (Winter et al., 2014), and driving in unfamiliar areas (Parkes et al., 1991), into intersections (Jahn et al., 2005) and in high traffic density (Teh et al., 2014). Frequent lane changing while driving on motorways (Teasdale et al., 2004), overtaking (Cantin et al., 2009) and high change rate of the road curvature (Richter et al., 1998) have also been found to increase mental workload. Completing secondary tasks while driving can be fatiguing too as found in a simulated driving study conducted by Matthews and Desmond (2002). In that study, two groups of participants were asked to complete the same drive (0 level of automation), but one of the groups completed a visually and mentally demanding secondary task that required paying attention to signs, letters, and numbers while driving. It was found that the latter group reported significantly higher levels of fatigue at the end of the drive and performed worse in the driving task (i.e. increased heading error).

It appears that the levels of task load experienced can make a difference in the mechanisms that cause fatigue. In a driving simulator study with conditions of either highly automated driving or driving of 0 level of automation in an environment with wind gusts, Saxby et al. (2013) found that only the highly automated driving (level 2 since the drivers were required to constantly monitor for a critical signal indicating an automation failure) caused disengagement from the driving task. Saxby et al.'s finding indicates that measuring the engagement with the driving task might be a good way of distinguishing between active and passive fatigue. Feldhutter et al. (2018) found that automated driving (level 3) can induce fatigue even after 15 minutes although inter-individual differences can affect the time needed before drivers' levels of fatigue increase. Level 3 of automation means that the vehicle drives itself without the drivers being required to constantly pay attention to the driving scene or keep their hands on the steering wheel and feet on the pedals. However, the drivers must take control over when requested. Compared to the studies with cars of 0 level of automation described above, the 15 minutes are much less, potentially because the low task load of highly automated driving causes disengagement from the task.

Some recent studies compared the speed of fatigue induction between driving of different levels of automation. In a driving simulator study of sixteen participants, Schömig et al. (2015) compared driver drowsiness between level 0 and level 2 (i.e. hands-free driving but paying attention to the driving scene was required) conditions with a drowsiness detection algorithm. The data were analysed only descriptively, but there was an indication that the levels of drowsiness increased sooner when driving was highly automated. The results of Vogelpohl et al.'s (2019) driving simulator study agreed with those by Feldhütter et al. and Schömig et al. since they suggested that highly automated driving induced fatigue (measured by trained raters) faster. In Vogelpohl et al.'s study, medium levels of fatigue were reached after 35 minutes of driving in the level 3 condition and remained moderate until the end of the drive. In contrast, the participants in the level 0 condition experienced low levels of fatigue even after 60 minutes of driving. Similar to Schömig et al. (2015), the ratings of fatigue in Vogelpohl et al.'s study were analysed only descriptively.

Besides fatigue induced by long driving to the airport, fatigue induced by flying is also of interest in this thesis because this can increase further the risk of aeroplane accidents and car crashes after duty. In some of the studies that explored the role of the time spent flying in fatigue in flight, longer times on task related to higher ratings of subjective fatigue in pilots that flew simulated (Morris and Miller, 1996) and real flights (Petrilli et al., 2006; Wright and McGown, 2001). Moreover, the adverse effects of long flights on fatigue have been inferred from deteriorations of flying performance at the end of a simulated flight in a study conducted by Morris and Miller (1996). More specifically, Morris and Miller found a lower accuracy of keeping the pre-defined altitude, heading, vertical velocity, and airspeed of the aeroplane as the flight progressed.

The role of the in-flight task load in pilots' fatigue has been mainly inferred from real flights. One of the factors that have been related to increased levels of fatigue is flying multiple sectors (Powell et al., 2007; Spencer and Robertson, 2002). A flying sector includes the travel between any two points, the takeoff and landing included. Moreover, the climb, descent and approach stages of flight (shown in Figure 2.3) usually are more demanding due to multi-tasking than the taxi and cruise stages. Therefore, these stages may be more likely to induce active fatigue due to high task load. Experiencing unanticipated problems in flight is another factor that can contribute to active fatigue because of the increased task requirements (Spencer and Robertson, 2002). For example, in case of an engine failure, pilots may need to read and follow standard operational procedures, communicate with the Air Traffic Control, and manually control the -otherwise flown by the autopilot- aeroplane. Moreover, studies have shown increased levels of fatigue when flying in adverse weather conditions (Rosekind et al., 2000) because the levels of task load increase.

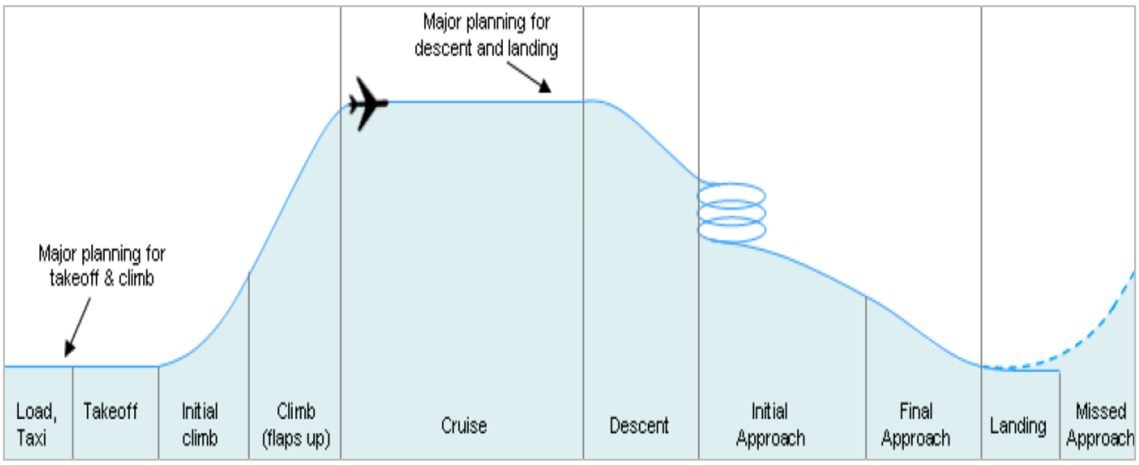


Figure 2.3 The stages of flight (Archer et al., 2012).

The low task load stages of flight can also induce fatigue. For example, Cabon et al. (1993) and Wright and McGown (2001) found that pilots experienced micro-sleeps (i.e. extremely short sleep episodes) more often during the cruise stage of flight. Although low task load could have just unmasked the effects of active fatigue caused by previous stages, it is also possible that cruise had induced fatigue. During cruise, pilots must mainly monitor the systems of the aeroplane; hence passive fatigue is more likely to be experienced due to low task load. However, active fatigue may sometimes also be induced during cruise if additional tasks are completed. For example, pilots may need to deal with an unexpected system failure.

Although all people have limited attentional resources, they may differ in how long they can handle a certain amount of task load before they get fatigued. Inter-individual differences can affect how quickly drivers get fatigued, as suggested by Nilsson et al. (1997) in a driving simulator study that explored how long participants would drive before they quit. That study found that the time spent on driving before quitting varied significantly amongst the participants (i.e. 90 - 240 minutes). Clearly identifying the role of specific inter-individual differences in fatigue induction is difficult. However, it appears that age may be one of them because the ability to divide attention declines with normal ageing (Holtzer et al., 2005). Motivation to continue investing effort in a task should also play a role based on Desmond and Hancock's theory of fatigue because that could either deplete the attentional resources more quickly in high task load conditions (i.e. active fatigue) or cause a quicker disengagement from the task (i.e. passive fatigue).

To sum up, the literature on drivers and pilots suggests that long drives and flights of either high or low task load can induce fatigue. Drawing from that literature, commuting pilots may experience a mix of active and passive fatigue in flight and while driving after duty. There is evidence that the mechanisms behind the induction of active and passive fatigue differ and this might play a role in fatigue experienced by commuting pilots. This considered, the study in chapter 5 will explore the role of the type of fatigue induced by a long drive (i.e. active or passive) in the levels of fatigue experienced in subsequent flying and driving tasks.

2.2 Theoretical frameworks

Understanding the mechanisms behind the induction of task-related fatigue (i.e. the focus of this thesis) and its adverse effects on performance is essential before exploring these links in experiments and testing intervention strategies. To this end, the relevant theories were reviewed.

2.2.1 Induction of task-related fatigue

Several models have attempted to explain the causation of fatigue based on task load and the time on task. For example, Brown (2001) stated that the mismatch between the task load and the processing capacity of individuals causes strain because individuals' attentional capacity is limited. Then, strain is experienced as fatigue if it persists. Brown's model explained fatigue causation in high task load conditions but failed to explain why completing long tasks of low task load can contribute to fatigue too (i.e. passive fatigue). Another model of fatigue was proposed by Hockey (1997) (i.e. the Compensatory Control Model). According to that, fatigue results from the strain caused by the effort to maintain performance high in conditions of high task load and lack of control. Hockey later (2011) proposed that individuals may decide to continue investing effort to a demanding task based on their goals. The continued investment of effort is expected to increase the levels of fatigue and impair performance after some time. Similar to Brown (2001), Hockey failed to explain fatigue in long tasks of low task load.

In contrast to the previous models, Desmond and Hancock (2001) explained why long tasks of either high or low task load can induce fatigue. They proposed that individuals attempt to modify their behaviour in response to environmental inputs by following attentional strategies that help them to maintain optimal performance. Individuals have limited attentional resources, which means that after some time of completing demanding driving or flying tasks, their ability to pay attention to the environment reduces. According to Desmond and Hancock, it is this reduction of attention that causes active fatigue. In turn, the depletion of attention associated with active fatigue leads to the reduction of sampling of the environment. This reduction may cause a lower frequency of steering and speed adjustments in driving or fewer control inputs to the joystick and thrust in flying. In contrast to conditions of increased task load, the proportion of time spent on paying attention to the environment decreases in low stimulation conditions or when individuals perceive a threat to performance or discomfort. In those cases, attention is mainly directed towards the self. This direction of attention causes fatigue (i.e. passive) because it is difficult to sustain the direction of attention towards the self for long periods. At the same time, investing fewer attentional resources towards the environment is expected to impair driving and flying performance in commuting pilots because important information may be missed.

Desmond and Hancock's model will be used in this thesis to explain why completing long and demanding drives to the airport can induce active fatigue and impair performance in flight and while driving after duty. Drawing from that model, long and demanding drives could deplete pilots' attentional resources. In turn, they might fly fatigued unless they reported their inability to start their duty or used an effective intervention strategy for fatigue before flight. Flying could, then, induce fatigue too, increasing the risk of aeroplane accidents and car crashes after duty. A long drive of low task load to the airport could also induce fatigue, but, in that case, pilots might reserve some attentional resources to invest in flight. This expected difference between active and passive fatigue will be explored in the experiment of chapter 5 in order to improve understanding of the role of the type of fatigue induced by driving in the levels of fatigue experienced in subsequent flying and driving tasks.

2.2.2 The link between fatigue and performance decrements

The need to manage fatigue in driving and flying stems from the fact that being fatigued increases the risk of performance decrements. Two of the theories that explain the link between fatigue and performance are described in this section. Human information processing theories treat human processing capacity as limited (Kahneman, 1973; Posner, 1980). Kahneman (1973) suggested a model of serial processing, where there is a single undifferentiated capacity from which the limited resources are dynamically shared between the stimuli that need to be processed during the completion of a task. A limitation of Kahneman's theory is the difficulty in explaining how two tasks can be perfectly time-shared when stimuli of different types (e.g. visual and auditory) need to be processed at the same time.

Wickens (1984) overcame the difficulty in explaining multi-tasking by proposing the multiple resource theory. Figure 2.4 shows a graphical representation of the human information processing model proposed by Wickens. In contrast to Kahneman, Wickens (1984) distinguished between the concepts of capacity and resources; capacity was defined as the upper limit of processing capability and resources as the mental effort applied to improve processing efficiency. In the multiple resource theory, the resources are limited but differentiated based on the modality of the stimuli (e.g. auditory or visual). That is, two concurrent tasks can be perfectly completed when resources of different modalities are needed. Since driving and flying require multi-tasking, this model will be used in this thesis to explain the effects of fatigue on commuting pilots' performance.

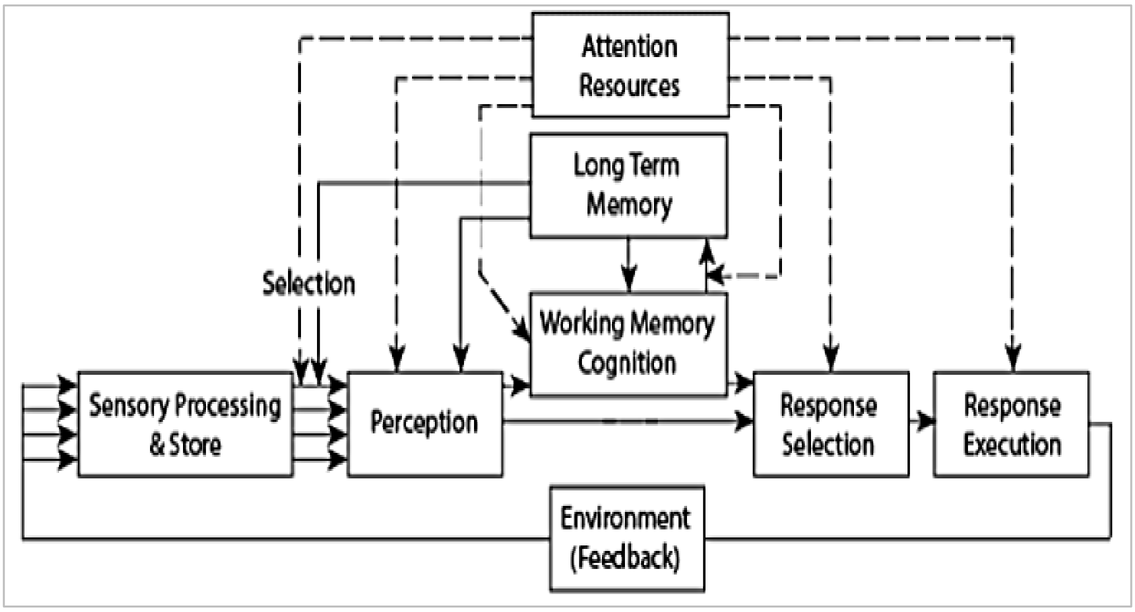


Figure 2.4 Wickens' model of human information processing (redrawn from Karanikas, 2016).

Drawing from Wickens's model, information processing starts with the detection of information from various sources (e.g. auditory and visual warnings). The impression of the information is stored briefly after it has ceased before it is identified and interpreted (perception). Then, the information that comes from the senses is stored in working memory before past experiences are drawn from the long-term memory to give it a meaning. This process is not followed in automatic processing, where response selection occurs directly after perception. Automatic and non-automatic perception are followed by the selection of responses among alternative options (i.e. decision making). The actions decided are, finally, executed and information processing continues based on the feedback received through a continuous feedback stream. As can be seen in Figure 2.4, attention affects all the components of the information processing system. Depending on the situation, attention can be driven by the salience of the stimuli (i.e. in bottom-up processing) or knowledge, goals, and expectations (i.e. in top-down processing). In driving and flying, bottom-up and top-down processing are often used interchangeably, and the one can activate the other.

Individuals' attentional resources are limited, so they need to be invested flexibly to achieve optimal performance. There are three types of attention used to optimise the investment of the attentional resources: selective, divided, and sustained. Selective attention is the ability to focus on certain information while filtering other. Selective attention in pilots and drivers can improve with training (Chapman et al., 2002; Crundall et al., 2012; Ottati et al., 1999). Divided attention is the ability to attend to more than one stimulus or task concurrently. In contrast to selective attention, where a part of the attentional resources is used while other information is automatically filtered, all the attentional resources are invested in divided attention. Divided attention is essential in driving and flying because these tasks require multi-tasking. As the attention needed for the task at hand exceeds a certain limit, performance deteriorates. Experience can help drivers and pilots to perform concurrent tasks because highly over-learned tasks that are completed automatically (e.g. emergency procedures in pilots) are less susceptible to mental fatigue than tasks that require the voluntary allocation of attentional resources (Boksem et al., 2005; Lorist et al., 2000; 2005).

Sustained attention (will be called vigilance in this thesis) is '*the degree of activation of the central nervous system*' (Campbell and Bagshaw, 2002 p.103), which affects '*the subject's readiness to detect rarely and unpredictably occurring signals over long periods of time*' (Sarter et al., 2001, p.146). Sustained attention is mentally demanding as supported by Warm et al. (2008), who found that brain activity and blood flow in the brain reduced after long periods of time. These reductions relate to fatigue (Ko et al., 2017). Research has also linked sleep deprivation to decrements in sustained attention (Lim and Dinges, 2008).

The literature consistently suggests that fatigue can impair divided (Boksem et al., 2005), sustained (e.g. Guo et al., 2016; Lorist et al., 2000), and selective attention (Harrison and Horne, 2000; Van der Linden and Eling, 2006). As can be seen in Wickens's model, attention plays a central role in information processing since it affects all the other components when processing is non-automated. Therefore, attentional issues caused by fatigue can impair performance (i.e. response execution in Wickens's model) not only directly but also indirectly by deteriorating the other components of the system.

Automated processing is less vulnerable to the effects of fatigue because of the reduced contribution of attention (Lorist and Faber, 2011). Therefore, training and experience might help a lot commuting pilots to avoid the adverse effects of fatigue on performance in highly automated flying tasks (e.g. making the pre-flight checks) but less in the non-automated ones (e.g. control of the joystick). Similar, driving training and experience could help pilots to prevent fatigue-related performance decrements while driving after duty because a great part of driving is completed automatically (e.g. changing gears). However, training and experience may not be that helpful when the driving tasks require the investment of a significant amount of attentional resources. For example, more attentional resources are needed when driving in high traffic density because paying attention to the position and speed of the car is needed to avoid crashing.

In summary, drawing from Wickens's model of human information processing, active fatigue induced by driving to the airport can impair pilots' performance in flight and while driving after duty by affecting all the components of the information processing system. Moreover, attention plays an important role in the link between fatigue and performance. Although automaticity could help commuting pilots to avoid fatigue-related performance decrements in certain tasks, the less automated ones could be affected more. The focus of the experiments described in chapters 5 and 6 is on the effects of fatigue on the non-automated tasks, such as completing turns while flying and maintaining a target speed while driving.

2.3 Self-initiated intervention strategies

If the literature described so far in this chapter applied to commuting pilots, they could end up being fatigued in flight and while driving after duty due to completing long and demanding drives to the airport and flying because these could deplete their attentional resources. One of the objectives of this thesis is to investigate if a self-initiated intervention strategy could help pilots to reduce their levels of fatigue in that case.

Several intervention strategies can be used to avoid high levels of fatigue in drivers and pilots. Some of them are technical solutions, such as in-car systems that detect high levels of fatigue by analysing steering inputs. Others are applied at an organisational level, such as the biomathematical models that are used by airlines to predict fatigue in-duty. Self-initiated intervention strategies for fatigue can be used too. This thesis focuses only on them and, more specifically, on the self-initiated strategies that pilots can use while at the airport. The reason is that certain factors can make the application of intervention strategies while flying difficult. For example, napping in flight may not be allowed by regulations and taking a break may be impractical in high task load stages of flight. Moreover, pilots may be at an increased risk of fatigue-related performance decrements in flight until they decide to use an intervention strategy. Intervention strategies during driving will not be tested in the experiment of chapter 6 either because drivers could be at an increased risk of fatigue-related car accidents until an intervention strategy was used.

Some studies suggest that driver sleepiness caused by high temperatures can reduce with cool air blowing for short periods (Anund et al., 2008; Schmidt and Bullinger-Hoffmann, 2017; Wyon et al., 1996) but others have questioned the effectiveness of this strategy (Reyner and Horne, 1995). Because of the inconsistent findings, the adjustment of the ambient temperature will not be used as an intervention strategy for fatigue in this thesis. Nevertheless, the room temperature will be controlled in the studies described in chapters 5 and 6 to control for its potential effect on fatigue. Similar, the potential effect of vibration will be controlled by not using a driving and a flight simulator with axial movement. No specific intervention will be used to control for the effect of humidity, but the studies will be run in a properly heated and ventilated office environment. It should be noted that controlling for the effect of vibration and humidity on fatigue are not intervention strategies that drivers and pilots could use in real life, so these will not be tested in the experiment of chapter 6.

The literature consistently suggests that naps can reduce individuals' levels of both task- and sleep-related fatigue (Hayashi et al., 2004; Tietzel and Lack, 2002) with the benefits of short naps (e.g. 15 minutes) including improved cognitive performance and reduced subjective sleepiness. The effectiveness of naps as an intervention strategy for fatigue has been supported in the context of driving (Horne and Reyner, 1996; Leger et al., 2009; Watling et al., 2014). For example, Horne and Reyner (1996) found in a study with simulated driving that a nap break reduced the subjective levels of sleepiness, the impairment of driving performance, and reduced the electroencephalographic (EEG) signs of sleepiness (see section 3.2.2.4 for information about the EEG). Although napping in flight (either in designated bunk facilities or in the cockpit) can also help pilots to reduce their levels of fatigue (Cabon et al., 1993; Petrie et al., 2004; Rosekind et al., 1994; Wright and McGown, 2001), in-flight intervention strategies are out of the scope of this thesis. Napping before and after flights will not be used in the experiment of chapter 6 either because that would require using sleep facilities to resemble the airport rest facilities. Such facilities were not available. In addition, napping could reduce both task- and sleep-related fatigue, so the benefits of this intervention strategy specifically for task-related fatigue induced by driving and flying could not be identified.

The literature on alertness-enhancing medications, such as modafinil, amphetamines, and caffeine pills, will not be described in this section because pilots are generally not allowed to use them. On the other hand, drinking caffeinated drinks (e.g. coffee and energy drinks) is a common way of keeping alertness high in both driving and flying when fatigue is caused by sleep-related factors (May and Baldwin, 2009). Drinking a cup of caffeinated coffee has been found to reduce the levels of fatigue in studies with sleep-deprived drivers (Mets et al., 2012; Reyner and Horne, 2000; 2002). For example, in a driving simulator study, Mets et al. (2012) found lower SDs of the lateral position and speed (i.e. indications of reduced levels of fatigue) and lower levels of subjective sleepiness compared to a placebo group when the participants had drunk coffee. Drinking caffeinated beverages has also been found beneficial for pilots (Caldwell et al., 2009; Caska and Molesworth, 2007). For example, Caska and Molesworth (2007) found that caffeine in low doses improved sleep-deprived pilots' performance when they completed simulated instrument landing approaches. Drinking coffee was not used as an intervention strategy for fatigue in this thesis because it is effective against sleep-related but not necessarily against task-related fatigue. That is, caffeine reduces sleepiness, but task-related fatigue is not always expressed as sleepiness.

Exposure to bright light at night is an intervention strategy for fatigue that has been mainly tested in pilots. There is evidence that exposure to light at night relates to increased vigilance and reduced levels of sleepiness (i.e. expressions of fatigue) (Badia et al., 1991; Cajochen et al., 2000). Exposure to light in the day can also be effective against sleep-related fatigue (Phipps-Nelson et al., 2003). It appears that specifically exposure to blue light reduces sleepiness because it can suppress melatonin and the phase changes associated with the circadian rhythms (Caldwell et al., 2009). Cajohen et al. (2000) found that subjective alertness increased when individuals were exposed to room light (90 – 180 lux). Although adjusting ambient light might help with sleep-induced fatigue, it is often impractical as an intervention strategy, especially during the day due to daylight. The adjustment of light will not be used as an intervention strategy in this thesis because this is effective only against sleep-related fatigue.

Taking a short break at the airport before and after flights could be a practical way of reducing fatigue induced by driving to the airport and flying. Pilots spend some time at the airport in order to get from their car to the location where they report for duty. Similar, they spend some time to walk to their car after duty. This time could be used to reduce their levels of fatigue and, thus, avoid impaired flying and driving performance. As will be explained in this section, commuting pilots might benefit from breaks that combine physical activity with postural change and mental disengagement. The relevant literature is described in the following paragraphs.

Transport organisations have highlighted the importance of rest breaks for the mitigation of the effects of fatigue. For example, the National Transportation Safety Board (2019[b]) recommends that drivers should take breaks that include stopping the car, stretching, or staying at a hotel when driving long distances. Aviation authorities have regulated the requirement for rest breaks before and after duty. For example, EASA (European Union Aviation Safety Agency) (Commission

Regulation, 2014, p.28) requires that pilots are given a minimum rest period of *at least as long as the preceding duty period, or 10 hours, whichever is greater*' between consecutive duties. Additionally, in some cases (e.g. on standby duty), airlines must provide suitable accommodation at the home base so that pilots get rest. The literature suggests that drivers often use breaks to reduce their levels of fatigue. For example, Anund et al. (2008) found in a study with more than 3000 drivers in Sweden that more than half of them stopped driving and went for a short walk to reduce their levels of sleepiness. Pilots can also have in-flight breaks depending on the length of flights (Commission Regulation, 2014).

The effectiveness of rest breaks on reducing fatigue has been suggested in studies with drivers and pilots. Although exploring the effectiveness of breaks used during driving and flying is out of the scope of this thesis, reviewing the relevant literature can help to also identify the potential benefits of breaks after driving and flying tasks. Additionally, if the study of chapter 6 suggested that the benefits of breaks before and after flying may not last for long enough, these breaks could be combined in the future with breaks during flying and driving.

To begin with drivers, Phipps-Nelson et al. (2010) examined whether breaks reduce non-sleep-deprived drivers' levels of task-induced fatigue in an experiment, where the participants had 1-hour breaks between four 2-hour simulated driving sessions. During the breaks, they completed self-reported scales of sleepiness and fatigue, and a Psychomotor Vigilance Task (PVT) while remaining seated. It was found that the SD of the lateral position of the car (i.e. a higher SD can indicate higher levels of fatigue) and the ratings of sleepiness and fatigue reduced after the breaks. As far as flying is concerned, Neri et al. (2002) explored whether breaks reduced airline pilots' levels of fatigue induced by flying. In that experiment, two groups of pilots flew a six-hour simulated flight. The experimental group had five 7-minute breaks during the flight, whereas the control group took only one. The pilots were taken to another room during their breaks, where they remained standing and discussed with an experimenter. Neri et al. found lower self-reported levels of sleepiness and quicker reaction times on a PVT in the group that took more breaks. Nonetheless, these positive effects were not evident after 40 minutes.

In some cases, taking breaks may be ineffective against fatigue. For example, Horne and Reyner (1996) conducted an experiment, where the participants had a 30-minute break (sitting on the wheel) in the middle of a 2-hour simulated drive. They found no differences in the ratings of self-reported sleepiness between before and after the break. This finding could be explained by the fact that the individuals remained seated, so they did not benefit from the alerting effect that changing posture can have.

Research suggests that maintaining wakefulness is more difficult when lying down compared to sitting or standing (Aeschbach et al., 1994; Bonnet, 2000). The effects of postural change on alertness are not fully understood, but it is believed that it affects the hemodynamic mechanisms and the body core temperature. Standing up appears to have an alerting effect because it increases blood pressure and heart rate, and activates the central nervous system (Caldwell et al., 2003; Goldstein et al., 1994). Caldwell and Roberts (2000) supported the link between body posture and alertness by finding that standing up reduced the theta and delta waves activity (known to relate to sleepiness) measured with EEG (see section 3.2.2.4) in sleep-deprived individuals. Body temperature has also been found to increase when standing upright. For example, Krauchi et al. (1997) found that higher body temperature when standing upright was accompanied by lower levels of subjective sleepiness measured on the Karolinska Sleepiness Scale (KSS) (see section 3.2.1.2). The literature on body posture considered, the participants in the experiments described in this thesis sat at an upright position to maintain wakefulness. Moreover, in the experiment of chapter 6, the participants stood up during the breaks between the simulated driving and flying tasks to increase their levels of alertness.

Besides changing posture from seated to standing, people can also benefit from physical exercise. This was suggested by Liang et al. (2009), who conducted a study with two groups that drove for 2 hours. One of the groups had a 15-minute break that included stretching and alternate foot stepping after the first hour of driving. The group that took the break reported being less fatigued after driving. This finding suggests that the effects of breaks that include exercise of low to medium intensity can last for up to 60 minutes. Loy (2013) supported the beneficial effect of breaks with exercise of low to moderate intensity and a duration of more than 20 minutes in their systematic review. Studies that used cycling as exercise suggested that performance in reaction time tasks can improve when the exercise causes a slight increase of heart rate (i.e. to between 90 and 120 beats per minute) (Hogervorst et al., 1996). However, exercise can impair performance if it is too intense (i.e. heart rate of more than 130 beats per minute) (McMorris and Keen, 1994). In another review, Abd-Elfattah et al. (2015) suggested that the relationship between cognitive performance and physical exercise is an inverted U curve. That is, the effectiveness of exercise depends on its duration and intensity with too long/too short durations and too high/too low intensiveness being ineffective against cognitive performance.

Increasing task load can help with task-related fatigue but only when this is caused by low task load (i.e. passive fatigue). Based on Desmond and Hancock's model (section 2.2.1), increased task load causes active fatigue because the limited attentional resources are depleted. Therefore, increasing task load further would not reduce the levels of fatigue in that case because no attentional resources would have been left to invest in the task. On the other hand, low task load conditions induce fatigue because individuals disengage from the task at hand. This means that some attentional resources are reserved, so increasing task load can make individuals feel alert again. The benefits of an increase in task load only in low task load conditions was suggested by

Oron-Gilad et al. (2002). According to them, when task load is high, improving drivers' fitness (e.g. with breaks) should be selected instead of an increase of task load.

The alerting effect of increases of task load in low task load conditions has been mainly supported in experiments with driving of 0 level of automation (see Figure 2.2) (Gershon et al., 2009; Oron-Gilad et al., 2002). For example, Oron-Gilad et al. (2002), conducted an experiment, where the participants completed a long drive on a straight rural road twice; while answering trivia questions and without any alertness maintaining tasks. They found that participants' heart rate variability increased significantly (i.e. an indication of increased alertness, see section 3.2.2.2) at the last stage of the drive only in the session with the trivia questions. Hence, they concluded that fatigue was suppressed by mental activation.

More recently, the role of secondary tasks on passive fatigue was explored in highly automated driving. Schömig et al. (2015) measured the number of eye-lid closures of participants who completed two simulated driving sessions in a car with automated lateral and longitudinal control (level 2). In one of the sessions, the individuals completed a quiz task while driving. The researchers found that the levels of drowsiness were lower in the session with the quiz. Similar, Jarosch et al. (2019) found in a study of 66 participants with simulated motorway driving that PERCLOS (Percentage of Eye Closure) (see section 3.2.2.1) and the ratings on the KSS (section 3.2.1.2) were lower for those who completed a quiz task while driving.

Turning on the radio is another common intervention strategy for fatigue in driving (Anund et al., 2008). Cummings et al. (2001) interviewed drivers and found that the risk of car accidents was lower for drivers who played the radio. The effectiveness of this countermeasure has also been suggested in studies with driving simulators and naturalistic driving (Reyner and Horne, 1998; Schwarz et al., 2012), but not in flight because this is not allowed. The beneficial effects of listening to the radio on fatigue in driving can be attributed to the increase of task load in low task load driving conditions (thus when fatigue is passive) (Oron-Gilad et al., 2002). This intervention strategy was not used in the study of chapter 6 because the focus of the thesis is on strategies that can help with active fatigue instead.

The literature on intervention strategies considered, it appears that taking breaks before and after flights might be a practical and effective way of reducing the levels of task-related fatigue experienced in the commuting cycle when pilots complete long and demanding drives to the airport. Although the effectiveness of short breaks is questionable, it may be boosted by combining them with changes in body posture and physical exercise of medium physical intensity. Increasing task load during these breaks is not expected to be effective for active fatigue induced by driving and flying. These conclusions will be used to design the break that will be tested in the experiment of chapter 6. This thesis explores the relationship between fatigue induced by long and demanding driving and pilots' fatigue in flight and while driving after duty. To achieve that, it is important to measure fatigue while controlling for the effect of factors other than the task-related ones. Chapter 3 provides a description of how this can be achieved.

Chapter 3. The comparison and contrast of the research regarding the methods that can be used to collect fatigue-related data.

Exploring task-related fatigue requires identifying the role of task load and time on task while controlling for the effect of other contributory factors. This will be achieved by following three approaches. The first one is to control for the effect of the contributors to fatigue through study design, for example, by running half of the trials in the morning and the others in the afternoon to counterbalance the effect of the time of the day. This approach will not be described in this section, but more details can be found in the methods section of chapters 5 and 6. The second approach is the collection of data about the contributors to fatigue. No methods of collecting data about the environmental contributors, the time on task, and the time awake since last sleep will be described in this section because these are controlled by study design or (about the time awake) by asking directly the individuals. The third approach is to measure the expressions of fatigue. The types of data that can be collected about the contributors to fatigue and its expressions are shown in Figure 3.1. More information about these data and the methods that can be used to collect them can be found in sections 3.1.1 to 3.2.3.2. Appendix A summarises the application of the three approaches in the thesis.

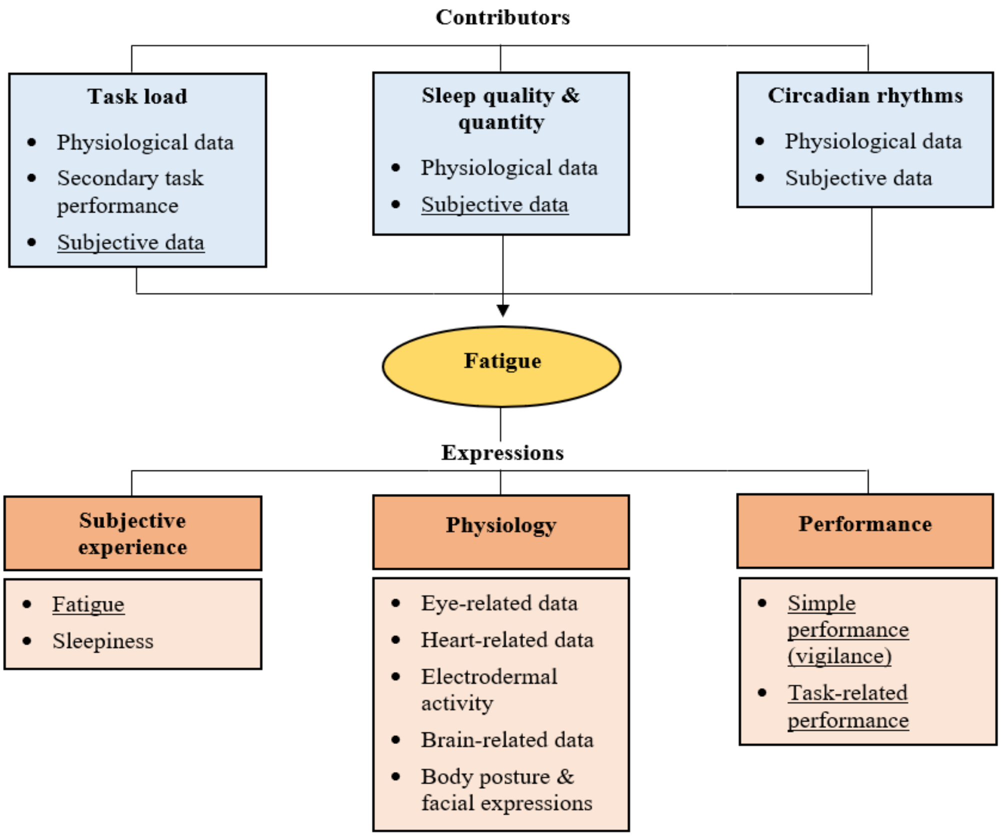


Figure 3.1 The types of data that can be collected about the contributors to and expressions of fatigue. The data that were collected and will be described in the thesis are underlined.

3.1 Data about the contributors to fatigue

Sections 3.1.1 and 3.1.2 include information about how researchers can collect data about task load and sleep-related factors respectively in order to explore how these contribute to individuals' levels of fatigue.

3.1.1 Task load

The driving and flying tasks that participants complete in studies (i.e. actual task load) are selected in the study design stage. However, it is essential in fatigue studies to collect data about the difficulty of individuals in completing tasks in order to explore the mechanisms behind fatigue induction and identify whether active or passive fatigue is caused. Then, appropriate intervention strategies for fatigue can be developed. This considered, this difficulty (called 'task load' in this thesis) will be measured in the studies described in chapters 4 and 5.

Changes in task load can be inferred by detecting changes in physiology (e.g. gaze dispersion - Louw and Merat, 2017) and performance in secondary tasks (e.g. answering questions) (Mehler et al., 2011). Although blinking behaviour and heart rate (which are affected by task load, see sections 3.2.2.1 and 3.2.2.2 respectively) will be measured in the experiments of this thesis, these metrics can also be affected by factors other than task load (e.g. fatigue and stress) and can vary significantly between participants. Therefore, using them to conclude about any differences in task load between or within groups would be difficult. Gaze dispersion will not be measured either because slippage of the eye-tracking device used in the study of chapter 5 could affect the accuracy of these data. Finally, using secondary tasks during the trials could potentially cause additional fatigue, thus, hindering the exploration of fatigue induction due to driving and flying.

The difficulty in inferring task load from physiology and performance considered, task load will be measured subjectively in this thesis. Subjective scales of task load are commonly used because they are quick and simple to administer. Moreover, in contrast to the physiological and performance data, they provide information about the difficulty of a task between two extremes. A disadvantage of these scales is that the ability to recall information can affect the ratings in long tasks. To this end, the subjective scale of task load will be administered often during the driving tasks of the study in chapter 5. Another disadvantage of the task load scales is that the perception of the difficulty of a task may vary amongst study participants (Da Silva, 2014).

Some of the task load scales are unidimensional, which means that they measure only one dimension of task load (e.g. mental workload or mental effort). A commonly used unidimensional task load scale is the Rating Scale Mental Effort (RSME) (Zijlstra, 1993), where individuals rate the amount of the mental effort they invested in a task on a line ranging from 0 to 150. The RSME is sensitive in measuring mental effort (Verwey and Veltman, 1996) and is increasingly used in traffic research (Da Silva, 2014). Another unidimensional task load scale is the Modified Cooper-Harper scale (MCH) (Wierwille and Casali, 1983), which measures task load in flight by using a decision tree. The MCH has shown good correlation with the NASA-TLX scale (described below) (Mansikka et al., 2019). Although there is some evidence that the unidimensional task load scales can give a valid measurement of task load, they are too simple to measure the complexity of task load (e.g. mental demand, physical demand, and effort). Therefore, a multidimensional scale of task load will be used in the experiment of chapter 5.

The most outstanding multidimensional task load scales are the SWAT (Subjective Assessment Technique) (Reid and Nygren, 1988), the Workload Profile (Tsang and Velazquez, 1996), and the National Aeronautics and Space Administration Task Load Index (NASA-TLX) (Hart and Staveland, 1988). The SWAT evaluates three task load dimensions: time load, mental effort load, and psychological stress load. It has good test-retest reliability, validity (as evidenced by comparing it to the NASA-TLX ratings), and sensitivity (Rubio et al., 2004). Sensitivity refers to '*the power to detect changes in task difficulty*' (Rubio et al., 2004, p.63). The SWAT does not measure physical task load, which is of interest too in the study of chapter 5. On the other hand, the Workload Profile includes eight sub-scales, on each of which individuals rate the amount of attentional resources they invested while completing a task. The Workload Profile has shown better sensitivity than the SWAT and the NASA-TLX (Freard et al., 2007; Rubio et al., 2004). However, the Workload Profile will not be used in this thesis because it has no dedicated effort item (that is needed to differentiate between active and passive fatigue).

The NASA-TLX evaluates six dimensions; physical demand, mental demand, temporal demand, effort, frustration, and performance. Subjects rate these dimensions on 100-point range scales with 5-point steps, where descriptions accompany the extremes (e.g. '*very low*' and '*very high*' mental demand). Then, individuals rate the importance of each dimension on their perceived task load and the ratings on each dimension are used to produce an overall workload score. Some researchers do not follow the weighting procedure due to complexity and prefer to calculate the total score by summing the ratings on the six items (i.e. Raw TLX). The Raw TLX has been found equally sensitive to the weighted scoring of the NASA-TLX (e.g. Moroney et al., 1995). It has also been suggested that researchers should examine the ratings on each of the items separately in order to get a more accurate picture of task load (Galy et al., 2017).

The NASA-TLX is sensitive in showing small variations of mental workload (Roskam et al., 2002), it has been validated against heart-related metrics and other subjective scales of task load (e.g. the Modified Cooper Harper, Wierwille and Casali, 1983) (Lee and Liu, 2003; Mansikka et al., 2019) and has good test-retest reliability (Xiao et al., 2005). It also allows the measurement of several aspects of task load, the physical, mental, and effort ones included. These aspects are of primary interest in the study described in chapter 5 to explore the fundamental aspects behind fatigue induction. This considered, the NASA-TLX scale and, more specifically, the raw NASA-TLX version (for increased practicality) will be used in that study. Besides the total scores calculated, the ratings on the effort scale will be examined separately to identify the type of fatigue induced by driving (i.e. active or passive) because a lower level of engagement with the driving task is expected in passive fatigue.

3.1.2 Sleep-related data

Collecting data about sleep duration, sleep quality, and circadian disruption is useful in fatigue studies because it can help to identify if participants experience sleep-related fatigue. Section 3.1.2 describes the methods that can be used to collect these data.

3.1.2.1 Sleep duration

Sleep duration can be measured by collecting physiological and/or subjective data. Polysomnography (PSG) is a test conducted in a sleep laboratory, where participants spend the night wired with electrodes and under the supervision of a sleep technician. This method is considered the 'gold standard' in the measurement of sleep duration. It is advantageous in that sleep and wake periods are inferred by collecting various types of data (i.e. EEG, muscle activity, heart-related metrics, respiratory function, and eye movements). However, participants must sleep in a sleep laboratory for one to two days, which can make recruitment difficult and distort their sleep patterns (Tonetti et al., 2008). This considered, PSG will not be used in the experiments of this thesis.

Actigraphs may also be used to identify if individuals are sleep deprived. These are usually wrist-worn devices (headbands also exist) that record body movements with an accelerometer. Participants need to wear actigraphs for one or more days before the experiments in order to collect data, which are analysed to infer periods of sleep and wake from activity and inactivity. Actigraphy is a validated (McCall and McCall, 2012; Van de Water et al., 2011) and non-intrusive method. Nonetheless, it demonstrates a lower specificity (i.e. the ability to detect wake correctly) than the PSG (De Souza et al., 2003; Marino et al., 2013) because staying awake and motionless can be interpreted as being asleep. At the same time, small body movements that generally occur while sleeping may be classified as being awake (Blackwell et al., 2008). Furthermore, using actigraphs to measure sleep duration before the experiments described in this thesis would require resources not available. Therefore, sleep duration will not be measured in this thesis with such devices.

Sleep diaries are self-completed daily logs that are used over several days or weeks to record sleep-wake patterns. They are simple to administer and studies have shown a good correlation between sleep duration measured with sleep diaries and physiological measures, such as actigraphy (Kawada, 2008) and PSG (Kushida et al., 2001). A disadvantage of sleep diaries is that the recorded sleep duration may not be accurate due to the retrospective collection of data since people often over-estimate sleep duration (Lauderdale at al., 2016). Underestimation of the sleep duration is also likely, especially in people with sleep disorders (Fernandez-Mendoza, 2011). Moreover, sleep diaries require a high level of engagement from the participants. This considered and because the main focus of the thesis is not on sleep-related fatigue, sleep diaries will not be used in the studies of this thesis.

As an alternative for the measurement of sleep quantity, the participants in the survey described in chapter 4 will be asked about their usual duration of sleep before their commutes to the airport. It has been suggested that single questions about the amount of sleep people usually get can be imprecise compared to sleep diaries and polysomnography (Miller et al., 2015; Silva et al., 2007). For example, Silva et al. found that people overestimated the duration of their night sleep by approximately 1 hour compared to polysomnography when they provided their habitual sleep duration. In the same study, a single question about sleep duration of a specific sleep period answered directly after it gave more accurate sleep durations but still approximately 15 longer than PSG. In contrast, Signal et al. (2005) suggested that a single question about sleep duration directly after a sleep period can measure sleep duration quite accurately by finding a good correlation between subjective estimates of sleep duration and PSG. Nevertheless, there were mean differences in sleep duration between the two methods that varied between 36 minutes less and 20 minutes more when measured subjectively than with PSG.

Although measuring sleep duration with a single question after sleep may be less accurate than sleep diaries and PSG, using a single question is quick and does not hinder recruitment. In addition, the aim of collecting sleep duration data in this thesis was to detect cases of sleep deprivation that could result in sleep-related fatigue during the trials rather than explore the accuracy of the measurement of sleep duration per se. Therefore, a single question will be used to collect data about the usual sleep duration in the survey of chapter 4 and sleep duration the night before the trials in the experiments of chapters 5 and 6.

3.1.2.2 Sleep quality

Ohayon et al. (2017, p.11) defined good sleep quality as sleep of '*shorter sleep latencies, fewer awakenings, reduced wake after sleep onset, and higher sleep efficiency*'. Sleep latency refers to the amount of time needed to go from being awake to sleeping, whereas sleep efficiency to the percentage of time spent on sleeping between the sleep onset and the final awakening when any time spent awake during this period is removed. PSG, actigraphy, and sleep diaries can be used to infer sleep quality by collecting data about the time spent in bed before and after sleep and awakenings. As mentioned in the previous sections, using these methods to collect sleep-related data will be avoided in the studies of this thesis. As an alternative, data about participants' sleep quality the night prior participation will be collected on the day of the trials in the experiments described in chapters 5 and 6. This method could help to identify if fatigue experienced in the experiments of this thesis was caused by poor quality of sleep.

Several sleep quality questionnaires can be used, but the Pittsburgh Sleep Quality Index (Buysse et al., 1989) and the Groningen Sleep Quality scale (Mulder-Hajonides Van Der Meulen et al., 1980) are probably the most commonly used ones. The Pittsburgh Sleep Quality Index will not be used in the studies of this thesis because it refers to sleep in the last month. As a result, it might not help to identify a high risk of fatigue specifically on the day of the trials. In contrast to the Pittsburgh Sleep Quality Index, the Groningen Sleep Quality scale refers to last night's sleep. It includes 15 items for each of which individuals need to select '*Yes*' or '*No*'. The Groningen Sleep Quality scale has been used in studies of healthy populations and individuals with sleep disorders (De Weerd et al., 2004; Jafarian et al., 2008). Nonetheless, it will not be used in the studies of this thesis because collecting in-depth sleep-related data is out of the scope of the thesis and the 'Yes/No' responses would provide limited flexibility in expressing perceived sleep quality.

As an alternative to the scales above, the St Mary's hospital sleep questionnaire (Ellis et al., 1981) was used in the studies described in chapters 5 and 6. The 14 items of this scale refer to last night's sleep and, more specifically, to sleep latency, nighttime waking, restlessness, and alertness in the morning. This questionnaire allows individuals to choose between six ratings on each item. The St Mary's hospital sleep questionnaire has been validated in clinical and non-clinical populations and has shown good test-retest reliability (Ellis et al., 1981). It has been mostly used to measure sleep quality in hospitalised patients but also in healthy populations (Freeman et al., 2015; Pien et al., 2008). Only one of the items of this questionnaire (i.e. '*How well did you sleep last night?*') was used in the experiments of this thesis in order to exclude those participants that reported sleep of poor quality and for increased practicality. This compromise might affect the psychometric properties of the questionnaire. However, a factor analysis suggested that the item selected loaded heavily on the factor '*sleep quality*' (Leigh et al., 1988). Additionally, Clarke et al. (2013) found a trend for a significant correlation of this item with sleep quality measured with actigraphy. Although more questions may give a better picture of sleep quality, there is evidence that even a single item can give a relatively good indication of sleep quality compared to longer questionnaires (Snyder et al., 2018).

3.1.2.3 Circadian rhythms

The circadian body clock controls alertness (reduced alertness is an expression of fatigue) in the day with the lowest levels of alertness typically being experienced early in the morning (i.e. 2 to 6 am) and in the afternoon (i.e. 2 to 4 pm). Experiments can be run on times outside these windows of circadian low to avoid these effects, but this approach may not necessarily be effective because of inter-individual differences. That is, some people are morning types (i.e. they like sleeping and waking up early) and others evening types (i.e. they like to stay awake late at night and wake up late) (Vink et al., 2001). Depending on ‘morningness’ and ‘eveningness’, the times of the circadian low differ. Therefore, running experiments at a time that is generally considered outside the circadian low (e.g. at 9 am) does not work equally well for everyone in terms of alertness. Researchers can control this misalignment by administering questionnaires that show morningness or eveningness (Horne and Ostberg, 1976), the results of which can be used to assign participants to slots. Alternatively, both morning and afternoon slots that participants can select from based on their preferences can be provided. The latter option will be offered to the participants in the studies described in chapters 5 and 6 because collecting detailed sleep-related data was out of the scope of the thesis.

Researchers also need to consider if participants’ circadian rhythms have been disrupted. The reason is that a time of the day with high levels of alertness for many people due to effect of the circadian body clock (e.g. 9 am) can for some people be on the circadian trough if they recently changed from morning to night shifts or flew from another continent. Collecting data before studies about morningness/eveningness, recent long-haul flights, and recent shift changes can reveal a disruption of the circadian rhythms that contributes to sleep-related fatigue. Besides asking individuals, a disruption of the circadian rhythms can also be detected by collecting physiological data. The circadian body clock is synchronised to the daylight and causes changes in melatonin (Cajochen et al., 2000) and body core temperature (Waterhouse et al., 2005). Therefore, the history of changes in melatonin, body core temperature, and exposure to light can show disruptions of the circadian rhythms. Melatonin can be measured by taking saliva or blood samples, but these two techniques can be impractical and require certain qualifications and equipment. Measuring body core temperature can be made with thermometers, whereas information about the exposure to light can be collected with questionnaires or interviews before studies. Collecting data with questionnaires and interviews is straightforward, but these methods rely on recall from memory.

To sum up, sleep duration the night before the trials will be measured in the experiments of this thesis by rating a single item. Similar, sleep quality will be evaluated with one item taken from St Mary’s hospital sleep questionnaire. No data will be collected about the potential disruption of the circadian rhythms because the focus of the thesis is on task-related fatigue rather than on collecting detailed sleep-related data.

3.2 The methods that can be used to collect data about the expressions of fatigue

Due to the absence of a ‘gold standard’ in fatigue measurement, using a combination of methods of collecting fatigue data should be preferred as this approach can be more informative (Netherlands Aerospace Centre and Netherlands Institute of Neurosciences, 2019; Sahayadhas et al., 2012). As explained in section 1.1.1, sleepiness and vigilance can be expressions of task- and sleep-related fatigue; therefore, they can be measured to infer it. Moreover, fatigue can affect performance, so the measurement of driving and flying performance can reveal increased levels of fatigue too. The following sections describe the methods that can be used to measure the expressions of fatigue (i.e. subjective experience, physiology, and performance).

3.2.1 Subjective experience

Research suggests that subjective ratings can be used to reliably show early increases in the levels of sleepiness and drivers are generally good at detecting sleepiness induction (Horne and Baulk, 2004; Nordbakke and Sagberg, 2007). There is also evidence that subjective ratings of sleepiness in sleep- and non-sleep-deprived individuals show a good correlation with EEG (see section 3.2.2.4), driving errors, and blink duration (i.e. a metric used to infer fatigue – see section 3.2.2.1) (Horne and Baulk, 2004; Ingre et al., 2006[a,b]; Philip et al., 2005[a]; Williamson et al., 2014). However, subjective ratings appear to be more sensitive in revealing early signs of fatigue in drivers than driving performance (e.g. lane deviations in simulator experiments) and blink duration (Ingre et al., 2006[a,b]). Although people can detect increased levels of fatigue before experiencing performance decrements, they may fail to predict falling asleep (Kaplan et al., 2007; Reyner and Horne, 1998). A disadvantage of the subjective scales is the potential lack of face validity (i.e. they may not measure what they were designed to) because the interpretation of the items can differ between respondents. Moreover, individuals may underestimate or overestimate their levels of fatigue or sleepiness and their ratings may not be accurate when they are asked to rate them retrospectively.

Fatigue and sleepiness scales can be used in fatigue studies depending on whether the focus is on the construct of fatigue or sleepiness (which can be an expression of fatigue). In practice, it is often difficult to identify which of the two each scale measures because of the use of items that refer to more than one construct (e.g. fatigue, tiredness, and sleepiness). The fatigue scales are broadly categorised as unidimensional or multidimensional. The former measure only a single, generic feeling of fatigue, whereas the latter multiple aspects of it (e.g. task engagement and alertness). It has been argued that the estimates of subjective constructs like fatigue can be inaccurate when only one item is used (Lee et al., 1991). On the other hand, the multidimensional scales provide more information, but they are more intrusive to primary tasks since they take longer to complete. Both types of fatigue scales can be useful depending on the research problem explored.

The Visual Analogue Scale (VAS) is a commonly used unidimensional fatigue scale. The VAS is a straight line usually with one word or phrase on each end. The word/phrase on the left end of the line represents the lowest levels of fatigue an individual can experience and the right end the highest. The line is divided into equal intervals that may be accompanied by words/phrases too (e.g. '*neither rested nor fatigued*'). Individuals are asked to put a mark on the line depending on how they feel that moment or rate their levels of fatigue over a longer period. A disadvantage of the VAS is that between-subject comparisons may be inaccurate because it is unclear what each interval represents. Moreover, its ratings are not accompanied by phrases that describe the degree to which performance has been impaired. Linking the levels of fatigue to the risk of performance decrements is essential in studies that aim to improve safety in driving and flying.

Another commonly used unidimensional scale of fatigue is the Crew Status Survey (CSS) fatigue scale (also called '*Samn-Perelli*' scale) (Samn and Perelli, 1982). This is a 7-point Likert item that ranges from 0 ('*fully alert, wide awake*') to 6 ('*completely exhausted, unable to function properly*'). Compared to the VAS, the CSS does not allow selecting any point on the rating item, but comparisons amongst individuals and studies are easier. Moreover, each of the 7 points on the scale is accompanied by phrases describing degrees of impaired performance, which helps to identify safety risks. The CSS has a high test-retest reliability (Miller and Narvaez, 1986) and has been widely used and validated in studies with pilots (Gander et al., 2015; Petrilli et al., 2006). Fatigue has also been measured with the CSS in studies with drivers (Belenky et al., 2012; Richter et al., 2005) and other professionals, such as train drivers (Kazemi et al., 2016), air traffic controllers (Kuo et al., 2017), and firefighters (Vincent et al., 2015).

The VAS has a multidimensional alternative (VAS-F) that includes 18 items about sleepiness, tiredness, drowsiness, energy, and concentration. Another multidimensional scale of fatigue is the Dundee Stress State Questionnaire (Matthews et al., 1999) with 10 Likert items that are used to collect data about three main concepts: task engagement, distress, and worry. Task engagement is of interest in this thesis because, according to Desmond and Hancock's (2001) model of fatigue, active fatigue is caused by the long investment of effort. However, similar to the VAS-F, the Dundee Stress State Questionnaire will not be used in this thesis because it would be too intrusive in the experiments and would reduce the response rates in the survey due to its length. Some multidimensional scales are used to measure fatigue during a longer period (e.g. the Profile of Mood States - Terry et al., 2003). This means that they are not suitable for studies that investigate the variation of the levels of fatigue during the trials (as in this thesis). Furthermore, using a scale like this in the survey of chapter 4 would not provide information about pilots' usual levels of fatigue during the commuting cycle because many pilots do not fly every week.

Both the unidimensional and multidimensional fatigue scales have advantages and disadvantages and choosing the right one depends on the aim and design of the study. The CSS is considered the most appropriate scale for the studies described in this thesis because it is quick to administer and straightforward, comparisons between individuals can be made more accurately, the variation of the levels of fatigue during trials can be investigated, and the ratings link directly to perceived deterioration of performance.

Besides using fatigue scales, researchers can use sleepiness scales to infer fatigue. The Karolinska Sleepiness Scale (KSS) (Horne and Reyner, 1995) is probably the most commonly used one in studies with drivers (Kecklund and Akerstedt, 1993) and pilots (e.g. Eriksen and Akerstedt, 2006). This scale is a 9-point Likert item on which people rate their levels of sleepiness in the 10 minutes before the administration. The KSS has shown high validity in measuring sleepiness compared to other techniques used to infer sleepiness like the Electroencephalography (EEG) (see section 3.2.2.4) and the Psychomotor Vigilance Task (PVT) (see section 3.2.3.1) (Kaida et al., 2006). In contrast to the KSS, the Epworth Sleepiness Scale (ESS) measures individuals' levels of sleepiness over the last week (Johns, 1992). This scale is widely used in clinical settings and has been validated against the EEG (Horne and Baulk, 2004). Additionally, the ESS measures individuals' trait levels of sleepiness (Shen et al., 2006), so it is not appropriate in studies that aim to measure participants' levels of sleepiness many times during trials.

The KSS may be preferred over the ESS to infer sleepiness in studies with repeated measures because it refers to sleepiness at the time of administration. However, the KSS and the SSS will not be used in the studies of this thesis. The reason is that, although their items include similar descriptions to those of the CSS, their highest ratings are accompanied by phrases that refer to falling asleep. In contrast, the CSS mentions performance decrements; thus, it provides more direct links between fatigue and the risk of accidents.

3.2.2 Physiology

As mentioned, a combination of subjective, physiological, and performance data can be more informative when exploring fatigue. For example, a limitation of the subjective scales is that fatigue cannot be measured continuously in order to identify small variations during driving and flying tasks. In contrast, this can be achieved by collecting physiological data that are known to be affected by fatigue or sleepiness (as an expression of fatigue). In addition, evaluating the results of the analysis of subjective and physiological data together can help researchers to reach more solid conclusions about participants' levels of fatigue.

As with the subjective scales, the physiological metrics used to infer fatigue have some limitations. That is, the equipment needed may be too intrusive and advanced technical knowledge may be required to use it. Moreover, there are trait-like inter-individual differences in the physiological metrics that can result in large inter-individual variations. Furthermore, it is often unclear what the cut-off scores for high levels of fatigue are when physiological metrics are used. Finally, all the physiological metrics that can be used to infer fatigue can also be affected by other factors (e.g. stress), so the data cannot be directly interpreted. The following sections describe the methods that can be used to infer fatigue from eye- and heart-related data, electrodermal activity, brain-related data, and body posture and facial expressions.

3.2.2.1 Eye-related metrics

Blink-related data, pupil diameter, and saccades are often used to measure sleepiness. Shiferaw et al. (2018) suggested that collecting eye gaze data to infer sleepiness is a promising method. However, there is not enough evidence that visual scanning in driving and flying can be affected by sleepiness and research has shown that eye gaze dispersion is affected by the level of automation of the vehicle and whether drivers' access to driving-relevant information is restricted (Louw and Merat, 2017). More specifically, Louw and Merat conducted a driving simulator study with 60 participants divided into 4 groups, each of which completed driving scenarios with a different screen manipulation (i.e. no fog, light fog, heavy fog, or heavy fog with a secondary task that included questions). Drivers' visual attention distribution was assessed during manual and level 2 automated driving on approach to a critical event (i.e. decelerating lead car). It was found that participants' horizontal gaze was generally less dispersed a) in manual driving and b) when the manipulations stopped and an automation uncertainty Human Machine Interface appeared. This appearance meant that the drivers should monitor the driving scene and decide whether to intervene or not. In contrast, the vertical gaze dispersion was higher in the heavy fog condition.

Regarding pupil diameter, studies suggest that it reduces when the levels of sleepiness increase (Morad et al., 2000; Wilhelm et al., 1998; Wilson et al., 2006) but increases with task load (Ahlstrom and Friedman-Berg, 2006) and is affected by light. The light conditions can be controlled in simulator studies, but the effect of task load on pupil diameter hinders the identification of the effect of sleepiness in low task load conditions.

As far as blinks are concerned, it is believed that blinks associated with sleep onset have unique characteristics that reflect changes in the Central Nervous System that relate to the sleep-wake process (Stern et al., 1984). These unique properties have been attributed to the proximity of the brain structures that relate to the sleep/wake process to the neural centres responsible for the facial nerves that control the eyes (Morris and Miller, 1996). Research suggests that higher levels of sleepiness relate to a higher blink frequency (i.e. more often blinks) (Fukuda et al., 2005; Schleicher et al., 2008; Summala et al., 1999). Nevertheless, there is also evidence that blink frequency reduces with increased task load (Ledger, 2013; Recarte et al., 2008; Veltman and Gaillard, 1998) and the level of engagement with a task (Fairclough and Venables, 2006). Additionally, blink data can be collected to calculate the blink duration, which refers to the time interval between fully closed and fully open eyes. Studies suggest that longer blink duration relates to increased levels of sleepiness (Caffier et al., 2003; Hakkanen et al., 1999; Van Orden et al., 2000; Verwey and Zaidel, 1999).

PERCLOS (i.e. Percentage of Eye Closure) is a metric often used to infer sleepiness. It refers to what percentage of a 1-minute time window the upper eyelid covers pupil by more than 80%, so both blink duration and frequency affect it. Since individuals tend to close their eyes less frequently in high task load conditions, PERCLOS may be affected not only by sleepiness but by task load too. Another disadvantage of using PERCLOS to infer fatigue is that compensation effects are possible in short tasks (e.g. drives and flights).

Researchers should be aware of the likelihood of paradoxical changes in blinking behaviour. Karrer et al. (2005) found in a driving simulator study that sleepy drivers may experience a state of 'absence' that seems like sleeping with the eyes open and is accompanied by lapses of attention. They called this phenomenon '*driving without awareness*'. It is also possible that individuals experience an increase in their levels of alertness when not expected. For example, drivers' levels of sleepiness may increase for a certain amount of time but then reduce. Schleicher et al. (2008) attributed this phenomenon to self-activation when drivers realise that they are making mistakes due to sleepiness. Similar, sleepiness may cause only minor increases of blink frequency to individuals with an already high blink rate, or individuals may show almost no blinks after an initial increase of blink frequency. That could result in an inverted U-shape link between sleepiness and blink frequency (Schleicher et al., 2008).

Saccades (i.e. the movement of the eyes between two fixations) are also used to infer sleepiness. It has been suggested that saccades are affected by sleep deprivation due to a dysfunction at the brainstem reticular formation when sleep deprived. It is believed that this formation plays an essential role in maintaining arousal (Zils et al., 2005). As far as the saccadic speed is concerned, research suggests that it reduces as the levels of sleepiness increase (Galley, 1993; Sirevaag and Stern, 2000). However, the saccadic speed also relates to task load with some studies showing a positive (Bodala et al., 2014) and others a negative relationship between the two (Di Stasi et al., 2013). These considered, measuring saccadic speed is not adequate to differentiate between the effects of sleepiness and task load. These effects might be better distinguished by also measuring the number and the duration of saccades as higher levels of task load correlate positively with the number and negatively with the duration of saccades (e.g. Rognin et al., 2004). Finally, although there is evidence that higher levels of sleepiness relate to longer fixation durations (i.e. the time intervals between two saccades) (Wang et al., 2018), this is not definite (Schleicher et al., 2008) because fixation duration can also increase with task load (O'Donnell and Eggemeier, 1986). In general, saccades might provide information regarding the levels of sleepiness, but sleepiness cannot be safely inferred by looking at saccades alone due to the potential effects of task load.

Eye-related metrics can be collected with electrooculography (EOG) and eye trackers. EOG is a technique where electrodes are placed above and below or on the left and right of the eyes to detect eye movement. This is achieved by measuring changes in the electrical potential. Although the EOG can be useful in studies that explore eye movement, it neither collects blink data (e.g. blink duration) nor measures pupil diameter. In contrast, these types of data can be measured with eye-trackers. Eye-tracking is a relatively non-intrusive method that has been used in studies with drivers and air traffic controllers (e.g. Ahlstrom and Friedman-Berg, 2006; Hwang et al., 2014). Eye-trackers are devices that record eye video or use infrared light to collect data about eye movement, blinks, and pupil diameter. Gaze data can also be collected via video recording. Depending on the study design, table- or head-mounted eye trackers can be used.

Despite the evidence that eye-related metrics can be used to measure sleepiness and, from that, infer task-related fatigue, inter-individual differences should be taken into consideration because these can affect blink duration (Hamada et al., 2003). Due to these differences, Schleicher et al. (2008) suggested that group means of eye-related metrics may be insufficient to reveal sleepiness. To sum up, blink-related data, pupil diameter, and saccades can show sleepiness, but they can also be affected by task load. Distinguishing between the two may be achieved by also collecting fatigue and task load data with subjective scales, using additional physiological metrics, and collecting performance data.

Since no table-mounted eye tracker was available for the research conducted in this thesis, gaze and blink data were collected in the experiments described in chapters 5 and 6 with a portable eye-tracker (Pupil Labs, 2019). However, these data will not be described in this thesis because the accuracy of the measurements dropped because the device often slipped during the trials. This was caused by sweat, the participants touching the device, and the movement between chairs in order to complete the driving and flying tasks. Pads were attached to the temples of the eye-tracker to stabilise it and the calibration procedure was performed in the two experiments both at the beginning and during the trials (i.e. between the driving and flying tasks) to control for the effect of that movement. However, it was not possible to stop the movement of the eye tracker during the tasks. Due to the slippage of the device, it was often noticed in the trials that participants' gaze was offset (i.e. it appeared constantly away from the screen when the participants were looking at the driving scene).

Pupil diameter was measured with eye cameras as the number of pixels of the image of the eye on the video recorded. The closer the eye cameras to the eyes, the more the pixels recorded. Therefore, the slippage of the device during the trials reduced the accuracy of the measurement of pupil diameter. The portable eye tracker that was used does not measure PERCLOS. Although the time spent between the start and end of blinks could be measured by that device, this metric has not been sufficiently validated. Furthermore, the movement of the device during the trials could have resulted in missing blinks due to the eyes not been detected at all times. Sample blink data were visually inspected during the preliminary study in order to identify whether the eye tracker detected the blinks that occurred. Nevertheless, it was not possible to safely conclude about the accuracy of measuring the time in blink state. This considered, no eye-related data will be described in the thesis.

3.2.2.2 Heart-related metrics

Heart Rate (HR) and Heart Rate Variability (HRV) are often used to measure fatigue. HR is the number of heart contractions per minute and is controlled by the sympathetic and parasympathetic nervous systems. The former prepares the body for physical activity (fight-or-flight response) and the latter often has 'opposite' actions that slow high energy functions. As far as heart is concerned, the sympathetic system increases HR and the parasympathetic decreases it. HR has been often used as the operational definition of the level of arousal, for example, in studies that found higher HR in people who completed arithmetic tasks (Allen et al., 1987), stood instead of lying down (Lechin et al., 1995), and completed flying tasks (Roscoe, 1993). In a truck driver naturalistic study, Apparies et al. (1998) found an increase of HR with the progression of the drives that they attributed to increased levels of fatigue. However, there is evidence that heart rate also increases with task load (Mehler et al., 2011), stress (Schubert et al., 2009), and exercise (Burton et al., 2004). HRV is the variation of beat-to-beat intervals. Studies have shown that HRV reduces with higher levels of fatigue caused by increased time-on-task (Apparies et al., 1998) but reduce with task load and stress (Schubert et al., 2009; Wierwille and Eggemeier, 1993). Since both the HR and HRV can also be affected by factors other than fatigue, higher levels of fatigue cannot be securely concluded based solely on these metrics.

HR and HRV can be measured with an electrocardiogram (ECG) or a wrist-worn device. The ECG is a test used to record parameters of the heart rhythm and electrical activity with sensors attached to the skin. The electrical signals produced by the heart are recorded and can, then, be analysed to deduce fatigue. Using ECG in experiments with simulated driving and flying may be impractical due to the need to attach the electrodes. Moreover, the accuracy of heart-related measurements may reduce during driving and flying simulation if the electrodes are detached due to head or body movements and sweat. Wrist-worn devices are an alternative to EEG. These measure HR and HRV with sensors that touch individuals' wrists and, thus, are more practical to use in simulated driving and flying than the ECG. Such devices infer HRV from the blood volume pulse. The blood volume pulse is the blood volume that passes through the tissues in a certain area with each heartbeat.

Inter-individual differences may affect HR and HRV. That is, research suggests that the resting HR (i.e. the HR when no activity is completed) differs amongst individuals. For example, Quer et al. (2020) measured HR in more than 90,000 people with wrist-worn trackers over more than

35 weeks and found that people may have a resting HR that is normal for them, but that can be very different in other individuals. That difference was as high as 70 beats per minute. Personality traits can also affect HR based on studies that found a relationship between lower resting HR and sensation seeking (Hammerton et al., 2018; Portnoy and Farrington, 2015). Inter-individual differences in HRV have also been found with studies relating HRV to genetic influences (Golosheykin et al., 2017), personality traits, such as depression and anxiety (Bleil et al., 2008), gender, and age (Kuo et al., 1999). These findings suggest that inferring differences in fatigue by comparing HR or HRV between groups may not be sufficient.

In summary, HR and HRV can be used to infer fatigue, but they can also be affected by factors other than fatigue. Similar to the eye-related metrics, heart-related data may be collected in conjunction with other types of data (i.e. eye-related, subjective, and performance) to deduce fatigue. EEG could be used to measure HRV data accurately but using it in the trials would be impractical because the participants moved between chairs (both experiments) and left the room during the breaks (experiment of chapter 6). As a more practical alternative, the Empatica E4 wrist-worn device (Empatica, 2019) was used in the studies of both chapters 5 and 6. This device detects changes in blood volume caused by the pressure pulse with photodiodes at a frequency of 64 Hz.

The accuracy of the E4 (64 Hz) in measuring HR and HRV has been compared to a variety of ECG devices (the ECG is considered the 'gold standard') that collected data at 1,000 (Bent and Dunn, 2020; Menghini et al., 2019; Ollander et al., 2016; Schuurmans et al., 2020) and 500 Hz (Milstein and Gordon, 2020). This literature suggests that the 64 Hz of the E4 are enough to achieve a high accuracy compared to the ECG in measuring HR in both static (e.g. when sitting) and dynamic conditions (e.g. during walking). In contrast, these studies suggest that the accuracy of the E4 at 64 Hz compared to the ECG is lower when measuring HRV, particularly when individuals move. This reduction of accuracy can be explained by the fact that blood volume pulse (which is used to deduce HRV) is affected by body movement.

Although the E4 measures HR accurately, HR is not a good metric when interested in fatigue because HR cannot be only affected by fatigue but also by a number of other factors, such as stress and physical activity. Similar to HR, HRV is affected by several factors besides fatigue. At the same time, the E4 does not measure HRV accurately in active conditions (as the ones in the experiments of this thesis). These limitations considered, no reference to heart-related data will be made in the remaining chapters of this thesis.

3.2.2.3 Electrodermal activity

Electrodermal activity (EDA) can also be measured to infer sleepiness. EDA is the skin conductance across two points of conduct on the skin. Skin conductance changes due to the production of sweat as a response to the preparation of the body for action by the sympathetic system (Picard, 2009). The research findings are contradictory regarding the relationship between sleepiness and skin conductance (Miro et al., 2002) because sweat can also be produced due to stress, high ambient temperatures, and physical activity. EDA is measured with two electrodes attached to the fingers or with wrist-worn devices. The attachment of electrodes can be intrusive, so measuring EDA with wrist-worn devices may be more practical in experiments with simulated driving and flying. EDA will not be measured in the experiments described in this thesis because of the inadequate scientific evidence supporting its link to fatigue.

3.2.2.4 Brain-related metrics

The functional magnetic resonance imaging (fMRI) and the EEG can be used to deduce sleepiness from brain-related data. The fMRI is a technique used to measure brain activity by detecting changes in blood flow. It can potentially detect changes associated with sleepiness (e.g. Chang et al., 2016), but it cannot be used during simulated driving and flying as people need to lie in a tube-shaped device during data collection. EEG, on the other hand, measures the cerebral arousal as indicated by changes in delta, theta, alpha, and beta wave frequencies. These changes are measured by recording voltage differences between parts of the brain with electrodes placed on the scalp. The literature suggests that increased activity of the delta waves indicates deep sleep, higher theta activity shows some but not deep sleep, alpha waves relate to wakefulness with low activation, and beta waves are associated with high alertness (Sahayadhas et al., 2012). EEG has been validated against self-report scales of sleepiness (Horne and Baulk, 2004).

Safely concluding that EEG changes have been caused by sleepiness is difficult because of the potential additional effect of task load. That is, there is evidence that task load correlates positively to activity in the theta band and negatively with activity in the alpha (Sterman et al., 1987) and delta bands (Wilson et al., 2006). Moreover, using EEG can be intrusive during simulated driving and flying tasks and large inter-individual differences in sleep variables measured with EEG (Tucker et al., 2007) can make the interpretation of the results difficult. To sum up, the fMRI and the EEG can be useful in inferring fatigue through sleepiness. However, the former cannot be used during simulated driving and flying tasks and the latter may be too intrusive. Therefore, neither of them will be used in the studies described in this thesis.

3.2.2.5 Body posture and facial expressions

Video recording can be used to capture postural changes and facial expressions (i.e. yawning, nodding, and face motion) that accompany increased levels of sleepiness. The recordings are then analysed either automatically with algorithms (Craye et al., 2015) or by trained raters (Wierwille and Ellsworth, 1994). The analysis by trained raters has good inter-rater and test-retest reliability (Wierwille and Ellsworth, 1994), and video recording is non-intrusive. Furthermore, collecting data about posture and recording facial expressions can increase objectivity compared to self-reports, especially when the analysis is performed automatically by algorithms. Nonetheless, small, continuous variations of the levels of fatigue cannot be identified. Moreover, the findings on the correlation between the analysis made by trained raters and alpha-wave activity measured with EEG (see section 3.2.2.4) are inconsistent (Lal and Craig, 2002). These limitations considered, the analysis of body posture and facial expressions will not be used in the studies described in this thesis because no algorithm was available to increase the objectivity of the analysis and because identifying small, continuous changes of sleepiness would not be possible.

3.2.3 Performance

As mentioned, the subjective scales of fatigue and the physiological data can be useful in exploring fatigue. However, they do not provide a direct link between fatigue and performance decrements, which is very important in studies that investigate safety in driving and flying. There is evidence that when individuals are fatigued, their performance can be impaired in both tasks of high and low task load (Pilcher and Huffcutt, 1996; Pilcher et al., 2007; Williamson et al., 2001). However, the effects of fatigue on performance may be mediated by the effect of inter-individual differences, such as experience (see top-down processing in section 2.2.2) and motivation. The potentially mediating role of motivation was suggested by Hockey et al. (1998). That is, they found in a study of 16 participants that completed an automated process control task that performance decrements due to fatigue were more likely in secondary tasks than in the primary ones. They attributed that difference in the fact that performance in the primary tasks is protected by deciding to allocate more attentional resources there. Because of the role of the inter-individual differences in performance, within-group study designs might be more effective in revealing fatigue-related performance decrements.

Researchers can measure two types of performance: the simple and the task-related one. The methods used to measure the former (e.g. PVT) do not require knowledge and experience, so fatigue is more likely to translate to performance decrements. Nevertheless, performance in those tests does not provide a direct link between fatigue and performance in real-world driving and flying tasks, as is the case with the measurement of task-related performance (i.e. simulated and naturalistic driving and flying). To this end, using a combination of measures of simple and task-related performance can help to draw more solid conclusions about fatigue.

3.2.3.1 Simple performance

In its simplest form, performance is measured with tests of simple reaction time. Completing these tests requires vigilance (see section 1.1.1), which is known to be impaired by fatigue (Abad and Guilleminault, 2012; Lim and Dinges, 2008). It should be noted that vigilance is not measured directly with these tests but deduced from the reaction times to a stimulus that appears on a screen and always requires the same response (e.g. pressing a button). Simple reaction time differs from choice reaction time in that the latter is measured in tasks with more than one stimulus, each of which must be responded with a different, unique behaviour. Since reaction time in choice reaction time tasks is not only affected by sustained attention but also by memory (Fishman et al., 2008), simple reaction time tasks are more appropriate to deduce decrements of vigilance. Simple reaction time can be measured during simulated driving and flying by presenting objects (e.g. rectangles) on a specific position on the screen on random times. In those cases, individuals need to respond by pressing a button when they see the object (Wu et al., 2014). This method will not be used in the experiments described in this thesis because that could distract the participants and, thus, distort their performance in the flying and driving tasks.

The predominant technique of measuring simple reaction time is the Psychomotor Vigilance Task (PVT). Individuals who complete the PVT need to press a button when they see an object (usually a ball or a box) on random times at the centre of a PC monitor. The usual duration of the PVT is 10 minutes. Loh et al. (2004) suggested that a 5-minute PVT version is affected by fatigue in a similar manner, but the 10-minute PVT is more sensitive to the effects of fatigue. The impact of sleepiness on performance in the PVT has been supported in studies, where extended time awake since last sleep and sleep loss increased simple reaction times (Baulk et al., 2006; Dinges et al., 1997) and lapses (Basner et al., 2011; Shattuck and Matsangas, 2014). Lapses are the reaction times in visual simple reaction times tests that are longer than 500 milliseconds (Matthews et al., 2017). The simple reaction times of healthy, non-sleep-deprived individuals in the PVT vary between 220 and 500 ms (Dreary and Der, 2005) with a mean of around 330ms (Woods et al., 2015). In a study of 1,930 participants between 16 to 63 years old, Dreary and Der found slower reaction times in older participants.

The PVT is simple to complete with minor practice effects (Basner et al., 2017; Dinges et al., 1997), it does not require training, and it has good test-retest reliability (Baulk et al., 2006). Additionally, it has a high signal-load, which means that a large amount of data is collected in a short period. Since using the PVT is intrusive, it is typically administered before or after the driving or flying tasks. It should also be noted that there is evidence of trait-like inter-individual differences in the PVT reaction times (Van Dongen and Belenky, 2009), so comparing group means may not reveal differences in vigilance.

Although the literature consistently suggests that vigilance decrements can be measured with the PVT, the evidence on whether PVT reaction times relate to task-related performance, such as driving and flying, is inconclusive. That is, some studies with sleep-deprived individuals showed that PVT reaction times were strongly associated with performance in real-world and simulated driving (Baulk et al., 2008; Kosmadopoulos et al., 2017), but others did not (e.g. Baulk et al., 2006). As Baulk et al. (2006) stated, measuring PVT reaction times is useful when exploring driver fatigue, but it cannot truly predict driving performance. To this end, they recommended combining PVT with driving simulation. Similar to driving, the link between performance in the PVT and flying tasks has been suggested only in some of the relevant studies. For example, Lopez et al. (2012) found in a study with simulated flights completed by military pilots that the PVT reaction time was a good predictor of simulator performance. In contrast, that link was not supported in O'Hagan et al.'s (2019) study. In general, even if performance in the PVT can predict task-related performance, it does not provide any information about which and how tasks may be impaired in real-world flying and driving.

To sum up, simple reaction time can be measured to deduce fatigue through decrements in vigilance. The literature suggests that PVT reaction times increase with fatigue, so this method will be used in the experiments of this thesis to infer fatigue before and after driving and flying tasks. However, performance in the PVT may not predict driving and flying performance in real-world settings and does not reveal how performance in certain real flying and driving can be affected by fatigue. Therefore, the PVT will be used in conjunction with task-related performance data. The methods that can be used to collect these data are described in the next section.

3.2.3.2 Task-related performance

The methods of measuring fatigue described so far (i.e. subjective scales, physiological data, posture and facial expressions, and simple reaction time) can help researchers to identify high levels of fatigue in participants and an increased risk of impaired performance. Nevertheless, they do not link fatigue directly to the deterioration of certain aspects of driving and flying performance. The only way to collect this information is by measuring performance during real-world or simulated tasks.

Performance should be ideally measured in real-world conditions to increase the ecological validity. This refers to the extent to which the methods used in a study resemble the real-world situation that is explored. However, there are some limitations associated with such measurements of performance in fatigue studies. First of all, real-world testing raises ethical issues because fatigue imposes a risk for safety. In addition, researchers may not be able to manipulate and control variables of interest (e.g. environmental conditions). Furthermore, it may be difficult to collect accurate performance data (e.g. the position of a car on the lane) and reproduce the conditions in future studies. Finally, real-world experiments with driving and flying can be costly to conduct.

As an alternative to real-world testing, simulated driving and flying tasks can be used. The literature suggests that driving simulator performance can be used to infer increases in individuals' levels of fatigue. That is, there is evidence that higher levels of fatigue in drivers relate to higher standard deviations of speed and lateral position of the car. Moreover, fatigued drivers are more likely to maintain shorter headways from the cars ahead and higher mean speeds, perform more frequent edge-line crossings, and respond less accurately to speed changes of a lead car (Akerstedt et al., 2005; Brookhuis et al., 1994; Du et al., 2015; Thiffault and Bergeron, 2003; Ting et al., 2008; Van der Hulst et al., 2001; Zhang et al., 2016).

Similar to driving simulators, studies suggest that increased levels of fatigue in pilots can be deduced from performance in flight simulators (Caldwell and Ramspott, 1998; Lopez et al., 2012; Morris and Miller, 1996; O'Hagan et al., 2019; Roach et al., 2006). These studies have found a relationship of higher levels of fatigue with a reduced accuracy in tracking tasks, slower reaction times to scale deviations and light changes, and more time-out errors in communications tasks. They have also linked fatigue in pilots to a lower accuracy of maintaining target airspeed, heading (i.e. the direction of movement), altitude, and vertical velocity (i.e. rate of climb and descent), more failures to detect and manage errors, and delays in completing mathematical operations.

Measuring driving and flying performance in simulators is safe, allows control over the conditions of the trials, and researchers can use equipment to collect data that can be difficult to set up in real-world conditions (e.g. EEG). Nevertheless, simulation has disadvantages too. That is, participants in simulator studies may experience motion sickness (Brooks et al., 2010) and become fatigued more quickly compared to real-world driving and flying. The latter has been suggested in studies that showed that participants had slower reaction times, rated their levels of sleepiness as higher, and crossed the lane more frequently when driving in a simulator than in the real world (Davenne et al., 2012; Philip et al., 2005[a]). Similar, two studies have suggested a more severe effect of fatigue on pilots' performance in simulators than in real flights. In the first one, Billings et al. (1975) found that hypnotic pills (i.e. pills that expedite sleep onset) affected flight performance (i.e. higher tracking error and higher airspeed variability) more in a flight simulator compared to real-world flying. More recently, Caldwell and Roberts (2000) found that stimulants had a stronger effect on flying performance (i.e. straight flying, turns, and changes of altitude) of helicopter pilots when flying in a simulator than in real-world conditions. The potentially quicker induction of fatigue in simulators might be attributed to the fact that individuals are more likely to compensate for the effects of fatigue on performance in the real world due to an increased motivation to perform well in order to avoid crashing. In contrast, completing simulated tasks is safe and individuals' motivations for participation can relate more to financial rewards and curiosity (Stunkel and Grady, 2011).

In addition to the potential lack of ecological validity, it is unclear how accurately changes in individuals' performance due to the manipulation of independent variables in studies with simulated tasks resemble performance changes in real-world tasks. Several studies have explored if simulated driving performance predicts real-world driving performance, but the evidence is inconclusive. For example, Blana and Golias (2002) found a higher lateral displacement of the car in a fixed-base simulator than in real-world driving and Reed and Green (1999) found a higher SD of the lateral position when driving in a simulator than on real roads. The possibility of finding differences in the ability to maintain lateral position between simulated and real-world driving may be attributed to the fact that simulators do not provide all the spatial information available in real-world driving. In turn, the lack of spatial information may result in an underestimation of distance (Saffarian et al., 2015). Nonetheless, the difference in the lateral position between simulated and real-world driving was not supported by Tornros (1998). The measures of speed may also differ between simulated and real-world driving. For example, Blana (2001) found lower average speeds in real-world compared to simulated driving. However, other studies did not find any difference in speed selection between simulated and real-world driving (Reed and Green, 1999).

The literature on whether flight simulator performance can predict performance in real flights is limited. Although some links can be drawn from studies that explored the transfer of training from simulators to real flights, it could be argued that performance in the two environments is not directly comparable due to the difference in the time spent practising in each of them. In a study that compared simulated to real-world flying operations, Magnusson (2002) found analogous psychophysiological reactions (i.e. HR, HRV, eye movements, and eye blinks) between the two conditions in fighter pilots. This finding suggests that pilots may respond in a similar way to both simulated and real tasks. However, performance was not directly compared. In a direct comparison of simulated to real-world flying performance of helicopter pilots, Schmeisser et al. (2008) did not find any correlation. Therefore, it is unclear if simulated flying performance can predict performance in real flights.

Besides the potential effect of differences in motivation, the level of fidelity (i.e. the degree to which simulators replicate reality) may also play a role in the performance differences found between real-world and simulated tasks. That is, the simulators should have the appropriate level of fidelity to maximise the possibility of predicting real-world performance. There are two types of fidelity; physical and functional. The former refers to how accurately the simulator replicates the physical elements (i.e. dashboard or flight deck) and feel of the real tasks. The latter is the level to which the simulated tasks replicate the skills (e.g. decision making and mental calculations) needed to drive or fly in the real world.

Jamson and Jamson (2010) compared performance in a high-fidelity driving simulator to performance in a lower fidelity one (i.e. the one used in the experiments of this thesis too). They found more often steering corrections and less accurate judgement of speed in the latter. This difference could be attributed to the less spatial information provided on the simulator of lower fidelity. Although studies that compared flying performance between simulators of different fidelity were not located, some links may be drawn from studies that explored the transfer of training. Alessi (1988) suggested that there is a certain point until which increasing fidelity increases the transfer of training, but after which, transfer increases to a gradually reduced rate. The transfer of training seems, however, to be affected by the simulator fidelity differently depending on the flight task. That is, using a flight simulator with motion results in better performance in disturbance tasks (e.g. engine failures) and helicopter flying in the real world, but the transfer is lower in manoeuvring training. In the latter case, simulators without motion could be equally helpful (De Winter et al., 2012).

To sum up, the literature is inconclusive regarding whether performance in driving and flight simulators can predict performance in real-world settings with some research suggesting that fatigue may have a more pronounced effect on simulated tasks. It is also unclear if using simulators of higher fidelity is better in terms of predicting performance decrements due to fatigue in the real world. Despite these limitations, driving and flying simulation will be used in the experiments of chapters 5 and 6 because measuring task-related performance can provide a direct link between fatigue and performance that no other methods can. The decisions about which level of simulator fidelity to use will be described in the methods section of chapters 5 and 6. The measurement of task-related performance will be accompanied by the measurement of simple performance because the latter is not affected by skills and experience.

Chapter 3 provided a description of the methods that can be used to control for the contributors to fatigue, collect data about them, and measure the expressions of fatigue. Chapter 4 describes an interview with a pilot that aimed to identify key aspects of the issue of fatigue experienced in the commuting cycle after long and demanding drives to the airport. An online survey that explored the extent of this issue among professional pilots worldwide is also included.

Chapter 4. Exploring the extent of the issue of fatigue in pilots who commute by driving

4.1 Background and justification of the study

Fatigue in flight has been a major concern for the aviation industry for decades because it can impair flying performance (Petrilli et al., 2007; Thomas and Ferguson, 2010) and contribute to incidents and accidents (Gokhale, 2010; National Transportation Safety Board, 2018). To reduce that risk, fatigue regulations require that pilots report fit for duty (e.g. Commission Regulation, 2014) and receive training on how to avoid and manage fatigue (Federal Aviation Administration, 2012b). In addition, airlines arrange pilots' work schedules following analyses with the use of biomathematical models that predict fatigue. These interventions are mainly based on research on fatigue caused by sleep deprivation, long time awake since last sleep, and operating during the circadian low (Le Duc et al., 1999; Paul et al., 2001; Wright and Mc Gown, 2001) (see section 2.1.2).

What has not been adequately considered in the current interventions is fatigue induced by driving to the airport. Aviation organisations (e.g. International Civil Aviation Organization, 2016) and a small number of studies (e.g. National Research Council, 2011) have suggested that pilots may fly fatigued due to commuting. At the same time, studies solely on drivers suggest that long and demanding drives can induce active fatigue (e.g. Ting et al., 2008) by depleting individuals' attentional resources (see section 2.2.1). If this was also the case in pilots who drive to the airport, then the likelihood of impaired flying performance would increase unless intervention strategies for fatigue were used before duty.

Based on the research solely on flying and driving, active fatigue induced by driving to the airport might impair not only flying performance but also pilots' driving performance after duty. This potential link between driving to and from the airport is based on studies that showed that flying can induce fatigue (Rosekind et al., 2000; Spencer and Robertson, 2002) and driver studies that suggest that people are at a higher risk of impaired driving performance when fatigued (e.g. Ting et al., 2008). When seen together, these studies indicate that the likely increased levels of active fatigue due to completing long and demanding drives to the airport might increase further in flight. If pilots did not manage fatigue while at the airport after duty, they would be at an increased risk of car accidents. The risks associated with driving fatigued after work have been researched in other professionals, such as nurses, medical residents, and air traffic controllers (e.g. Barger et al., 2005). In those studies, fatigue after work was attributed to sleep loss, work schedules, and long drives back home. Therefore, the effects of driving to work on driving after work have not been explored.

In high-hazard environments like aviation, it is essential to know the likelihood of a factor contributing to incidents and accidents. In the case of the risk of impaired performance due to active fatigue induced by driving to the airport, there is inadequate evidence of this likelihood. This considered, the research described in this chapter aimed to explore the extent to which fatigue induced by long and demanding drives to the airport affects negatively professional pilots' fatigue in flight and while driving after duty. An interview with a pilot was, first, conducted to identify the key aspects of the issue of fatigue in pilots who commute by driving. Its findings were then used to design the survey that aimed to investigate the extent of the issue.

4.2 Interview with an airline pilot

Since there was no literature on fatigue in pilots who commute by driving, it was decided to conduct an interview with a professional pilot in order to identify some aspects of the issue that could be then investigated in a survey to draw conclusions regarding the extent of the issue. Therefore, the aim of the interview was not to test specific hypotheses or explore the issue in-depth. This considered and because access to experienced professional pilots that could have an overview of the profession of pilots in various countries was limited at the beginning of this research, one interview was conducted. Since there was only one interviewee, no statistical analyses will be described in this section.

The interview was conducted via Skype (Skype, n.d.). The interviewee was a professional airline pilot with more than 10000 flight hours and a home base in the United Kingdom. The interview lasted one hour. The questions of the interview were based on the literature about pilot commuting and fatigue in driving and flying. Since this literature was limited, it was decided to conduct an unstructured interview. The main theme of the interview was whether and why driving to and from the airport is a safety issue amongst professional pilots.

The interviewee expressed concerns that commuting by driving a car can increase pilots' levels of fatigue in flight. He believed that commuting to the airport can be fatiguing irrespective of the means of transport used. However, he found driving the most fatiguing type of commuting because pilots cannot sleep during the drives. According to the pilot, driving is fatiguing when it is too long and the traffic density is high. He identified those commutes as the ones that mostly contribute to pilots flying fatigued. The reference to these characteristics of driving indicates the induction of active fatigue. The interviewee believed that fatigue in flight can contribute to errors in communicating with the Air Traffic Control and the other pilot in the cockpit, omission of important information coming from the instruments, slower reaction times, and increased difficulty in staying awake in flight.

The pilot admitted that his performance in flight had at some point been impaired due to fatigue caused by driving to the airport. When he was asked to describe what had happened on that day, he stated: '*That was not the first time I drove a long distance to get to the airport, but that day I spent about 1 hour and a half driving because of traffic. When I arrived at the airport, I was tired, so I had a coffee. I thought I would be fine, but after some time of flying, I started feeling tired again and realised that I could not think clearly*'. The interviewee believed that many pilots have flown fatigued due to driving long distances to the airport. When asked how commonly that happens, he stated that it is not easy to identify if fatigue in flight is caused by driving to the airport or other factors like sleep deprivation or the flight itself.

In real life, sleep- and task-related factors of fatigue often interact. For example, pilots may be fatigued when reporting for duty due to inadequate sleep the previous night, a long drive to commute to the airport, and pre-flight tasks that are completed at the airport (e.g. review flight information). Then, flying multiple sectors and dealing with unanticipated problems in flight can increase pilots' levels of fatigue even further. In occasions like this, it may be difficult to identify what contributed to fatigue and to what extent.

In the question about whether fatigue induced by driving to the airport can impair driving performance after duty, he stated that he and other pilots had at some point found it difficult to stay awake while driving after duty. In the interviewee's opinion, being fatigued while driving after duty is something that happens frequently, but '*it is difficult to say where fatigue comes from because flights are also tiring*'. According to the interviewee, commuting-related fatigue is worth exploring further in order to develop science-based training, new policies that will protect pilots against commuting-induced fatigue, and scheduling that will consider driving-induced fatigue. He stated that changes in the flight time limitations might need to be considered based on research findings on commuting, but he was concerned about regulating pilots' personal time. To sum up, this interview indicated that fatigue induced by long and demanding driving to the airport is a safety issue. The role of such drives in fatigue experienced in flight and while driving after duty were further explored in the survey and the experiments of this thesis.

4.3 Survey with professional pilots

An international, online survey was conducted to explore the extent to which fatigue induced by long and demanding drives to the airport affects pilots' levels of fatigue negatively in flight and while driving after duty. Since the data were collected by asking the pilots, the survey only dealt with perceived, reported fatigue. As mentioned in section 2.1, fatigue is usually the output of the interaction of task- with sleep-related factors. Therefore, although the focus of the survey was on active fatigue induced by driving, the role of two sleep-related factors (i.e. usual sleep duration and frequency of early morning drives) on fatigue and performance was investigated too. Other sleep-related factors (e.g. time awake) can also affect fatigue. However, these were not explored in this survey because the aim was not to investigate sleep-related fatigue in-depth but rather to identify the potential interactive effect of task- and sleep-related factors in commuting pilots' fatigue.

4.3.1 Research questions

The following research questions were formed to achieve the aim of this study;

- How common is commuting by driving among professional pilots worldwide?
- How commonly do professional pilots undertake long and demanding driving commutes?
- How do professional pilots' self-reported levels of fatigue during the commuting cycle relate to the investigated task- and sleep-related factors?
- How does professional pilots' self-reported difficulty in staying awake in flight and while driving after duty relate to the investigated task- and sleep-related factors?
- How does professional pilots' frequency of experiencing impaired performance in flight and while driving after duty relate to the investigated task- and sleep-related factors?

4.3.2 Materials and Methods

Ethical approval was obtained from the AREA Faculty Research Ethics Committee of the University of Leeds (reference number: LTTRAN-071). The survey was conducted in February 2017 by sending an online questionnaire to 580 pilots in the LinkedIn network of the author. The minimum sample size was calculated using the sample size calculator provided by QualtricsXM (2016). For a population of 130000 airline pilots worldwide (Women in Aviation, 2012), a level of confidence of 95%, and a margin of error of 5%, data should be collected from 383 pilots. The only criterion for participation was to be a professional pilot. The contacted pilots varied in the type of their employer, flight experience, and home base. The anonymous questionnaires were completed through the Bristol Online Survey platform (Bristol Online Survey, 2017).

4.3.2.1 Questionnaire design

The questionnaire was first tested in a preliminary study with three professional airline pilots (captains). These were recruited by contacting pilots in the researcher's LinkedIn network. The pilots were asked to complete the online survey and provide feedback regarding its structure and content: what they liked and did not like, how engaging the survey was, and what changes they thought were needed to improve the questionnaire. The pilots found the questionnaire generally straightforward, but they also recommended some changes. One of them suggested adding a definition of the word 'domicile' because many pilots use the term 'home base' instead and another recommended clarifying that the flight hours during the PPL should be counted too in the question about pilots' total flight hours. The feedback that was received was considered to improve the questionnaire.

The first page of the questionnaire provided information about the aim and structure of the survey. It was also stated that the participants had the right to stop completing the survey at any point and withdraw their data. In addition, the participants were informed about the storage of, access to and use of the data collected. Finally, it was mentioned that the survey went through the University of Leeds ethics review and the individuals were asked to give their consent to participate.

Data were collected regarding pilots' age ('*21-30*', '*31-40*', '*41-50*', '*51-60*', '*61-70*', '*over 70*'), gender ('*male*', '*female*', '*prefer not to say*') and total flight hours ('*Less than 2,500*', '*2,500 to 5,000*', '*between 5,000 and 7,500*', '*between 7,500 and 10,000*', '*more than 10,000*'). The participants were also asked about their rank ('*first officer / co-pilot for the military*', '*captain / pilot for the military*'), types of employers ('*mainline airline*', '*regional airline*', '*charter airline*', '*business aviation*', '*ground instructor*', '*cargo airline*', '*military*', '*retired*'), and country of home base. Data about the means of transport used to commute to and from the airport were collected (Figure 4.1) to answer the research question about how common commuting by driving is among professional pilots worldwide.

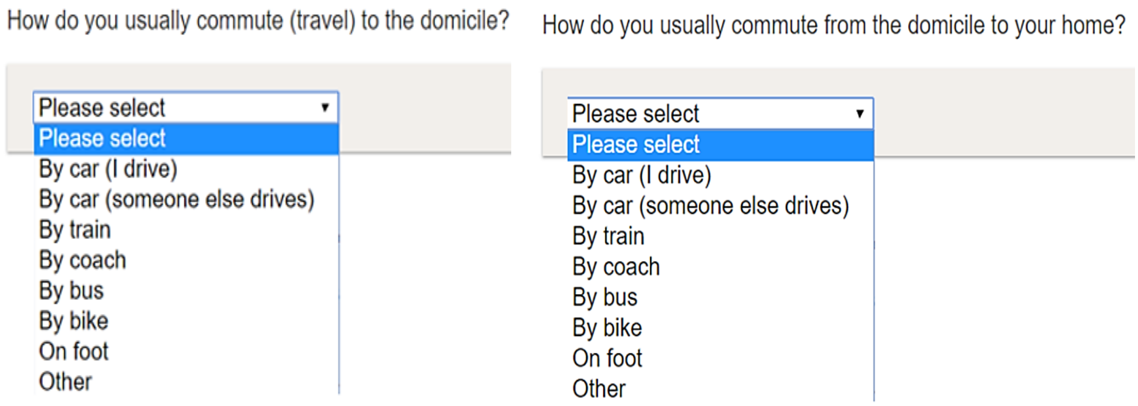


Figure 4.1 The questions about the means of transport used to commute.

The participants were asked about the usual duration (Figure 4.2) and mental demand (Figure 4.3) of their commutes to the airport to explore how these related to the reported levels of fatigue in the commuting cycle, the difficulty in staying awake, and performance decrements in flight and while driving after duty. The item that was used to collect information about individuals' usual levels of mental demand was developed for this survey to focus specifically on commuting to the airport. The definition of 'mental demand' used in this thesis is the one used in the NASA-TLX scale (Hart and Staveland, 1988). According to that, mental demand refers to how much mental and perceptual activity (e.g. thinking and searching) is required to complete a task. Although no definition of mental demand was provided to the participants in this survey, the preliminary study suggested that the participants found the term easy to understand with one of the individuals in that study stating that pilots are familiar with the term mental demand due to their training.

How long does it usually take you to commute from:

	0 - 15 minutes	16 - 30 minutes	31 - 45 minutes	46 - 60 minutes	61 - 75 minutes	76 - 90 minutes	91 - 105 minutes	106 - 120 minutes	More than 120 minutes
home-to-domicile?	☐	☐	☐	☐	☐	☐	☐	☐	☐

Figure 4.2 The question about the usual duration of the commutes to the airport.

How mentally demanding is your usual home-to-domicile commute for you?

	0 (not at all demanding)	1	2	3	4	5	6	7	8	9	10 (extremely demanding)
Mental demand	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

Figure 4.3 The question about the levels of mental demand during the commutes to the airport.

The participants provided information about their usual duration of sleep (Figure 4.4) before commuting to the airport and the frequency of driving to the airport between 2 and 6am (Figure 4.5) in order to investigate if these sleep-related factors related to the reported fatigue, the difficulty in staying awake, and impaired performance. As mentioned in section 2.1.2, there are two circadian troughs in a 24-hour day: one early in the morning and another in the afternoon. However, data were only collected about the early morning commutes in this survey since the aim was to search for indications of an interactive effect of task- and sleep-related contributors to fatigue of commuting pilots rather than explore sleep-related fatigue in-depth.

The data about the usual duration of sleep were collected with a five-point Likert item because it was expected that the participants would find it more difficult to provide an exact number of usual hours of sleep since that may vary between days. In contrast, the categories of the Likert item could help to capture this variation by using ranges of hours. Moreover, the focus was on identifying cases of usual sleep deprivation among the participants rather than calculating sleep duration accurately.

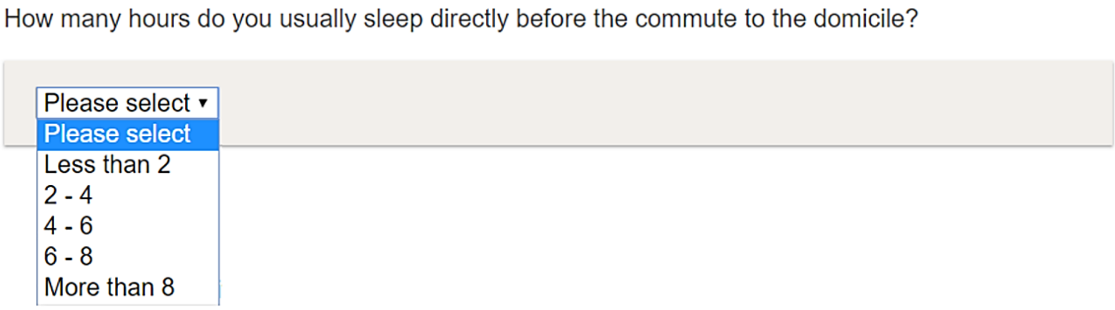
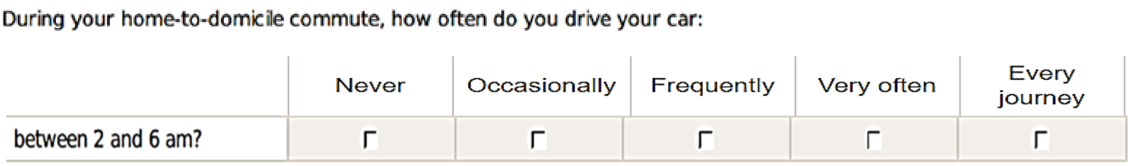


Figure 4.4 The question about the hours of sleep before the commutes to the domicile.



During your home-to-domicile commute, how often do you drive your car:

	Never	Occasionally	Frequently	Very often	Every journey
between 2 and 6 am?					

Figure 4.5 The question about the frequency of commuting between 2 and 6 am.

Fatigue in the commuting cycle was measured by asking the pilots to rate retrospectively their levels of fatigue on the time points shown in Figure 4.6. These questions are shown in Figures 4.7 to 4.9. Since the aim was to explore the extent of the issue of fatigue in pilots who commute by driving, the pilots were asked about their usual levels of fatigue rather than fatigue on a certain day. These levels of fatigue were rated on the CSS fatigue scale. More information about the CSS can be found in section 3.2.1.1.

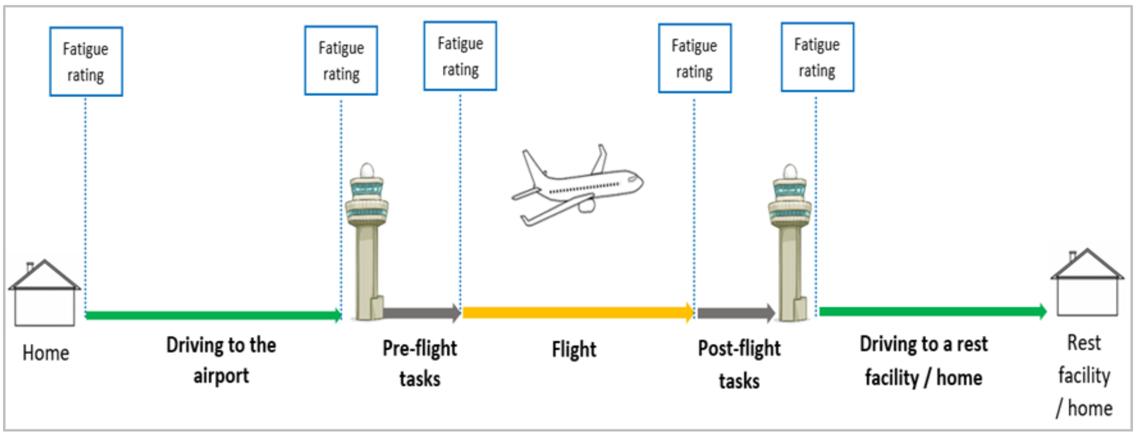


Figure 4.6 The time points of the fatigue ratings during the commuting cycle (blue lines).

- **Fatigue before flight**

Choose the statement which describes how you usually feel:

	1. Fully alert, wide awake	2. Very lively, responsive, but not at peak	3. Okay, somewhat fresh	4. A little tired, less than fresh	5. Moderately tired, let down	6. Extremely tired, very difficult to concentrate	7. Completely exhausted, unable to function properly
just before you start your home-to-domicile commute	☐	☐	☐	☐	☐	☐	☐
when you arrive at the domicile after having commuted from your home	☐	☐	☐	☐	☐	☐	☐

Figure 4.7 The question about the usual levels of fatigue before flight.

- **Fatigue in flight**

Choose the statement which describes how you usually feel:

	1. Fully alert, wide awake	2. Very lively, responsive, but not at peak	3. Okay, somewhat fresh	4. A little tired, less than fresh	5. Moderately tired, let down	6. Extremely tired, very difficult to concentrate	7. Completely exhausted, unable to function properly
just before you enter the aircraft to start your flight as aircrew	☐	☐	☐	☐	☐	☐	☐
when you leave the aircraft after your flight as aircrew	☐	☐	☐	☐	☐	☐	☐

Figure 4.8 The question about the usual levels of fatigue in flight.

- **Fatigue while driving from the airport**

Choose the statement which describes how you usually feel:

	1. Fully alert, wide awake	2. Very lively, responsive, but not at peak	3. Okay, somewhat fresh	4. A little tired, less than fresh	5. Moderately tired, let down	6. Extremely tired, very difficult to concentrate	7. Completely exhausted, unable to function properly
just before you start your commute home from the domicile	☐	☐	☐	☐	☐	☐	☐
when you arrive home after having commuted from the domicile	☐	☐	☐	☐	☐	☐	☐

Figure 4.9 The question about the usual levels of fatigue while driving from the airport.

Data about the difficulty in staying awake were collected with two questions (Figures 4.9 and 4.10). The first one asked about the difficulty in remaining awake in non-rest periods of flights because pilots are under certain conditions allowed to sleep while flying as aircrew either in their seat or in designated bunk facilities in the aeroplane.

- **Staying awake in flight**

How often have you had difficulty remaining awake while flying as aircrew (in non-rest periods of flights) after having commuted to the domicile by driving a car? *

	Never	Occasionally	Frequently	Very often	Every journey	N/A
Frequency	▢	▢	▢	▢	▢	▢

Figure 4.10 The question about the difficulty in staying awake in flight.

- **Staying awake while driving after duty**

How often have you had difficulty remaining awake while commuting from the domicile to your home when driving a car?

	Never	Occasionally	Frequently	Very often	Every journey
Frequency	▢	▢	▢	▢	▢

Figure 4.11 The question about the difficulty in staying awake while driving after duty.

The participants were asked to rate their agreement or disagreement with the statements about performance decrements experienced in flights and while driving after duty (Figures 4.11 and 4.12). The statement about performance decrements while driving after duty referred to the effects of flying rather than those of driving to the airport. Therefore, it did not capture the direct potential link between fatigue induced by driving to the airport driving performance after duty. This decision was made because the interview suggested that pilots may find it difficult to identify that fatigue after duty relates to fatigue before it.

- **Performance decrements in flight**

"Driving the home-to-domicile commute by car has at some point negatively affected your flying performance". Please rate your level of agreement or disagreement with this statement. *

	Totally disagree	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree	Totally agree	N/A
Agreement / Disagreement	▢	▢	▢	▢	▢	▢	▢	▢

Figure 4.12 The question about performance decrement in flight.

- **Performance decrements while driving after duty**

"Your driving performance while commuting home from the domicile directly after a flight duty period has at some point been negatively affected by the flight time of that duty period". Please rate your level of agreement or disagreement with this statement. *

	Totally disagree	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree	Totally agree	N/A
Agreement / Disagreement	☐	☐	☐	☐	☐	☐	☐	☐

Figure 4.13 The question about performance decrement while driving after duty.

Data were also collected about other characteristics of the commutes (i.e. their frequency, the traffic density, and weather conditions), the number of sectors usually flown, and the usual duration and levels of mental demand experienced in flights. The results of the analysis of these data will not be presented in this chapter because it was decided to focus on the identification of the extent of the issue of commuting-related fatigue in the survey and explore the contributors to it in more depth in the study of chapter 5.

4.3.2.2 Data analysis

Descriptive statistics were used to answer how common commuting by driving among professional pilots is and how commonly pilots undertake long and demanding driving commutes. The time point of the commuting cycle (Figure 4.6), the usual levels of mental demand experienced during the drives to the airport, the usual duration of those drives, the frequency of driving to the airport between 2 and 6 am, and the usual duration of sleep before commuting were the independent variables. The dependent variables were the reported usual levels of fatigue and the ratings on the questions about the difficulty in staying awake and the performance decrement in flight and while driving after duty. All these variables were treated as ordinal because their categories were ordered but the intervals on the relevant items were unequal. The reported levels of mental demand were also treated as ordinal, although the intervals on that item were equal. This decision was made because there is no dedicated test for correlations between continuous and ordinal variables.

The correlations of the investigated task- and sleep-related factors with the reported usual levels of fatigue, the difficulty in staying awake, and impaired performance in flying and driving after duty were tested with Spearman's tests. The reported usual levels of fatigue were compared between the time points of the commuting cycle with a Friedman test and Wilcoxon tests for pairwise comparisons (Bonferroni correction). These tests would show how pilots' levels of fatigue usually changed due to driving to the airport, flying, and spending time at the airport before and after their duties. The associations of the investigated task- and sleep-related factors with reported fatigue, the difficulty in staying awake, and performance decrements were tested with Spearman's tests. The statistical analyses described in this chapter were performed with the IBM SPSS Statistics 24.0 (IBM, 2016).

4.3.3 Results

4.3.3.1 Demographics

Four hundred thirty-four pilots flying from 52 countries participated (74.8% response rate). Three of them reported usually not using the same means of transport for both their commutes to and from the airport, so their data were removed. Moreover, three pilots working for the military, two in retirement, and seven ground instructors participated. The data of the retired pilots were removed because the focus of the study was on pilots currently working in commercial aviation. The data of the military pilots and the ground instructors were removed too because, as explained in chapter 1, factors not related to a certain flight are likely to often mediate those pilots' levels of fatigue in the commuting cycle. The demographics of the 419 pilots left are shown in Table 4.1 and the percentages of participants per means of transport in Figure 4.14. As depicted, there were similar percentages of pilots in all the age groups and the ranks. Nevertheless, most participants were males, very experienced, worked for a mainline airline, and had a home base in Europe.

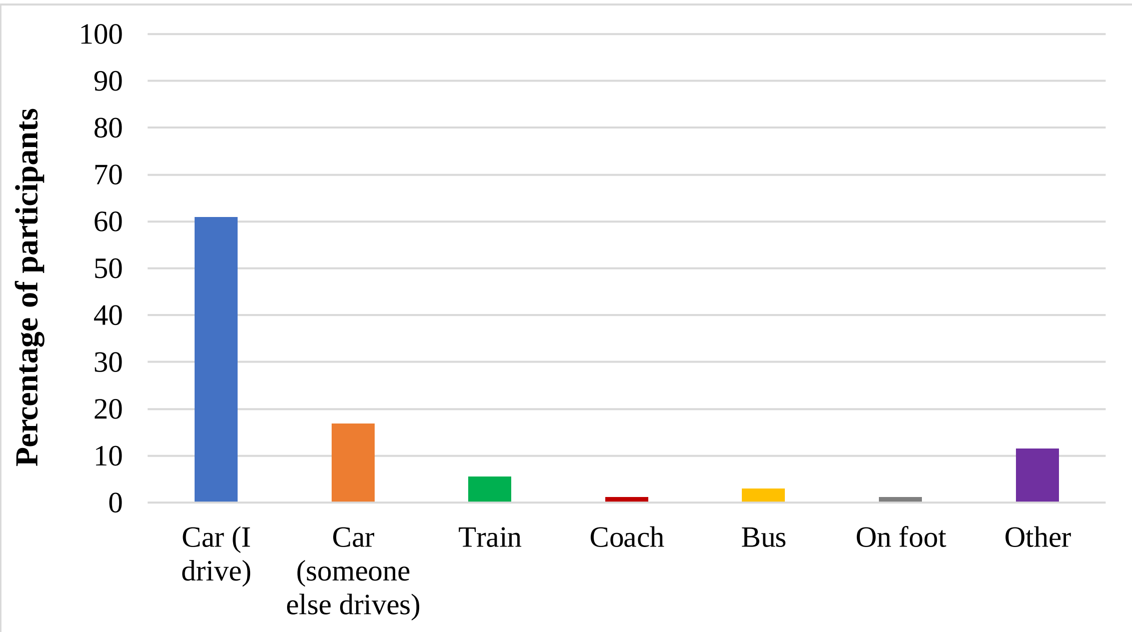


Figure 4.14 The percentage of participants per means of transport usually used to commute to and from the airport (N=419).

Table 4.1 Demographics of the participants in the survey (N=419).

Characteristic	Category	N (%)
Age	21 – 30	20.6
	31 – 40	30.4
	41 – 50	27.2
	51 - 60	21.8
Gender	Male	97.2
	Female	2.8
Total flight hours	<2,500	18
	2,500 – 5,000	16.1
	Between 5,000 and 7,500	13.1
	Between 7,500 and 10,000	13.4
	>10,000	39.4
Rank	First officer	47.5
	Captain	52.5
Employer	Mainline airline	63.6
	Regional airline	11.1
	Charter aviation	9.4
	Business aviation	8.2
	Cargo	7.7
Home base	Europe	58.1
	Asia	23.9
	America	16
	Africa	1
	Oceania	1

The demographics of the 255 pilots who commuted by driving are shown in Table 4.2. The percentages of pilots were similar in the age groups between 21 and 50 years old, but there were fewer pilots over the age of 50. Half of the pilots were captains and had more than 7500 flight hours, the majority were males, and most worked for a mainline airline and had a domicile in Europe. This sample is representative of the population of pilots in terms of age (Civil Aviation Authority - CAA, 2018; Federal Aviation Administration, 2018) and gender (Women in Aviation International, 2016). All the analyses that follow in this chapter refer to the 255 pilots who commuted by driving.

Table 4.2 Demographics of the pilots who usually commuted by driving (N=255).

Characteristic	Category	N (%)
Age	21 – 30	24.9
	31 – 40	29.4
	41 – 50	29.4
	51 - 60	16.3
Gender	Male	97
	Female	3
Total flight hours	<2,500	23.7
	2,500 – 5,000	17.6
	Between 5,000 and 7,500	14.9
	Between 7,500 and 10,000	9.5
	>10,000	34.3
Rank	First officer	49.6
	Captain	50.4
Employer	Mainline airline	55.7
	Regional airline	13
	Charter aviation	11.8
	Business aviation	10.8
	Cargo	8.7
Home base	Europe	68.7
	Asia	16
	America	13.4
	Africa	0.8
	Oceania	1.1

4.3.3.2 Task- and sleep-related factors

The pilots were asked about the usual duration and mental demand of their drives to the airport, their usual duration of sleep, and the frequency of driving between 2 and 6 am. The results are shown in Figures 4.15 to 4.18. These data were needed to explore how the reported levels of fatigue before and after flights, the difficulty to stay awake, and performance decrements in flight and while driving after duty related to the investigated task- and sleep-related factors.

As shown in Figure 4.15, around two out of three pilots reported usually completing short drives to the airport (i.e. less than 45 minutes) and only about 20% of the pilots reported usually experiencing high levels of mental demand (i.e. six or more). According to the National Sleep Foundation (Hirshkowitz et al., 2015), adults should get 7 to 8 hours of sleep every night. This considered, approximately 35% of the pilots reported usually not getting enough sleep before duty (figure 4.17). The percentage of usually sleep-deprived pilots may, however, be higher than 35% because some of those who selected the '*6-8 hours*' option might have referred to sleep of fewer than 7 hours. Regarding the frequency of driving early in the morning, approximately two out of three pilots reported driving between 2 and 6 am frequently, very often, or in every journey.

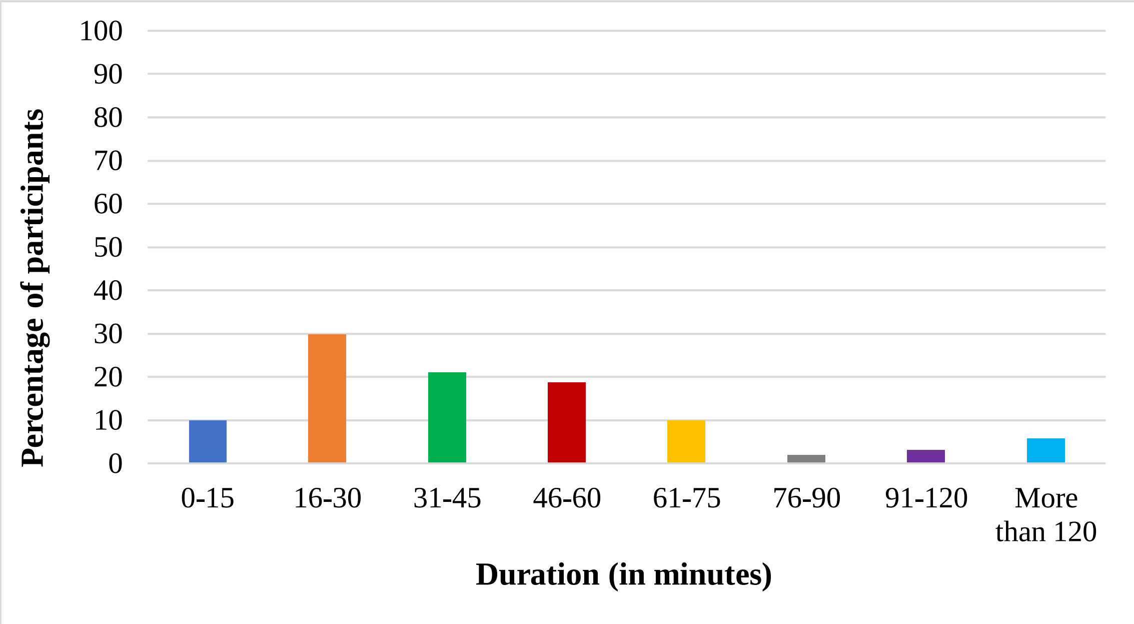


Figure 4.15 Usual duration (minutes) of the drives to the airport.

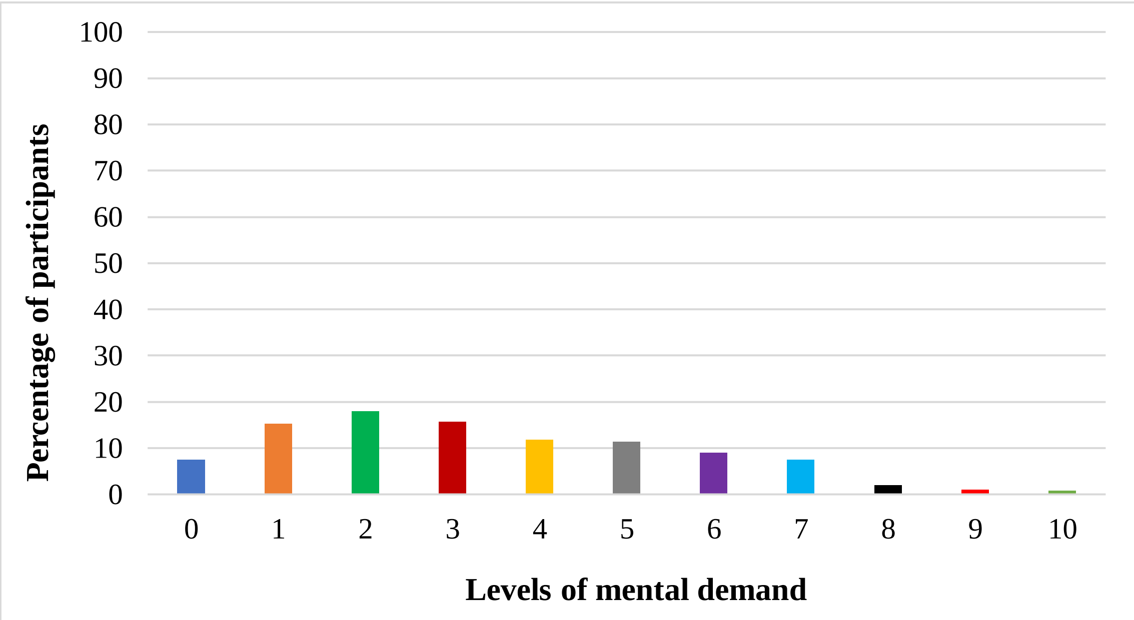


Figure 4.16 Usual levels of mental demand while driving to the airport (0 = '*not demanding at all*', 10 = '*extremely demanding*').

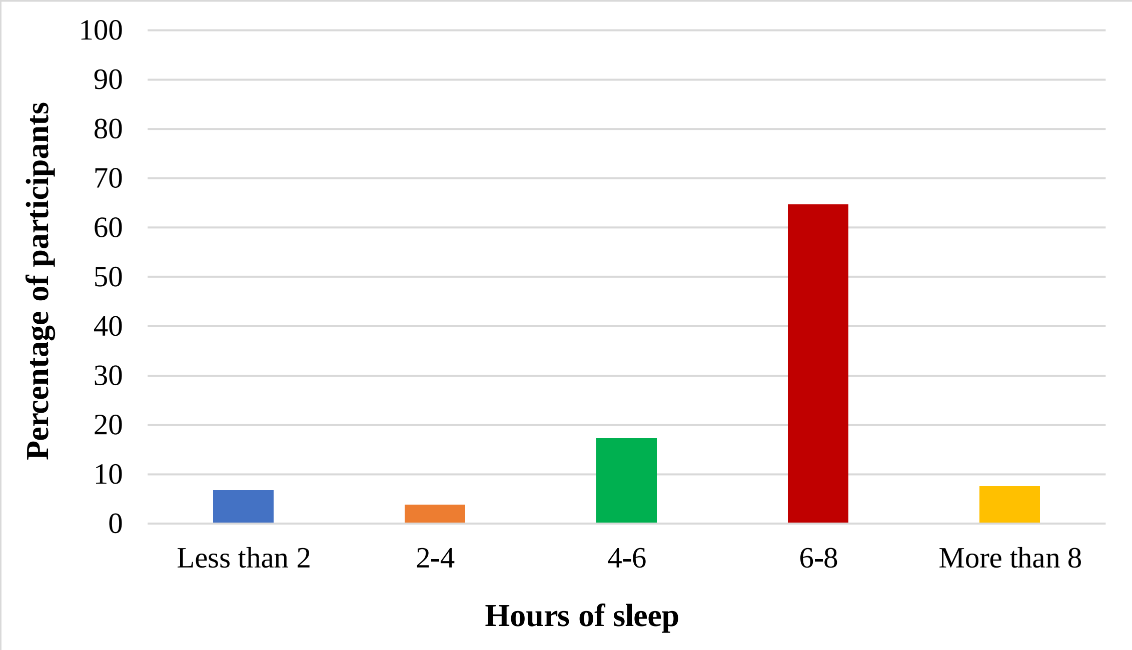


Figure 4.17 Usual sleep duration before driving to the airport (hours).

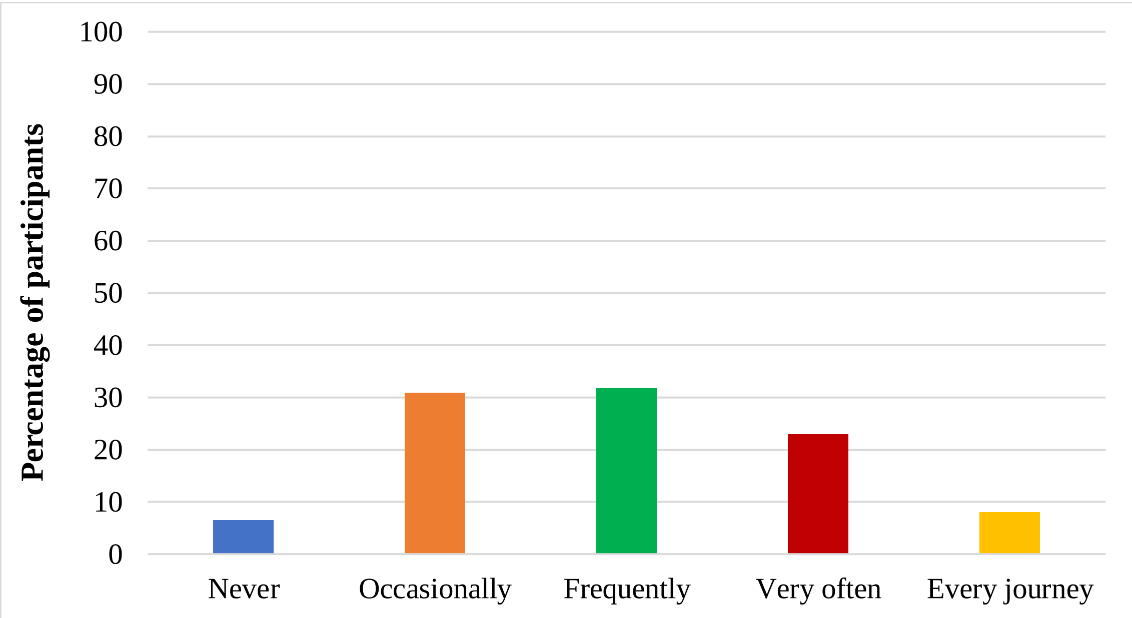


Figure 4.18 Frequency of driving between 2 and 6 am to the airport.

4.3.3.3 Reported usual levels of fatigue during the commuting cycle

The pilots were asked to rate their usual levels of fatigue on five time points of the commuting cycle to explore how they usually changed due to driving, flying, and spending time at the airport. At the same time, the correlations of the investigated task- and sleep-related factors with fatigue before and after the drives and flights could be investigated. Figure 4.19 depicts the medians (with 95% confidence intervals) of the reported usual levels of fatigue during the commuting cycle for the pilots who usually commuted by driving (n=255). It should be noted that the bars 'Before driving to the airport' and 'When they enter the aeroplane' look the same because the medians are depicted. Although the ratings in the y axis between these two measurements could differ, that comparison was not made because the interest was in the changes of the ratings of fatigue between consecutive time points in the commuting cycle rather than on differences between any measurements. As shown in Figure 4.19, the levels of fatigue before driving to the airport, upon arriving at the airport before duty, and when entering the aeroplane were relatively low (3 = '*okay, somewhat fresh*'). In contrast, the median of fatigue when exiting the aeroplane (5 = '*moderately tired, let down*') and just before driving after duty (4 = '*a little tired, less than fresh*') were moderate.

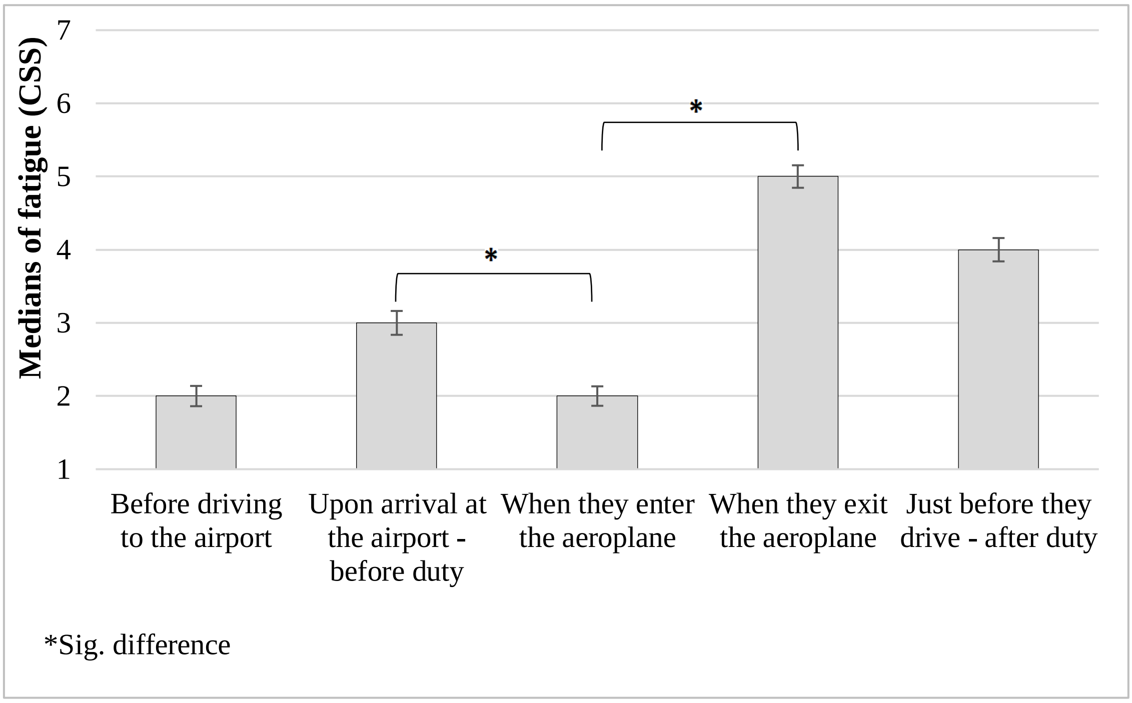


Figure 4.19 The CSS ratings during the commuting cycle of the pilots that usually commuted by driving.

The self-reported usual levels of fatigue of the pilots that commuted by driving were compared between the consecutive time points of the commuting cycle with a Friedman test with Bonferroni corrections. This test was significant, χ^2 (255) = 529.474, p < .001, Kendall's coefficient of concordance = .51. Follow-up pairwise comparisons were conducted with a Wilcoxon test with a

Bonferroni correction in order to reveal differences between consecutive measurement time points of the commuting cycle. It was found that the usual levels of fatigue;

- Upon arrival at the airport to report for duty (median = 3) were not significantly different than before driving to the airport (median = 2) (p = .396).
- When entering the aeroplane (median = 2) were significantly lower than upon arrival at the airport to report for duty (median = 3) (p < .001).
- When exiting the aeroplane (median = 5) were significantly higher than upon entering it (median = 2) (p < .001).
- Just before driving after duty (median = 4) were not significantly different than when exiting the aeroplane (median = 5) (p = 1).

Spearman's tests that were performed to explore how the investigated task- and sleep-related factors associated with fatigue during the commuting cycle in all the pilots that commuted by driving. As shown in Table 4.3, usually completing longer and more mentally demanding drives to get to the airport and usually sleeping less before those commutes related to significantly higher self-reported levels of fatigue upon arriving at the airport to report for duty.

Table 4.3 The correlations of the investigated task- and sleep-related factors with fatigue directly after driving to the airport for the pilots who usually commuted by driving.

	Usual duration of driving to the airport	**Usual mental demand during driving to the airport**	**Usual sleep duration before driving to the airport**	**Frequency of driving to the airport 2-6 am**
Fatigue directly after driving to the airport	r_s = .237, p < .001*	r_s = .356, p < .001*	r_s = -.218, p < .001*	r_s = .077, p = .221

As shown in Table 4.4, usually completing longer and more mentally demanding drives to commute to the airport, usually sleeping less before driving to the airport, and commuting to the airport more often between 2 and 6 am related to significantly higher self-reported levels of fatigue when entering the aeroplane. Fatigue when exiting the aeroplane was also related to the mental demand of the drives, sleep duration, and driving early in the morning, but there was only a trend for a correlation with the duration of the drives.

Table 4.4 The correlations of the investigated task- and sleep-related factors with fatigue in flight for the pilots who usually commuted by driving.

	Usual duration of driving to the airport	**Usual mental demand during driving to the airport**	**Usual sleep duration before driving to the airport**	**Frequency of driving to the airport 2-6 am**
Fatigue when entering the aeroplane	$r_s = .166$, $p = .008$*	$r_s = .313$, $p < .001$*	$r_s = -.224$, $p = .001$*	$r_s = .144$, $p = .022$*
Fatigue when exiting the aeroplane	$r_s = .112$, $p = .074$	$r_s = .289$, $p < .001$*	$r_s = -.224$, $p < .001$*	$r_s = .157$, $p = .012$*

As depicted in Table 4.5, the usual reported levels of fatigue just before driving after duty did not relate to longer and more often early-morning drives to the airport. In contrast, reporting higher usual levels of mental demand during driving to the airport and shorter usual sleep duration related to significantly higher self-reported levels of fatigue just before driving after duty.

Table 4.5 The correlations of the investigated task- and sleep-related factors with fatigue just before driving after duty for the pilots who usually commuted by driving.

	Usual duration of driving to the airport	**Usual mental demand during driving to the airport**	**Usual sleep duration before driving to the airport**	**Frequency of driving to the airport 2-6 am**
Fatigue just before driving after duty	$r_s = .081$, $p = .197$	$r_s = .261$, $p < .001$*	$r_s = -.167$, $p = .007$*	$r_s = .032$, $p = .606$

4.3.3.4 Difficulty in staying awake

The pilots were asked to rate the frequency of experiencing a difficulty in staying awake in flight and while driving after duty to explore how these related to the investigated task- and sleep-related factors. As shown in Figure 4.20, more than 75% of the pilots reported that they had experienced a difficulty in staying awake in flight after driving to the airport occasionally, frequently or very often. Furthermore, more than 85% of the pilots had experienced a difficulty in staying awake while driving after duty occasionally, frequently, very often or on every journey (Figure 4.21).

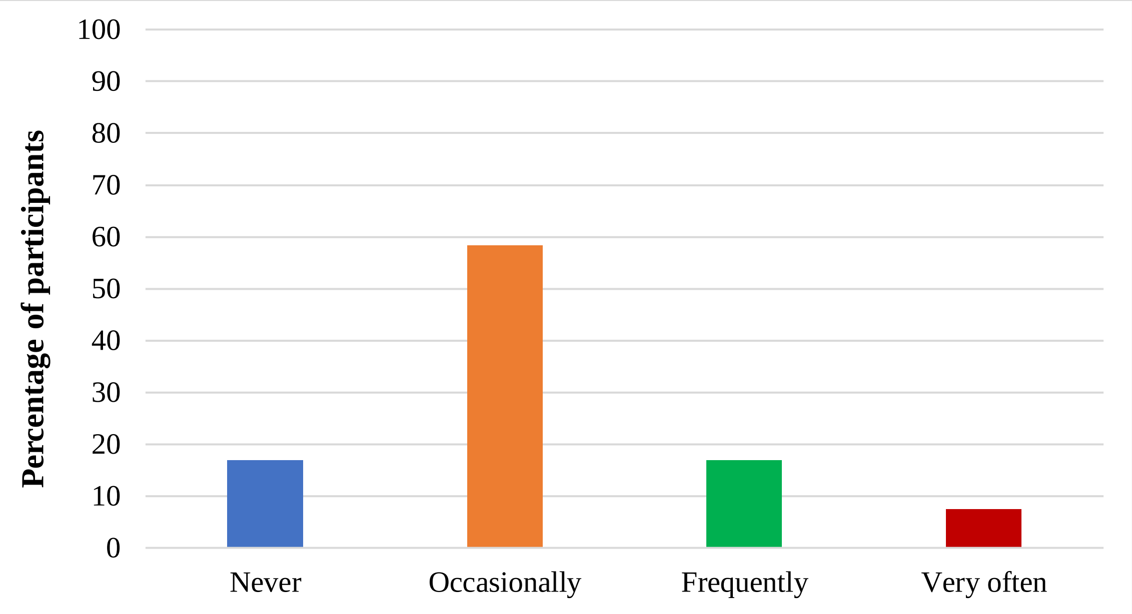


Figure 4.20 Frequency of experiencing a difficulty in staying awake in flight.

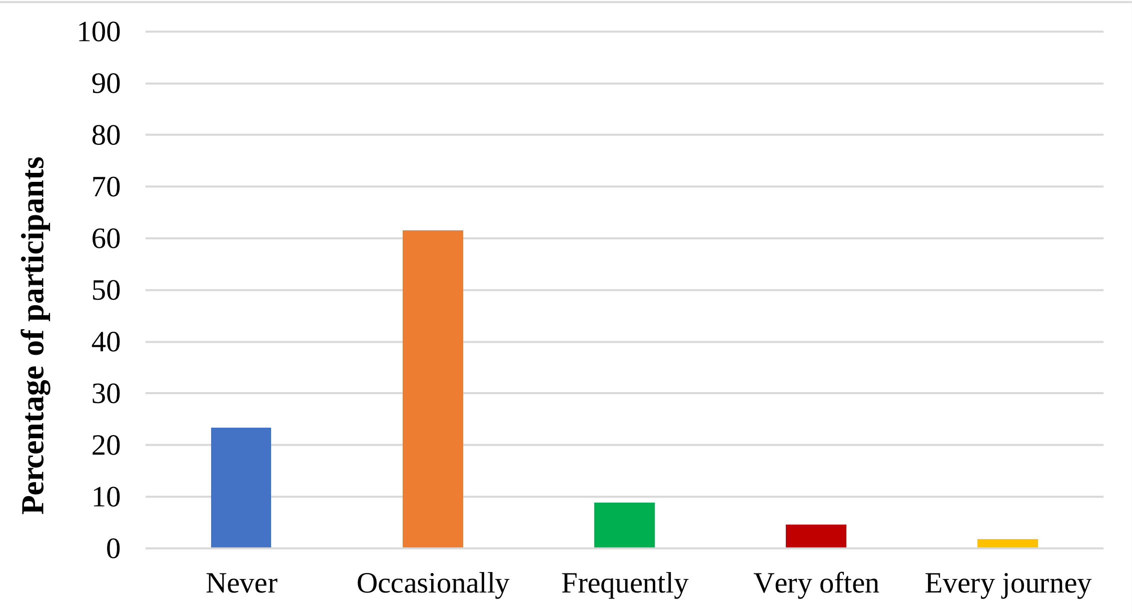


Figure 4.21 Frequency of experiencing a difficulty in staying awake in driving after duty.

Table 4.6 depicts the results of the Spearman's tests that were run to explore how the investigated task- and sleep-related factors related to the frequency of experiencing a difficulty in staying awake in flight and while driving after duty in the pilots who commuted by driving. As depicted, usually completing longer drives to the airport and experiencing higher levels of mental demand during those drives related to a significantly more frequently experienced difficulty in staying awake in flight and while driving after duty. In contrast, usually sleeping more before the commutes to the airport related to a significantly less frequently experienced difficulty in staying awake in flight and while driving after duty. Finally, driving to the airport more often between 2 and 6 am related to a significantly more frequently reported difficulty in staying awake in flight, but not while driving after duty.

Table 4.6 The correlations of the investigated task- and sleep-related factors with the frequency of experiencing a difficulty in staying awake in flight and while driving after duty.

	Usual duration of driving to the airport	**Usual mental demand during driving to the airport**	**Usual sleep duration before driving to the airport**	**Frequency of driving to the airport 2-6 am**
Difficulty in staying awake in flight	$r_s = .125$, $p = .047$*	$r_s = .265$, $p < .001$*	$r_s = -.292$, $p < .001$*	$r_s = .204$, $p = .001$*
Difficulty in staying awake while driving after duty	$r_s = .313$, $p < .001$*	$r_s = .265$, $p < .001$*	$r_s = -.157$, $p = .012$*	$r_s = .112$, $p = .075$

4.3.3.5 Performance decrements

The relationship of the investigated task- and sleep-related factors with performance decrements in flight and while driving after duty was explored by asking the participants to rate their agreement with two relevant items. As depicted in Figure 4.22, more than 2 out of 3 pilots agreed at least to a degree that driving to the airport had at some point affected negatively their flying performance. Furthermore, more than 80% of the pilots (Figure 4.23) agreed at least partially that their performance while driving after duty had at some point been impaired by the flight time of that duty period. The data were analysed with Spearman's tests, the results of which are shown in Table 4.7. As depicted, usually completing longer drives to the airport and experiencing higher levels of mental demand during those drives related to a significantly higher agreement with both statements. In contrast, no significant correlations were found between the agreement with these statements and the investigated sleep-related factors.

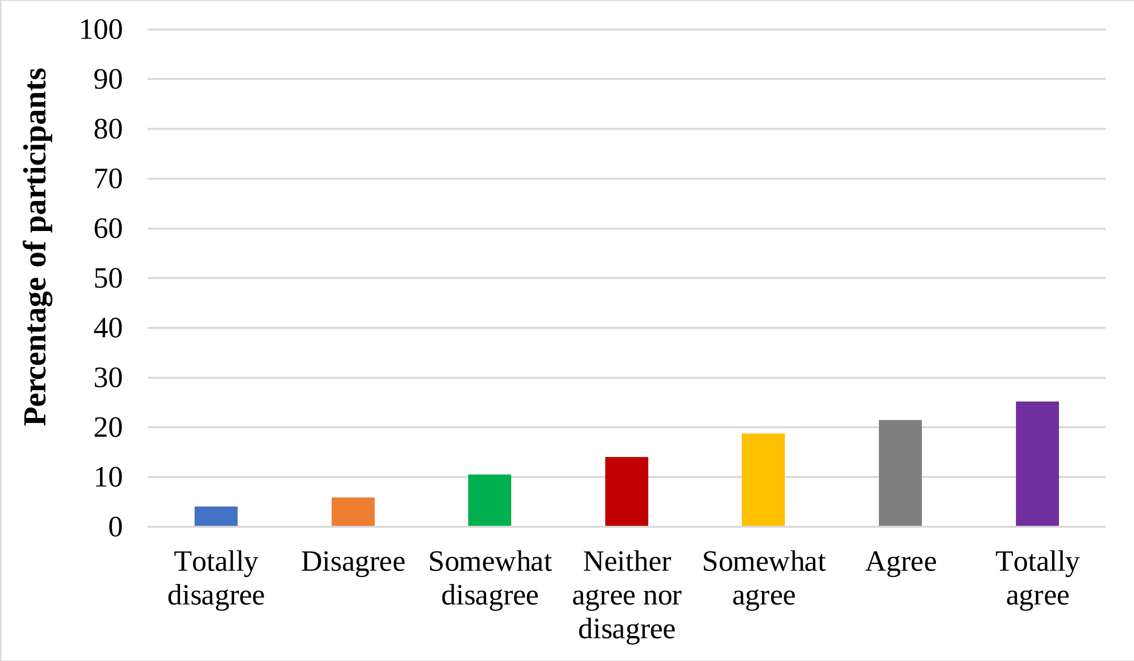


Figure 4.22 Agreement with the statement about impaired flying performance.

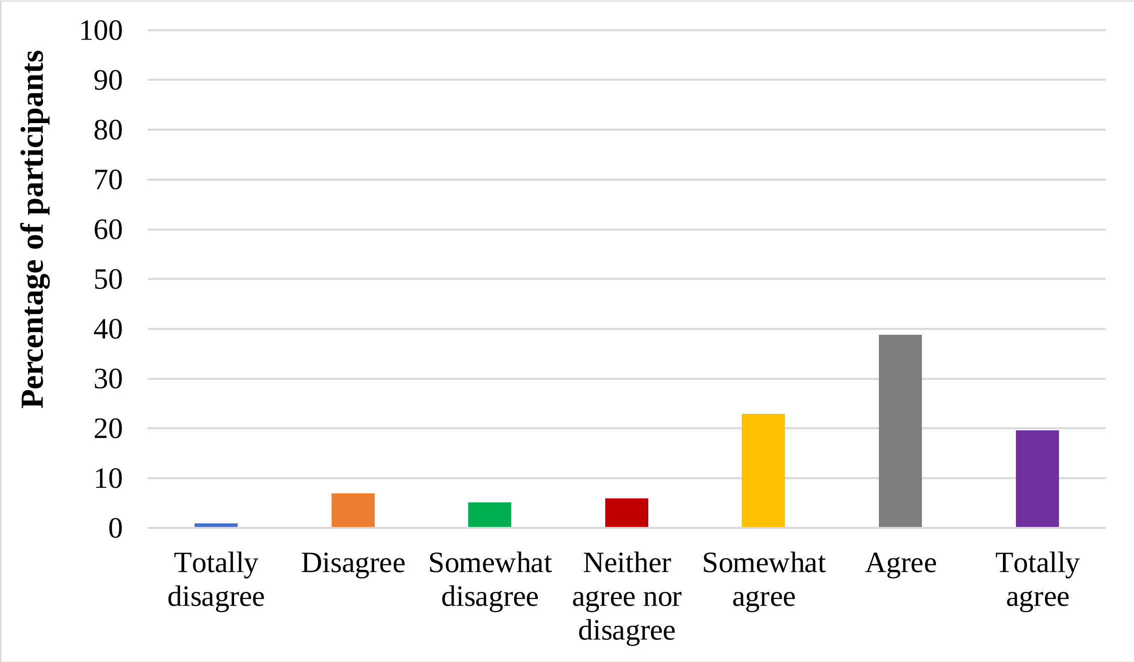


Figure 4.23 Agreement with the statement about performance decrements in driving after duty.

To summarise, the analysis showed that most pilots usually commuted by driving but those drives were usually not very demanding and long. Their reported levels of fatigue were significantly lower when entering the aeroplane than directly after the drive to the airport, and significantly higher when exiting than entering the aeroplane. The investigated task- and sleep-related factors correlated significantly to the usual reported levels of fatigue in the commuting cycle, and the difficulty in staying awake and the performance decrements in flight and while driving after duty. These results are discussed in the next section.

Table 4.7 The correlations of the investigated contributors to fatigue with the reported performance decrements in flight and while driving after duty.

	Usual duration of driving to the airport	**Usual mental demand during driving to the airport**	**Usual sleep duration before driving to the airport**	**Frequency of driving to the airport 2-6 am**
Performance decrements in flight after driving	r_s = .198, p = .001*	r_s = .398, p < .001*	r_s = -.094, p = .135	r_s = .044, p = .479
Performance decrements while driving after duty	r_s = .280, p < .001*	r_s = .342, p < .001*	r_s = -.091, p = .148	r_s = .015, p = .815

4.4 Discussion

Aviation organisations have identified fatigue induced by commuting to the airport as a potential risk for flight safety. Drawing from driver studies, that risk might be particularly high when pilots commute to the airport by completing long drives of high task load because these could deplete their attentional resources. Since flying can be fatiguing too, the adverse effects of driving to the airport might be intensified, increasing the risk of aeroplane crashes and car accidents after duty. Everyday life is full of risks and high-hazard industries like aviation decide which ones to tackle by assessing their likelihood and impact on safety. In the case of fatigue induced by long and demanding drives to the airport and flying, these are unknown. As a result, aeroplane accidents and car crashes after duty may occur because this risk has not been properly identified. This considered, the survey described in this chapter aimed to explore the extent to which fatigue induced by long and demanding drives to the airport affects negatively pilots' levels of fatigue in flight and while driving after duty.

Due to the lack of literature specifically on fatigue in pilots who commute by driving, an interview with an airline pilot was conducted before the survey in order to identify the key aspects of the issue. The interviewee stated that fatigue induced by driving to the airport had resulted in many pilots (including himself) flying fatigued and experiencing impaired performance in flight. The long drives in high traffic density were found particularly fatiguing and were identified as the ones that are more likely to contribute to pilots flying fatigued. Therefore, he indirectly suggested that active fatigue induced by driving affects performance in flight negatively. He also mentioned that some pilots have difficulties in staying awake while driving after duty, which he believed could have been caused by fatigue induced by driving to the airport and flying. These findings indicate that fatigue (especially active) caused by driving to the airport and flying is an extensive safety risk among professional pilots. They also highlight the need for further investigation of the issue before developing appropriate interventions.

Following from the interview, an online survey of 419 professional pilots was conducted to explore the extent of the issue with the focus being on the role of the task load and the duration of the drives to the airport. Data were collected about the means of transport the pilots usually used for their commutes and their perceived levels of fatigue before and after driving and flying. Moreover, they reported how often they had experienced difficulty in staying awake in flight after having commuted to the airport by driving a car and while driving after duty. They also rated their agreement with a statement about performance deterioration in flight due to driving to the airport and impaired performance while driving after duty due to the flight time of that duty period. Data about the usual duration of sleep before commuting to the airport and the frequency of commuting between 2 and 6 am were collected as well to identify the potential interactive effect of task- and sleep-related factors on pilots' fatigue in the commuting cycle.

The aim of the survey was achieved by finding that most pilots usually commuted by driving and that approximately one in ten of them usually completed long and mentally demanding drives to the airport. Moreover, usually completing long and mentally demanding drives to the airport related to higher reported levels of fatigue in the commuting cycle, experiencing more often a difficulty in staying awake in flight and while driving after duty, and more frequently reported performance decrements in flight and while driving after duty. Sleep deprivation and driving early in the morning were linked to fatigue and performance decrements too. Furthermore, the survey suggested that fatigue induced by driving to the airport usually increases further in flight. Finally, it was indicated that the time spent at the airport before flights usually helps pilots to reduce their levels of fatigue, but this is usually not the case with the time spent at the airport after duty. These results are discussed in detail in this section.

Since most pilots usually commuted by driving to and from the airport (60.9%), the findings of the study represent a great proportion of the sample. The fact that driving was the most common commuting option is in line with the literature on other professionals (e.g. Department for

Transport, 2015). Nevertheless, it disagrees with the results of the studies on pilots conducted by the National Research Council (2011) and Kleinfehn (2016). It should be noted that the data in these two studies were collected in the US, which is a country with a big geographical spread. In contrast, more than half of the pilots in this survey had a home base in Europe and only 16% of them were based in the US. This difference suggests that regional effects can play a role in pilots' commuting habits. The percentage of pilots who commute by driving at least sometimes is expected to be higher than the 60.9% found in this study (Figure 4.1) because the participants were asked about their usual commutes only.

The reported usual levels of fatigue directly after the drives to the airport were not significantly higher than those before starting the commutes to work. This finding might be explained by the fact that most pilots reported usually not completing long and mentally demanding drives to work. According to the literature solely on drivers, it is specifically those drives that induce active fatigue (e.g. Gimeno et al., 2006). However, in line with that literature, reporting longer and more mentally demanding drives to the airport related to higher usual levels of fatigue upon arriving at the airport to report for duty. Therefore, although the survey suggests that fatigue induced by driving to the airport is not usually a safety issue for most pilots, it indicates that pilots who complete long and demanding drives are at an increased risk of reporting fatigued for duty.

Regarding the proportion of the population of pilots that this increased risk could apply to, that could be quite high because about 10% of the participants reported usually completing long and mentally demanding drives to the airport. It should be noted that this percentage does not capture those pilots who complete long and demanding drives only in some of their commutes, for example, because they drive in peak traffic. This considered, more than 10% of the professional pilots worldwide could be at an increased risk of experiencing active fatigue upon arriving at the airport due to driving.

This survey extends the existing literature not only by suggesting that many pilots are at an increased risk of arriving fatigued at the airport due to completing long and demanding drives but also by indicating that the duration and mental demand of those drives relate to fatigue in flight. That is, the pilots who reported usually completing longer (a trend was found for that) and more mentally demanding drives to the airport also reported usually starting and finishing their flights more fatigued. The medians of fatigue were 3 when entering the aeroplane ('*okay, somewhat fresh*') and 5 ('*moderately tired, let down*') when exiting it, which means that most pilots usually did not experience high levels of fatigue in flight after driving to the airport. However, the association of fatigue in flight with the characteristics of the drives to the airport indicates an increase in the risk of aeroplane accidents after long and mentally demanding drives. In real life, the first and last stages of flights are usually the most demanding ones and, therefore, require the investment of more attentional resources than the cruise stage. As a result, even a small induction of active fatigue by driving to the airport might contribute to accidents, particularly in those stages by causing a mismatch between the resources available and those needed.

The potential association of driving to the airport with the risk of aeroplane accidents was also supported by finding that the pilots who reported usually completing longer and more mentally demanding drives to the airport stated that they had experienced more often a difficulty in staying awake in flight. They also agreed more with the statement that their flying performance had at some point been negatively affected by driving to the airport. It is concerning that approximately 80% of the participants who commuted by driving agreed at least to a degree with the statement about the difficulty in staying awake in flight and about 65% with the one about the decrements of flying performance. When seen together with the finding that most pilots reported usually not completing fatiguing drives to the airport, these percentages suggest that fatigue induced by driving to the airport could have affected most pilots in the survey only on some of their duties. Based on the significant correlation of the duration and mental demand of the drives with fatigue in flight, the pilots might have completed long and demanding drives on those days. Those cases are of particular interest because that is when the risk of accidents would increase.

It should be noted that the statements about the difficulty in staying awake and performance impairment in flight did not refer specifically to the effects of fatigue induced by driving. Hence, other factors might have also contributed to those decrements. For example, Friesacher and Greaves (2016) attributed pilots' reported performance decrements in flight to stress caused by commuting. In addition, the correlations between driving and flying do not imply causation. Moreover, the survey did not reveal if and how often pilots had fallen asleep in flight or which aspects of performance and how frequently had been impaired. However, it indicated that fatigue caused by driving increased the risk of aeroplane accidents and, therefore, highlighted the need to explore commuting-related fatigue further in the future.

Nodding off in flight and experiencing deterioration of flying performance has been linked to sleep-related factors in previous studies (Le Duc et al., 1999; Thomas and Ferguson, 2010). This association was also supported in this survey because it was found that the pilots who reported usually sleeping less before their commutes also reported usually being more fatigued upon arriving at the airport and when they started their flights. Sleep deprivation (i.e. sleep of fewer than 8 hours according to the FAA, 2009) was found to be common among the participants with approximately 30% of them usually sleeping less than 6 hours per night. The percentage of usually sleep deprived pilots was probably even higher because about 65% of the participants reported sleeping between 6 and 8 hours. Personal factors might have played a role in sleep deprivation (Bourgeois-Bougrine et al., 2003). Another contributor to sleep loss might have been driving to the airport early in the morning because about two-thirds of the sample completed drives between 2 and 6 am at least frequently. Besides sleep duration, higher usual levels of fatigue at the beginning of the flights related to commuting more often to the airport between 2 and 6 am. These findings suggest that sleep-related fatigue can coexist with fatigue induced by driving to the airport and, thus, the effect of the one on performance might be intensified by the other.

The relationship found between the duration of the drives to the airport with fatigue, the difficulty in staying awake, and performance decrements in flight supports the suggestions made by aviation organisations that long commutes to the airport increase the risk of impaired flying performance (e.g. FAA, 2012[a]). This risk has also been suggested in studies with non-pilots (Azwar et al., 2018) and pilots (Zakariassen et al., 2019). Although the findings of this survey contradict those by Brown and Whitehurst (2011) and Roma et al. (2012), Roma et al. also believed that commuting to the airport can be fatiguing for pilots when it is performed by driving. This study extended the existing knowledge in pilots and other professionals by supporting the association between driving to the airport and work performance based on ratings of fatigue. It also identified the potential role of task load during the commutes to the airport in the risk of aeroplane accidents. In addition, in contrast to the existing literature in pilots, this survey explored fatigue in flight specifically when pilots commute by driving.

In contrast to aviation organisations and the studies in pilots and other professionals, this survey suggested a link between active fatigue induced by driving to the airport and the risk of car crashes after duty. That is, the pilots who reported usually completing drives of higher levels of mental demand to get to the airport also reported usually starting the drives after duty more fatigued. The median of the usual levels of fatigue when starting the drives after duty was 4 (‘*a little tired, less than fresh*’), which suggests usually experiencing moderate levels of fatigue. The correlation between mental demand and fatigue does not mean that the one caused the other but indicates an increased risk of fatigue-related car crashes after duty when the drives before flight are mentally demanding. Experiments with simulated driving and flying could provide a better insight into this link.

Some more evidence about the link between fatigue induced by driving to the airport and the risk of car crashes after duty was provided by analysing the data about the participants’ difficulty in staying awake in driving and impaired driving performance. That is, it was found that the pilots who reported usually completing longer and more mentally demanding drives to the airport also reported more frequent difficulties in staying awake while driving after duty. They also agreed more with the statement about deteriorated driving performance. Approximately 75% of the pilots reported a difficulty in staying awake while driving after duty and 80% agreed at least to a degree with the statement about impaired driving performance. These percentages suggest that the risk of car crashes when commuting after duty is extensive among professional pilots. However, the full extent of this issue was not revealed because the participants were not asked how often these adverse effects had been and which aspects of driving performance had been impaired.

Experiments could help to explore in more depth the link between fatigue induced by driving before flight and fatigue-related performance decrements while driving after flight.

It is worth noting that most pilots stated that flying had contributed to a deterioration of their driving performance after duty. This risk has not been identified by aviation organisations and studies in pilots. Although there is evidence from studies in other professionals that the risk of car crashes after work increases due to extended and night shifts (e.g. Barger et al., 2005), the survey described in this chapter is the first that suggests that the risk of fatigue-related crashes when driving after work is also high in pilots. Furthermore, in line with the literature in other professionals (Dorrian et al., 2008; Scott et al., 2008), it was found that usually sleeping less related to reporting usually starting the drives after duty more fatigued and experiencing a higher difficulty in staying awake during those drives. Additionally, driving to the airport more often between 2 and 6 am related to fatigue and the difficulty in staying awake in flight. The time of driving to work has not been identified as a contributor to fatigue experienced while commuting after work in the existing literature in pilots and other professionals. These results suggest that pilots who commute by driving experience a mix of task- and sleep-related fatigue during the commuting cycle. Hence, it is possible that the effects of the one intensify the effects of the other. Experiments could improve understanding of this interaction by controlling for the effect of some of the contributors to fatigue while manipulating others. Then, intervention strategies specifically for the type of fatigue experienced could be tested.

Another contribution of this survey is the identification that active fatigue caused by driving to the airport is usually mediated by the time spent at the airport before duties. That is, it was found that the reported usual levels of fatigue when entering the aeroplane were significantly lower than upon arriving at the airport before duty. No data were collected about pre-flight activities and no literature was sourced on pre-flight intervention strategies for fatigue. Nevertheless, this finding indicates that the pilots who drive to the airport usually use effective self-initiated intervention strategies to manage fatigue before flying. On the other hand, regulations require that pilots report fit for duty, so it is unclear if some participants avoided reporting that they usually started their flights fatigued. Moreover, it should also be noted that the pilots were asked about their usual levels of fatigue during the commuting cycle. Therefore, it is possible that, on some days, pilots do not manage before their flights any fatigue induced by driving to the airport. However, this was not captured in this survey. Although the focus of this study was on the extent of the issue of commuting-related fatigue, those cases are of interest because that is especially when driving to the airport is expected to increase the likelihood of impaired flying performance and performance decrements after duty.

In contrast to the time spent at the airport before flights, the survey showed that the participants did not usually manage to reduce their levels of fatigue while at the airport after duty. That is, the usual reported levels of fatigue just before driving after duty were not significantly different from those when exiting the aeroplane. If this finding applied to the population of professional pilots, then applying self-initiated intervention strategies for fatigue after duty would be needed to reduce the risk of car crashes. No data were collected to investigate why the time spent at the airport usually did not help the participants and no literature was sourced on how post-flight activities affect pilots' fatigue. Future studies could explore why pilots do not manage to reduce their levels of fatigue before commuting back home by focusing on factors, such as the rush to return home and the lack of awareness regarding the risk associated with driving fatigued after duty.

As explained in chapter 1, it is expected that any active fatigue induced by driving to the airport might increase further due to flying, thus, contributing to an increased risk of car crashes after duty. In line with the literature solely on pilots (e.g. Petrilli et al., 2006), the reported usual levels of fatigue when exiting the aeroplane were significantly higher than when entering it. The median of the usual levels of fatigue when exiting the aeroplane was 5 ('*moderately tired, let down*') compared to 2 ('*very lively, responsive, but not at peak*') at the beginning of the flights. Although the median of fatigue at the end of the flights was moderate, the increase from the beginning of the flights indicates a higher risk of impaired performance at the end of the flights.

It is unclear if the increased levels of fatigue at the end of flights were caused by the duration of the flights (e.g. short-haul versus long-haul flights), flying tasks (e.g. dealing with unanticipated problems), sleep-related factors (e.g. the time awake since last sleep), or any combination of them. Moreover, pilots can use countermeasures for fatigue in flight, but these were not investigated as this was out of the scope of this study. For example, sleeping in designated bunk facilities in the aeroplane is a generally accepted practice that assists pilots with staying alert in long-haul flights. More than two pilots are on duty on those flights in order to allow one of them to leave the cockpit in order to rest. In flights with two pilots on duty, it may be allowed to take controlled rest (Commission Regulation, 2014). Controlled rest is not planned before-flight and pilots do not leave the cockpit. Instead, they take a nap in-seat for about 40 minutes. Controlled rest is not permitted in all the countries (e.g. not in the US). Longer naps and better performance have been found when pilots slept in bunks than in the cockpit (Eriksen et al., 2006), but controlled rest can help to reduce fatigue too (Petrie et al., 2004). Future experiments could provide better insights into the effectiveness of sleep during flights in terms of fatigue in commuting pilots.

4.5 Conclusion and implications for the next study

Based on the interview, fatigue induced by driving to the airport is a safety issue because it has affected negatively many professional pilots' fatigue in flight and, potentially, their fatigue while driving after duty. The interviewee highlighted that long and demanding drives are the most fatiguing ones and recommended further research on the topic in order to develop intervention strategies. The extent of the issue was explored with the survey. Based on that, fatigue induced by driving to the airport is not usually an issue for pilots. This might be attributed to the fact that they usually avoid fatiguing commutes and reduce their levels of fatigue before flight. However, the study indicated that many professional pilots worldwide are at an increased risk of aeroplane accidents and car crashes after duty when they complete long and demanding drives to the airport. The sub-sample of pilots who reported commuting by driving was representative of the population of pilots in terms of age (CAA, 2018; FAA, 2018) and gender (Women in Aviation International, 2016). Nevertheless, generalising the findings of this survey to the population of professional pilots should be made carefully due to the lack of extensive data on pilots' commuting habits and the overrepresentation of pilots with a home base in Europe in the sample.

The findings of the survey suggest that active fatigue induced by driving to the airport is an issue for many professional pilots worldwide. Nevertheless, it is essential to explore in more depth if the type of fatigue induced by driving (i.e. active or passive) subsequently affects the levels of fatigue experienced in flying and driving. This will not only improve understanding of the fundamental aspects of the issue but also help commuting pilots to select effective self-initiated intervention strategies. This considered, the experiment described in chapter 5 was conducted.

Chapter 5. The effect of the type of fatigue induced by driving on the levels of fatigue experienced in subsequent flying and driving

5.1 Background and justification of the study

Aviation organisations suggest that the risk of flying fatigued may increase after long commutes to the airport. Drawing from the literature solely on drivers, that risk might be particularly high when pilots commute to the airport by completing long and demanding drives because their attentional resources could be depleted. If pilots did not manage fatigue while at the airport, they would be at an increased risk of fatigue-related performance decrements in flight. Flying might then increase their levels of fatigue even further. As a result, fatigue induced by driving to the airport and flying might deteriorate their driving performance while commuting after duty.

The survey in chapter 4 suggested that professional pilots are at an increased risk of aeroplane accidents and car crashes after duty when they complete long and demanding drives to the airport. However, the potentially adverse effects of active fatigue induced by those drives were based on self-reports about the levels of fatigue usually experienced and the characteristics of pilots' usual commutes. At the same time, the effects of sleep-related factors and self-initiated intervention strategies on fatigue in the commuting cycle were not controlled. Therefore, it was not possible to explore in depth the impact of the type of fatigue induced by long drives on fatigue experienced in subsequent flying and driving. Without improving understanding of that impact, it would difficult to suggest intervention strategies as some of them are effective against active and others against passive fatigue (section 2.3). This considered, the experiment described in this chapter aimed to investigate if the type of fatigue induced by driving (i.e. active or passive) subsequently affects the levels of fatigue experienced in flying and driving. Three objectives were set;

- Explore how completing a long drive affects individuals' levels of fatigue in subsequent flying and driving tasks. Conditions with and without a long drive will be used to achieve this objective.
- Investigate how low and increased levels of task load in a long drive affect the levels of fatigue experienced in subsequent flying and driving tasks. Drawing from Desmond and Hancock's theory of fatigue (section 2.2.1), the levels of fatigue experienced in flying and driving tasks after a long drive might be higher when that preceding drive is demanding because the continuous investment of effort depletes individuals' attentional resources. In contrast, when the task load of driving is low, the fatigue caused is passive. Passive fatigue relates to disengagement from the driving tasks and, thus, some attentional resources are reserved. As a result, individuals may have some attentional resources left to invest in subsequent flying and driving.

- Develop a methodology that could be used in the study of chapter 6 to reveal the effects of a long and demanding drive on individuals' fatigue in subsequent flying and driving tasks.

This experiment included a long, simulated drive followed by simulated flying tasks and a simulated short motorway drive. The control group completed flying tasks and motorway driving tasks of level 0 of automation (Figure 2.2) only. The manual group completed a long and demanding drive of 0 level of automation followed by the same flying and motorway driving tasks to the control group. The automated group completed the same driving and flying tasks to the manual but drove a highly automated car (level 2 of automation) during the initial long drive. Highly automated driving was selected to induce passive fatigue due to the low levels of task load. Instead of changing the level of automation of the car, the task load in the long drive could have been changed by modifying the driving scene (e.g. the traffic density). However, highly automated driving was used to increase the likelihood of inducing passive fatigue by minimising the physical task load while putting drivers into a monitoring state. The expressions of fatigue were measured with subjective and performance data collected in the flying and driving tasks, whereas task load in the long drive was measured with a subjective scale.

5.2 Hypotheses

Based on the existing literature, a number of alternative hypotheses were formed;

Hypothesis 1. The CSS ratings, the PVT reaction times, and the number of PVT lapses of the manual and the automated group will be significantly higher at the end than at the beginning of the 60-minute drive. Testing this hypothesis would show if the long drives induced fatigue due to the extended time on task.

Hypothesis 2. The total task load and the ratings on each of the NASA-TLX during the 60-minute drive will be significantly lower in the automated than in the manual group. Testing this hypothesis would show if the fatigue caused was active or passive.

Hypothesis 3. The control group will be less fatigued than the manual and the automated in the MATB-II tasks as evidenced by significantly lower CSS ratings and PVT reaction times, significantly fewer PVT lapses, and significantly better performance in the tracking, system monitoring and communication tasks.

Hypothesis 4. The automated group will be less fatigued than the manual in the MATB-II tasks as evidenced by significantly lower CSS ratings and PVT reaction times, significantly fewer PVT lapses, and significantly better performance in the tracking, system monitoring and communication tasks.

Hypothesis 5. The control group will be less fatigued than the manual and the automated ones in the 14-minute drive as evidenced by significantly lower CSS ratings and significantly better driving performance (i.e. lower SDs of lateral position and speed).

Hypothesis 6. The automated group will be less fatigued than the manual in the 14-minute drive as evidenced by significantly lower CSS ratings and significantly better driving performance (i.e. lower SDs of lateral position and speed).

If it was found that the control group was less fatigued than the others during the flying and the motorway driving tasks, that would suggest that fatigue induced by long driving increases the likelihood of being fatigued in subsequent flying and driving tasks. In addition, it would indicate that this likelihood increases with long drives irrespective of whether they induce active or passive fatigue. If that finding applied to professional pilots, they would be at a higher risk of aeroplane accidents and car crashes after duty due to fatigue induced by driving to the airport. Finding that the manual group was more fatigued in the flying and the motorway driving tasks than the automated would suggest that whether a preceding drive induces active or passive fatigue can make a difference in the levels of fatigue experienced in subsequent flying and driving tasks. If that finding applied to professional pilots, they would need to avoid particularly those drives to the airport that induce active fatigue.

5.3 Materials and methods

Ethical approval was obtained from the ESSL, Environment and LUBS (AREA) Faculty Research Ethics Committee of the University of Leeds (ethics reference number: AREA 17-067).

5.3.1 Participants

The focus of this study was on the mechanisms behind the effects of the type of fatigue induced by driving, which are expected to be the same for everyone irrespective of profession. Moreover, access to pilots was limited. Hence, the participants were non-pilots. A non-probability-based convenience sampling technique was used due to time constraints. All the participants were adults. Not having a full driving license was an exclusion criterion for increased ecological validity. People with training in piloting an aircraft were excluded too to control for the effects of flying experience on performance. The software used to simulate flying is not commercially available, so people with experience of playing flight simulation games were not excluded. Individuals with a hearing impairment or dyslexia did not participate because of the need to complete communication tasks and read rating scales under time pressure. People with a diagnosis of sleep apnoea could not take part either because this disorder can cause sleepiness. In addition, the participants should not have consumed alcohol, coffee or any other substance that can affect alertness in the 8 hours before the trials and should not have completed intense physical exercise in the 4 hours prior to participation to avoid starting the sessions on a plateau of fatigue.

Participants who reported on the personal information sheet that they had slept less than 7 hours the night before the experiment were excluded to control for fatigue caused by sleep deprivation. The 7-hour threshold was selected based on the recommendation of Hirshkowitz et al. (2015) and not the (considered too strict) threshold of 8 hours set by the FAA (2009). Finally, individuals who rated at the beginning of the study their levels of fatigue on the CSS fatigue scale (see section 3.2.1.1 for a description of the CSS) as 4 (‘*a little tired, less than fresh*’) or higher were excluded too. This decision was made to increase the likelihood that the participants would not have reached a plateau of fatigue before starting the session.

The participants were approached through posters in the libraries of the University of Leeds, posts on official University of Leeds Facebook groups, and through University databases. An email with the participant information sheet was sent to the individuals who were interested to give them enough time to decide whether they wanted to participate. Twelve individuals were recruited for the preliminary study that was conducted to investigate the feasibility of the crucial steps of the main study. Sixty individuals participated in the main study. The sample size of the main study was calculated using the G*Power software (Heinrich Heine Universität Dusseldorf, 2017) for a level of confidence of 90% and a margin of error of 10%. This sample size would be adequate to reveal effect sizes of 0.8.

5.3.2 Pre-study documents

The participant information sheet included information about the purpose of the study, the exclusion criteria, and why the participant was approached. Moreover, there was a description of what the experiment would include and information about confidentiality, voluntary participation, and the right to withdraw. The potential benefits and disadvantages/risks were outlined, and the researcher’s contact details were provided. Furthermore, the participants were informed about who would have access to their data and where the data would be stored. Finally, information was provided about the possibility of publishing the results of the study to a conference or a scientific journal and about who had organised, funded, and reviewed the study. Besides reading the participant information sheet, the participants read and signed a consent form. That form asked them to confirm that they agreed with the participant information sheet and had the opportunity to make questions. Moreover, the consent form included information about the voluntary participation, confidentiality, and withdrawal, access and storage of data.

The personal information sheet was used to collect information about demographics (i.e. age and gender) that would help to describe the sample. The literature suggests that experienced drivers may perform better in simulated driving tasks than the less experienced ones. This may occur, for example, because they are more likely to perform better in lane-keeping tasks (Hollopeter, 2011) and look more at the front view than the dashboard (Nabatilan et al., 2011). This considered, the participants were also asked when their driver’s license was issued to explore if there was a difference in the driving experience between the groups.

Since the focus of this experiment was on task-related fatigue, sleep-related data were collected with the personal information sheet to control for sleep-induced fatigue. That is, the participants were asked about the duration of their sleep and sleep quality the night prior participation. Sleep quality was rated on a multiple-choice question taken from the St Mary’s Hospital Sleep Questionnaire (see section 3.1.2.2). The St Mary’s Hospital Sleep Questionnaire includes items about sleep quantity and evaluates various aspects of sleep quality with several items. Exploring sleep quality in-depth was out of the scope of this study, so only one item was used (‘*how well did you sleep last night*?’). The participants selected amongst six options that ranged from ‘*very badly*’ to ‘*very well*’.

5.3.3 The long drive that was used to induce fatigue

A 60-minute drive was designed to induce fatigue in the manual and the automated group. This was completed on a Z400 PC running Windows 7 with a 40" monitor and Logitech G27 dual-motor force feedback steering wheel and pedals (Figure 5.1).

Figure 5.1 The equipment used to run the simulated drives in experiment 1.

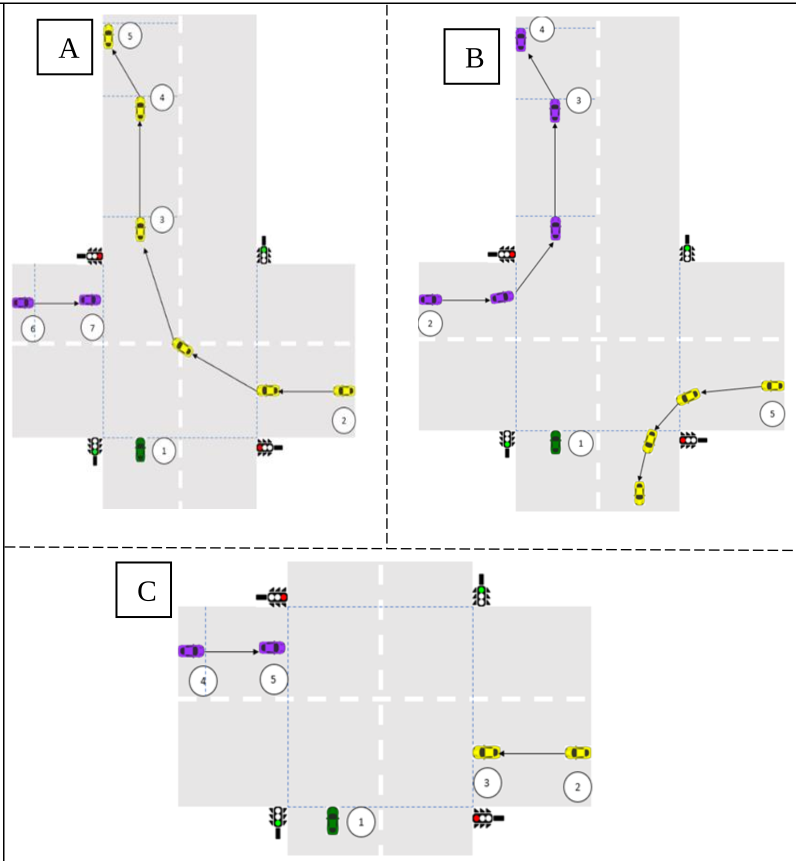
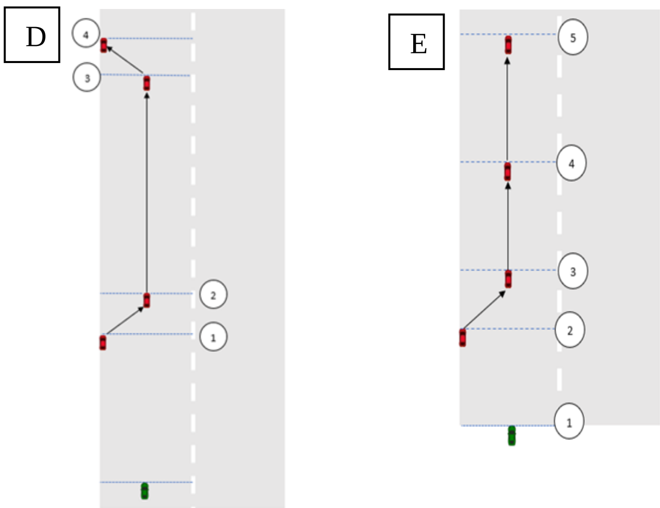
The manual group completed the 60-minute drive by driving a car of level 0 of automation and the automated by driving a car of level 2 of automation (Figure 2.2). The level 2 drive was a video recording of the car being driven by the researcher as safely as possible on urban and rural roads. The participants in the automated group were not aware that a video was shown. Level 2 was used in order to maintain some engagement with the driving task. Engagement would be achieved by requiring the participants to continuously monitor the driving scene and keep their hands on the steering wheel and feet on the pedals at all times because they would be asked by the car to take control over at some point (that did not finally happen). If the participants were allowed to disengage from the task completely, they might not experience passive fatigue because non-driving tasks could activate them (Jarosch et al., 2019; Schömig et al., 2015) and sleeping would reduce their levels of fatigue. Thus, it would not be possible to explore the role of the type of fatigue induced by driving on fatigue experienced in subsequent flying and driving tasks.

It was decided to design a demanding drive that would have a length similar to that of the long drives that pilots complete in real life. According to the survey described in chapter 4, about 20% of the pilots usually completed drives between 45 to 60 minutes to commute to the airport. A 60-minute drive was designed to resemble that duration. In addition, the duration of 60 minutes was selected to increase the likelihood of inducing fatigue based on other driving simulator studies. In those studies, participants' levels of fatigue in manual driving increased after 35 (Vogelpohl et al., 2019), 40 (Thiffault and Bergeron, 2003), and 45 (Oron-Gilad and Ronen, 2007) minutes. Thirty five minutes were also enough to induce fatigue when driving a highly autonomous vehicle in Vogelpohl et al.'s study (2019). The 60-minute drive started with a section of 4.5 km of urban road followed by 4 km of rural road. This 8.5 road section was repeated until the participants had driven for 60 minutes. The surroundings of the urban road sections were houses, trees, and parked cars. The surroundings of the rural road sections were a valley with trees.

The 60-minute drive would need to be demanding enough to induce active fatigue. The first method used to increase task load was adding road curves on both the urban and rural road sections (Richter et al., 1998). Moreover, overtaking and driving into intersections have been linked to increased levels of mental workload (Cantin et al., 2009; Jahn et al., 2005). Hence, several events of cars pulling in front of the participant's car and intersections with unpredictable movements of cars were added throughout the drive (Table 5.1). The participants in the preliminary study found the events used non-predictable, so they were used in the main study too. The literature suggests that mental workload increases when the traffic density is high (Teh et al., 2014), so oncoming traffic was also added to urge the participants to pay attention to cars that travelled on the opposite lane and decide if they were safe to overtake.

The colours and the models of the cars varied between the 8.5 km sections. Moreover, the first 8.5 km road section was the same in terms of the events used to the third, the fifth, and the seventh section. This design aimed to prevent the participants from predicting the events. The second, fourth, sixth, and eighth 8.5-km sections included a different version of the event with the parked car that pulled in front of participant's car (i.e. longer distance travelled before pulling out). All the other events were the same in all the sections, but they were activated at different locations of the drive in some sections. The driving environment was the same for the manual and the automated group, but the level of automation of the car driven differed (0 and 2 respectively).

Table 5.1 The events that were activated in the 60-minute drive (green icon: participant's car). The participants were asked to drive as in real life.

Traffic light events (urban road) The yellow car (panel A) and the purple car (panel B) passed the traffic light illegally, decelerated to 10mph, travelled at that speed for 5 seconds, and then pulled out. In the event shown in panel C, the purple and yellow cars passed the traffic light only when it was green.	
Parked car events (panel D – urban road, panel E – rural road) In both events, the parked car pulled in front of the participant's car. In the event shown in panel D, the car decelerated to 10mph, travelled at that speed for 5 seconds, and pulled out. In the event of panel E, the car decelerated to 30mph and then accelerated to 60mph until it was overtaken.	

5.3.4 The methods that were used to measure the expressions of fatigue

The methods that were used to measure the expressions of fatigue are shown in Figure 5.2. Three baseline scores were collected; a rating of fatigue on the CSS (see section 3.2.1.1), and the reaction time and number of lapses in the PVT (see section 3.2.3.1). The CSS baseline data were collected to exclude individuals already near a plateau of fatigue. At the same time, all the baselines would be used to explore if the groups differed in their levels of fatigue experienced before the trials. The induction of fatigue in the 60-minute drive was measured with CSS ratings during the drive, and the PVT reaction times and lapses before and directly after the drive. Fatigue was not inferred during the 60-minute drive by measuring driving performance because that would be affected by the events used. Perceived task load was rated every four minutes during the 60-minute drive with the NASA-TLX scale (see section 3.1.1). The manual and the automated group rated their levels of fatigue on the CSS scale again after the PVT that followed the 60-minute drive. For the control group, the baseline CSS rating was used as the rating of fatigue at the beginning of the MATB-II tasks. The same data were collected for all the groups in the rest of the trials. These included the CSS ratings and performance in the flying and motorway driving tasks. The PVT reaction times and lapses were measured again between the flying and the motorway driving tasks and fatigue was rated on the CSS again just before the motorway drive.

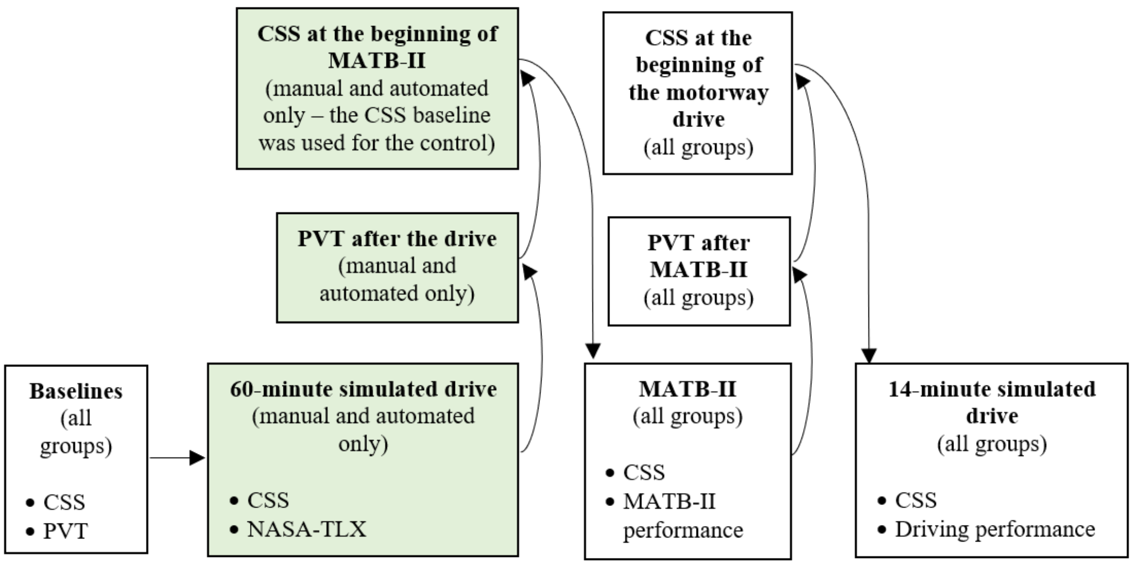


Figure 5.2 The methods that were used to measure the expressions of fatigue in experiment 1. Green boxes: data collected only for the manual and the automated group.

5.3.4.1 The simulated flying tasks

The simulated flying tasks were completed with the Multi-Attribute Task Battery II (MATB-II) (Figure 5.3) software on the PC used for the driving simulator with a mouse and a Logitech G

X52 Saitek joystick (Figure 5.4).

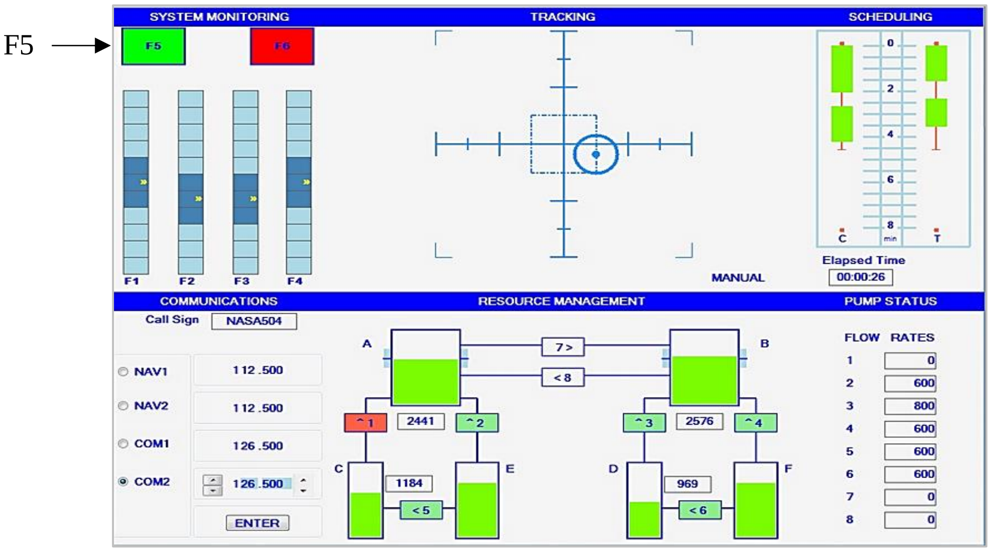


Figure 5.3 The user interface of the Multi-Attribute Task Battery-II (MATB-II) (Santiago-Espada et al., 2011).


Figure 5. 4 The equipment that was used for the MATB-II flight tasks.

MATB-II was developed by NASA (Santiago-Espada et al., 2011) to evaluate pilots' performance and mental workload with tasks that resemble those usually completed by pilots in flight. However, the MATB-II tasks can also be completed by non-pilots because no knowledge of flying an aeroplane is needed. Participants can concurrently complete tracking, monitoring, communication, and resource management tasks (Figure 5.4). Researchers can select which of the four types of tasks (i.e. tracking, system monitoring, communications, and resource management) will be completed and, thus, if single- of multi-tasking will be required. Moreover, the number and timing of the events can be controlled. For example, they can select when the green light (shown as F5 in Figure 5.4) will switch off and on.

The MATB has been used in studies with pilots because of its high face validity and sensitivity of performance to fatigue (Caldwell and Ramspott, 1998; Caldwell et al., 2004; Lopez et al., 2012). MATB and MATB-II are different versions of the same software but include the same tasks. In a study with pilots and pilot trainees, Caldwell and Ramspott (1998) found that increased time-on-task related to reduced accuracy in the tracking task, more time-out errors in the communications task and slower reaction times to the scale deviations and light changes of the system monitoring task. The sensitivity of the MATB to sleep-induced fatigue has been supported in studies that showed that sleep-deprived pilots completed less accurately the tracking task (Caldwell et al., 2004; Lopez et al., 2012).

The aim in the tracking task (Figure 5.5) is to keep the circle for as much as possible at the centre of the cross by moving the joystick appropriately. The blue circle moves continuously and drifts away from the centre of the cross. The researcher can increase the need for corrective inputs or switch the tracking task to auto mode (Figure 5.5 – panel B). Completing the tracking task requires sustained attention (i.e. vigilance) to detect small movements of the blue circle and respond quickly to them. The ability to detect small environmental changes can be negatively affected by fatigue (Thiffault and Bergeron, 2003).

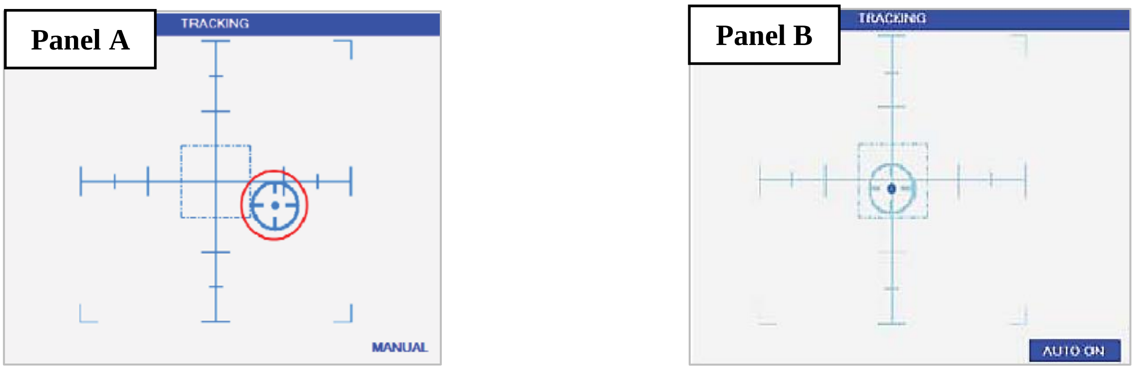


Figure 5.5 Panel A: the tracking task with the blue circle away from the centre of the cross in manual mode. Panel B: the tracking task in auto mode.

The tracking task resembles the manual handling of the joystick (Figure 5.6 – panel A) of a real aeroplane. In real flights, pilots use an instrument called ‘attitude indicator’ (Figure 5.6 – panel B) to fly the aeroplane straight (i.e. without turning) and level (i.e. without changing altitude). The joystick is moved appropriately to keep the centre of the horizontal line of the cross on the horizon line (see yellow arrow). When the aeroplane is turning, these two lines are not parallel, whereas when the aeroplane is changing altitude, the one line is above the other. Although pilots often engage the autopilot to avoid flying the aeroplane by moving the joystick, sometimes they fly by controlling it. For example, that can happen in takeoff and landing or during an emergency (e.g. engine failure). The difficulty in controlling the joystick increases when flying in turbulence and through strong crosswinds.

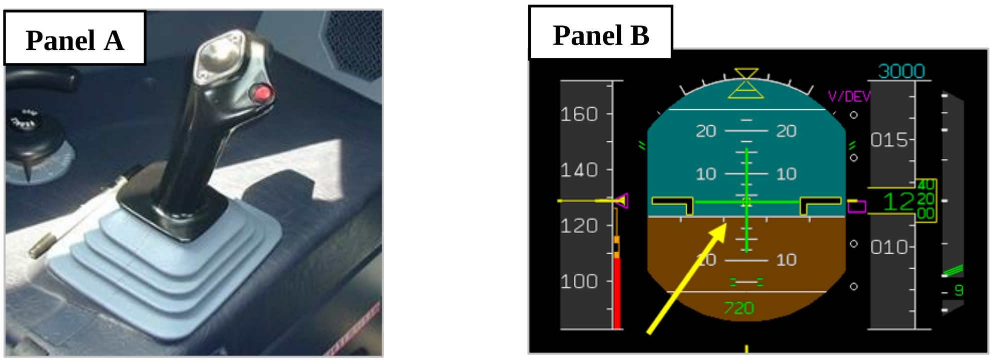


Figure 5.6 Panel A: the joystick of an Airbus A320 (Airliners, 2019). Panel B: the attitude indicator on an Airbus A320 (StackExchange, 2019).

The system monitoring task includes the light and the scale tasks (Figure 5.7). In the lights task, the left light is normally green and the right grey (Figure 5.8 – panel A). The researcher determines the times when the green light will switch off and on (Figure 5.8 – panel B). The one sub-task of the lights task is to detect when the green light turns grey and click on it as quickly as possible before it turns green again. Furthermore, the researcher determines when the grey light will turn red and grey again (Figure 5.8 – panel C). This is the second sub-task of the lights task, where users must detect the red light and click on it as quickly as possible before it turns grey again. If the correct light is not clicked within the pre-determined time or the users click the correct light on time, the lights return to their normal colour.

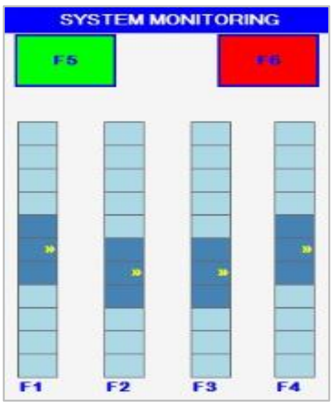


Figure 5.7 The system monitoring task of the MATB-II.

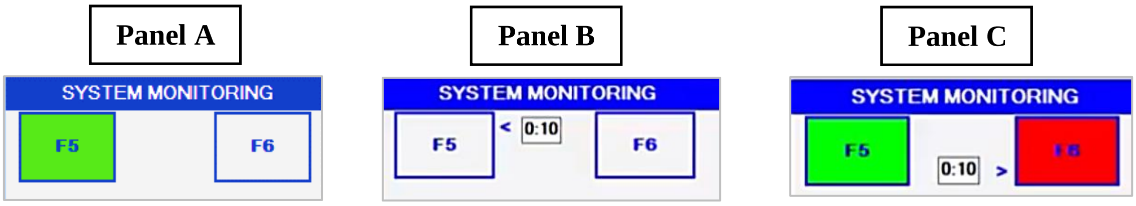


Figure 5.8 Panel A: the green light is switched on. Panel B: the green light is switched off. Panel C: both the green and red lights are switched on.

In the scales task, all the dark blue squares move constantly around the middle of the columns but not more than one rectangle upwards or downwards (Figure 5.9 – panel A). When a dark, blue square moves too high or too low (Figure 5.9 – panel B), users need to click anywhere on the relevant scale as quickly as possible to bring it back to its normal position. The System Monitoring task resembles the monitoring of lights and scales on the instrument panel in real aeroplanes (Figure 5.10). The main principle in both situations is that sustained attention is needed to detect changes in the lights and scales.

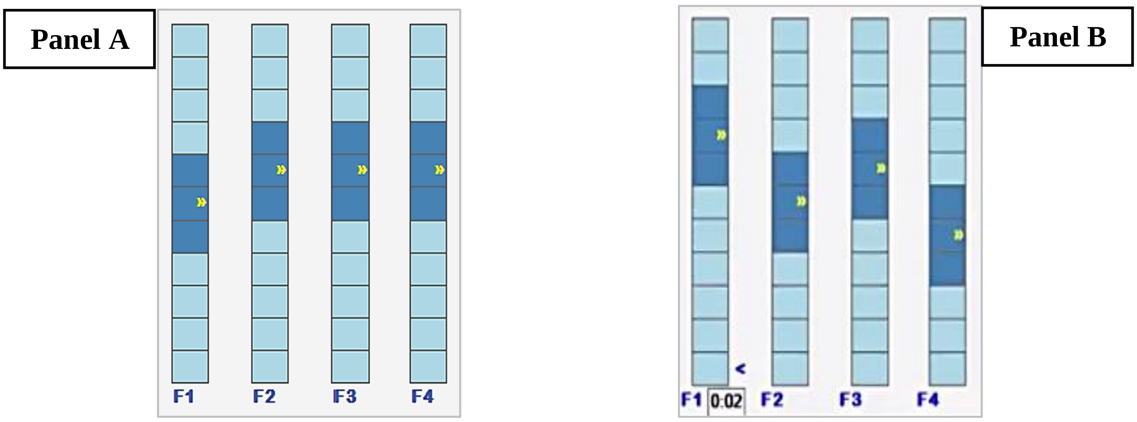


Figure 5.9 Panel A: the normal position of the dark blue boxes on the scales. Panel B: the dark blue square of scale F1 has moved too high on the scale.

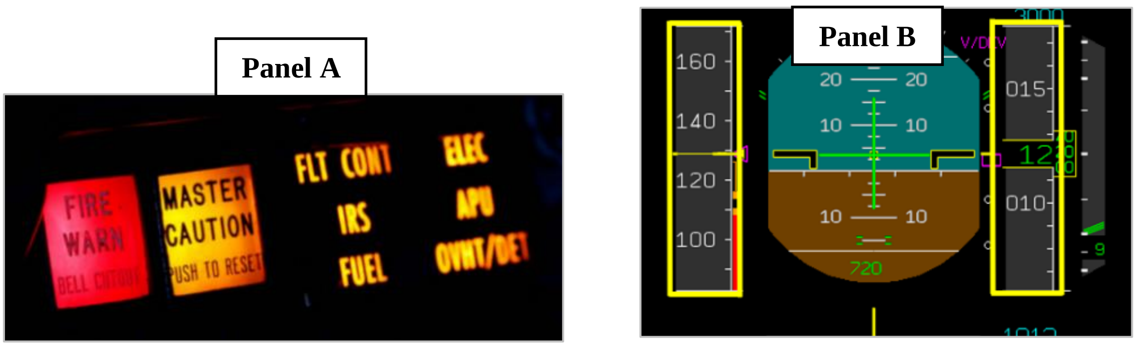


Figure 5.10 Panel A: Boeing 737 master caution (Flaps2Approach, 2019). Panel B: Airbus A320 moving scales (in the yellow boxes) (StackExchange, 2019).

The aim of the communications task (Figure 5.11) is to select the correct radio and change the frequency shown so that it matches the frequencies heard. Working memory is needed to remember the radio frequencies and rehearse this information. An example voice message in the MATB-II is the following; *'NASA five zero four, NASA five zero four, turn your NAV one radio to frequency one-one-four, five-five-zero'.* When users believe that the frequency shown is the frequency in the ATC message, they must click the 'ENTER' button.

The communications task of the MATB-II resembles the communication of pilots with the ATC in real flights. Although the design of the MATB-II communications panel is simpler than that of real aeroplanes, the messages heard are similar to those sent by the ATC in real flights and the task that users complete resembles what pilots do in real operations. To change the radio frequency in an Airbus A320, pilots push a radio selection button (e.g. VHF1) (Figure 5.12), rotate the knob to change the frequency, and confirm it by pressing the green arrow.

Figure 5.11 The communications panel in MATB-II.

Figure 5.12 The radio management panel of the Airbus A320.

The Resource Management task (Figure 5.13) resembles the fuel management task completed by pilots in flight by asking participants to move fuel between the pumps in order to keep the fuel levels between the pre-determined limits in tanks A and B. This task remained inactive in the trials because the preliminary study showed that the participants needed at least 15 minutes of practice only for this task. That could result in starting the trials at a plateau of fatigue.

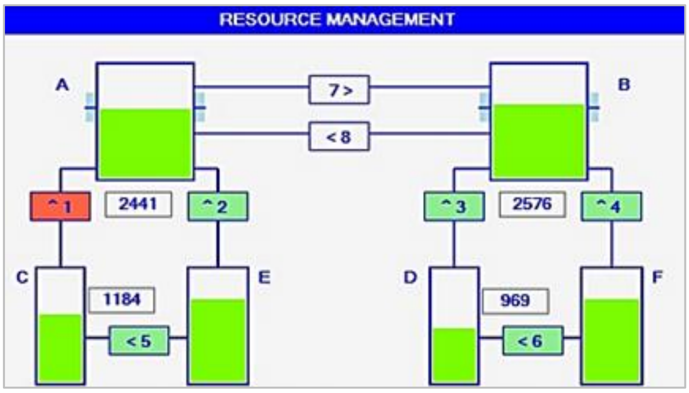


Figure 5.13 The MATB-II resource management task.

The scheduling section of the MATB-II interface (Figure 5.14) shows the scheduled tracking and communications tasks. The participants in this study were advised to use this information in order to be prepared for the next tasks.

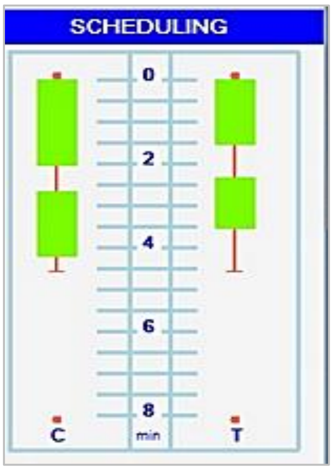


Figure 5.14 The MATB-II scheduling section.

The MATB-II scenario that was used in the main session was designed to measure performance in the tracking, monitoring, and communications task. The total duration of the scenario was 32 minutes in order to simulate the duration of a short domestic flight and was divided into five blocks. The duration of the first, second, fourth and fifth blocks of tasks was four minutes. In contrast, the third block was sixteen minutes long to simulate the cruise stage, which is usually longer than the other stages in real flights. The tasks that participants needed to complete in each block can be found in Appendix B**.** The levels of task load were higher in the second and fourth blocks to simulate the differences in the task load experienced between different stages of flight. That is, the first, third and fifth blocks required performing only the tracking task to simulate the lower levels of task load experienced during the taxi and cruise stages. In contrast, the second and fourth blocks required completing the tracking, monitoring, and communication tasks at the same time in order to simulate the higher levels of task load usually experienced during the takeoff, climb, descent, and landing stages of flight.

The duration of the practice session was 16 minutes, which, according to the preliminary study, was enough to familiarise with the MATB-II tasks. The practice session was divided into two 4-minute blocks that included only the tracking task and two that included the tracking, monitoring, and communications tasks. The events of the system monitoring task in the practice session differed from the ones in the main session to avoid predictions. This was achieved by having red light events where there were green light ones in the main session and vice versa, triggering different scales, and using different frequencies in the communications task.

5.3.4.2 The 14-minute motorway drive

A 14-minute motorway drive was designed to infer high levels of fatigue through its effects on driving performance. The equipment described in section 5.3.3 was also used here. This drive was completed on a car of level 0 automation (see Figure 2.2) by the manual, the automated, and the control groups. In contrast to the 60-minute drive, this drive was not used to induce fatigue but to unmask its effects on performance. Therefore, a shorter duration than the 60 minutes was selected for increased practicality. No simulator studies that explored fatigue in driving after completing flying tasks were sourced. However, driving simulator studies that compared fatigue between sleep- and non-sleep-deprived individuals suggested that between-group differences in the effects of fatigue on subjective experience, driving performance and physiology can be detected by using drives with a duration of 15 minutes (Akerstedt et al., 2010; Lenne et al., 1998; Vakulin et al., 2007). In regard to the adverse effects of fatigue on driving performance, there is evidence that higher levels of fatigue relate to higher standard deviations of speed and lateral position, shorter headways from the cars ahead, higher mean speeds, more frequent edge-line crossings, and less accurate responses to speed changes of a lead car (Akerstedt et al., 2005; Brookhuis et al., 1994; Du et al., 2015; Thiffault and Bergeron, 2003; Ting et al., 2008; Van der Hulst et al., 2001; Zhang et al., 2016).

The participants were told that the lead car would always drive at 70 mph in the middle of the middle lane. They were instructed to follow that car by driving as much as possible at 70 mph in the middle of the middle lane. It was expected that not changing lanes would keep the task load low, thus, unmasking any effects of fatigue on performance by not engaging the individuals. To keep task load low, most of the road was straight. For increased realism, a car that drove at 80 mph on the right lane decelerated to 70mph, entered the middle lane far in front of the lead car (see Figure 5.15) and stayed there for five seconds before moving back to the right lane. This event was activated when the distance between positions 1 and 2 in Figure 5.15 became 100 metres. This event was triggered approximately once every minute.

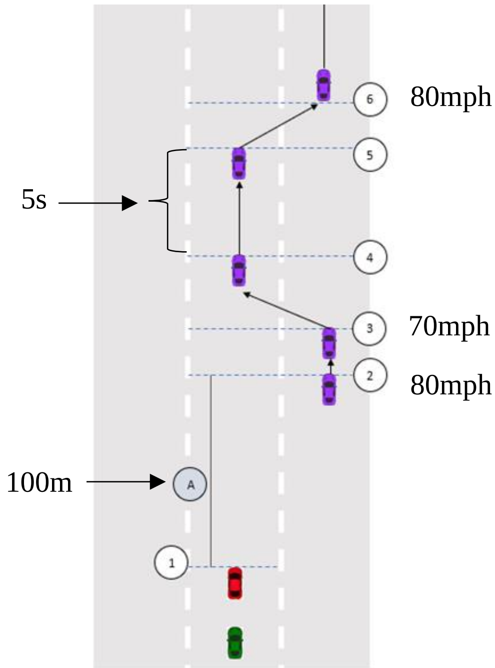


Figure 5.15 The lane change event during the 14-minute drive.

After 14 minutes of driving, the lead car started changing its speed based on the speed of the participant's car to achieve a Time-To-Collision (TTC) of two seconds. TTC was defined by Vogel (2003) as '*the time span left before two vehicles collide if nobody takes evasive action*'. The TTC of two seconds was selected based on a review of the brake reaction times of non-fatigued drivers in unexpected braking events that was conducted by Green (2000). When the TTC of two seconds was achieved, the lead car braked hard and stopped. The participants had not been informed that this event would occur. The preliminary study showed that some participants drove too far from or too close to the lead car just before the 14 minutes of driving. That resulted in delays of the software in achieving the two-second TTC soon after the 14 minutes and, in some cases, a failure to achieve the target TTC in order to activate the braking event. To ensure that the braking event would be activated the soonest after the 14 minutes in all the trials (so that the time on task would be the same), messages were shown above the dashboard (Figure 5.16). These messages reminded the participants of the necessity to remain within a certain distance from the lead car (i.e. not too far, not too close). The behaviour of the lead car and the tasks that the participants completed during the 14-minute drive are shown in Figure 5.17.

Figure 5.16 Panel A: the message that appeared when the lead car was too far. Panel B: the message that appeared when the lead car was too close.

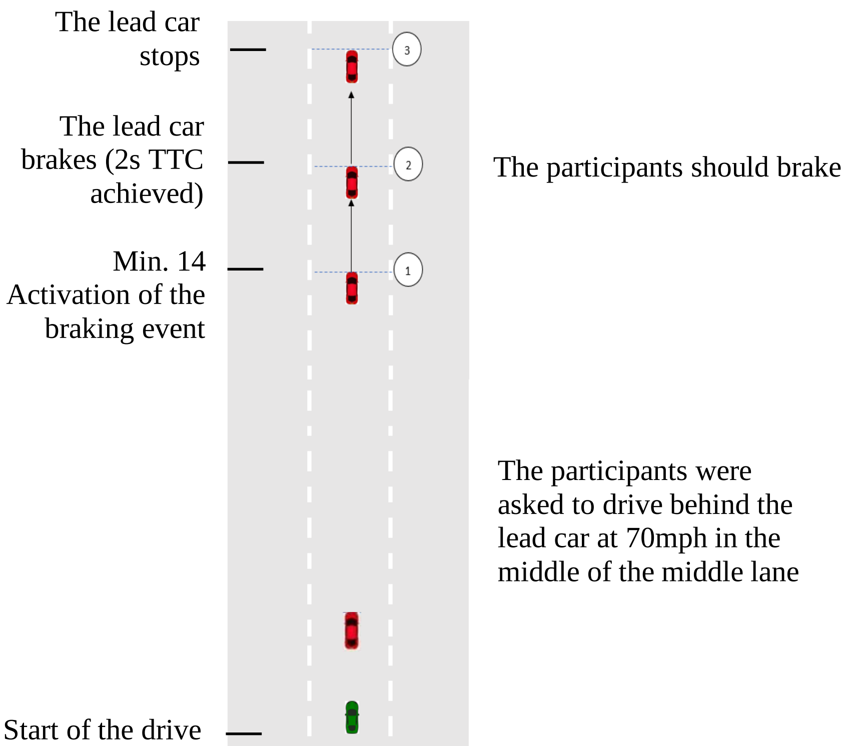


Figure 5.17 The behaviour of the lead car in the 14-minute drive and the tasks that were completed by the participants.

5.3.4.3 The CSS fatigue scale

Subjective fatigue was measured with the CSS (see section 3.2.1.1), which is sensitive in detecting increases of the levels of fatigue due to time on task (Gander et al., 2015; Petrilli et al., 2006) and drivers (Belenky et al., 2012; Richter et al., 2005). The participants rated their fatigue just before the session and every 4 minutes during the 60-minute drive (only the manual and automated group). Fatigue was also rated after 4, 8, 24, 28, and 32 minutes in the MATB-II tasks, and after

4 and 9 minutes during the 14-minute motorway drive. The CSS was presented during the drives on the top left corner of the screen (Figure 5.18) and the participants should rate their levels of fatigue aloud. The CSS was shown on the screen during the MATB-II scenario when the flying tasks were stopped. Again, the participants provided a rating aloud. The participants practised rating their levels of fatigue on the CSS during the practice drive.

Figure 5.18 The CSS scale that appeared during simulated drives.

5.3.4.4 PVT reaction times and lapses

The PVT (section 3.2.3.1) reaction times and lapses (i.e. reaction times longer than 500 milliseconds) were measured to infer fatigue from vigilance. The impact of sleepiness on PVT performance has been supported in studies, where extended time awake since last sleep and sleep loss increased the PVT reaction times (Baulk et al., 2006; Dinges et al., 1997) and the number of lapses (Basner et al., 2011; Shattuck and Matsangas, 2014). In addition, some studies have found an effect of the time on task on PVT reaction times in participants who completed simulated tasks. For example, the PVT reaction times after a 30-minute simulated drive were significantly longer than before the drive in Graaumans et al.'s (2010) study.

The reaction times were measured with the PEBL 2.0 software (PEBL, 2018), which has precise timing (Mueller and Piper, 2014) and has been used in research with pilots (Hoermann et al., 2015; Karlen et al., 2010). The PVT was completed on a laptop running on Windows 10 64-bit, an i7 processor at 2.9GHz, and 8 GB RAM. The standard duration of the PVT is 10 minutes, but shorter versions can also be used (Basner et al., 2011; Lamond et al., 2008). A 5-minute PVT was used in this study for practical reasons. The signal presented in PVT was a pink ball that always appeared at the centre of a black screen. The participants were asked to press the spacebar as quickly as possible when the ball appeared. The intervals between the appearances of the ball varied between 2 and 10 seconds. A 1-minute practice preceded the baseline PVT.

5.3.5 The method that was used to measure task load in the 60-minute drive

Task load was subjectively measured during the 60-minute drives with the NASA TLX scale to investigate its role on fatigue induction. The participants rated each of the six items of the NASA TLX scale aloud by saying a number from 1 (‘*very low*’) to 20 (‘*very high*’). These items appeared every 4 minutes during the 60-minute drive above the dashboard (Figure 5.19) one after the other when the CSS disappeared. The titles of the items were heard upon their appearance.

Figure 5.19 One of the NASA TLX sub-scales (mental demand) that appeared during the 60-minute drives in experiment 1.

5.3.6 Procedure

Figure 5.20 shows the setup in the room of the equipment used for the simulated tasks. The monitor used for the experiment was moved appropriately depending on whether driving or MATB-II tasks were completed. The PVT was completed on the same desk on a laptop.

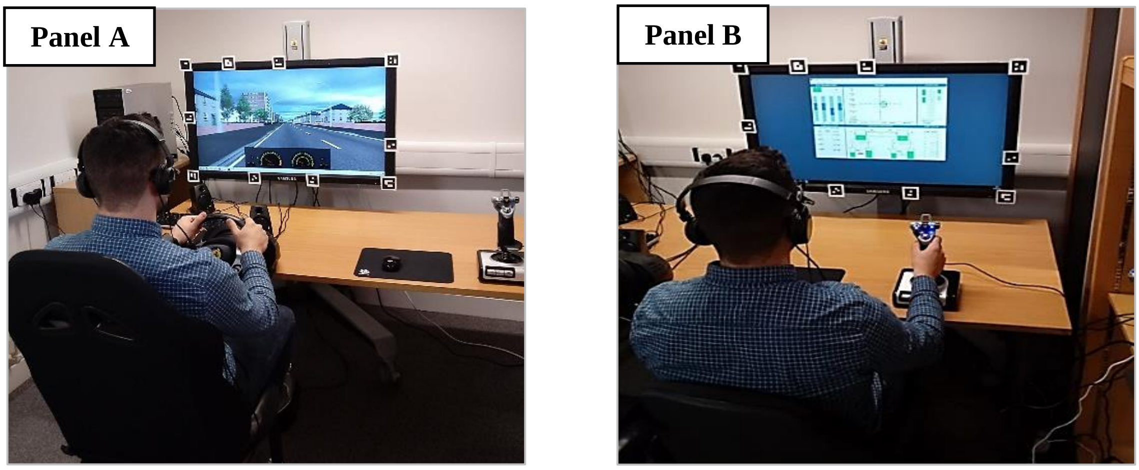


Figure 5.20 Panel A: completion of the simulated driving task. Panel B: completion of the MATB-II task.

The participants paid one visit to the researcher at the Institute for Transport Studies. Half of the experiments started at 9.00 and the others at 13.00 to counterbalance the effects of the circadian rhythms on fatigue. When they arrived, they were given a reference code and they were assigned to one of the three groups; every first participant was assigned to the control group, every second to the manual, and every third to the automated. Then, they read the participant information sheet, signed the consent form, and completed the personal information sheet.

The participants were explained the CSS and NASA-TLX scales and what they would do in the practice drive. After completing the practice drive, they were given instructions for the MATB-II and started the practice flight scenario. At the end of each block of tasks, the researcher stopped the tasks remotely and showed the CSS on the screen. The participants rated the scale aloud and the tasks continued. When the MATB-II practice finished, the participants rated their levels of fatigue on the CSS (baseline) and completed the baseline PVT.

Before starting the 60-minute drive, the participants in the manual group were told that they would drive on urban and rural roads for 60 minutes. They were asked to drive as they would do in real life and rate the scales that would appear on the screen. Then, the researcher started the drive. The instructions given before the 60-minute drive were different for the automated group. Those participants were also told that they would drive on urban and rural roads for 60 minutes, but the car would drive itself. Therefore, they would not have to control the steering wheel and the pedals constantly. However, they were informed that their car would drift away from the road twice in the 60 minutes. When that happened, they would have to brake and move the steering wheel appropriately as quickly as possible to bring the car back onto the road. Then, the car would take over control again. They were advised to have their hands on the steering wheel continuously and their feet on the pedals to respond quickly. Then, the 60-minute drive started.

After the 60-minute drive, the manual and the automated group completed a PVT. The main MATB-II scenario followed. The control group started the main session from the MATB-II tasks. All the groups were informed about the duration of the scenario and the tasks that should be completed. The flying tasks were stopped remotely at the end of each block to show the CSS on the screen. All the groups completed a PVT after the MATB-II tasks and, then, started the 14-minute drive. They were given the same instructions as in the practice drive and no interventions were made by the researcher while driving. When the 14-minute drive finished, the participants were thanked, debriefed, and paid £20.

5.3.7 Methodological considerations

Since fatigue is a subjective construct, a within-groups design should be ideally used. This was avoided, first, because it could not be assured that the participants would start the sessions on the same levels of fatigue as fatigue may not develop linearly (Schömig et al., 2015). Second, the braking event in the motorway drive would be predictable after the first session. Third, recruitment might be difficult because each participant would need to visit the researcher three times. A between-groups study design with baselines of subjective fatigue and performance in the PVT was used as an alternative.

Perceived task load was measured on the NASA-TLX during the 60-minute drives. The data collected on each of the NASA-TLX items were analysed separately from the total task load score to identify if the 60-minute drives induced active or passive fatigue. According to Desmond and Hancock (2001), active fatigue relates to the investment of resources (thus high effort) and passive with the disengagement from the task (thus low effort). Galy et al. (2017) suggested the analysis of the data collected on each of the NASA-TLX items separately when researchers are interested in certain aspects of task load. The weighting procedure was not used because increasing task load every four minutes in the automated group could increase participants' levels of alertness. As a result, it would not be possible to induce passive fatigue. The Raw NASA-TLX TLX has been found equally sensitive to the weighted scoring (Moroney et al., 1995).

Performance in the tracking task of the MATB-II scenario was one of the metrics used to infer fatigue. This task required keeping a drifting ball at the centre of the cross by moving the joystick appropriately. Therefore, deviation from the centre of the cross was calculated. The deviations of lateral position and speed were also used to measure performance in the motorway drive. However, the Root Mean Square Deviation (RMSD) was calculated for the tracking task and the Standard Deviation (SD) for the lateral position and speed in driving. This decision was made because the SD is sensitive to outliers since it gives the spread of the data around the mean. This sensitivity is important when exploring driving performance because it is those outliers that indicate a higher risk of crashing. In contrast, the RMSD is calculated by squaring the differences between the predicted and the target values and averaging these squared values. Therefore, the effect of any outliers is smaller than with the SD. The effect of a large momentary deviation of the flight parameters (e.g. of the altitude change rate) is less probable to have an adverse effect on safety than, for example, a large deviation of the lateral position of a car due to large separations between aeroplanes.

5.3.8 Statistical analyses

The independent variables of the study were the group (i.e. control, manual, and automated) and the time point of measurement during the trials. The dependent variables were the CSS ratings, the NASA-TLX ratings, performance in the PVT (i.e. reaction time and lapses), performance in the MATB-II (i.e. tracking RMSD, omissions of lights and scales, errors in the communications task, and reaction time to lights and scales), and performance in the motorway drive (i.e. SD of the lateral position, SD of the speed, and reaction time in the braking event).

Data analysis was performed with the IBM SPSS Statistics 24.0 (IBM, 2016). R version 3.5.1 (R-Project, 2018) was also used. Mixed-model analyses were performed for those metrics that data were collected more than once. Two-way mixed ANOVAs (i.e. Analyses of Variance) were used for the analyses when there was one between-groups factor (i.e. the group) and one within-groups factor measured in the continuous level (e.g. tracking RMSD). If a two-way mixed ANOVA showed that a simple main effect of the measurement time point on a variable was significant, this was further investigated with one-way repeated measures ANOVAs (when comparing more than two time points) or paired samples, two-tailed t-tests (when comparing two time points). When the assumption of sphericity was not met in the one-way repeated measures ANOVAs, the degrees of freedom were corrected with the Greenhouse-Geisser estimates of sphericity. A significant simple main effect of the group was further explored with independent samples, two-tailed t-tests for the comparisons for which data only from the manual and the automated group were collected (i.e. in the 60-minute drive). One-way ANOVAs were run to explore further a simple main effect of the group when comparing between the three groups (i.e. in the MATB-II and motorway driving tasks).

The data were transformed when the assumptions of a two-way mixed ANOVA were violated or there were many outliers in the data. Logarithmic transformations were used when the data were strongly positively skewed and square root transformations when they were moderately positively skewed (Tabachnick and Fidell, 2007). If the transformation was not successful, a non-parametric alternative analysis was run with the nparLD package (Noguchi et al., 2012) in R. The nparLD package is a rank-based method used to allow performing non-parametric analysis of longitudinal data in factorial experiments. The benefit of using a rank-based method is that there is no requirement for normally distributed data and the data analysed can be continuous, ordinal, or dichotomous (Konietschke et al., 2010). The nparLD package was also used when there was one between-groups factor (i.e. the group) but the within-groups factor was measured in the ordinal level. If the analysis with the nparLD showed that a simple main effect of the measurement time point was significant, Friedman's tests were then run. In contrast, a significant simple main effect of the group found with the nparLD package was further investigated with Kruskal-Wallis H tests. Subsequently, pairwise comparisons were performed using Dunn's (1964) procedure with a Bonferroni correction.

One-way ANOVAs were used when there was one between-groups factor (i.e. the group) and the dependent variable was measured once on the continuous level. In contrast, Kruskal-Wallis tests were run to compare ordinal variables measured once and the baseline CSS ratings between the three groups. Furthermore, a chi-square test of homogeneity was performed to compare dichotomous variables between the groups. Although the CSS and PVT data were collected throughout the sessions, it was decided to analyse the data separately per phase rather than inserting all the data into a single mixed-effects model for each metric. This decision was made, first, because data were collected in the 60-minute drives only for the manual and automated groups. Second, it would be difficult to identify which phase (i.e. 60-minute drive, flying, or motorway drive) the interaction was significant at. Wilcoxon signed-rank tests were used for the comparisons of the CSS ratings at the end of the 60-minute drive with those just before the MATB-II. Moreover, Wilcoxon signed-rank tests were performed to compare the CSS ratings at the end of the MATB-II with those just before the motorway drive. These comparisons would show if the transition itself from one type of task to another affected fatigue.

Before testing the hypotheses, it was explored if there were any significant differences between the three groups in factors other than the 60-minute drives that could have affected the expressions of fatigue in the trials. These factors were age (see section 2.1.2), driving experience (because more experienced drivers may perform better in driving - Hollopeter, 2011; Nabatilan et al., 2011), sleep-related factors of fatigue (see section 2.1.2), and the fatigue baselines. The between-groups comparisons of age, the years of driving experience, sleep duration, and the baseline PVT reaction times and lapses were performed with one-way ANOVAs. In contrast, the between-groups comparisons of sleep quality and the baseline CSS ratings were conducted with Kruskal-Wallis H tests. When the data were transformed due to not initially meeting the assumptions of the analyses performed, the descriptive statistics of the transformed data will be depicted in the Figures of the results section. The 95% CIs will be depicted when the data were treated as ordinal. On the other hand, the standard errors will be shown when the data were treated as continuous.

Hypothesis 1 stated that the CSS ratings and the PVT reaction times and number of lapses of the manual and the automated group will be significantly higher at the end than at the beginning of the 60-minute drive. In contrast, according to hypothesis 2, the total task load and the ratings on all the items in the NASA-TLX during the 60-minute drive will be significantly lower in the automated than in the manual group. The data that were analysed to test hypothesis 1 are shown in Figure 5.21.

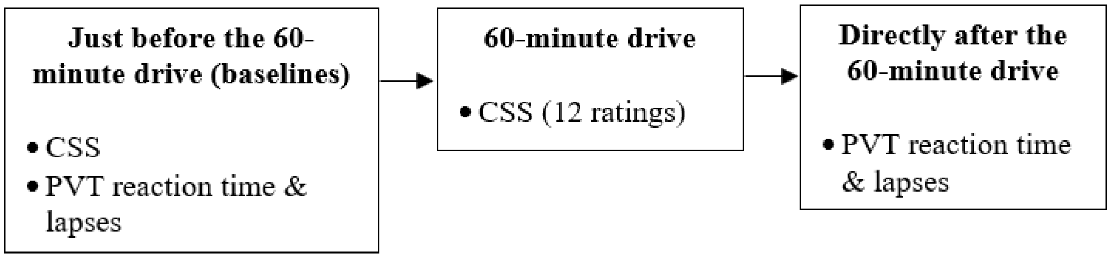


Figure 5.21 The data used to test hypothesis 1 in experiment 1.

The following analyses were performed to test hypothesis 1:

- Two-way mixed ANOVAs for the reaction times and the lapses in the two PVTs collected in the 60-minute drive.
- Analyses with the nparLD package for the CSS ratings before, during, and after the 60-minute drive.

Hypothesis 2 was tested by running two-way mixed ANOVAs for the NASA-TLX data (i.e. total score and ratings on each of the items) collected in the 60-minute drive (12 ratings for each).

According to hypothesis 3, the control group will be significantly less fatigued than the manual and the automated in the MATB-II tasks, whereas hypothesis 4 stated that the automated group will be significantly less fatigued than the manual in the MATB-II tasks. These hypotheses were tested by analysing the data shown in Figure 5.22.

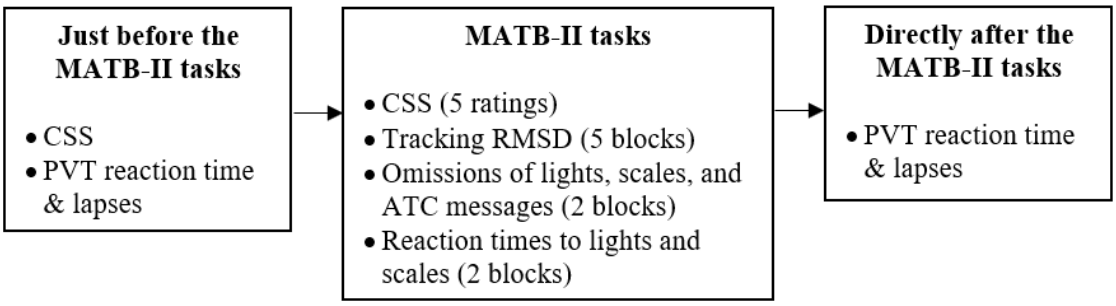


Figure 5.22 The data used to test hypotheses 3 and 4 in experiment 1.

The following analyses were performed to test hypotheses 3 and 4:

- Two-way mixed ANOVAs for the PVT reaction times and the lapses before and after the MATB-II tasks, the tracking RMSD, the omissions of lights, scales, and communication tasks, and the reaction times to the lights and scales.
- Analyses with the nparLD package for the CSS ratings just before and in the MATB-II tasks.

According to hypothesis 5, the control group would be significantly less fatigued than the manual and the automated in the 14-minute drive, whereas according to hypothesis 6, the automated group would be significantly less fatigued than the manual in the 14-minute drive. These was tested by analysing the data shown in Figure 5.23. The reaction times and the outcomes (i.e. crash or not) in the braking event were finally not analysed because some participants braked too hard, causing the brake pedals to slide on the floor. Therefore, the braking data were not accurate.

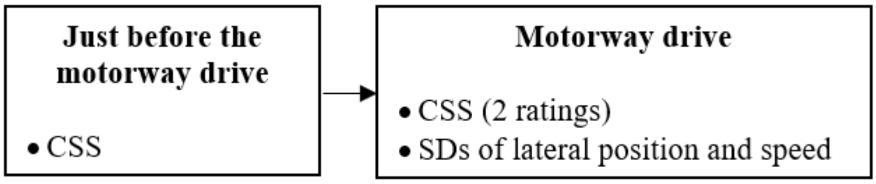


Figure 5.23 The data used to test hypotheses 5 and 6 in experiment 1.

The following analyses were run to test hypotheses 5 and 6:

- Two-way mixed ANOVAs for the SDs of the lateral position and speed.
- Analyses with the nparLD package for the CSS ratings just before and during the motorway drive.

5.4 Results

5.4.1 Age, driving experience, sleep-related factors, and fatigue baselines

Sixty individuals took part in this study. The percentages of females were 65% in the control group, 55% in the manual, and 60% in the automated. Before testing the hypotheses formed, it was explored if there were differences between the groups in the age, driving experience, the investigated sleep-related factors of fatigue, and the fatigue baselines. These comparisons would show if any differences in fatigue found between the groups during the flying and the motorway driving tasks could be attributed to factors other than the completion of the long drives. As explained, being older, sleeping less, getting sleep of lower quality (section 2.1.2), and starting the session more fatigued could affect the fatigue-related data collected in the sessions. The means of participants' age, years of driving experience, and hours of sleep the night before the trials are shown in Table 5.2. The percentages of participants that selected each of the categories in the sleep quality item selected from the St. Mary's hospital sleep questionnaire and the baselines of the PVT and CSS are also depicted in Table 5.2.

A one-way ANOVA showed that the difference in age between the groups was not statistically significant, $F(2,57) = 2.219$, $p = .118$. Moreover, a one-way ANOVA showed that the differences in the years of driving experience were not significant between the groups, $F(2,57) = 2.430$, $p = .097$. A one-way ANOVA did not show a significant difference between the groups in the hours of sleep the night before the trials, $F(2,57) = 0.211$, $p = .810$. Finally, a Kruskal-Wallis test showed that the difference in the reported sleep quality was not significantly different between the groups, $x^2(2) = 1.514$, $p = .469$.

The one-way ANOVA that was performed to compare the PVT reaction time baselines between the three groups did not show a significant difference between the groups, $F(2,57) = 1.70$, $x^2(2) = 1.703$, $p = .191$. The reaction times in the PVT that are longer than 500 milliseconds are categorised as lapses (Matthews et al., 2017). Similar, the one-way ANOVA that compared between the groups the frequency of lapses in the PVT baseline did not show a significant difference between the groups, $F(2,57) = 1.198$, $p = .309$. Finally, the Kruskal-Wallis test did not show a significant difference in the CSS baseline ratings between the groups, $x^2(2) = 1.853$, $p = .396$.

Table 5.2 Participants' age, years of driving experience, hours of sleep, sleep quality, and fatigue baselines.

		Control	**Manual**	**Automated**
Age		28 (SD=4.96)	30 (SD = 6.53)	26.6 (SD = 3.62)
Driving experience (years)		7.4 (SD = 4.81)	10.2 (SD = 6.84)	6.5 (SD = 6.5)
Hours of sleep		7.7 (SD = 0.52)	7.6 (SD = 0.44)	7.5 (SD = 0.32)
Sleep quality (% of participants)	**Very well**	30	10.3	10.2
	Well	35	65.1	45.2
	Fairly well	29.9	14.9	39.6
	Fairly badly	5.1	9.7	5
PVT baseline	**Reaction time (ms)**	334.11 (SD = 13.00)	315.62 (SD = 8.08)	310.61 (SD = 6.07)
	Frequency of lapses	1.50 (SD = 0.38)	1.45 (SD = 0.40)	0.80 (SD = 0.27)
CSS baseline		2 (95% CI = 0.08)	2 (95% CI = 0.10)	2 (95% CI = 0.08)

The results described in this section showed that there were no significant differences between the groups in age, driving experience, the investigated sleep-related factors, and the fatigue baselines. Fully controlling for the effects of sleep-related factors and inter-individual differences on fatigue is difficult, but these results suggest that any differences in the expressions of fatigue found between the groups in the trials could be attributed to the tasks completed.

5.4.2 Fatigue in the 60-minute drive

According to hypothesis 1, the CSS ratings, the PVT reaction times, and the number of PVT lapses of the manual and the automated group would be significantly higher at the end than at the beginning of the 60-minute drive, whereas according to hypothesis 2, the total task load and the ratings of effort in the NASA-TLX during the 60-minute drive would be significantly lower in the automated than in the manual group. This was tested by analysing the CSS, PVT (reaction time and lapses), and NASA-TLX data collected in this drive. Reaction times in the PVT are recorded as lapses when they are longer than 500 milliseconds (Matthews et al., 2017). The simple reaction times of healthy, non-sleep-deprived individuals in the PVT vary between 220 and 500 ms (Dreary and Der, 2005) with a mean of around 330ms (Woods et al., 2015).

5.4.2.1 CSS

The medians of the baseline CSS ratings and the CSS ratings collected during the 60-minute drive are shown in Figure 5.24. It should be noted that the CSS ratings appear the same on several measurement time points in the Figure because of the use of the medians (instead of the means). However, the significant differences of interest for the hypotheses tested are described in the next paragraph and shown as asterisks in Figure 5.24.

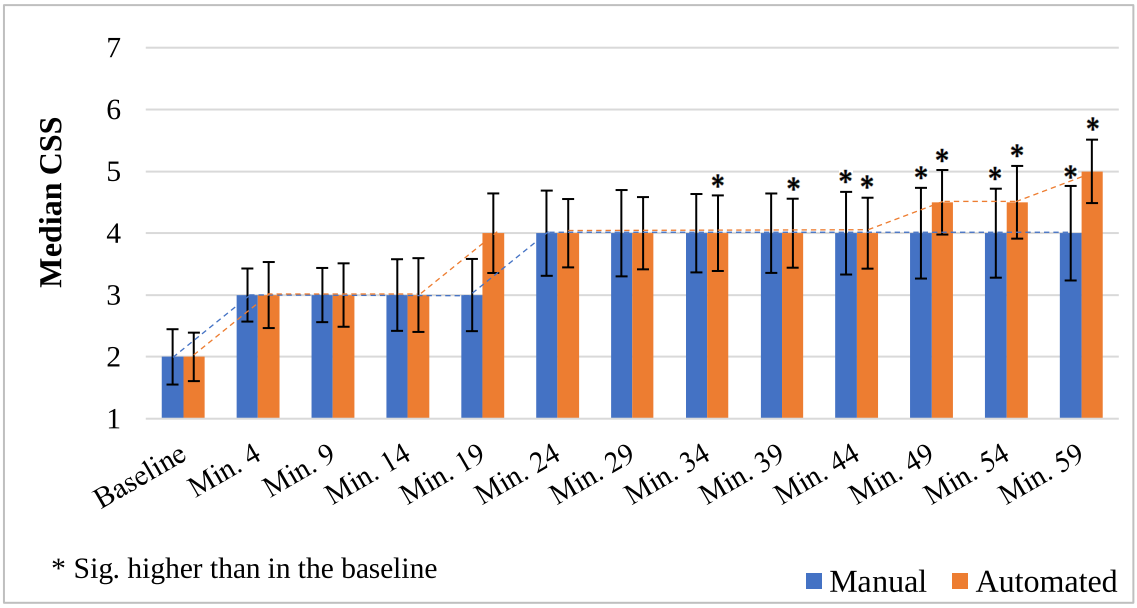


Figure 5.24 The CSS ratings before and during the 60-minute drives in experiment 1.

The analysis with the nparLD package showed no significant interaction between the group and the measurement time point and no significant simple main effect of the group. The simple main effect of the measurement time point was significant, Wald chi-square (1, N=40) = 3.101, *p* <

.005, η_p^2 = .134. This was further explored with Friedman tests. For the manual group, the difference between the measurement time points was significant, $\chi^2(12)$ = 114.177, p < .005, Kendall's coefficient of concordance = .476. The post-hoc analysis revealed that the ratings of fatigue were significantly higher than the baseline (median = 2) on all the measurement time points from minute 44 (median = 4) ($\chi^2(12)$ = 4.243, p = .002, r = .512) onwards to minute 59 ($\chi^2(12)$ = 6.600, p < .005, r = .430) for the manual group. Similar to the manual group, the CSS ratings of the automated group were statistically significantly different between the measurement time points during the 60-minute drive, $\chi^2(12)$ = 167.256, p < .005, Kendall's coefficient of concordance =.69. The post-hoc analysis revealed a statistically significant increase in the ratings of fatigue of the automated group on the CSS scale from the baseline (median = 2) to all the measurement time points from minute 34 (median = 4), $\chi^2(12)$ = 4.650, p = .012, r = .181, onwards to minute 59, $\chi^2(12)$ = 9.400, p < .005, r = .212.

5.4.2.2 PVT performance

The mean PVT reaction times before and after the 60-minute drives are shown in panel A of Figure 5.25. A two-way mixed ANOVA did not show a significant interaction or simple main effect of the group but showed a significant simple main effect of the measurement time point, $F(1, 38)$ = 23.698, p < .005, η_p^2 = .384. This was further investigated with paired samples t-tests. For the manual group, the PVT reaction times increased significantly from the baseline (mean = 315.63 milliseconds, SE = 8.08) to the measurement directly after the 60-minute drive (mean = 340.11, SE = 10.76), $t(19)$ = 3.812, p = .001, d = 0.85. Similar to the manual group, the PVT reaction times of the automated group increased significantly from the baseline (mean = 310.66, SE = 6.07) to the measurement after the 60-minute drive (mean = 322.52, SE = 7.51), $t(19)$ = 3.115, p = .006, d = 0.69.

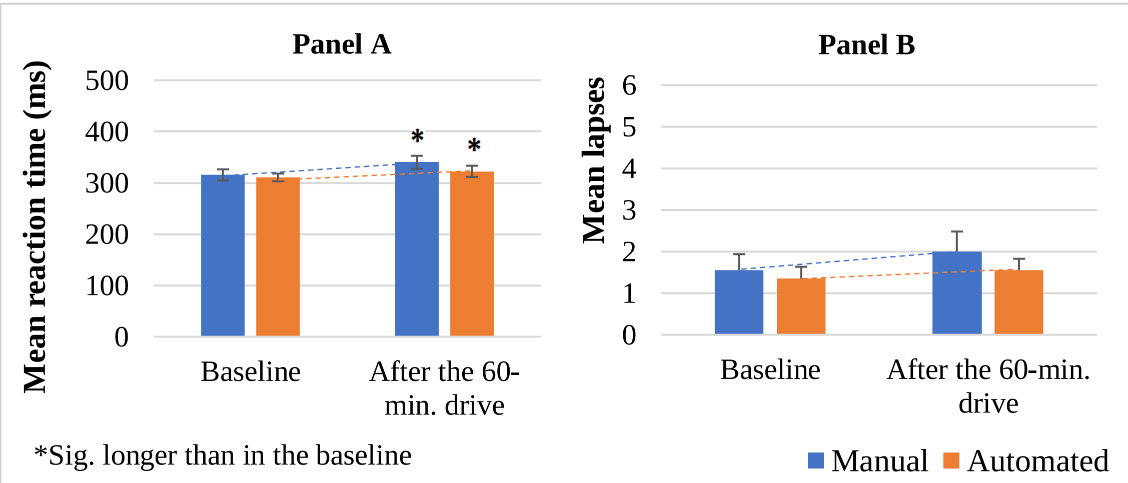


Figure 5.25 The PVT reaction times (panel A) and lapses (reaction > 500 ms) (panel B) before and directly after the 60-minute drive of experiment 1.

The mean frequency of the lapses before and after the 60-minute drives are shown in panel B of Figure 5.25. The two-way mixed ANOVA did not show a significant interaction between the group and the measurement time point. In contrast to the simple main effect of the group, the simple main effect of the measurement time point was significant, $F(1, 38) = 44.564$, $p < .005$, $\eta_p^2 = .54$. This was further investigated with paired samples t-tests, which were not significant.

5.4.3 Task load in the 60-minute drive

The NASA-TLX total scores and the ratings on each of its items during the 60-minute drive were analysed to identify if the 60-minute drives induced active or passive fatigue.

5.4.3.1 Total task load

The means of the total NASA-TLX scores are depicted in Figure 5.26. 600 is the maximum total task load score that can be given on the NASA-TLX scale because there are six items rated on a 100-point scale each (with 5-point steps). The assumption of sphericity of the two-way mixed ANOVA with the total NASA-TLX scores was violated, so the degrees of freedom were corrected with the Greenhouse-Geisser estimates of sphericity ($\varepsilon = 0.379$). A new two-way mixed ANOVA showed that the interaction between the measurement time point and the group was not significant. In contrast, the simple main effects of the measurement time point and the group were significant $F(4.172, 158.531) = 7.440$, $p < .005$, $\eta_p^2 = .16$ and $F(1, 38) = 23.564$, $p < .005$, $\eta_p^2 = .38$ respectively.

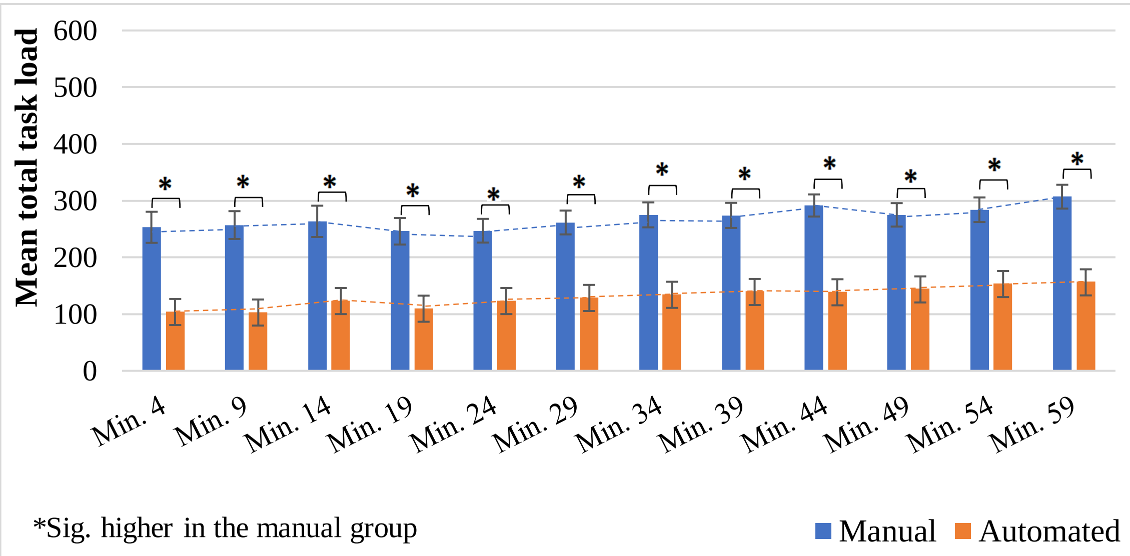


Figure 5.26 The total NASA-TLX scores in the 60-minute drive of experiment 1.

The simple main effect of the measurement time point was further investigated with one-way repeated measures ANOVAs with a Bonferroni correction for multiple comparisons. For the manual group, the assumption of sphericity was violated (degrees of freedom corrected with the Greenhouse-Geisser estimates of sphericity, $\varepsilon = 0.446$). The total TLX scores of the manual group were statistically significantly different between the measurement time points, $F(4.905, 93.197) = 3.389$, $p = .008$, $\eta_p^2 = .15$. Post-hoc analyses revealed a significant increase from minute 19 to minute 59 ($p = .034$, $r = .212$) for the manual group. Similar to the manual group, the assumption of sphericity was not met for the automated group (degrees of freedom corrected with the Greenhouse-Geisser estimates of sphericity, $\varepsilon = .171$). The total TLX scores differed significantly between the measurement time points for the automated group, $F(1.876, 35.643) = 5.780$, $p = .008$, $\eta_p^2 = .23$. The post-hoc analysis, however, did not show any significant differences. The simple main effect of the group was further investigated with a series of independent samples t-tests. The analysis showed that the total task load scores were significantly higher in the manual group on all the measurement time points during the 60-minute drive (all $p < .005$, all $d > 1.34$).

5.4.3.2 Mental demand

The means of the ratings on the mental demand item during the 60-minute drives are depicted in Figure 5.27. The assumption of sphericity of the two-way mixed ANOVA was not met, so the degrees of freedom were corrected with the Greenhouse-Geisser estimates of sphericity ($\varepsilon = 0.298$). The new two-way mixed ANOVA showed that the interaction between the measurement time point and the group was not significant. In contrast, the simple main effects of the measurement time point and the group were significant, $F(3.491, 339.235) = 6.662$, $p < .005$, $\eta_p^2 = .23$ and $F(1, 38) = 24.342$, $p < .005$, $\eta_p^2 = .39$ respectively.

The simple main effect of the measurement time point was further investigated with one-way repeated measures ANOVAs with a Bonferroni correction. For the manual group, the assumption of sphericity was violated (degrees of freedom corrected with the Greenhouse-Geisser estimates of sphericity, $\varepsilon = 0.312$). The ratings of the manual group on the mental demand item were statistically significantly different between the measurement time points, $F(5.399, 87.331) = 4.701$, $p = .032$, $\eta_p^2 = .08$. Post-hoc analyses revealed a significant increase from minute 24 to minutes 44 ($p = .022$, $r = .12$), 49 ($p = .030$, $r = .23$), and 59 ($p = .017$, $r = .30$).

The assumption of sphericity was violated for the automated group too (Greenhouse-Geisser estimates of sphericity, $\varepsilon = .334$). The ratings on the mental demand item differed significantly between the measurement time points for the automated group, $F(2.330, 29.331) = 4.090$, $p < .005$, $\eta_p^2 = .29$. The post-hoc analysis showed a significant increase from minute 9 to minutes 34 onwards (all $p < .009$, all $r > 0.19$). The simple main effect of the group was further investigated with a series of independent samples t-tests. The analysis showed that the mental demand scores were significantly higher in the manual group on all the measurement time points during the 60-minute drive (all $p < .005$, all $d > .05$).

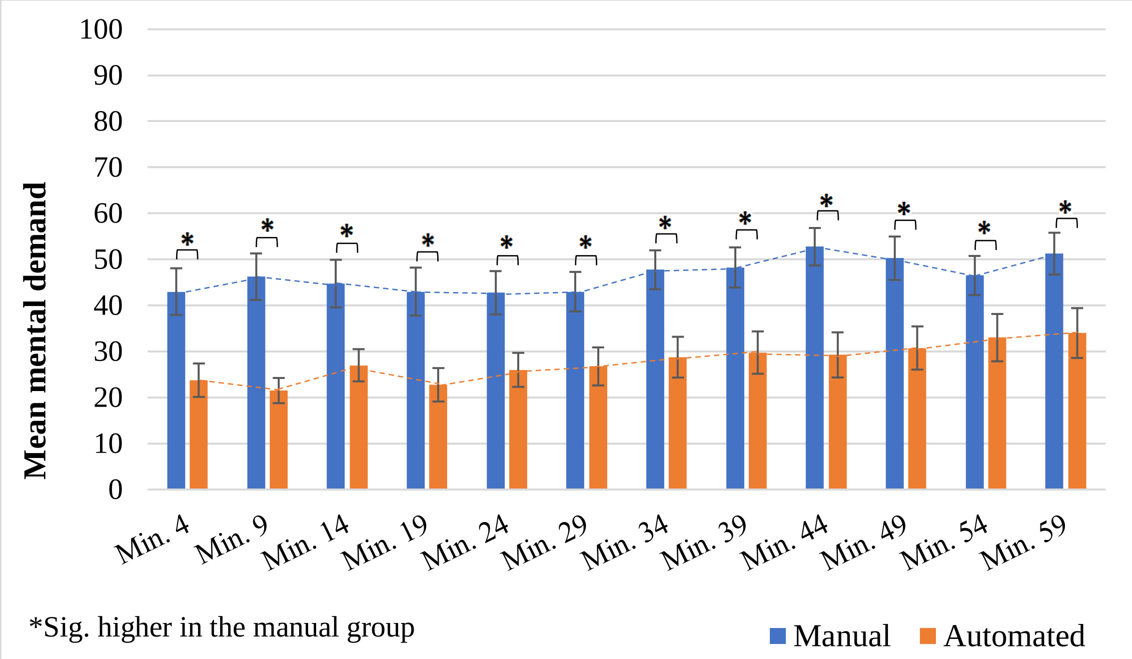


Figure 5.27 The ratings of mental demand (NASA-TLX) in the 60-minute drive (experiment 1).

5.4.3.3 Physical demand

The means of the ratings on the physical demand item of the NASA-TLX are shown in Figure 5.28. The assumption of sphericity of the two-way mixed ANOVA was violated, so the degrees of freedom were corrected with the Greenhouse-Geisser estimates of sphericity (ε = 0.256). The new two-way mixed ANOVA showed that the interaction between the measurement time point and the group was not significant. In contrast, the simple main effects of the measurement time point and the group were significant, $F(4.222, 143.551) = 6.543$, $p = .022$, $\eta_p^2 = .09$ and $F(1, 38) = 21.312$, $p < .005$, $\eta_p^2 = .26$ respectively.

The simple main effect of the measurement time point was further explored with one-way repeated measures ANOVAs with a Bonferroni correction for multiple comparisons. For the manual group, the assumption of sphericity was not met (degrees of freedom corrected with the Greenhouse-Geisser estimates of sphericity, ε = 0.401). The ratings of the manual group on the physical demand item were statistically significantly different between the measurement time points, $F(5.254, 77.309) = 2.344$, $p = .012$, $\eta_p^2 = .19$. Post-hoc analyses showed that the ratings of physical demand increased significantly from minute 19 to minutes 39 ($p = .026$, $r = .19$), 44 ($p = .012$, $r = .23$), 49 ($p < .005$, $r = .15$), 54 ($p < .005$, $r = .33$), and 59 ($p < .005$, $r = .40$). The assumption of sphericity was violated for the automated group too (Greenhouse-Geisser estimates of sphericity, ε = .334). The ratings on the mental demand item differed significantly between the measurement time points for the automated group, $F(1.749, 29.440) = 6.121$, $p < .005$, $\eta_p^2 = .29$. The post-hoc analysis showed a significant increase from minute 9 to minutes 34 onwards (all *p*

< .009, all $r > 0.16$). The simple main effect of the group was further investigated with a series of independent samples t-tests. The analysis showed that the mental demand scores were significantly higher in the manual group on all the measurement time points during the 60-minute drive (all $p < .005$, all $d > .05$).

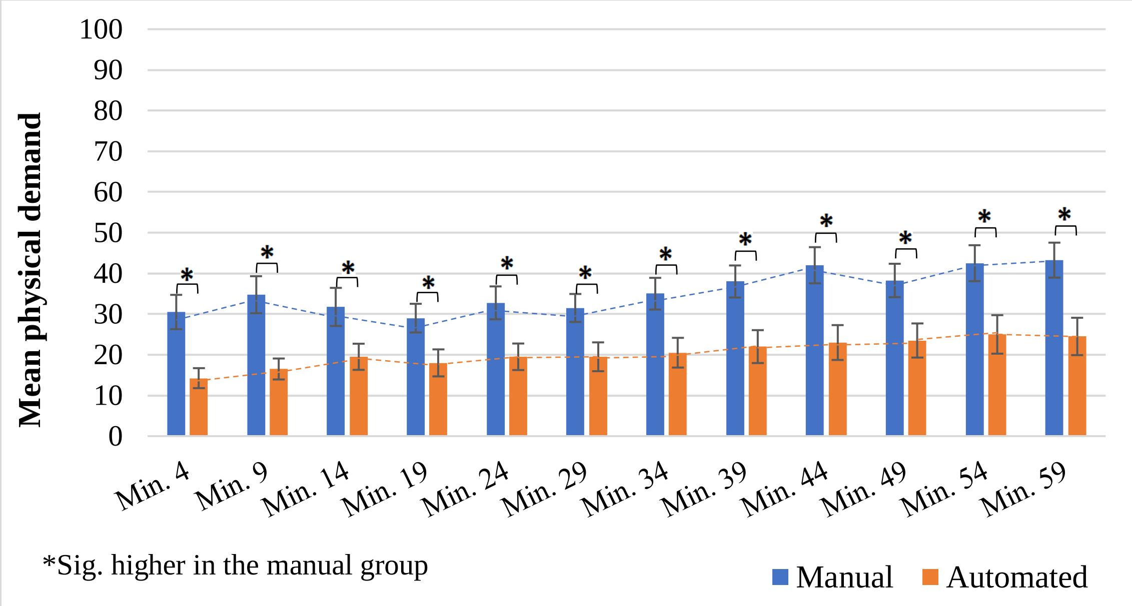


Figure 5.28 The ratings of physical demand (NASA-TLX) in the 60-minute drive (experiment 1).

5.4.3.4 Temporal demand

Figure 5.29 depicts the means of the ratings on the temporal demand item.

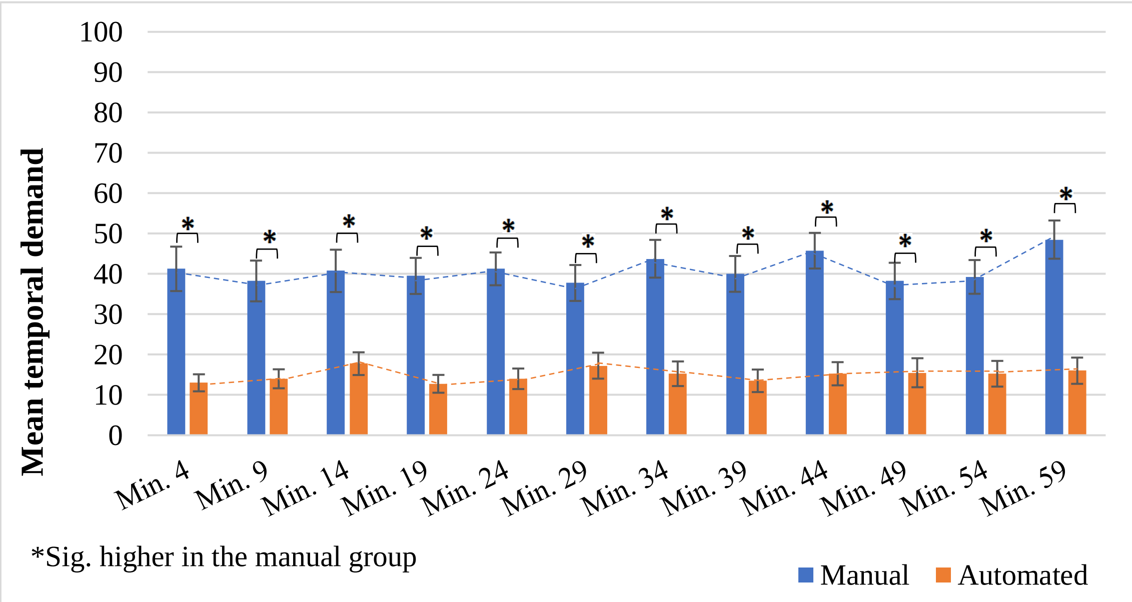


Figure 5.29 The ratings of temporal demand (NASA-TLX) in the 60-minute drive (experiment 1).

The assumption of sphericity of the two-way mixed ANOVA was violated, so the degrees of freedom were corrected with the Greenhouse-Geisser estimates of sphericity (ε = 0.254). A new two-way mixed ANOVA showed that the interaction between the measurement time point and the group and the simple main effect of the measurement time point were not significant. In contrast, the simple main effect of the group was significant, $F(1, 38) = 20.555$, $p < .005$, $\eta_p^2 = .24$. This was further explored with independent samples t-tests. The analysis showed that the temporal demand scores were significantly higher in the manual group on all the measurement time points during the 60-minute drive (all $p < .005$, all $d > .85$).

5.4.3.5 Performance

Figure 5.30 shows the means of the ratings on the performance item.

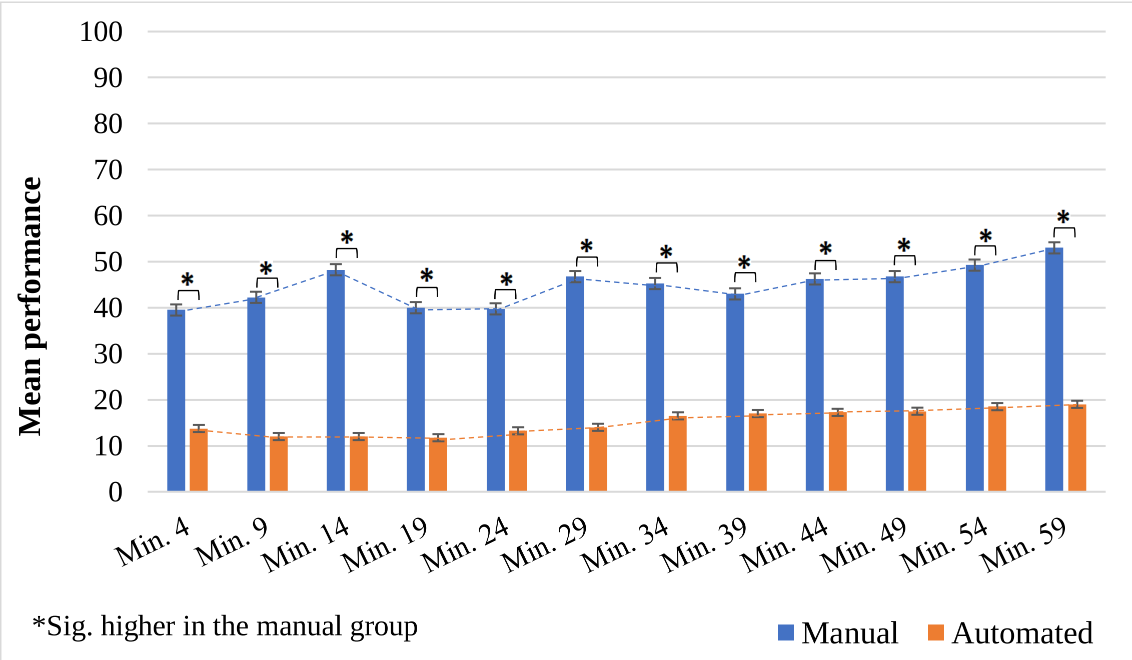


Figure 5.30 The ratings of performance (NASA-TLX) in the 60-minute drive (experiment 1).

The assumption of sphericity of the two-way mixed ANOVA was not met (the degrees of freedom were corrected with the Greenhouse-Geisser estimates of sphericity (ε = 0.199). A new two-way mixed ANOVA showed that the interaction between the measurement time point was not significant. The simple main effect of the group was significant, $F(1, 38) = 17.660$, $p < .005$, $\eta_p^2 = .35$. This was further investigated with independent samples t-tests. The analysis showed that the temporal demand scores were significantly higher in the manual group on all the measurement time points during the 60-minute drive (all $p < .005$, all $d > .63$). Although the simple main effect of the measurement time point was not significant, there was a trend for an increase in the ratings on the performance item (p = .062).

The simple main effect of the measurement time point was significant, F(3.288, 397.423) = 5.745, p = .038, η_p^2 = .09. This was further explored with one-way repeated measures ANOVAs with a Bonferroni correction. For the manual group, the assumption of sphericity was violated (degrees of freedom corrected with the Greenhouse-Geisser estimates of sphericity, ε = 0.277). The ratings of the manual group on the performance item were statistically significantly different between the measurement time points, F(5.002, 89.519) = 3.268 p = .012, η_p^2 = .19. Post-hoc analyses showed that the ratings of performance increased significantly from minute 4 to minutes 14 (p = .017, r = .22), 29 (p = .024, r = .19), 44 (p = .033, r = .29), 49 (p = .012, r = .20), 54 (p = .015, r = .32), and 59 (p < .005, r = .40). The assumption of sphericity was violated for the automated group (Greenhouse-Geisser estimates of sphericity, ε = .299). The ratings on the performance item did not differ significantly between the measurement time points for the automated group, but there was a trend for an increase in the ratings on the performance item (p = .070).

5.4.3.6 Effort

Figure 5.31 shows the means of the ratings on the effort item during the 60-minute drives. 100 was the highest rating that could be given.

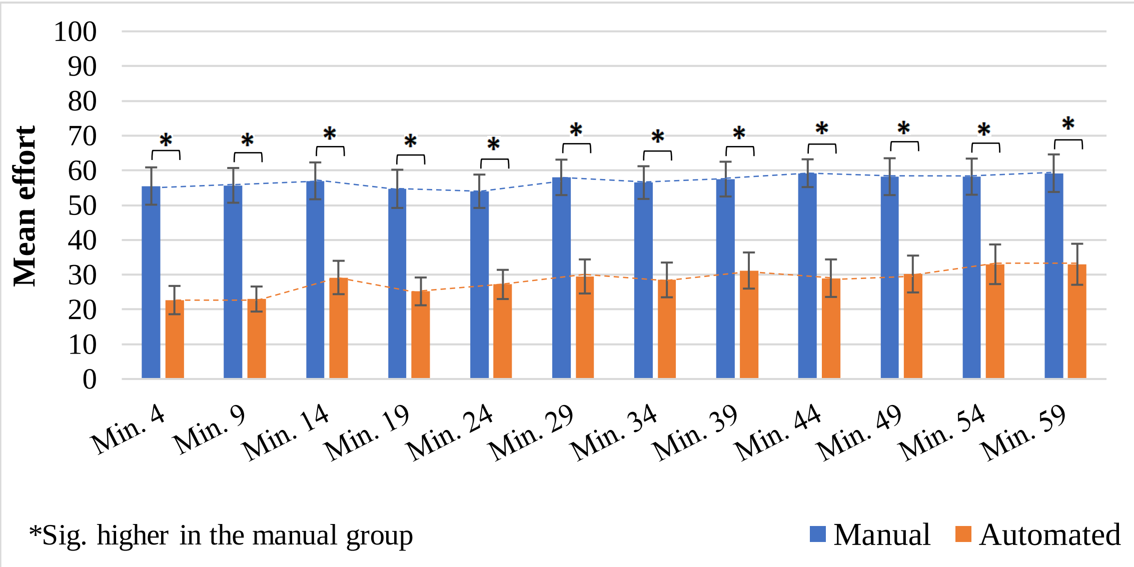


Figure 5.31 The ratings of effort (NASA-TLX) in the 60-minute drive (experiment 1).

The assumption of sphericity of the two-way mixed ANOVA that was performed to analyse these data was violated, so the degrees of freedom were corrected with the Greenhouse-Geisser estimates of sphericity (ε = 0.332). A new two-way mixed ANOVA showed that the interaction between the measurement time point and the group and the simple main effect of the measurement time point were not significant. In contrast, the simple main effect of the group was significant, $F(1, 38) = 19.558$, $p < .005$, $\eta_p^2 = .34$. This was further investigated with a series of independent samples t-tests. The analysis revealed that the effort scores were significantly higher in the manual group on all the measurement time points during the 60-minute drive (all $p < .005$, all $d > 1.02$). The simple main effect of the measurement time point was not significant, but there was a trend for an increase in the ratings on the effort item ($p = .057$).

5.4.3.7 Frustration

The means of the ratings on the frustration item of the NASA-TLX are shown in Figure 5.32.

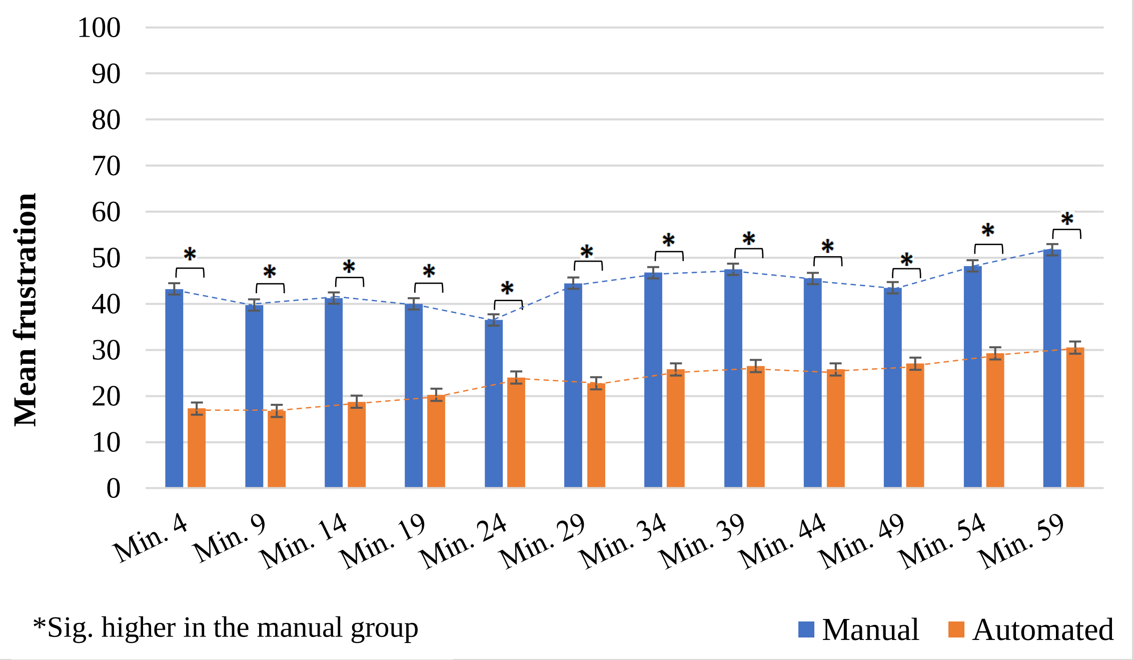


Figure 5.32 The ratings of frustration (NASA-TLX) in the 60-minute drive (experiment 1).

The assumption of sphericity of the two-way mixed ANOVA was violated, so the degrees of freedom were corrected (Greenhouse-Geisser estimates of sphericity, $\varepsilon = 0.411$). The new two-way mixed ANOVA showed that the interaction between the measurement time point and the group was not significant. In contrast, the simple main effects of the measurement time point and the group were significant, $F(5.831, 398.760) = 5.441$, $p = .021$, $\eta_p^2 = .08$ and $F(1, 38) = 23.542$, $p < .005$, $\eta_p^2 = .29$ respectively.

The simple main effect of the measurement time point was further investigated with one-way repeated measures ANOVAs with a Bonferroni correction. For the manual group, the assumption of sphericity was violated (degrees of freedom corrected with the Greenhouse-Geisser estimates of sphericity, $\varepsilon = 0.401$). The ratings of the manual group on the frustration item were statistically significantly different between the measurement time points, $F(4.743, 88.434) = 2.098$, $p = .022$,

η_p^2 = .11. Post-hoc analyses showed that the ratings of frustration increased significantly from minute 24 to minutes 39 (p = .012, r = .17), 54 (p < .005, r = .32), and 59 (p < .005, r = .44). The assumption of sphericity was violated for the automated group (Greenhouse-Geisser estimates of sphericity, ε = .188). The ratings on the mental demand item differed significantly between the measurement time points for the automated group, F(2.468, 29.420) = 4.239, p < .005, η_p^2 = .16. The post-hoc analysis showed a significant increase from minute 4 to minutes 24 onwards (all p < .012, all r > 0.09). The simple main effect of the group was further investigated with independent samples t-tests. The analysis showed that the frustration scores were significantly higher in the manual group on all the measurement time points during the 60-minute drive (all p < .005, all d > .04).

5.4.4 Fatigue in the MATB-II tasks

According to hypothesis 3, the control group would be significantly less fatigued than the manual and the automated in the MATB-II tasks, whereas hypothesis 4 stated that the automated group would be significantly less fatigued than the manual in the MATB-II tasks. The CSS, PVT, and MATB-II performance data were analysed to test these hypotheses.

5.4.4.1 CSS

The medians of the CSS ratings during the MATB-II tasks are shown in Figure 5.33. The ratings at the end of the 60-minute drives are also depicted for comparison purposes. Some of the differences between the measurement time points cannot be seen in this Figure because the use of the medians (instead of the means) resulted in some bars looking the same. However, the significant differences of interest for the hypotheses that were tested are described in the next paragraph and shown as asterisks and triangles in Figure 5.33.

A mixed model analysis (nparLD) showed that the interaction between the group and the measurement time point was significant, Wald chi-square (10, N=60) = 51.753, p < .005, η_p^2 = .125. Therefore, the ratings of subjective fatigue changed differently over the MATB-II tasks for the groups. The simple main effect of the group was significant, Wald chi-square (2, N=60) = 27.498, p < .005, η_p^2 = 0.1169. The Kruskal-Wallis tests revealed that the CSS ratings differed significantly between the groups on all the time points (all p < .005, all ε^2 > .132). Subsequently, pairwise comparisons were performed with Dunn's (1964) procedure with a Bonferroni correction. The control group reported significantly lower levels of fatigue than the manual and the automated just before the MATB-II tasks (p = .001, η^2 = .122 and p < .005, η^2 = .148 respectively) and on minutes 4 (p = .003, η^2 = .102 and p = .002 respectively, η^2 = .121), 8 (p < .005, η^2 = .162 and p = .023 respectively, η^2 = .094), and 24 (p < .005, η^2 = .133 and p = .018, η^2 = .088 respectively). However, only the manual group reported significantly higher levels of fatigue than the control on minutes 28 (p < .005, η^2 = .110) and 32 (p < .005, η^2 = .178). Moreover, it was found that the ratings of fatigue were significantly lower in the automated group than in the manual on minutes 28 and 32 (both $p < .005$, $\eta^2 = .056$ and $\eta^2 = .070$ respectively).

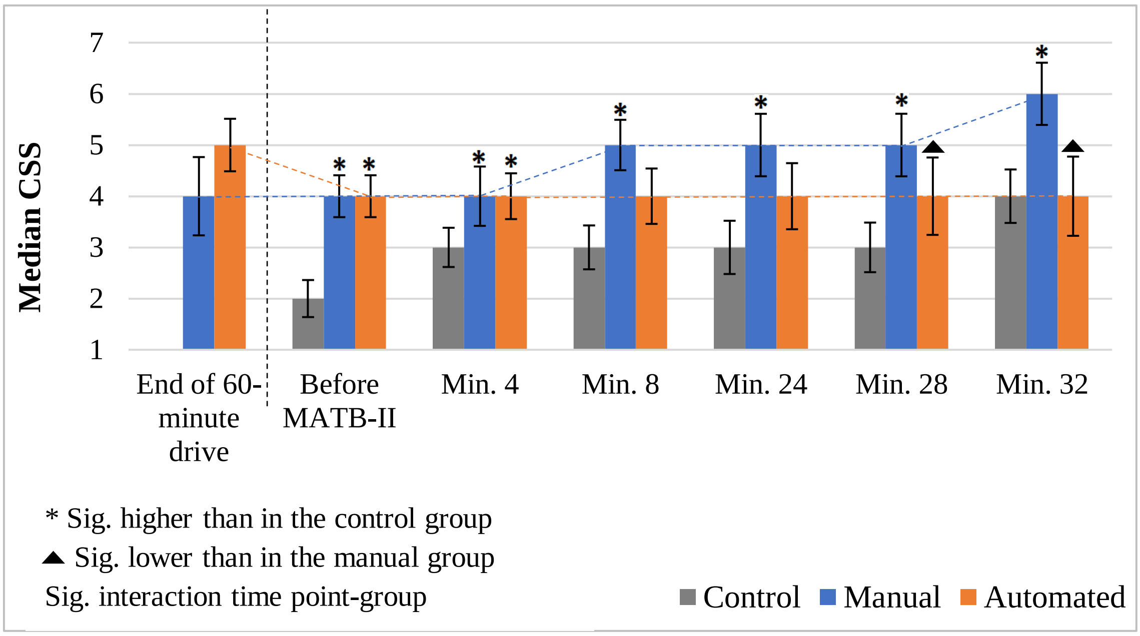


Figure 5.33 The CSS ratings in the MATB-II.

The simple main effect of the measurement time point was significant, Wald chi-square (5, N=40) = 86.951, $p < .005$, $\eta_p^2 = .212$. This was further investigated with Friedman tests. For the control group, the CSS ratings were statistically significantly different between the measurement time points, $\chi^2(5) = 41.774$, $p < .005$, Kendall's coefficient of concordance = .41. The post-hoc analysis revealed a statistically significant increase in fatigue from just before MATB-II (median = 2) to minutes 24 (median = 3), $\chi^2(5) = 3.296$, $p = .015$, $r = .127$, and 32 (median = 4), $\chi^2(5) = 4.775$, $p < .005$, $r = .189$. Moreover, there was a trend for significantly higher levels of fatigue than just before the MATB-II tasks on minute 28 ($p = .057$). The comparisons between consecutive measurement time points showed a significant increase from minute 8 to 24 ($p = .040$, $r = .092$).

For the manual group, the CSS ratings were statistically significantly different between the time points, $\chi^2(5) = 81.916$, $p < .005$, Kendall's coefficient of concordance = .81. The post-hoc analysis revealed a statistically significant increase in the ratings of fatigue from just before MATB-II to minute 8, $\chi^2(5) = 3.085$, $p = .031$, $r = .194$, minute 24, $\chi^2(5) = 5.494$, $p < .005$, $r = .212$, minute 28, $\chi^2(5) = 4.606$, $p < .005$, $r = .176$, and minute 32, $\chi^2(5) = 6.423$, $p < .005$, $r = .302$, The comparisons between the consecutive measurement time points did not show any significant differences. Similar to the other groups, the ratings of fatigue were statistically significantly different between the measurement time points for the automated group, $\chi^2(5) = 14.180$, $p = .015$, Kendall's coefficient of concordance = .14. The post-hoc analysis revealed no significant differences. However, a trend was found for higher reported levels of fatigue on minute 8 compared to fatigue just before the MATB-II tasks (p = .070).

Table 5.3 shows the number of participants that reported high levels of fatigue in the flying tasks. A rating of six means '*extremely tired, very difficult to concentrate*', whereas a rating of seven refers to '*completely exhausted, unable to function properly*'. As can be seen, more participants reported being very fatigued in the last blocks of the flying tasks in the manual compared to the other groups.

Table 5.3 The frequency of high ratings of fatigue (CSS) in the MATB-II tasks.

	Rating	**Just before the MATB-II**	**Min. 4**	**Min. 8**	**Min. 24**	**Min. 28**	**Min. 32**
Control	**6**	-	-	-	-	-	-
	7	-	-	-	-	-	-
Manual	**6**	-	-	-	6	4	6
	7	-	-	-	-	-	-
Automated	**6**	1	1	2	1	1	1
	7	-	-	1	1	1	1

Wilcoxon signed-rank tests were run to explore if transitioning from the 60-minute drive to the MATB-II tasks by completing the PVT affected the reported levels of fatigue. The analysis did not show a significant difference between the CSS ratings at the end of the drive and just before the flying tasks in the manual group. In contrast, the automated group reported significantly lower levels of fatigue just before the MATB-II tasks compared to the end of the 60-minute drive, z = -3.213, p = .001, ε^2 = .523.

5.4.4.2 PVT and MATB-II performance

Figure 5.34 (Panel A) shows the means of the PVT reaction times collected before and after the MATB-II tasks. A two-way mixed ANOVA showed that the interaction between the group and the time and the simple main effect of the group were not significant. In contrast, the simple main effect of the measurement time point was significant, $F(1, 57) = 25.898$, $p < .005$, $\eta_p^2 = .312$. This was further investigated with paired samples t-tests. These showed that the reaction times after the MATB-II tasks were significantly slower than before them: $t(19) = 3.707$, $p = .001$, $d = 0.82$ in the control group, $t(19) = 4.598$, $p < .005$, $d = 1.02$ in the manual group, and $t(19) = 4.383$, $p <$

.005, *d* = 0.98 in the automated group. The mean number of lapses before and after the MATB-II are shown in panel B of Figure 5.34. The two-way mixed ANOVA showed that only the simple main effect of the measurement time point was significant in the number of lapses, $F(1, 57) = 4.174$, $p = .046$, $\eta_p^2 = .06$. The paired samples t-tests did not show any significant differences between before and after the MATB-II. However, there was a trend for significantly more lapses after the MATB-II compared to before it in the automated group ($p = .056$).

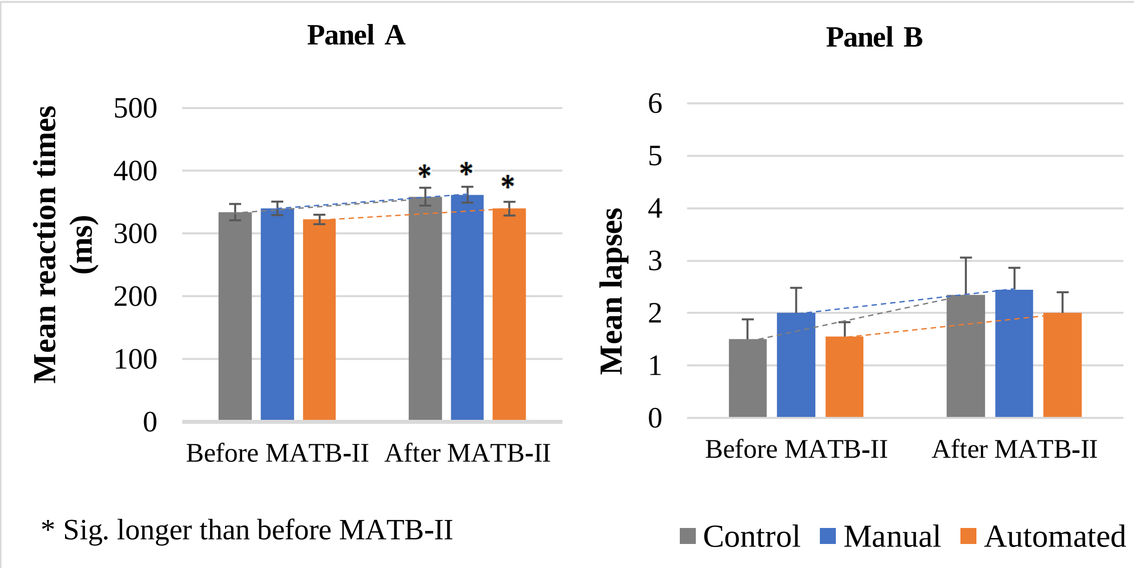


Figure 5.34 The reaction times (panel A) and frequency of lapses (reaction > 500 ms) (panel B) on the PVT before and after the MATB-II tasks.

The mean RMSDs of tracking in the five MATB-II blocks of tasks are shown in Figure 5.35.

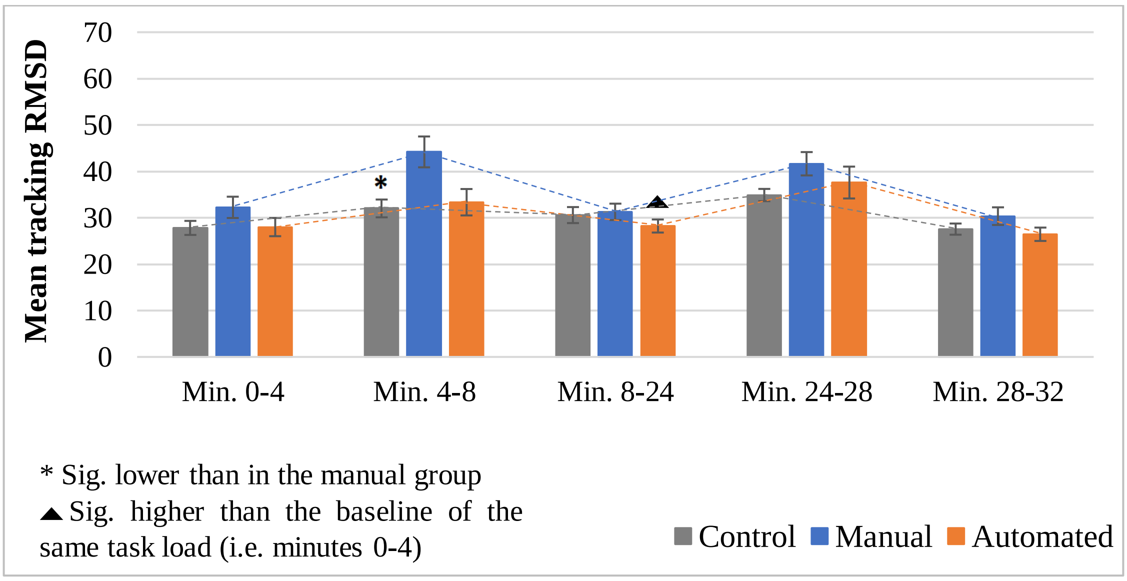


Figure 5.35 The tracking accuracy in the MATB-II tasks.

A two-way mixed ANOVA was performed to analyse these data but the assumption of homogeneity of variances was violated. The logarithmic transformation was not successful, so a mixed-model analysis was performed with the nparLD package. The two-way mixed ANOVA showed that the interaction and the simple main effect of the group were not significant. In contrast, the simple main effect of the block was significant, Wald chi-square (4, N=60) = 105.994, p < .005, η_p^2 = .119 It was expected that the participants would perform worse in the multi-tasking blocks (i.e. minutes 4-8 and 24-28) than in the single-tasking ones (i.e. minutes 0-4, 8-24, and 28-32) because of the requirement to complete the monitoring and the communication tasks besides the tracking one. Therefore, any differences between the single- and the multi-tasking blocks in the tracking RMSD were not considered to infer differences in the levels of fatigue.

Friedman tests and pairwise comparisons were performed with a Bonferroni correction to investigate the simple main effect of the block on tracking accuracy separately for the single- the and multi-tasking blocks. In the control group, the tracking RMSD was statistically significantly different between the blocks, $\chi^2(4)$ = 35.480, p < .005, Kendall’s coefficient of concordance = .44. However, the post-hoc analysis showed no significant difference between the blocks of the same levels of task load (i.e. single- or multi-tasking). Similar to the control group, the tracking RMSD was statistically significantly different between the blocks for the manual group, $\chi^2(4)$ = 33.800, p < .005, Kendall’s coefficient of concordance = .42. However, the post-hoc analysis did not show any significant differences between blocks of the same task load. Finally, the RMSD of tracking was statistically significantly different between the blocks in the automated group, $\chi^2(4)$ = 66.236, p < .005, Kendall’s coefficient of concordance = .76. In contrast to the other groups, the post-hoc analysis in the automated group showed a statistically significant increase from block 1 (i.e. minutes 0-4) to block 3 (i.e. minutes 8-24) (p = .001, r = .187).

The simple main effect of the group in the tracking accuracy was significant, Wald chi-square (2, N=60) = 6.197, p = .045, η_p^2 = .092. This was further investigated with Kruskal-Wallis tests that showed a significant simple main effect of the group in block 2 (i.e. minutes 4-8), $x^2(2)$ = 11.282, p = .004, ε^2 = .215. Subsequently, pairwise comparisons were performed using Dunn's (1964) procedure with a Bonferroni correction for multiple comparisons. The analysis showed that the tracking RMSD in the manual group was significantly higher than in the control group in block 2 (p = .014, η^2 = .113).

The mean frequencies of errors in the communications task are shown in Figure 5.36. A two-way mixed ANOVA was performed but the assumption of homogeneity of variances was violated. The square root transformation was not successful, so a mixed-model analysis was performed with the nparLD package. The two-way mixed ANOVA showed that the interaction between the measurement time point and the group were not significant. In contrast, the simple main effect of the block was significant, Wald chi-square (1, N=60) = 13.375, $p < .005$, $\eta_p^2 = .092$. This was further investigated with Wilcoxon signed-rank tests. For the control group, there was a statistically significant median increase of communication errors from block 2 (i.e. minutes 4-8) to block 4 (i.e. minutes 24-28), $z = 1.998$, $p = .046$, $r = .230$. For the manual group, the within-groups difference was not significant. In contrast, for the automated group, there was a significant median increase of the communication errors from block 2 (i.e. minutes 4-8) to block 4 (i.e. minutes 24-28), $z = 2.586$, $p = .010$, $r = .298$.

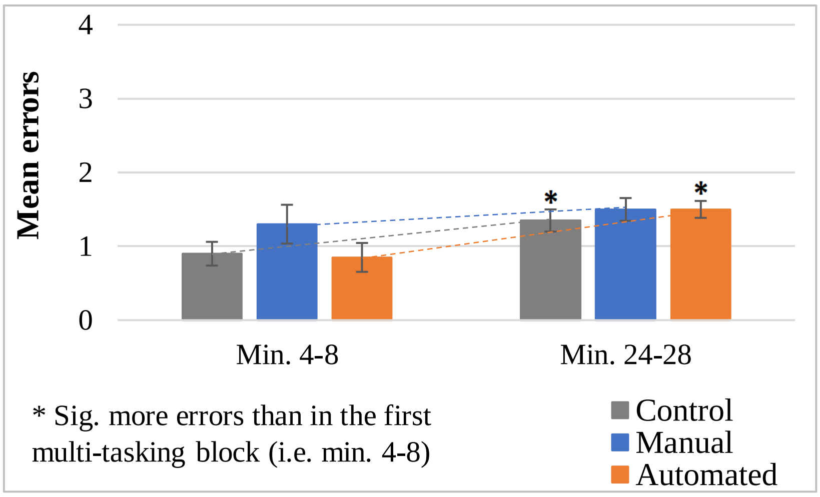


Figure 5.36 The frequency of errors in the communications task of the MATB-II. Each of the blocks shown included four messages.

The mean frequencies of omitted responses at the red light, green light, and scale tasks are shown in Figure 5.37.

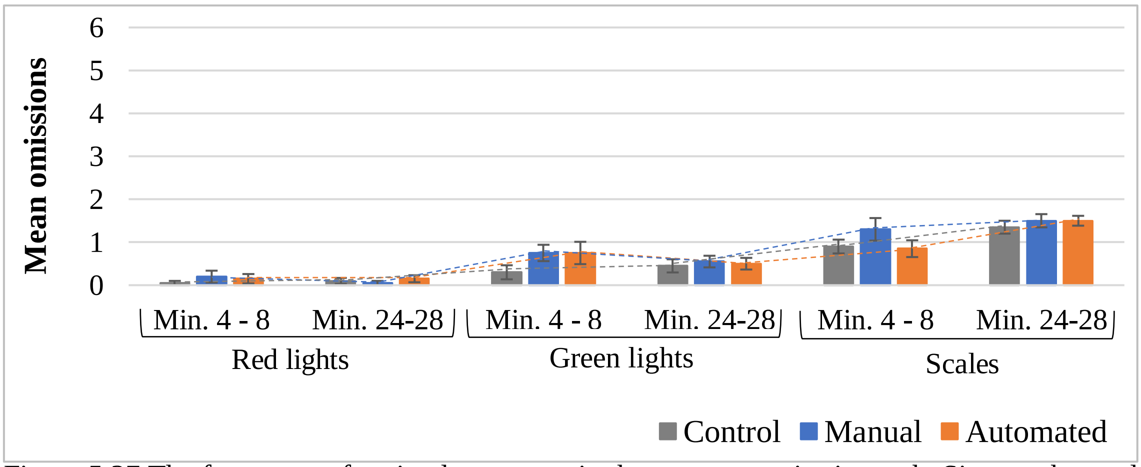


Figure 5.37 The frequency of omitted responses in the system monitoring task. Six was the total number of events of each type of task presented per block.

The two-way mixed ANOVAs showed no significant interaction between the group and the block and no significant simple main effects of the group and the block in the number of omitted responses in the monitoring task. The means of the response times to the red light, green light, and scales tasks are shown in Figure 5.38. When a light or a scale was omitted, the maximum time allowed to respond (i.e.10 seconds) was recorded as the response time. The two-way mixed-ANOVAs performed to explore response times to the red lights, the green lights, and the scales showed no significant interaction between the group and the block and no significant simple main effects of the group and the block in the reaction times to the events.

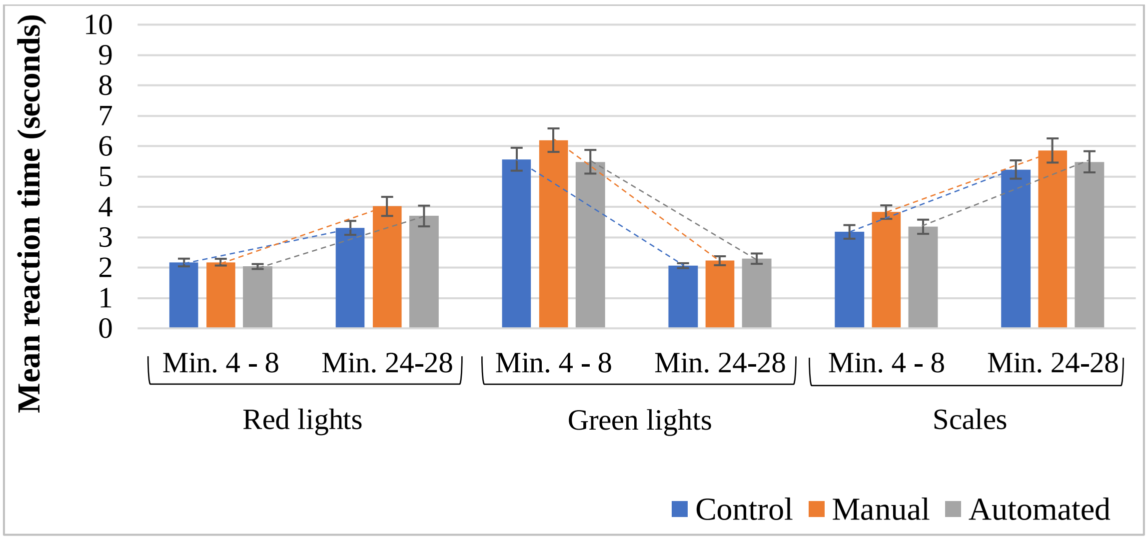


Figure 5.38 The reaction times in the systems monitoring task.

5.4.5 Fatigue in the 14-minute drive

According to hypothesis 5, the control group would be significantly less fatigued than the manual and the automated in the 14-minute drive in the short drive, whereas according to hypothesis 6, the automated group would be significantly less fatigued than the manual in the 14-minute drive. These hypotheses were tested by analysing the CSS and driving performance data.

5.4.5.1 Self-reported fatigue

The medians of the CSS ratings for the three groups during the motorway drive are shown in Figure 5.39. The ratings at the end of the MATB-II tasks are also depicted for comparison purposes. Some of the differences between the measurement time points cannot be seen in this Figure because the use of the medians (instead of the means) resulted in some bars looking the same. However, the significant differences of interest for the hypotheses tested are described in the next paragraph and shown as asterisks and triangles in Figure 5.39.

Wilcoxon signed-rank tests were performed to investigate if the transition from the MATB-II tasks to the motorway drive affected the reported levels of fatigue. No significant difference was found for the control and the manual group. In contrast, the ratings of fatigue just before the motorway drive were significantly lower than at the end of the MATB-II tasks for the automated group, z = -2.121, p = .034, r = .319. An analysis with the nparLD package showed that the interaction between the group and the measurement time point and the simple main effect of the measurement time point were not significant. In contrast, the simple main effect of the group was significant, Wald chi-square (2, N=60) = 4.931, $p < .005$, $\eta_p^2 = .221$. This was further investigated with Kruskal-Wallis tests, which revealed a significant difference between the groups before the motorway drive, $\chi^2(2) = 8.704$, $p = .013$, $\varepsilon^2 = .211$, on minute 4, $\chi^2(2) = 15.322$, $p < .005$, $\varepsilon^2 = .335$, and on minute 9, $\chi^2(2) = 11.486$, $p = .003$, $\varepsilon^2 = .298$. Pairwise comparisons (Dunn's, 1964 procedure with Bonferroni correction) showed that the control group reported significantly lower levels of fatigue before the motorway drive and on minutes 4 and 9 than the manual ($p = .018$, $\eta^2 = .242$, $p = .001$, $\eta^2 = .209$, and $p = .003$ $\eta^2 = .188$, respectively). Moreover, the automated group reported significantly lower levels of fatigue than the manual on minutes 4 ($p = .014$, $\eta^2 = .102$) and 9 ($p = .047$, $\eta^2 = .119$). No significant difference was found between the control and the automated group.

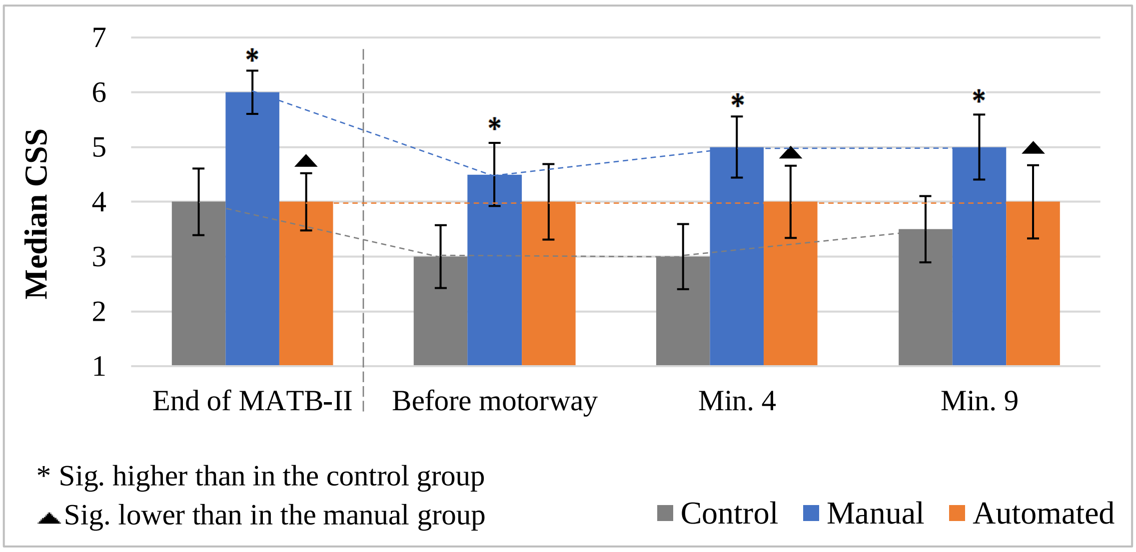


Figure 5.39 The CSS ratings in the motorway drive of experiment 1.

Table 5.4 shows the number of participants who reported high levels of fatigue during the motorway drive. A rating of six means '*extremely tired, very difficult to concentrate*' and seven '*completely exhausted, unable to function properly*'. As can be seen, more participants in the manual group reported being very fatigued than in the control and the automated.

Table 5.4 The frequency of high ratings of fatigue (CSS) in the motorway drive of experiment 1.

	Rating	Just before the motorway drive	Minute 4	Minute 9
Control	**6**	-	-	-
	7	-	-	-
Manual	**6**	5	4	1
	7	-	-	3
Automated	**6**	2	2	2
	7	-	-	-

5.4.5.2 Performance

The mean SD of the lateral position per block of the motorway drive is shown in Figure 5.40. A two-way mixed ANOVA was performed but the assumption of homogeneity of variances was violated. The logarithmic transformation was not successful, so a mixed-model analysis was performed with the nparLD package. The two-way mixed ANOVA showed that simple main effects of the measurement time point and the group were not significant. In contrast, the interaction between the group and the block was significant, Wald chi-square (4, N=60) = 12.659, p = .013, η_p^2 = .315. Figure 5.40 suggests a trend for higher SDs of the lateral position on minutes 10-14 in the manual group.

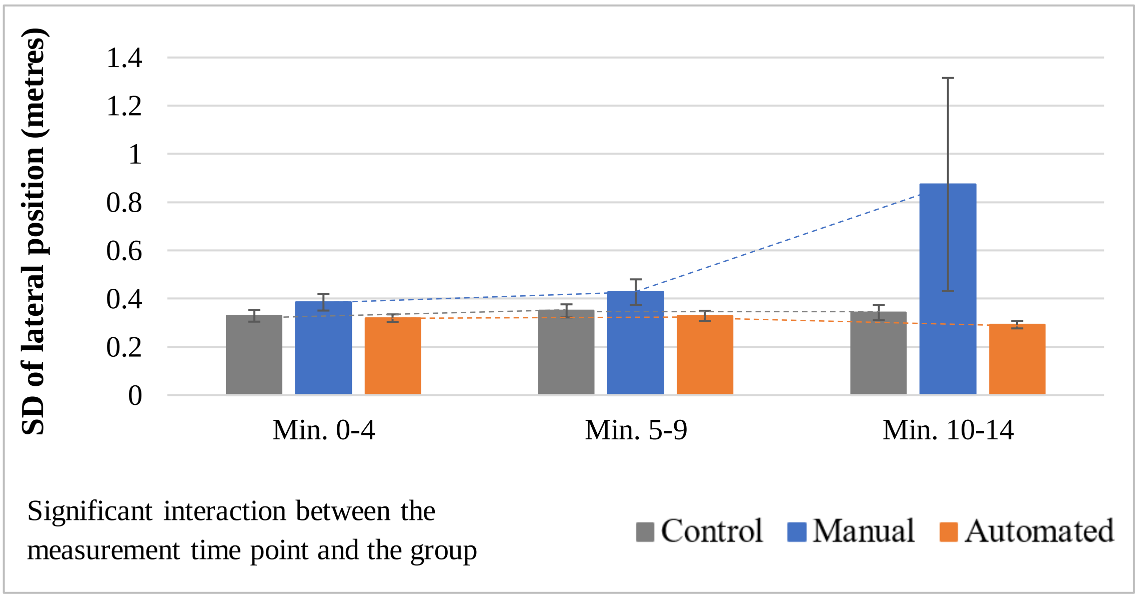


Figure 5.40 The SDs of the lateral position in the motorway drive of experiment 1.

The ability of the individuals to drive on the target speed was evaluated by calculating the SDs of speed. The means of these SDs per block in the motorway drive are shown in Figure 5.41. A two-way mixed ANOVA was performed to analyse the SDs of speed. The assumption of homogeneity of variances was violated. The logarithmic transformation was not successful, so the data were analysed with the nparLD package. The analysis showed that the interaction between the measurement time point and the group and the simple main effect of the measurement time point were not significant. In contrast, the simple main effect of the group was significant, Wald chi-square (2, N=60) = 8.695, p = .012, η_p^2 = .167. This was further explored with Kruskal-Wallis tests, which showed that the distributions of the SDs of speed were statistically significantly different between the groups on minutes 10-14. Subsequently, pairwise comparisons were performed using Dunn's (1964) procedure and a Bonferroni correction for multiple comparisons. The analysis showed a statistically significantly higher SD of speed in the manual group than in the automated (p = .009, η^2 =.127) on minutes 10-14.

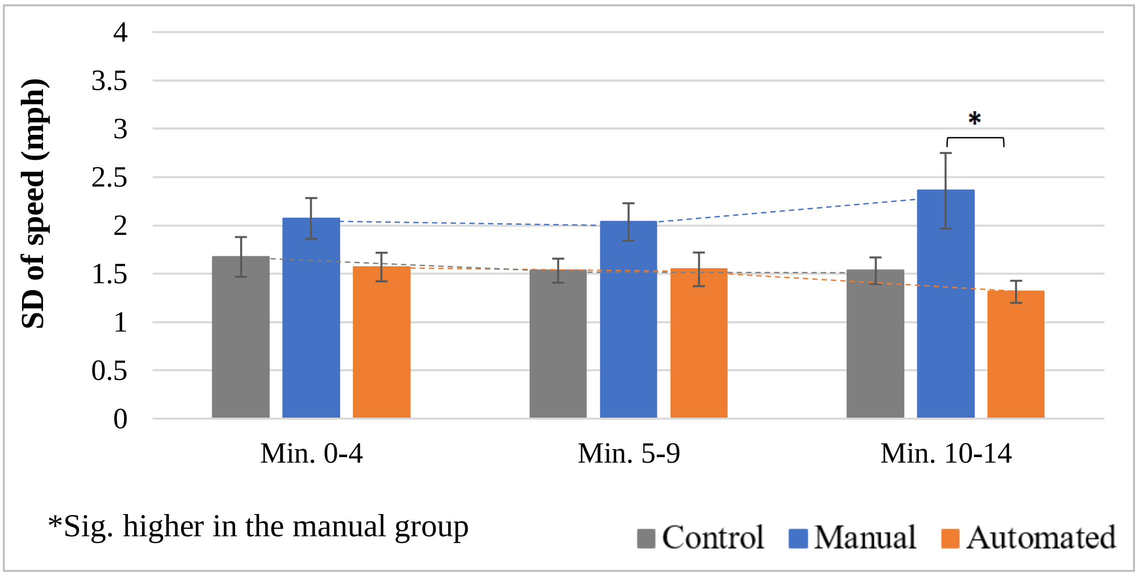


Figure 5.41 The SDs of the speed in the motorway drive of experiment 1.

5.4.6 Summary of main results

60-minute drive:

- The CSS ratings increased significantly in both the manual and the automated group.
- The PVT reaction times after the 60-minute drives were significantly longer than the baselines in the manual and the automated group.
- The perceived total task load and the ratings on each of the NASA-TLX were significantly lower in the automated group than in the manual.

MATB-II tasks:

- The CSS ratings just before and during the MATB-II tasks were significantly higher in the manual than in the control group.
- The CSS ratings just before and after 4 minutes of MATB-II tasks were significantly higher in the automated than in the control group.
- The CSS ratings in the last minutes of the MATB-II tasks were significantly higher in the manual than in the automated group.
- The CSS ratings changed significantly differently over the MATB-II tasks between the groups with a significant increase found only in the manual and the control group.
- The PVT reaction times after the MATB-II tasks were significantly longer than just before them in all the groups.
- The RMSD of the MATB-II tracking in the manual group was significantly lower than in the control group in the first multi-tasking block.
- The number of errors in the MATB-II communications task increased significantly during the MATB-II tasks in the control and the automated group.

14-minute motorway drive:

- The CSS ratings in the motorway drive were significantly higher in the manual than in the control and the automated group.
- The SD of speed was significantly higher in the manual group than in the automated in the last 4 minutes of the motorway drive.

5.5 Discussion

Drawing from studies solely on drivers and pilots, the limited literature on commuting pilots, and documents published by aviation authorities, long and demanding drives to the airport might contribute to increased levels of fatigue and impaired performance in flight (e.g. Dorrian et al., 2008). Flying might then increase pilots' levels of fatigue further, resulting in car crashes after duty. The survey described in chapter 4 suggested that many professional pilots worldwide are at an increased risk of aeroplane and car accidents after completing long and demanding drives to the airport. The study described in this chapter explored if the type of fatigue induced by driving (i.e. active or passive) subsequently affects the levels of fatigue experienced in flying and driving.

This experiment included simulated driving and flying tasks that were completed by 60 non-pilots divided into three groups (i.e. control, manual, and automated). All the groups completed flying tasks with the MATB-II software followed by a 14-minute simulated motorway drive. The manual and the automated group also completed a 60-minute simulated drive before the MATB-II tasks. The long drive completed by the manual and the automated group differed on the level of task load because, drawing from Desmond and Hancock's theory of fatigue, the type of fatigue induced by the drive (i.e. active or passive) could make a difference in the levels of fatigue experienced in the subsequent flying and driving tasks. This was achieved by asking the participants to drive a car of level 0 of automation in the manual group and level 2 in the automated group. Subjective and performance data were collected to infer fatigue.

The aim of this experiment was achieved because it was found that participants' levels of fatigue in the flying and the 14-minute driving tasks differed not only based on whether a preceding long drive was completed, but also based on the type of fatigue that drive induced (i.e. higher levels of fatigue when fatigue was active). The findings of the study described in chapter 5 are discussed in the following sections.

5.5.1 The role of participant age, driving experience, the investigated sleep-related factors, and the baselines of fatigue in fatigue experienced in the trials

5.5.1.1 Participant age

It was explored if any differences found in the expressions of fatigue between the three groups could be attributed to factors other than the completion of one of the two types of the 60-minute drive used. Age was one of these factors because there is evidence that older age can relate to higher levels of fatigue in drivers (McCartt et al., 2009) and pilots (Van Drongelen et al., 2017). Moreover, older individuals may find it more difficult to adjust their circadian rhythms to changes in the wake/sleep cycle (Harma et al., 1994) and their sleep may be shorter and of lower quality than in younger people (Ohayon et al., 2004). As a result, they are more likely to experience sleep-related fatigue. The analysis showed that the groups did not differ significantly in age, so age differences should not have contributed to the significant between-group differences in fatigue found during the simulated flight and motorway drive. It should be noted that it is possible that age affects fatigue differently in commuting pilots in real life. That is, the mean ages in this study were 28 years old for the control group, 30 for the manual, and 26.6 for the automated. These means are approximately 15 years lower than the mean age of airline pilots in Europe (i.e. 45.67 years) (Civil Aviation Authority, 2018). Moreover, only 31% of the airline pilots in the USA are younger than 35 years old (FAA, 2018). Therefore, it might be speculated that older age in real flights could result in pilots experiencing higher levels of sleep-related fatigue in driving and flying than seen in this experiment. Answering this question requires further research.

5.5.1.2 Driving experience

Performance in the motorway drive was used to infer fatigue. Since studies suggest that more experienced drivers are more likely to perform better in driving tasks than younger ones (e.g. Hollopeter, 2011), the participants stated when their driving license had been issued. The means of the years of driving experience were between 6.5 and 10.2 years for the groups, so most of the participants held their driving licenses for a fair amount of time. Moreover, the groups did not differ significantly in the years of driving experience. Therefore, it is believed that driving experience should not have contributed to the between-group differences in the expression of fatigue found in the trials. However, as mentioned, a great percentage of the population of airline pilots is older than the participants in this study. As potentially more experienced drivers, professional pilots' driving performance after flight may be better than that of the participants in this experiment. Future studies could explore if pilots with longer driving experience compensate for fatigue while driving after duty better than the less experienced ones.

5.5.1.3 Sleep duration

Sleep duration the night before the trials was compared between the groups because sleep deprivation impairs driving performance (e.g. Connor et al., 2001). Sleep deprivation can impair several aspects of flying performance too, such as reaction times and threat and error management (e.g. Le Duc et al., 1999). The optimal sleep duration per night may differ between individuals, but generally adults are recommended to sleep at least seven hours every night (Hirshkowitz et al., 2015). This threshold was an eligibility criterion in this study and resulted in mean durations of sleep of more than 7.5 hours in all the groups. These means are very close to the eight-hour guidelines for pilots provided by the FAA (2009). The groups did not differ significantly in the reported sleep duration the night before the trials. Although sleep duration was not measured by collecting physiological data during sleep, studies suggest that subjective estimations of sleep duration show good correlation with actigraphy and polysomnography (Kawada, 2008; Kushida et al., 2001). The effect of the sleep-related factors on fatigue is difficult to control, but it is believed that the between-group differences in fatigue found in the trials are likely not attributable to differences in sleep duration.

The participants reported adequate sleep before the trials, but it is unknown if they had experienced long-term sleep deprivation. If this was the case, sleeping 7 or 8 hours the night prior participation might not have been enough to recover. Moreover, there are trait differences in how individuals respond to sleep deprivation (Rupp et al., 2012; Van Dongen et al., 2004), but these were not captured in this study. Nevertheless, the participants rated their levels of fatigue at the beginning of the trials, completed a baseline PVT, and those who had already been fatigued based on their CSS baseline were excluded. Hence, any severe adverse effect of all the factors that could have contributed to fatigue before participating (the long-term sleep deprivation included) should have been controlled at least to the degree possible.

5.5.1.4 Sleep quality

Sleeping enough does not necessarily mean that people are well-rested when they wake up because sleep quality plays a role too as shown in studies with drivers and pilots (Reis et al., 2016; Stutts et al., 2003). Sleep quality was measured with an item taken from the St. Mary's Hospital Sleep Questionnaire, which asked how well the participants had slept the night before the session. Most than 90% of the participants in all the groups reported '*fairly good*' or even better quality of sleep (Table 5.2). Moreover, the difference in the reported sleep quality between the groups was not significant. However, it is possible that the subjective ratings of sleep quality may sometimes not represent sleep quality accurately enough. In addition, the St. Mary's Hospital Sleep Questionnaire evaluates several aspects of sleep quality, but only one of its items was used. Moreover, due to the focus being on task-induced fatigue, sleep quality data were not collected for a more extended period before the trials. It is likely that the participants might not have recovered from long-term sleep of poor quality by sleeping well the night before participating. Nonetheless, any such adverse effects should have been controlled by asking the participants to rate their levels of fatigue at the beginning of the trials, completing the baseline PVT, and excluding those who were already tired.

5.5.1.5 Baselines of fatigue

The measures of baseline fatigue (i.e. PVT reaction times and lapses and the CSS ratings) did not differ significantly between the groups. Furthermore, all the groups started the trials with low levels of subjective fatigue since all the medians of the baseline CSS ratings were 2 (i.e. '*very lively, responsive, but not at peak*'). Fatigue is a complex concept and individuals differ in how they are affected by the factors contributing to it. However, the absence of significant differences in age, the driving experience, the investigated sleep-related factors of fatigue, and the baselines suggests that any differences in the expressions of fatigue found in the simulated tasks are likely attributable to the tasks completed in the trials.

5.5.2 Fatigue induction by the 60-minute drive

According to hypothesis 1, the CSS ratings, the PVT reaction times, and the number of PVT lapses of the manual and the automated group would be significantly higher at the end than at the beginning of the 60-minute drive the CSS ratings, whereas according to hypothesis 2, the total task load and the ratings on each of the NASA-TLX items during the 60-minute drive would be significantly lower in the automated than in the manual group. The 60-minute drive was designed so that task load would be high when driven on a car of level 0 of automation due to the need to steer, brake, overtake, drive into intersections, pay attention to unpredictable movements of other cars, and make decisions about how to deal with traffic situations (Cantin et al., 2009; Jahn et al., 2005; Richter et al., 1998; Teh et al., 2014). Completing these tasks for an extended period was expected to induce active fatigue by depleting the participants' attentional resources (Desmond and Hancock's model of fatigue). Although the same 60-minute drive was completed by the automated group too, those drivers were not required to manually control the vehicle and make decisions at all times. In contrast, they were only required to continuously monitor the driving scene because they would be asked to take control over at some point (level 2 of automation). The level 2 drive was expected to induce passive fatigue because the required monitoring of the driving scene without manually controlling the vehicle would cause a partial disengagement from the driving task. This disengagement would be evidenced by low total task load and ratings on each of the NASA-TLX items (Matthews and Desmond, 2002; Saxby et al., 2013; Schömig et al., 2015; Vogelpohl et al., 2019).

Hypothesis 1 was tested by analysing the CSS and the PVT data collected during the first section of the trials (i.e. before the MATB-II tasks), whereas hypothesis 2 by analysing the total task load and ratings on the effort item of the NASA-TLX. Task load data were collected during the 60-minute drives with the NASA-TLX scale.

5.5.2.1 NASA-TLX

The hypothesised difference between the manual and the automated group in the type of fatigue induced (hypothesis 2) is supported by the finding that the total NASA-TLX scores and the ratings on each of the NASA-TLX items were significantly higher in the manual than in the automated group throughout the 60-minute drive (Figures 5.26 – 5.32).

All the means of the total task load were moderate in the manual group (higher than 250) and low in the automated (just above 100). These scores suggest that the 60-minute drive induced active fatigue in the former and passive in the latter in line with the study by Saxby et al. (2013) that found that active fatigue related to high and passive with low levels of effort and mental workload (Saxby et al., 2013). The fact that the participants in the automated group were not required to manually control the steering wheel and the pedals and make decisions (e.g. about overtaking or braking) resulted in moderate means of physical and mental demand in the manual group (higher than 40 and 30 respectively) and low to moderate in the automated (lower than 35 and 25 respectively). Although significantly lower, the absolute values of mental and physical demand in the automated group were not much lower than those in the manual. This could be attributed to the fact that, in contrast to higher levels of automation, level 2 required continuous monitoring of the driving scene and keeping the hands on the steering wheel and the feet on the pedals at all times.

The participants were told before the study that the 'mental demand' item referred to how much thinking, deciding, calculating, remembering, looking, and searching they had to do to complete the task. The individuals in the manual group should have referred to all these activities when rating the mental demand item. In contrast, the ratings of the individuals in the automated group should have only referred to looking at and searching in the driving scene because monitoring was the only task they were asked to complete. The amount of attentional resources that were invested by the automated group should have been affected by how much each participant kept engaged with the monitoring task. Irrespective of the level of that engagement, the level 2 drive was less demanding mentally than the manual drive because the drivers were required to monitor the driving scene but not scan it actively for information that would help them to avoid crashing (e.g. cars that passed traffic lights illegally). As a result, the frequency of sampling the visual environment should have been lower than in manual driving. Moreover, the participants in the automated group should have paid attention to the speedometer less often than those in the manual because they were told that the vehicle would always comply with the speed limits. Finally, the drivers in the automated group were not required to complete any mental calculations (e.g. if there was enough time to overtake when there was oncoming traffic) and make decisions (e.g. overtake or wait). As a result, mental demand was lower than in the manual group.

All the means of the ratings of effort were lower than 25 for the automated group and higher than 50 for the manual, where 0 means the least effort someone can invest on a task and 100 the most. This suggests that the automated group engaged less with the driving task than the manual with disengagement being one of the characteristics of passive fatigue according to Desmond and Hancock's model of fatigue (2001). Due to that disengagement, it could be speculated that the participants reduced their visual scanning of the driving scene, which resulted in reserving some attentional resources. It was noticed during the trials that most of the participants in the automated group often did not pay attention to the driving scene, which is a sign of disengagement from the driving task. However, the disengagement of the automated group should have been partial because of using level 2 of automation. That is, the requirement in the automated condition to continuously monitor the driving scene and keep a certain body posture in order to react quickly when taking control over was requested should have resulted in maintaining some engagement with the driving task. This engagement could explain why the absolute difference in the ratings of effort between the manual and the automated group were not big.

The requirement to monitor the driving scene without having any control over the performance of the car could explain why level 2 driving caused some frustration in line with Tops et al.'s (2004) effort-reward imbalance theory. The participants might have tried to combat these feelings by reducing their effort (Korber et al., 2015; Saxby et al., 2008). Nevertheless, fully disengaging from the task was not possible because level 2 of automation required that the participants monitored the driving scene continuously and kept their hands on the steering wheel and their feet on the pedals at all times. As a result, the absolute differences between the manual and the automated group in the ratings of frustration were not big.

5.5.2.2 CSS

To begin with the results of the analysis of the CSS data, the participants were asked to rate aloud their levels of fatigue every 4 minutes while driving. The manual and the automated group started the 60-minute drives with low levels of subjective fatigue (i.e. median = 2: '*very lively, responsive, but not at peak*'), but after 44 and 34 minutes respectively their levels of fatigue became significantly higher than the baselines with a median of 4 ('*a little tired, less than fresh*') for both. The levels of fatigue remained significantly higher than the baselines until the end of the drives, where the median was still 4 for the manual group and 5 ('*moderately tired, let down*') for the automated. Therefore, the analysis of the CSS data suggests that both drives induced fatigue in line with hypothesis 1. The medians of the CSS ratings at the end of the drives were moderate. However, the significant increase of the reported levels of fatigue indicates an increased risk of experiencing fatigue in flying and driving tasks after a long drive irrespective of whether it is demanding or not.

The ratings of fatigue were not significantly different between the groups before and during the 60-minute drives, so both drives induced a similar amount of fatigue and the groups finished the drives on similar levels of fatigue. It is believed that the CSS ratings are an accurate representation of participants' levels of fatigue in the 60-minute drives of this study because drivers are generally good at detecting increased levels of fatigue (e.g. Horne and Baulk, 2004). The absence of a significant difference in the levels of fatigue during the 60-minute drive between the manual and the automated group might be attributed to the use of level 2 of automation instead of higher levels of automation. That is, the participants did not control the vehicle manually, but were required to continuously monitor the driving scene and keep a certain body posture. Therefore, they also experienced some mental and physical fatigue that resulted in reducing the difference in the CSS ratings between the manual and the automated group.

As far as driving of level 0 of automation is concerned, Oron-Gilad and Ronen (2007) found that driving for 45 minutes was fatiguing as evidenced by increased subjective ratings of fatigue. This duration is the same as the 44 minutes found in this study, but Oron-Gilad and Ronen used motorway driving with sections of different task load (i.e. high, moderate, or low). Therefore, fatigue in that case might have been a mix of active and passive rather than only active as in the manual group in this study. Vogelpohl et al. (2019) concluded that fatigue was induced after 30 minutes of driving a car of 0 level of automation, which is less than the 44 minutes found from the analysis of the CSS data in this study. Finally, the time needed for the levels of fatigue to increase as found by the CSS ratings was very close to the 40 minutes found by Thiffault and Bergeron (2003) when they explored fatigue in driving a car of level 0 automation. Subjective fatigue in these studies was not measured with the CSS, but the findings of the experiment described in this chapter suggests that future researchers who explore how quickly fatigue in driving is induced might benefit from using the CSS.

It should be noted that the task load of the drives and the approaches used to measure fatigue in Vogelpohl et al.'s (2019) and Thiffault and Bergeron's (2003) studies differed a lot from the current study because they used motorway driving of low task load. As a result, fatigue in those cases should have been passive, which is known to be elicited sooner than active (as was the case in the manual group). This matches the fact that the participants reached their plateau of fatigue after approximately 30 minutes in Vogelpohl et al.'s study. In Thiffault and Bergeron's study, fatigue was measured subjectively only at the end of the drive (i.e. 40 minutes), whereas there were trends for earlier impairment of driving performance. Therefore, passive fatigue might have been induced sooner than the 40 minutes in their study too.

The significant increases in the manual group of the CSS after 45 minutes of driving suggest a sooner induction of fatigue when driving a car of 0 level of automation than the 80 minutes found by Ting et al. (2008). Ting et al. measured fatigue subjectively only at the beginning and end of the drive and inferred it during the drive from driving performance and reaction times to dots that appeared on the screen. Therefore, the delay in detecting increased levels of fatigue might be attributed in that case to not collecting subjective ratings of fatigue during driving and the fact that the levels of fatigue might not have been high enough to cause performance decrements earlier during driving. In general, drivers can identify increased levels of fatigue before their driving performance deteriorates (Ingre et al., 2006[a,b]; Williamson et al., 2014), so subjective ratings might reveal fatigue induction earlier than performance metrics. Therefore, the 45 minutes found by analysing the CSS data is accepted as a more accurate timing of when the levels of fatigue increased significantly for the manual group.

Simulated driving can induce fatigue sooner than real driving (e.g. Davenne et al., 2012). This considered, if the findings of this study applied to professional pilots, they might not arrive fatigued at the airport after 45 minutes of demanding driving as suggested for individuals in this experiment. It should be noted, however, that there are inter-individual differences in how quickly driving induces fatigue (Nilsson et al., 1997) and the motivation to drive safely (and thus invest more attentional resources) is expected to be higher in pilots who commute to the airport. Therefore, some pilots may be fatigued sooner than others.

According to the CSS data, highly automated driving caused a significant increase in the levels of fatigue after 35 minutes, which is what Vogelpohl et al. (2019) also found. It should be noted that Feldhütter et al. (2018) found that the time needed for the induction of fatigue in highly automated driving varied between 15 and 35 minutes and attributed this variation to individual factors. Since the time needed to induce fatigue may differ between individuals, the focus should be more on the finding that fatigue in this study was induced sooner in the highly automated driving condition rather than on exactly when that happened.

The literature suggests that highly automated driving (levels 2 and 3) induces fatigue sooner than the driving of 0 level of automation (Schömig et al., 2015; Vogelpohl et al., 2019). This was supported in this study since the CSS ratings increased 10 minutes sooner in the automated group. The difference in the speed of fatigue induction between the groups might be attributed to the fact that the drive induced active fatigue in the manual group and passive in the automated. This is supported by the finding that the total task load was rated as intermediate in all the blocks of the drive by the manual group and low by the automated group with the ratings given by the former being more than double those of the latter (Figure 5.26).

5.5.2.3 PVT

The PVT data supported hypothesis 1 since the reaction times after the 60-minute drive were significantly longer than the baselines for both groups. Since the CSS showed that driving induced fatigue, the significant increases of the reaction times indicate that the 5-minute version of the PVT can show increases of the levels of fatigue in line with Loh et al. (2004). The increase of the PVT reaction times due to fatigue agrees with the existing literature in sleep-deprived individuals (Baulk et al., 2006; Dinges et al., 1997). A similar increase of reaction times in the PVT was also found in Graaumans et al.'s study (2010), where the duration of the simulated drive was 30 minutes. Based on the existing literature (Oron-Gilad and Ronen, 2007; Thiffault and Bergeron, 2003) and the CSS results of this study, although simulated drives shorter than 60 minutes may also induce fatigue, future researchers might increase the likelihood of inducing fatigue by using scenarios longer than 40 minutes.

Although some studies have shown a link between PVT reaction times and performance in driving and flying tasks (Baulk et al., 2008; Kosmadopoulos et al., 2017; Lopez et al., 2012), others have not (Baulk et al., 2006; O'Hagan et al., 2019). Therefore, the increased PVT reaction times were accepted in this study as an indication of a potentially higher risk of impaired performance in flying tasks after a long drive rather than a definite link. The absence of a between-groups difference in the PVT reaction times after the 60-minute drive could be attributed to the fact that both groups finished driving on similar levels of fatigue. This is suggested by the absence of a significant between-group difference in the CSS ratings at the end of the 60-minute drive and could be attributed to comparing a drive of level 0 of automation to a level 2 drive. That is, in contrast to higher levels of automation, the level 2 drive induced some mental and physical fatigue due to the requirement to constantly monitor the driving scene and keep the hands on the steering wheel and the feet on the pedals. This induction of fatigue in the automated group should have reduced the difference in the levels of fatigue between the manual and the automated group.

To sum up section 5.5.2, in line with hypothesis 1, the analyses showed that both 60-minute drives induced fatigue as evidenced by increased CSS ratings as the drives progressed and longer PVT reaction times after the drives compared to the baselines. The 60-minute drive induced active fatigue in the manual group and passive in the automated as evidenced by the lower total perceived task load and ratings on each of the NASA-TLX items in the automated group (hypothesis 2 accepted). As shown in the survey described in chapter 4, completing drives to the airport of a duration of 45 to 60 minutes one-way is often amongst pilots. The PVT reaction times increased after 60 minutes of driving and the CSS ratings after 45 minutes. If these findings applied to pilots, many could be at an increased risk of reporting fatigued for duty even after 45 minutes of driving. Nonetheless, it should be noted that driving simulators can induce fatigue sooner than real driving and inter-individual differences affect fatigue induction. Based on the findings described in this section, long drives can induce fatigue irrespective of whether they are completed on a car of level 0 or level 2 of automation. However, whether driving induces active or passive fatigue makes a difference in how quickly the levels of fatigue increase.

5.5.3 The effect of the type of fatigue induced by the 60-minute drive on the levels of fatigue experienced in the flying tasks

According to hypothesis 3, the control group would be significantly less fatigued than the manual and the automated in the MATB-II tasks, whereas hypothesis 4 stated that the automated group would be significantly less fatigued than the manual in the MATB-II tasks. The CSS, PVT, and MATB-II performance data were analysed to test these hypotheses. The levels of fatigue in the control group were expected to be lower than in the others because the 60-minute drives were predicted to induce fatigue due to the extended time on task. The automated group was predicted to be less fatigued than the manual because highly automated driving could help the participants to reserve some attentional resources due to the disengagement from the driving task (Desmond and Hancock, 2001). These resources could be then invested in the flying tasks producing an alerting effect. In contrast, the high task load in the 60-minute drive was expected to deplete the participants' attentional resources. Consequently, the manual group should be more fatigued than the automated one in the flying tasks. Hypotheses 3 and 4 were tested by analysing the CSS data collected after the end of each of the five blocks of the MATB-II tasks (i.e. minutes 4, 8, 24, 28, and 32), the PVT reaction times and lapses before and after the MATB-II tasks, and performance in the MATB-II tasks. The between-group differences found in the self-reported levels of fatigue and the performance in the flying tasks supported hypotheses 3 and 4.

5.5.3.1 CSS

The participants had no breaks between the 60-minute drive and the flying tasks but completed a 5-minute PVT in-between. Mental disengagement from the driving task could have helped the participants to start the MATB-II tasks less fatigued than they finished the long drive (Neri et al., 2002). Nevertheless, the PVT had an alerting effect only for the automated group as evidenced by the significantly lower levels of fatigue just before the MATB-II tasks compared to the end of the 60-minute drive (section 5.4.2.1). This finding suggests that the long drive induced passive fatigue because research has shown that completing secondary tasks in low task load conditions in highly automated driving can have an alerting effect (Gershon et al., 2009; Jarosch et al., 2019; Oron-Gilad et al., 2002; Schömig et al., 2015). In contrast, completing the PVT was not helpful for the manual group because increasing task load is not an effective intervention strategy for active fatigue (Oron-Gilad et al., 2002). The absence of a beneficial effect of the PVT in the manual group highlights the importance of individuals using self-initiated intervention strategies after driving that are effective specifically for the type of fatigue they experience. This considered, the experiment described in chapter 6 will explore the effectiveness of taking a short break between driving and flying because the same 60-minute drive used in the manual group in this study will be used there too.

The analysis of the CSS data showed that the control group reported starting the MATB-II tasks significantly less fatigued (i.e. a median of 2 = '*very lively, responsive, but not at peak*') than the manual and the automated group (i.e. a median of 4 = '*a little tired, less than fresh*') (Figure 5.33). In contrast, the difference in the CSS ratings was not significant just before the flying tasks between the manual and the automated group. As far as the self-reported levels of fatigue during the MATB-II tasks are concerned, the control group reported being less fatigued than the manual until the end of the flying tasks. In contrast, the automated group reported higher levels of fatigue than the control only after the first block of the flying tasks (i.e. minute 4) and, then, that difference disappeared. In addition, the manual group reported higher levels of fatigue than the automated only after the last two blocks of the flying tasks (i.e. minutes 28 and 32).

The significant between-group differences and the significant interaction between the group and the measurement time point in the CSS ratings during the MATB-II tasks confirm hypotheses 3 and 4 by suggesting that not only the time spent on driving can affect fatigue in subsequent flying tasks but also whether that preceding drive induces active or passive fatigue. It is believed that these results show actual differences in the levels of fatigue because previous studies suggested that subjective ratings are a reliable way of detecting increases of the levels of fatigue (e.g. Horne and Baulk, 2004).

The analysis of the CSS data suggests that the participants' levels of fatigue in the control group during the MATB-II tasks first increased after 24 minutes and remained higher until the end of the flying tasks. Despite the induction of fatigue, the participants in the control group generally reported low to moderate levels of fatigue (three = '*okay, somewhat fresh*' or four = '*a little tired, less than fresh*') during the flying tasks with none of them reporting being very fatigued (i.e. six

= '*extremely tired, very difficult to concentrate*' or seven = '*completely exhausted, unable to function properly*') during the flying tasks (Table 5.3). The induction of fatigue due to the completion of flying tasks as evidenced by increased CSS ratings agrees with previous studies with pilots (Gander et al., 2015; Petrilli et al., 2006).

Similar to the control group, the self-reported levels of fatigue of the manual group increased significantly during the flying tasks compared to the ratings provided just before them. However, that increase started sooner than in the control group (i.e. after only 8 minutes in the simulated flight). It is believed that the attentional resources of the participants in the manual group had already reduced before the flying tasks because the 60-minute drive that had induced active fatigue. Therefore, the manual group should have had fewer attentional resources than the control group to resist fatigue induction in the flying tasks. As the manual group invested effort to complete the flying tasks, their levels of fatigue continued to increase. In contrast to the control group, the reported levels of fatigue on minute 24 (i.e. end of the long single-tasking block) were not significantly higher than on minute 8 (i.e. start of that block). This could be an indication that the manual group had already reached near their plateau of fatigue on minute 8; thus, they might have been less sensitive in the induction of further fatigue than the control group. Moreover, approximately one in four participants in the manual group reported being very fatigued after 24 minutes of flying tasks (i.e. a rating of 6 or 7). This can be attributed to the fact that flying added to their already increased by the 60-minute drive levels of fatigue.

As hypothesised, the flying tasks had a very different effect on individuals' levels of fatigue in the automated group compared to the manual and the control ones. The CSS ratings were not significantly different between the manual and the automated group at the end of the 60-minute drive. Therefore, someone might expect that flying would induce fatigue similar to the manual group because of the extended time on the preceding driving task. Nevertheless, the automated group did not report being more fatigued during the flying tasks than just before them. This finding suggests that the flying tasks had an alerting effect on the automated group that counterbalanced the induction of fatigue during the flying tasks.

The higher resilience in fatigue induction that was found in the automated group could be attributed to the fact that highly automated driving could have put the participants into a passive role causing disengagement from the driving task. Although the level 2 drive required some engagement with the driving task (i.e. monitoring and keeping a certain body posture), the individuals should have reserved some attentional resources that they invested in the flying tasks. In contrast to the manual group, that investment made a difference in their performance in line with the 'effort/reward imbalance' theory (Tops et al., 2004). The alerting effect of increasing task load in task load conditions has been supported in studies with driving (e.g. Gershon et al., 2009). This finding extends that literature by suggesting that increasing task load may not only help during but also after drives of low task load. As a result of the alerting effect, fewer participants in the automated group reported very high levels of fatigue in the flying tasks than in the manual group. Although the CSS ratings were significantly higher during the MATB-II tasks in the manual than in the automated group, this between-group difference of the ratings as absolute values might have been bigger if the car driven by the automated group was of level 3 or higher. The reason is that higher levels of automation would reduce the engagement with the driving task more than a level 2 drive resulting in the reservation of more attentional resources.

If the findings of the analysis of the CSS data applied to professional pilots, they would be at an increased risk of aeroplane accidents after drives to the airport that induced active fatigue because higher levels of fatigue in flight relate to a higher likelihood of impaired flying performance (e.g. Morris and Miller, 1996). Based on the CSS data, this increased risk could be evident from the beginning to the end of flights. The manual group generally reported moderate levels of fatigue during the flying tasks, but even those might cause a mismatch between the attentional resources required and those available to maintain flying performance high. This mismatch would be more likely to occur in the demanding stages of flight, such as the takeoff and landing. On the other hand, commuting pilots' attentional resources might suffice to maintain performance in the low task load stages of flight (e.g. cruise), but task-related fatigue could contribute to falling asleep if expressed as sleepiness (Jen et al., 2009).

In commuting pilots, the difference found in this study in the CSS ratings between the manual and the automated group would mean that commuting to the airport as a passenger in a car, by any public means of transport, or by highly automated vehicles in the future might reduce the risk of flying fatigued. Those pilots might still be at a higher risk of impaired flying performance than the ones that would not commute at all just before their duties, for example, because of staying at a hotel near the airport overnight. Nevertheless, they might have more attentional resources to invest in flight than those who completed long and demanding commutes to the airport by driving a car. It should be noted that the duration of the flying tasks in this study was shorter than the duration of many real-world flights. However, the higher resilience to fatigue induction in the MATB-II in the automated than in the manual group suggests that, even in longer flights, experiencing passive (rather than active) fatigue due to driving to the airport might help pilots to resist fatigue induction. If avoiding long and demanding commutes to the airport as drivers was not an option for pilots, then pre-flight self-initiated intervention strategies might be applied. One of these strategies will be tested in the study described in chapter 6.

5.5.3.2 PVT

Hypotheses 3 and 4 were not supported by the analysis of the PVT data since the PVT reaction times and lapses after the MATB-II tasks did not differ significantly between the groups. This finding contradicts the literature that suggests that the PVT reaction times are sensitive to the reductions of vigilance (as an expression of fatigue) (Baulk et al., 2006; Dinges et al., 1997; Loh et al., 2004). An explanation for the absence of significant between-group differences could be that the 5-minute PVT test may be less effective in revealing the effects of fatigue on vigilance than the 10-minute version (Loh et al., 2004). The detection of increased PVT reaction times compared to baselines has been used to infer inductions of fatigue due to long flights (Lopez et al., 2012). However, maybe the between-groups difference in vigilance in the study described in this chapter was not big enough to counterbalance any masking effects of inter-individual differences on simple reaction time (Van Dongen and Belenky, 2009). The absence of a big enough between-group difference in vigilance (and thus PVT performance) after the MATB-II tasks might be attributed to the fact that the level 2 drive did not allow the participants to disengage fully from the driving task. Although disengaging partly from the drive should have helped the individuals to reserve some attentional resources, the requirement to monitor the driving scene and keep a certain body posture induced some mental and physical fatigue respectively.

5.5.3.3 MATB-II performance

It was expected that the manual and the automated group would perform worse than the control in the flying tasks, whereas the manual group would also perform worse than the automated. Nonetheless, no significant between-group differences were found in the systems monitoring and communications tasks. These findings contradict those in previous studies that had found longer response times to the lights and scales sub-tasks and more time-out errors in the communications task with higher levels of fatigue (Caldwell and Ramspott, 1998; Wilson et al., 2006). Maybe the absence of significant differences in performance in the systems monitoring and communications tasks between the automated group and the others resulted from the fact that the automated drive induced some fatigue. Since level 2 driving required paying attention to the driving scene, some attentional resources should have been invested causing mental fatigue. In addition, keeping the hands on the steering wheel and the feet on the pedals should have contributed to physical fatigue in the automated group. As a result, the difference between the automated and the other groups in the levels of fatigue experienced in the MATB-II tasks was not big enough to detect significant differences in performance in these tasks.

The absence of a big enough between-group difference in vigilance (and thus PVT performance) after the MATB-II tasks might be attributed to the fact that level 2 driving did not allow the participants to disengage fully from the driving task. Although disengaging partly from the drive should have helped the individuals to reserve some attentional resources, the requirement to monitor the driving scene and keep a certain body posture induced some mental and physical fatigue respectively.

An explanation for the absence of between-group differences in the systems monitoring and communications task might be that these tasks were completed with the non-dominant hand. Although the participants had completed the practice session, it is possible that slight delays in clicking the correct areas of the screen occurred because of the hand used. Thus, any effect of fatigue on the systems monitoring and communications tasks could have been masked. The absence of significant between-group differences in performance in the systems monitoring and the communications tasks might also be attributed to differences between the individuals in the strategies followed to complete the tasks in the multi-tasking blocks. The participants were briefed that all the MATB-II tasks had the same priority. Nevertheless, it is possible that some prioritisation occurred. Since performance is more likely to be protected in the primary tasks (Hockey et al., 1998), different prioritisations could have masked the effects of any differences in fatigue on performance in the monitoring and the communications tasks. It is believed that the activation of the systems monitoring and the communications tasks could have drawn participants' attention away from the tracking task. That could potentially have happened because of the boredom and the visual strain caused by continuously focusing on the moving ball in the tracking task in the previous blocks. However, exploring task switching behaviour was out of the scope of this thesis, so the participants were not asked if and how they had decided to prioritise certain tasks over others.

The control and the automated group made more mistakes in the communications task of the second block (i.e. minutes 24-28) than in the first one (i.e. minutes 4-8), which suggests that the flying tasks induced fatigue. The increase of the levels of fatigue during the flying tasks in the participants of these groups was suggested by the CSS data. The increase of the number of errors in the communications task of the MATB-II contradicts the findings in Caldwell et al.'s (2004) study, where the frequency of errors in that task did not increase significantly over a period of 37 hours without sleep. In contrast, the result of the study described in this chapter aligns with Caldwell and Ramspott (1998) who had found that increased time-on-task related to more time-out errors in the communications task. In contrast to the control and the automated group and in agreement with Caldwell et al., the number of communication errors did not increase with the time on task in the manual group. This result might be attributed to the fact that the participants in the manual group had already reached their plateau of fatigue in the first multi-tasking block as suggested by the results of the analysis of the CSS data.

It was expected that the manual and the automated group would perform less accurately the tracking task than the control group. Furthermore, the manual group was predicted to be less accurate in the tracking task than the automated group based on previous literature on the effects of fatigue on performance in this task (e.g. Caldwell and Ramspott, 1998; Caldwell et al., 2004; Lopez et al., 2012). In line with that literature and the results of the analysis of the CSS data, the manual group performed worse than the control in the tracking task in the first multi-tasking block (i.e. minutes 4-8). Nonetheless, no other between-group difference was found in the tracking task (Figure 5.35). Therefore, hypothesis 3 was partly confirmed and hypothesis 4 was rejected by the analysis of the tracking performance data. The level 2 drive allowed the participants to disengage from the driving task only partly. Therefore, some fatigue was induced in the automated group too. As a result, the difference in the levels of fatigue in the MATB-II tasks between the automated and the other groups was not big enough to detect significant differences.

Perceiving the tracking task as secondary in the multi-tasking blocks but as primary in the single-tasking ones might explain why a significant between-group difference in the tracking accuracy was found in one of the multi-tasking blocks but not in any of the single-tasking ones. It is possible that the single-tasking blocks were not demanding enough to unmask the effects of fatigue on primary performance (i.e. on the tracking task). In other words, despite any higher levels of fatigue in the manual group during the MATB-II tasks, the participants in that group likely had enough attentional resources available to invest in maintaining tracking performance in the single-tasking blocks. In contrast, the multi-tasking blocks required investing more attentional resources and dividing attention between concurrent tasks. As the attentional resources of the participants in all the groups started to reduce during the MATB-II tasks, they could have selected to focus more on the monitoring and the communications tasks, which were perceived as primary. The control and the automated group might still have enough attentional resources to maintain performance high in concurrent tasks. In contrast, the resources of the manual group might not suffice for all the tasks. Since the tracking task could have been perceived as secondary due to the more engaging nature of the other ones, tracking performance deteriorated in the manual group resulting in the significant between-group difference. An explanation for the absence of other significant between-group differences in the tracking performance might be that the manual group had likely reached a plateau of fatigue earlier in the flying tasks as suggested by the CSS data. If that was the case and the flying tasks continued to induce fatigue in the control and the automated group, the significant between-group difference in the tracking accuracy in minutes 4-8 disappeared with the progress of the MATB-II tasks.

To sum up, hypothesis 3 was confirmed by finding significant between-groups differences during the MATB-II tasks in the CSS ratings and the tracking accuracy, whereas hypothesis 4 was accepted based on the results of the analysis of the CSS data. Moreover, the CSS data, the PVT reaction times, and the MATB-II performance data suggested that the flying tasks induced fatigue. However, the automated group showed a higher resistance in fatigue induction than the others. The results discussed in this section suggest that the levels of fatigue during flying tasks relate to how fatigued individuals start these tasks. At the same time, the risk of being more fatigued in flying tasks increases when a preceding drive induces active rather than passive fatigue because that depletes individuals' attentional resources. If the findings of this study applied to professional pilots, they could be at an increased risk of aeroplane accidents after completing long drives to the airport, especially when those drives induced active fatigue.

5.5.4 The effect of the type of fatigue induced by the 60-minute drive on the levels of fatigue experienced in the 14-minute drive

According to hypothesis 5, the control group would be significantly less fatigued than the manual and the automated in the 14-minute drive in the short drive, whereas according to hypothesis 6, the automated group would be significantly less fatigued than the manual in the 14-minute drive. The control group was expected to be less fatigued than the others due to not completing the 60-minute drive. In contrast, the automated group was expected to be less fatigued than the manual because highly automated driving could have helped the participants to reserve attentional resources. This hypothesis was tested by analysing the CSS data collected just before and twice during that drive and driving performance data in the three 4-minute blocks of the motorway drive. Hypothesis 5 was confirmed by the results of the analysis of the CSS data and hypothesis 6 by the results of the analysis of the driving performance data.

5.5.4.1 CSS

Completing the PVT after the flying tasks helped only the automated group to reduce their reported levels of fatigue just before the motorway drive as evidenced by the significantly lower ratings of fatigue just before the motorway drive than at the end of the MATB-II tasks (section 5.4.4.1). This result agrees with the significant reduction of the CSS ratings of the automated group after the PVT completed when the 60-minute drive finished. In both occasions, it is believed that completing secondary tasks had an alerting effect because the automated group had experienced passive fatigue due to the 60-minute drive. This beneficial effect of increasing task load in low task load conditions has been supported in the literature (e.g. Gershon et al., 2009). The flying tasks likely induced a mix of active and passive fatigue in all the groups, but only the automated group had experienced additional passive fatigue due to the 60-minute drive. Since increasing task load is effective only against passive fatigue (Oron-Gilad et al., 2002), the mental activation of the PVT after the MATB-II tasks benefited only the automated group. Since the manual group was not helped by the PVTs before and after the flying tasks, breaks that will include physical activity will be used in the experiment described in chapter 6. As explained in section 2.3, these could be more effective against active fatigue.

Hypothesis 5 was supported by the higher CSS ratings just before and during the motorway drive in the manual than in the control group (Figure 5.39). Moreover, as expected, the manual group reported higher levels of fatigue than the automated during the same drive. In contrast, the reported levels of fatigue did not differ significantly between the automated and the control group.

This absence of a significant difference might be attributed to the fact that, although the level 2 drive required monitoring the driving scene and keeping a certain body posture, it also allowed the participants to partially disengage from the task because the speed and lateral position of the car were controlled automatically.

The higher CSS ratings during the motorway drive in the manual than the control group suggest that fatigue induced by completing a long drive cannot only affect subsequent flying tasks but also make a difference in fatigue while driving after that flight. This might be attributed to the fact that, despite the induction of fatigue during the flying tasks, the participants in the automated group had likely reserved some attentional resources by disengaging from the 60-minute drive. Having available resources allowed the participants in the automated group to cope with fatigue better than the manual group during the motorway drive. In contrast, the difference in the reported levels of fatigue between the manual and the control group was expected because the participants in the control group had not invested any attentional resources in a preceding long drive.

Although the medians of the ratings of fatigue were low to moderate for all the groups in the motorway drive (Figure 5.39), the between-group differences in the levels of self-reported fatigue indicate differences in the risk of impaired driving performance. Moreover, approximately one in four participants in the manual group reported high levels of fatigue (i.e. a rating of 6 or 7), but none in the control group and only a couple in the automated (Table 5.4). Furthermore, the adverse impacts of active fatigue induced by the 60-minute drive were evident from the beginning of the motorway drive by finding significantly higher reported levels of fatigue in the manual group compared to the other groups on minute 4. If these findings applied to professional pilots, they might be at an increased risk of car crashes even during short drives after duty when they completed long and demanding drives to commute to the airport.

5.5.4.2 Driving Performance

Hypothesis 6 was accepted by the analysis of the driving performance data. Although no significant differences in driving performance were found between the control and the other groups, the manual group maintained the target speed less accurately than the automated at the end of the motorway drive (Figure 5.41). This finding agrees with the significantly higher reported levels of fatigue in the manual group during the motorway drive and the literature, where being more fatigued related to a higher SD of speed (Lenne et al., 2007; Ting et al., 2008). Moreover, it indicates a higher risk of crashing after flying tasks when a preceding long drive induces active rather than passive fatigue. Although this finding supports the expected difference in fatigue between the manual and the automated group, it is also likely that the automated group performed better due to the potentially higher engagement with the motorway driving task mentioned earlier.

Based on the existing literature in driver fatigue, it was predicted that higher levels of fatigue would relate to a higher SD of the lateral position (Ting et al., 2008). In contrast to hypotheses 5 and 6, the SD of the lateral position did not differ significantly between the groups. In regard to the absence of a significant difference in the SD of the lateral position between the manual and the automated group, it might be also attributed to the fact that the level 2 drive required some engagement with the driving task, thus, induced some fatigue too. As a result, the difference in fatigue at the end of the 60-minutte drive between the manual and the automated group was not big enough to contribute to a difference in the SD of the lateral position in the motorway drive.

Despite the absence of a significant between-group difference in the SD of the lateral position, that changed differently between the groups over the drive (Figure 5.40). That is, the ability of the participants to keep the car in the middle of the middle lane was similar during the motorway drive in the control and the automated group but showed a steep increase in the last driving block in the manual group. This finding indicates a lower resilience against fatigue induction for the participants that had completed the 60-minute drive of high task load. For professional pilots, such a lower resilience would mean a higher risk of car crashes after duty.

The significant interaction of the measurement time point with the group in the SD of the lateral position and the significant difference in the SD of speed between the manual and the automated group should be carefully interpreted due to the potential effect of the 'too close' and 'too far' messages that appeared above the dashboard. These messages were shown when the participants' car was outside the pre-determined range from the lead car in order to ensure that the braking event would be triggered the soonest after the end of the drive. The participants were familiar with these messages from the practice drive. However, it is possible that, when these messages appeared on the screen, the participants got distracted from the driving task. In addition, the visual, mental, and physical task load increased because they had to pay additional attention to their headway distance from the lead car, decide what to do to bring the car within the target range, and operate the gas and brake pedals respectively.

The distraction and the increased task load when these messages appeared could have affected the participants' speed and lateral position. Since some participants saw these messages more often than others, that effect could have masked any effect of fatigue on performance differently among the participants. Although this effect could have made a difference in the absolute values of the SD of speed and lateral position, its extent should be similar in all the groups. Therefore, that should not have played a significant role in the results. The lateral position and speed of the car should have been affected less by the CSS and NASA-TLX scales that appeared on the screen because these were accompanied by auditory messages too (thus, they should have been less distracting) and did not require an adjustment of driving behaviour.

To sum up, hypothesis 5 was supported by finding that the manual group reported significantly higher levels of fatigue than the other two groups in the motorway drive and hypothesis 6 by finding that the SD of speed was significantly higher in the manual group than in the automated one.

5.6 Conclusion and implications for the next study

The study described in this chapter suggests that fatigue induced by completing long drives can increase individuals' levels of fatigue in subsequent flying and driving tasks. In addition, this experiment improved understanding of the role that the type of fatigue (i.e. active or passive) induced by long drives plays in fatigue experienced in subsequent flying and driving tasks. In line with Desmond and Hancock's theory of fatigue (2001), this study suggests that increased levels of task load in driving can deplete individuals' attentional resources (i.e. active fatigue) resulting in increased levels of fatigue in subsequent flying and driving tasks. Although a long drive of low levels of task load can also be fatiguing (i.e. passive fatigue), individuals may reserve some attentional resources that can help them to resist fatigue induction in flying and driving tasks that follow. If this finding applied to professional pilots, the risk of accidents would be higher when driving to the airport induced active fatigue. This considered, the study described in chapter 6 explored the effectiveness of a self-initiated intervention strategy for fatigue experienced in the commuting cycle after a long and demanding drive.

Despite having non-pilots as participants, the findings of this study could apply to professional pilots too. This could be particularly the case in real-flight tasks that are not completed automatically due to and, thus, require the investment of many attentional resources. The MATB-II scenario was designed so that it resembled flying tasks that require professional pilots to invest effort (i.e. flying the aeroplane by manually controlling the joystick, monitoring several flight parametres, and communicating with the Air Traffic Control). Despite the equipment and systems used in real aeroplanes for these tasks being more complex than in the MATB-II, the basic principle behind their use is that investing attentional resources is needed. When the available resources do not suffice, performance will start degrading irrespective of whether someone is a pilot or not because this is a limitation of human cognition. Therefore, the increased risk of impaired performance in flying tasks due to fatigue induced by driving that was found in this study could also apply to professional pilots, especially in the phases of flight with high task load.

In contrast, the findings of this study may not apply to professional pilots in those tasks that are completed automatically (e.g. recalling the emergency procedures) because complex production rules are then developed in the long-term memory (Salvucci and Taatgen, 2008) and minimum attentional resources are required. Performance in such tasks is less susceptible to the effects of fatigue than performance in non-automated tasks (Boksem et al., 2005; Lorist et al., 2000; 2005). Moreover, pilot training and experience can help professional pilots to use intervention strategies for fatigue before, during, and after flying. For example, they can sleep or drink coffee. Pilots can also fly the aeroplane by engaging the autopilot, so fewer attentional resources are required in that case. All these factors can contribute to reducing the risk of the fatigue-related performance decrements found in this study.

As mentioned, there are some differences between the participants in this study and professional pilots that could result in pilots being more protected against the effects of fatigue while completing flying tasks under certain conditions. Nevertheless, there are no differences in driver training between professional pilots and non-pilots. Therefore, an induction of fatigue similar to that found in the 60-minute manual drive of this study could also be experienced by pilots who commute to the airport and do not apply intervention strategies during driving or while at the airport. Similar, the adverse effects of fatigue found in the drive after the MATB-II tasks could also be experienced by professional pilots who drive after duty. Moreover, as mentioned in section 5.5.1, some professional pilots are older than the participants in this study, which could result in those pilots experiencing higher levels of sleep-related fatigue in driving and flying than in this study.

The study described in this chapter showed that the methodology used would be appropriate to reveal the effects of a long and demanding drive on individuals' fatigue in the study of chapter 6. This objective was achieved because the long and demanding drive was found fatiguing enough to make a difference in fatigue in the flying and motorway driving tasks. Although the performance metrics used in the trials did not show all the between- and within-group differences found with the CSS, they provided useful insights regarding fatigue. Therefore, a similar combination of metrics will be used in the study described in chapter 6.

A within-subject design in the experiment of chapter 6 could also control for the effect of inter-individual differences in flying and driving performance. Moreover, using flying tasks of higher fidelity in the next study could provide more direct links between active fatigue induced by driving and impaired flying performance. That is, although the MATB-II resembled the main flight tasks, the participants did not have to fly an aeroplane by processing data, such as airspeed and altitude (see Table 6.1). Furthermore, although some effects of fatigue were found on motorway driving performance, that drive was designed with low task load in order to unmask the effects of fatigue by causing sleepiness. In real life, pilots may complete drives after duty with both sections of low and high task load. This considered, the motorway drive in the study of chapter 6 will include both low and high task load sections. Finally, completing the PVTs did not reduce individuals' levels of fatigue in the manual group because most of the fatigue they experienced was probably active due to completing the 60-minute drive and the flying tasks. Therefore, in agreement with the literature on intervention strategies described in section 2.3, the next study will test the effectiveness of short bursts of physical activity in reducing the levels of fatigue experienced in the commuting cycle after a long and demanding drive.

To sum up, this study suggests that;

- Individuals' levels of fatigue during flying and subsequent driving tasks can be higher when a long drive is completed before the flying tasks and, especially, when that drive induces active fatigue.
- The methodology used in this study is generally appropriate for the study described in chapter 6, so a similar methodology will be used there.

Chapter 6. Using short bursts of physical activity to reduce the levels of fatigue experienced in the commuting cycle after a long and demanding drive

6.1 Background and justification of the study

The survey described in chapter 4 suggested that many professional pilots worldwide are at an increased risk of fatigue-related accidents in flight and car crashes after duty when driving to the airport is long and mentally demanding. The study of chapter 5 explored this risk with experimental work and suggested that individuals are more likely to fly and drive fatigued after a long drive, especially when that induces active fatigue. If the findings of the research described in chapters 4 and 5 applied to professional pilots, intervention strategies should be developed to manage the risk of fatigue-related performance decrements in the commuting cycle after long and demanding drives to the airport.

Some of the intervention strategies for fatigue can be applied at an organisational level (e.g. training), but the study described in this chapter focuses on the self-initiated intervention strategies. There is a variety of such strategies; however, not all of them are effective against task-related, active fatigue (see section 2.3), which is of main interest in this thesis. Pilots can use intervention strategies in flight. For example, they can reduce their workflow and extend their breaks while flying to mitigate the adverse effects of the time-on-task (Dawson et al., 2012). They may also sleep in the cockpit or in designated bunk facilities in the aeroplane and walk in the cabin. Nonetheless, using such strategies may not be possible due to high task load (e.g. during descent and when systems fail) or because this is not allowed by regulations (i.e. sleeping). Moreover, even if pilots used an intervention strategy while flying, active fatigue induced by driving to the airport might impact their performance until that point. This considered, the focus of this study is on an intervention strategy for fatigue that pilots could use while at the airport before and after their flights. Using an intervention strategy only before flight might not reduce the risk of car crashes after duty because its beneficial effects could dissipate by then. Hence, using intervention strategies both before and after duties may be needed.

No literature was sourced on how the time spent at the airport affects commuting pilots' levels of fatigue. Nevertheless, the survey described in chapter 4 suggested that the time spent at the airport before flight usually helps pilots to reduce fatigue induced by driving to the airport. It should be noted that the participants were asked about their usual commutes in that study, so it is likely that sometimes pilots do not manage their driving-induced fatigue before flying. Moreover, that survey suggested that many professional pilots usually do not manage to reduce their levels of fatigue while at the airport after their duties. Hence, pilots might benefit from using an intervention strategy for fatigue both before and after their flights. Providing evidence of the effectiveness of the intervention strategy tested in this study might urge airlines, regulators, and pilots to explore how any adverse effects of active fatigue caused by driving to the airport can be managed.

The aim of this study was to explore if short bursts of physical activity can reduce the levels of fatigue experienced in the commuting cycle after a long and demanding drive. This strategy was selected because the literature suggests that the combination of postural change, physical exercise, and mental disengagement can help individuals to reduce their levels of active fatigue (Abd-Elfattah et al., 2015; Anund et al., 2008; Hogervorst et al., 1996; Liang et al., 2009; Loy, 2013; Neri et al., 2002; Phipps-Nelson et al., 2010). In real life, pilots could use this strategy just before starting their flights and drives after duty.

To achieve the aim of this study, 15 participants with a Private Pilot's License participated in two sessions each. In both sessions, they completed the 60-minute simulated drive of experiment 5 by driving a car of level 0 of automation. That was followed by a 52-minute simulated flight and a 20-minute simulated motorway drive. In one of the sessions, the participants took two 6-minute breaks that included postural change, physical exercise, and mental disengagement. The first break was taken just before the simulated flight and the second directly after it. Within-subject comparisons of subjective and behavioural data were performed to infer fatigue.

6.2 Hypotheses

Three alternative hypotheses were formed;

Hypothesis 1. The CSS ratings, the PVT reaction times, and the number of PVT lapses will be significantly higher at the end than at the beginning of the 60-minute drive in both the 'break' and 'no-break' sessions. Testing this hypothesis would help to explore if the long and demanding drive increased individuals' levels of fatigue before investigating whether the breaks tested reduced them. Based on the study of chapter 5, this drive was expected to induce active fatigue.

Hypothesis 2. In the simulated flight, the participants will report significantly lower levels of fatigue and will perform better (i.e. lower RMSDs of heading, altitude, altitude change rate, rate of turn, and airspeed, and fewer errors in the communications task) in the 'break' than the 'no-break' session. This hypothesis was formed to explore if the break taken before the flying tasks would be effective for fatigue in the flying tasks.

Hypothesis 3. In the 20-minute simulated drive, the participants will report significantly lower levels of fatigue and will perform better (i.e. lower SDs of the lateral position, heading, and speed) in the 'break' than the 'no-break' session. This hypothesis was tested to investigate if the break taken after the flying tasks would help the participants to reduce the levels of fatigue experienced in the motorway drive.

6.3 Materials and methods

Ethical approval was obtained from the ESSL, Environment and LUBS (AREA) Faculty Research Ethics Committee of the University of Leeds (ethics reference number: AREA 18-020).

6.3.1 Participants

In contrast to the study of chapter 5, the participants had pilot training for increased realism. Due to the difficulty in recruiting professional pilots, it was decided to recruit participants with a PPL that studied towards a BSc in Aviation Technology with Pilot Studies at the University of Leeds. From the 40 students with a PPL available, 15 agreed to participate. A power analysis with the G*Power software (Heinrich Heine Universität Dusseldorf, 2017) showed that this sample size was adequate for a level of confidence of 90%, a margin of error of 10%, and a 0.8 effect size. The participants should hold a full driver license because driving performance would be measured. Similar to the study of chapter 5, individuals were excluded if they had consumed alcohol, coffee or any other substance that can affect alertness in the 8 hours before the trials or had completed intense physical exercise in the 4 hours before the sessions. Moreover, they could not participate if they had been diagnosed with sleep apnoea, had slept less than 7 hours the night before the sessions, reported average or poor general overall health or stated that a short exercise of medium intensity would harm their health. Finally, individuals were excluded if they rated their levels of fatigue on the CSS (section 3.2.1.1) prior to the sessions as 4 ('*a little tired, less than fresh*') or higher. This decision was made to avoid the participants starting the sessions on a plateau of fatigue. Individuals were approached in one of their classes and those who were interested were sent the participant information sheet. Each participant received £60 after the second session.

6.3.2 Pre-study documents

The participant information sheet included information about the aims of the study, the voluntary participation, and the right to withdraw, and a description of why the participants had been approached. It was also explained that not participating would not have any implications for their course. Moreover, descriptions of the sessions, the participation criteria, the possible advantages and benefits, anonymity, and potential sharing of data were included. Finally, information about who had organised and funded the study and about who had reviewed the participant information sheet was provided. Before starting their first session, the participants read and signed the consent form. That form stated that they had understood the participant information sheet, had the chance to make questions, and knew that participation was voluntary. In addition, they confirmed that they were aware of their right to withdraw their data during or after the sessions and decline to answer any questions. Finally, the participants confirmed their agreement with the policies about data access and storage.

The personal information sheet collected information about participants' age, gender, and years of driving experience in order to describe the sample. In addition, the participants were asked to rate their general overall health from '*excellent*' to '*poor*' and state if they had any medical conditions that could affect their ability to complete a short exercise of medium intensity. Moreover, they confirmed that, to the best of their knowledge, that exercise would not harm their health. Furthermore, the participants were asked if they had received any medications or substances that could affect their alertness in the session. This study focused on task-induced fatigue, so sleep-related fatigue was controlled by collecting information about sleep-related contributors to it. This was achieved with questions about sleep duration the night before participating and the number of hours of being awake since the last sleep. Data were also collected about sleep quality the night before the session by using one item from the St Mary's Hospital Sleep Questionnaire (see section 3.1.2.2), similar to the study of chapter 5. The participants completed two sessions, so the questions about medications, substances, and sleep-related factors were answered twice.

6.3.3 The simulated 60-minute drive

The driving simulator was the one used in the study of chapter 5 (Figure 6.1).

Figure 6.1 The setup of the driving simulator for experiment 2.

The first section of the sessions included the 60-minute drive of level 0 of automation (see Figure 2.2) that was used by the manual group in the experiment of that chapter. This drive was used to induce active fatigue, which was expected based on the findings of the study described in chapter 5. Performance in the 60-minute drive was not measured because the events that were added would affect driving behaviour (i.e. lateral position, speed, headway from lead cars) by urging the participants to steer and brake.

6.3.4 The methods that were used to measure the expressions of fatigue

The methods that were used to measure fatigue in the trials are depicted in Figure 6.2. Similar to the experiment of chapter 5, CSS and PVT baselines were collected. A 1-minute practice preceded the baseline PVT and the participants practised rating their levels of fatigue on the CSS during the practice drive. Fatigue during the trials was measured with CSS ratings and performance in the PVT (PEBL PVT, section 5.3.4.4). The CSS ratings were collected every four minutes during the 60- and 20-minute drives and after each of the three stages of the flight. This scale was shown on the top left corner of the simulator screen during the drives (Figure 5.18) and on the laptop (Figure 6.3) during the flying tasks. Performance in flight (i.e. heading, rate of turn, rate of altitude change, airspeed, and errors in the communications task) and during the driving tasks was measured too. The lateral position, the heading from the lead car, and the speed of the car were measured in the 20-minute motorway drive because the literature suggests that these can be affected by increased levels of fatigue (e.g. Ting et al., 2008). The flight and driving tasks that were used are described in detail in the following sections.

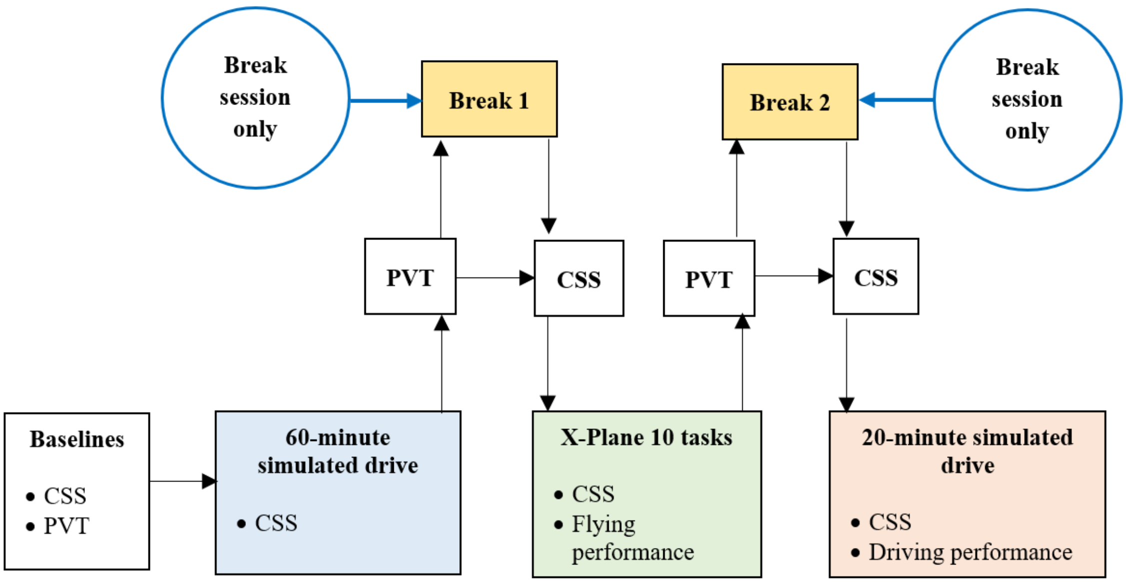


Figure 6.2 The methods used to measure the expressions of fatigue in experiment 2. The CSS was rated directly after the PVT only in the no-break session.

6.3.4.1 The simulated flying tasks

The flying tasks were completed in a simulator for increased safety and convenience. The basic principles behind the tasks completed in the MATB-II software used in chapter 5 are the same to real flying. However, researchers cannot simulate flying tasks that require pilot training, such as reading flight instruments to complete manoeuvres. Using such tasks could help to explore how fatigue induced by driving and flying can affect the completion of manoeuvres. Flight simulators with axial movement are expensive to use and simulators of lower fidelity are a good alternative when interested in pilots' ability to complete flight manoeuvres (De Winter et al., 2012). This considered, the X-Plane 10 professional flight simulation software (Laminar Research, 2014) was used. The X-Plane 10 simulates the visual and flight dynamics accurately and the version used in the trials is certified as a training tool for pilots (United States Department of Transportation, 2018). The software was run on a PC with Windows 8.1 Pro, a 24-inch monitor, a 17-inch laptop monitor, and the Thrustmaster Hotas Warthog joystick and throttle (Figure 6.3). A 24-inch monitor (Figure 6.3, top) was used to display the simulated flying tasks and a laptop monitor (Figure 6.3, left) showed instructions for the tasks that participants should complete (Figure 6.10).

Figure 6.3 The equipment used to complete the X-Plane 10 flight tasks.

Users can fly a variety of aeroplanes on the X-Plane 10. The Cessna 172 (Figure 6.4) was used in this study because this was the aeroplane that most of the participants had flown when they received training for the Private Pilot License. Although some participants had received their training on a Piper Warrior PA-28, the basic instruments of the two aeroplanes (which the participants needed to use to complete the flight manoeuvres in the study) are almost identical (Figure 6.5). Moreover, all the participants who had received their flight training on the PA-28 confirmed that they felt comfortable in flying the Cessna 172 on the X-Plane 10. Commercial aeroplanes provide the same information to the Cessna 172 but integrated into a primary flight display (Figure 6.6).

Figure 6.4 The Cessna 172 (Wikipedia, 2019).

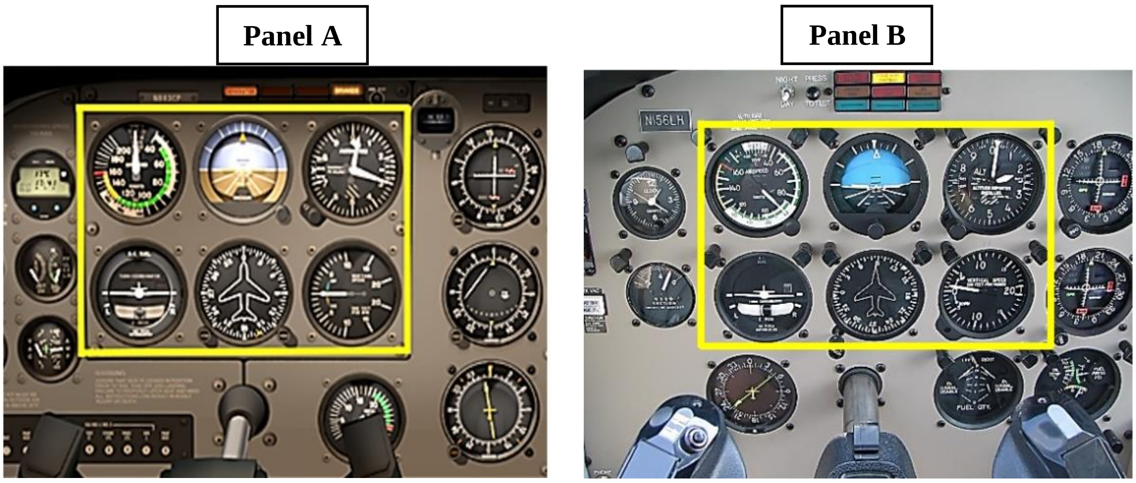


Figure 6.5 Panel A**:** the basic instruments of the Cessna 172. Panel B: the basic instruments of the PA-28 (Laminar Research, 2014).

Figure 6.6 The primary flight display of a Boeing 737 (Wikipedia, 2020)

The duration of the X-Plane 10 scenario was 52 minutes to simulate the duration of an average real-world domestic flight. For increased realism, the levels of task load varied between the blocks of tasks similar to real flights. As mentioned in section 2.1.3, task load is usually higher during the initial and final stages of flight than during cruise. This considered, the flight started with a block, where the participants completed manoeuvres and a communications task with arithmetic calculations. The next block included only straight flying to simulate the low levels of task load of the cruise, whereas the last block was similar to the first one in terms of task load. The taxi, takeoff, and landing stages were not simulated (Figure 6.7) because;

- If participants completed take off, they might not all be at the same altitude with the same airspeed and heading when the instructions to complete the first manoeuvre were given.
- Landing would require providing additional briefing before flight, for example, about navigating to a specific airport. That would be impractical timewise.

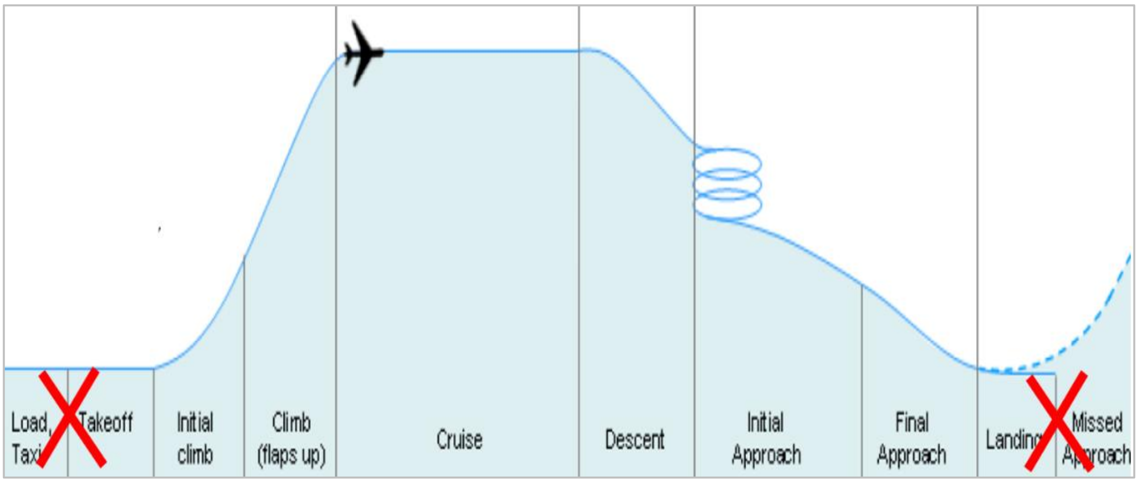


Figure 6.7 The flight stages simulated in experiment 2 (adapted from Archer et al. 2012).

All the sessions started with the aeroplane flying at 6,000 feet (ft), on 80 knots (i.e. unit of speed equal to 1 nautical mile per hour), and a heading of 360°. The heading is the direction of the compass in which the nose of the aeroplane is pointed. The duration of the climb and descent stages was 14 minutes, whereas that of cruise 22 because cruise is usually longer in real flights. Although the autopilot is often used to fly aeroplanes, pilots may control the joystick and the throttle manually (e.g. in system failures). Completing turns, changing altitude and heading, and responding to ATC messages are more likely during the busy stages of flight, such as the climb and descent. Therefore, the cruise stage (called '*stage 2*' in this study) required only flying straight and level. In contrast, the climb (called '*stage 1*') and descent (called '*stage 3*') required changing altitude and heading and completing arithmetic calculations based on ATC messages. All the turns were 360° in this experiment to collect more performance data during the turns. A detailed description of the X-Plane 10 scenario can be found in Appendix C.

The joystick in real flights is adjusted ('trimmed') so that not too many inputs are needed from the pilots. However, the participants flew the aeroplane without any trimming to avoid differences in flight performance due to differences in that. The scenario simulated night flying without any ground lights (Figure 6.8) to reduce distractions. Completing the manoeuvres required using information from six instruments (yellow box, Figure 6.9). A description of these instruments follows in Table 6.1. The participants completed the same manoeuvres in both sessions and heard the same ATC messages. However, the numbers that should be used in the calculations differed between the sessions.

Figure 6.8 The cockpit as it appeared in the study (Laminar Research, 2014).

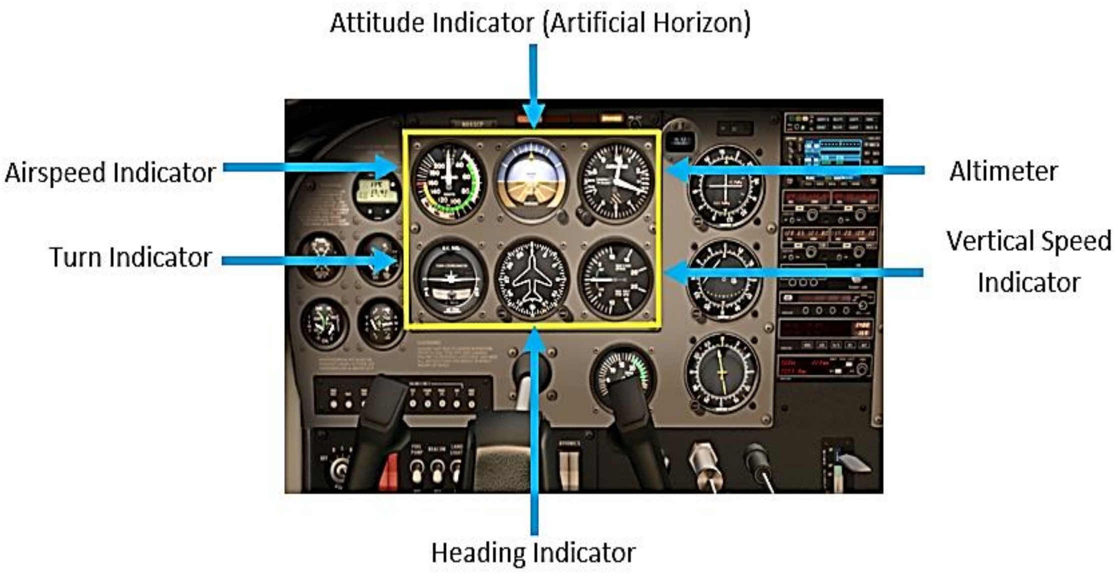


Figure 6.9 The instruments that the participants should use information from in the X-Plane 10 flight (Laminar Research, 2014).

Table 6.1 Description of the flight instruments used to complete the manoeuvres in the X-Plane 10 sessions. Photos taken from Laminar Research (2017).

Attitude Indicator 	The orientation in relation to the horizon	**Altimeter** 	The altitude measured in feet (ft)
Airspeed Indicator 	The airspeed measured in Knots (nautical miles/hour)	**Vertical Speed Indicator** 	The ascend and descend rate in ft/minute
Turn Indicator 	The rate of turn	**Heading Indicator** 	The heading of the aeroplane

The instructions for the manoeuvres and the arithmetic calculations were presented automatically on a laptop. A countdown clock was shown when a new instruction appeared to prepare the participants for the next manoeuvre. When the 10-second countdown finished, a bell rang, and the countdown cycle and the instructions box turned green. Moreover, the word '*Start*' appeared (Figure 6.10). The countdown cycle and the box remained like this until the next instruction appeared on the screen.

The participants had to respond to ATC messages while completing the manoeuvres by making simple arithmetic calculations in their mind and saying the results aloud. In real life, pilots need to process the ATC messages mentally and, sometimes, change the flight parameters accordingly. The ATC messages heard were taken from the MATB-II software (see section 5.3.4.1). However, the participants were not asked to change the radio or the frequency but use the frequency heard to do the arithmetic calculation shown in a yellow box (Figure 6.10). For example, when the frequency heard was '*126.725*' and the yellow box with the phrase '*Add 2*' appeared, the participants had to add 2 to the last digit of the first part of the frequency and say the result aloud (in this example '128.725'). The CSS scale was shown on the laptop screen at the end of stages 1 (i.e. minute 14), 2 (i.e. minute 37), and 3 (i.e. minute 52) and the participants had to rate their levels of fatigue aloud.

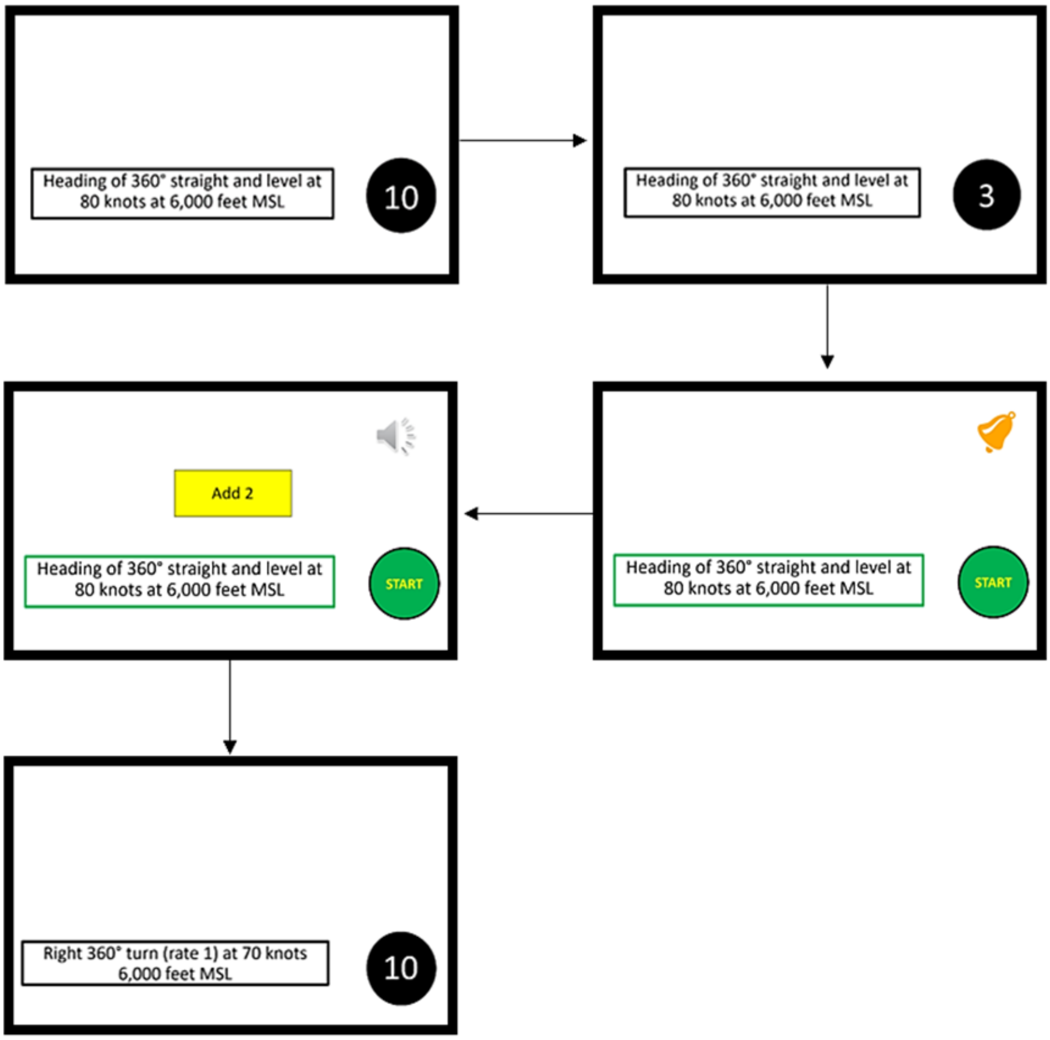


Figure 6.10 An example of the presentation of instructions and the arithmetic calculation task in the X-Plane 10 scenario.

6.3.4.2 The 20-minute motorway drive

Besides flying performance, performance was also measured in the 20-minute simulated drive. This was a drive of level 0 of automation on a 3-lane motorway. In contrast to the motorway drive used in the experiment of chapter 5, no curves were used to facilitate the analysis of the data. In contrast to the study of chapter 5, where the duration of the drive after the flying tasks was 14 minutes, the duration of the motorway drive in this study was increased to 20 minutes. This duration was expected to be long enough to unmask the effects of fatigue on the subjective fatigue and driver performance metrics (Akerstedt et al., 2010; Lenne et al., 1998; Vakulin et al., 2007), but it would also allow using both sections of low and high task load. The 14-minute drive in experiment 1 consisted only of a low task load section to avoid activating the participants mentally. However, pilots' commutes after duty may include a combination of low and high task load sections. If demanding driving after duty activated pilots, it could delay the adverse effects of fatigue. Thus, it could have a protective role against fatigue caused by commuting to the airport and flying especially when the commutes after flight were short. To test that, the first 9 minutes of the motorway drive in this experiment were more demanding than the motorway drive in the study of chapter 5, but the second part of the drive was the same to the one in that chapter. The event that was used in the experiment of chapter 5 with the car moving from the right to the middle lane (Figure 5.17) was used in this drive too. The CSS scale appeared on the top left corner of the screen (Figure 5.18) on minutes 4, 9, 14, and 19 and had to be rated aloud. A description of the two parts of the motorway drive follows;

- **Minutes 00:00 – 08:59**:
  The participants were asked to always drive in the middle of the middle lane while keeping a two-car distance from the lead car. The lead car completed four cycles, each of which included an increase of speed, travelling at a steady speed, and slowing down. The distance travelled by the lead car at steady speed was the same in the second and fourth cycles but longer in the first and third ones. These differences were used to avoid the prediction of the behaviour of the lead car.

- **Minutes 10:00 – 18:59**:

  The lead car always drove in the middle of the middle lane at 70 mph. The participants were asked to drive behind it and as much as possible in the middle of the middle lane at 70 mph. To remind them of the requirement to keep a target distance from it, the message shown in Figure 6.11 was displayed for 50 seconds on minutes 10:05 and 15:05 of the drive.

Figure 6.11 The message about keeping a constant speed from the lead car.

6.3.5 The breaks that were taken before and after the flight

Individuals can use a number of self-initiated intervention strategies for fatigue, such as resting, napping, and drinking coffee. However, as explained in section 2.3, taking short breaks that combine physical activity of medium intensity, mental disengagement, and an upright body posture might be one of the most practical (e.g. sleeping facilities are not needed as is the case with napping) and effective ways of reducing the levels of active fatigue. This considered, the participants in this study took a break after the 60-minute drive and another after the X-Plane 10 flight in one of their sessions (called the '*break session*'). This break was developed for this study based on the literature on the effectiveness of physical exercise, breaks, postural change, and mental disengagement for active fatigue (see section 2.3).

The duration of each break was 6 minutes because completing physical exercise that is not long enough might be ineffective, whereas, if physical activity is completed for too long, it can induce fatigue (Abd-Elfattah et al., 2015). Moreover, taking long breaks might be impractical before and after real flights. For example, taking a 20-minute nap (see section 2.3 for the benefits of napping) would take about 30 minutes or more (the time spent in bed before falling asleep included). However, pilots may not have that much time available because of completing pre-flight tasks. The preliminary study (section 6.3.7) suggested that the 6-minute break helped the individuals to reduce their levels of fatigue. Thus, the 6 minutes was considered an appropriate duration for the break tested. The exercises that were completed during the breaks are shown in Figure 6.12. The walking and stair climbing exercises were completed in the building where the sessions were conducted but outside the room with the simulators. In contrast, the other exercises were completed in the room with the simulators. Walking and stair climbing were selected because this is what many pilots do to go from the terminal to the aeroplane and from the aeroplane to their car. The jog in place and the jumping jacks could be completed by pilots in any room while at the airport.

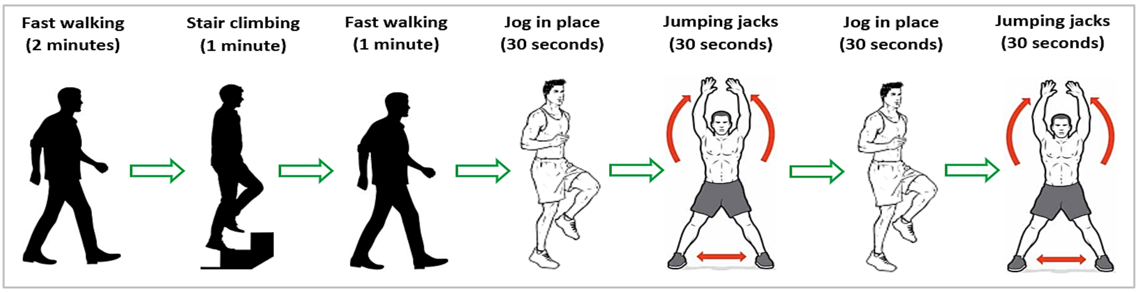


Figure 6.12 The exercises completed during the breaks.

6.3.6 Procedure

The participants visited the researcher twice, once per day in the same week. Eight of the participants were scheduled for sessions starting at 09.00 and seven for sessions starting at 13.00 to control for the effects of the time of the day on fatigue. Both sessions started at the same time for each participant. Eight participants completed the no-break session on their first day of participation and seven on their second day to avoid ordering effects. Based on the counterbalancing plan, when the participants arrived for their first session, they were either told that they would have two or no breaks during that session. Then, they read the participant information sheet, read and signed the consent form, and completed the personal information sheet. They were explained the CSS scale and the driving controls and started the practice drive. The practice drive consisted of the first 5 minutes of the 20-minute drive. That was taken from the version used in the second session for each participant to prevent them from predicting the behaviour of the lead car. The participants were asked to drive in the middle of the middle lane. They were also told that the lead car would decelerate and accelerate and that their task was to modify their speed accordingly to keep a 2-car distance from it.

After the practice drive, the participants were explained what they would have to do in the practice flight. The practice flight consisted of the first 10 minutes of flying of the second session of each participant (Appendix C). The practice included straight flying, one 360° turn, one climb, and arithmetic calculation tasks based on ATC messages. After the practice flight, the participants were given instructions about the PVT and completed the 5-minute version. They were informed that they would drive for 60 minutes on urban and rural roads and would need to rate aloud their levels of fatigue while driving on the CSS. The participants were asked to drive as they would do in real life. After the drive, the participants completed another 5-minute PVT. In the break-session, they took a 6-minute break directly after this PVT. In both sessions, a CSS was completed just before starting the simulated flight. The next phase of the trials included the simulated flight. The participants were told that they would complete a 52-minute flight similar to the practice and that they would have to rate their levels of fatigue on the CSS. After the flights, another PVT was completed. In the break-session, the participants had a 6-minute break directly after this PVT and fatigue was rated again on the CSS just before the 20-minute drive in both sessions.

The last phase of the trials included the motorway drive. The participants were told that they would drive on a motorway for 20 minutes and would need to rate their levels of fatigue on the CSS. They were asked to drive as much as they could in the middle of the middle lane for the whole drive. Moreover, they were informed that the lead car would accelerate and decelerate for the first 9 minutes and they would have to adjust their speed to keep a constant 2-car distance from it. In contrast, they were told that the lead car would constantly travel at 70 mph after the second time they would see the scales on the screen. Their task from that point onwards was to drive as much as they could at 70 mph irrespective of their distance from the lead car. The sessions finished with the end of this drive.

In their second session, the participants completed the Personal Information Sheet again. They were told that the tasks would be the same as in the previous session with the difference that they would (or would not) have breaks. No practice took place and no other instructions were given. The participants completed the second version of the long drive, the same flight, and the second version of the motorway drive. At the end of the second session, they were thanked for participating, they had the chance to make questions about the experiment and received £60 each.

6.3.7 Preliminary study

Three preliminary trials were completed before running the experiment. Two of the participants were students of the University of Leeds with a PPL and one participant was an airline pilot. The students completed the whole session, whereas the airline pilot completed only a part of the simulated flight. The wording of some of the instructions for the manoeuvres was modified based on their feedback. Moreover, the arithmetic calculations were changed so that they were neither too difficult nor too easy to complete while performing the flight manoeuvres. That is, only numbers smaller than 10, additions, and subtractions were used for the calculations. The three participants believed that the flying tasks could be completed by participants with a PPL and found flying without trimming difficult. The participants who completed the whole session found the breaks refreshing and the duration of the practice flight long enough to familiarise with the manoeuvres of the main scenario and the controls.

6.3.8 Methodological considerations

- As mentioned, pilots may use intervention strategies for fatigue while flying but these will not be explored in this study. Therefore, the participants were not allowed to sleep, stand up (unless they did not feel well), and manage the workflow (e.g. by skipping tasks) during the flight.
- In contrast to many real flights, the participants were not allowed to fly the aeroplane by using the autopilot because the aim was to reveal any adverse effects of fatigue on their ability to control the aeroplane manually.
- Pilot training results in completing many flying tasks without the conscious investment of effort (i.e. automaticity due to practised habit) because complex production rules are developed in the long-term memory (Salvucci and Taatgen, 2008). Over-learned, automated performance is less susceptible to the effects of mental fatigue than performance in tasks that require the voluntary allocation of attentional resources (Boksem et al., 2005; Lorist et al., 2000; 2005). Therefore, participants' performance in highly automated tasks (e.g. recall of emergency procedures) was not evaluated. Instead, the tasks used (i.e. manoeuvres and communication task) were expected to require the investment of more attentional resources

than the highly automated ones, thus, should be more susceptible to the effects of fatigue.

- Decision-making was not evaluated to avoid any inter-individual differences in the ability to make decisions masking the effects of fatigue on performance and to keep the flight parameters at the beginning of the manoeuvres similar between the participants.
- As mentioned, having all the participants starting both sessions at the same level of fatigue could not be achieved. Hence, CSS and PVT baselines were collected before each session to identify if there was a between-session difference in the baselines of fatigue.
- Similar to the study described in chapter 5, SD was used to calculate deviation from the pre-determined target values (e.g. speed of 70 mph) in simulated driving. The reason for that decision was that even momentary large deviations in the lateral position and speed of the car can in real life contribute to car accidents. In contrast, RMSDs of flight parameters were calculated because instant large deviations (e.g. of the rate of turn) are less probable to contribute to an accident due to the large separations between aeroplanes.

6.3.9 Statistical analyses

The independent variables of the study were the session (i.e. break and no-break) and the measurement time point during the trials. The dependent variables were the CSS ratings, performance in the X-Plane 10 flight (i.e. RMSDs of heading, altitude, altitude change rate, rate of turn, and airspeed, and errors in the communications task), performance in the motorway drive (i.e. SDs of the lateral position, heading, and speed), and performance in the PVT (i.e. reaction time and lapses).

The analyses were performed with the IBM SPSS Statistics 24.0 (IBM, 2016), but R version 3.5.1 (R-Project, 2018) was also used for non-parametric mixed-method analyses. Mixed-model analyses were run for the metrics that data were collected more than once. Two-way repeated measures ANOVAs were run for the analyses with two within-group factors, one of which dichotomous (i.e. the session) and the other continuous. If a two-way repeated measures ANOVA showed a significant simple main effect of the measurement time point, this was further investigated with one-way repeated measures ANOVAs (to compare between three time points or more) or with paired samples, two-tailed t-tests when only two time points were compared. When the assumption of sphericity was not met in the one-way repeated measures ANOVAs, the degrees of freedom were corrected with the Greenhouse-Geisser estimates of sphericity. On the other hand, if a simple main effect of the session was significant, this was further explored with paired-samples, two-tailed t-tests.

When the assumptions of a two-way repeated measures ANOVA were not met or there were many outliers, the data were transformed. Square root transformations were used when the data were moderately positively skewed and logarithmic transformations when they were strongly positively skewed (Tabachnick and Fidell, 2007). Where the data were transformed, the descriptive statistics of the transformed data will be depicted in the figures. The 95% CIs will be shown in the figures when the data were ordinal and the standard errors when they were continuous. If a transformation was not successful, the analysis was then performed with the ARTool package (Wobbrock et al., 2011). The ARTool facilitates testing for interactions of non-parametric data between two within-group factors by aligning data before averaging ranks. Then, ANOVA procedures can be used. The ARTool package was also used when there were two within-group factors, one of which dichotomous (i.e. the session) and the other ordinal.

If the analysis with the ARTool package showed that a simple main effect of the measurement time point was significant, this was further investigated with Friedman's tests (to compare between three time points or more) or Wilcoxon signed-rank tests (to compare between two time points). Wilcoxon signed-rank tests were also performed if a simple main effect of the session was significant. Paired-samples, two-tailed t-tests were run when there was only one within-groups variable (i.e. break or no-break session) and the dependent variable was measured on the continuous level. In contrast, a Wilcoxon signed-rank test was performed when the session was the only within-groups factor and the dependent variable was ordinal.

The data were analysed separately per stage (i.e. 60-minute drive, flight, and 20-minute drive) in order to identify where any significant interaction between the time point and the session had occurred in the subjective ratings of fatigue. Before testing the hypotheses formed, it was decided to investigate if there were significant differences between the break- and the no-break sessions in factors (besides the 60-minute drives) that could have had an impact on fatigue experienced in them. These factors were the sleep duration and quality of sleep the night before each session, the time awake since the last sleep, and fatigue baselines (i.e. CSS and PVT reaction times and lapses). Paired-samples, two-tailed t-tests were run for the sleep duration, the times awake since the last sleep, and the baselines of the PVT reaction time and lapses. Wilcoxon signed-rank tests were run for the sleep quality (one item from the St. Mary's Hospital Sleep questionnaire - section 3.1.2.2) and the CSS baseline.

According to hypothesis 1, participants' levels of fatigue would be significantly higher at the end than at the beginning of the 60-minute drive in both the 'no-break' and 'break' sessions. This was tested by analysing the data shown in Figure 6.13. Two-way repeated measures ANOVAs were run for the reaction times and the lapses in the PVT before and after the 60-minute drive. Factorial analyses with the ARTool package were performed for the CSS ratings (baseline, during the 60-minute drive, and after the drive).

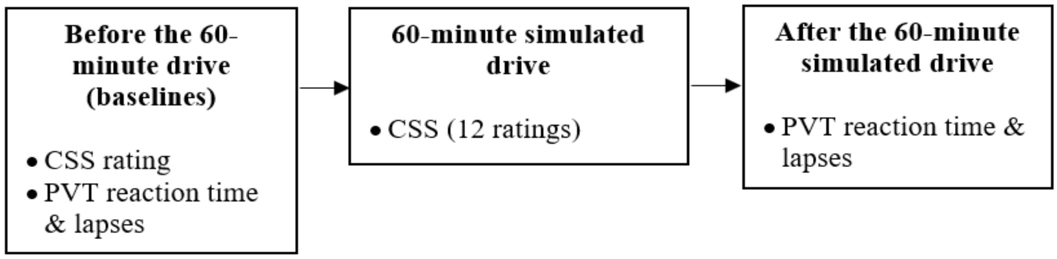


Figure 6.13 The data analysed to test hypothesis 1 in experiment 2.

According to hypothesis 2, in the simulated flight, the participants will report significantly lower levels of fatigue and will perform better in the 'break' than the 'no-break' session. This was tested by analysing the data shown in Figure 6.14.

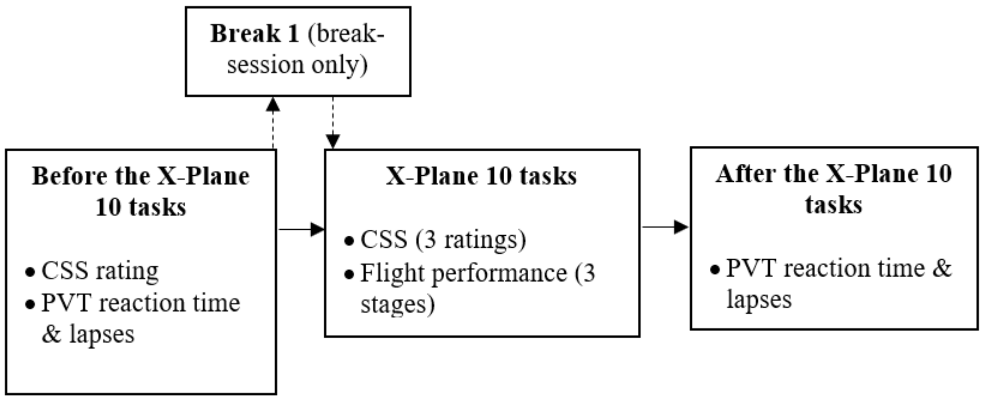


Figure 6.14 The data analysed to test hypothesis 2 in experiment 2.

Two-way repeated measures ANOVAs were performed for the PVT reaction times and lapses before and after the X-Plane 10 flight. Two-way repeated measures ANOVAs were also used to analyse the flight performance data. Factorial analyses with the ARTool package were run for the CSS ratings just before, during, and directly after the simulated flight.

The RMSD of the heading was only calculated for the sections of flight that required straight flight because the participants were asked to change their heading during the manoeuvres that included turns. Separate RMSDs of the heading were calculated for the sections of straight and level flying and those of straight flying with a change of altitude because of the difference of these manoeuvres in task load. One RMSD of the heading was calculated per stage of flight (i.e. stage 1 = climb, stage 2 = cruise, and stage 3 = descent) for the straight and level flying and another for the straight flying with changes of altitude to increase the power of the analysis by reducing the number of the categories. The heading data collected during the straight flying were organised into one category and those during the changes of altitude into another because the difference in task load between these manoeuvres could mask any effects of fatigue on flying performance.

The rate of turn was only calculated for the turns because the participants were instructed to fly straight and level in the rest of the flight. The RMSD of the altitude was not calculated for the climbs and the descents because the participants were instructed to change it. In contrast, the RMSD of the altitude was calculated separately for those periods, when the participants were instructed to fly straight and level or complete turns. This approach was followed to increase the power of the analysis and because the task load differed between these manoeuvres. The RMSD of the airspeed was also calculated separately for the straight and level flying, the changes of altitude, and the turns for increased power of analysis and due to differences in the task load. The RMSD of the altitude change rate was only calculated for the manoeuvres that required changing the altitude (i.e. climb and descent). Finally, the errors in the communications task were counted per stage (i.e. stages 1 and 3), since this task was not active in stage 2.

A Wilcoxon signed-rank test was performed for each session to compare the CSS ratings at the end of the 60-minute drive with those just before the flying tasks. This comparison would show if the first break helped the participants to reduce fatigue induced by the 60-minute drive and, therefore, help to explain any differences in the expressions of fatigue found in the flying tasks.

According to hypothesis 3, in the 20-minute simulated drive, the participants will report significantly lower levels of fatigue and will perform better in the 'break' than the 'no-break' session. The data shown in Figure 6.15 were analysed to test this hypothesis.

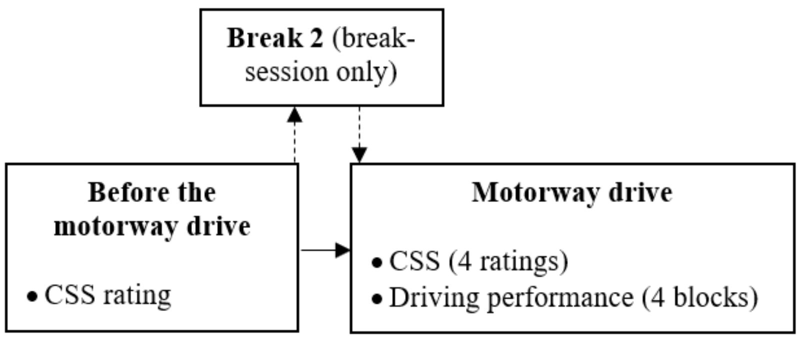


Figure 6.15 The data used to test hypothesis 3 in experiment 2.

The tasks that the participants were asked to complete differed between the first (i.e. minutes 0-9) and the second half (i.e. minutes 10-19) of the 20-minute motorway drive. This considered, a different approach was followed to analyse the data collected in each of the two halves;

- **Minutes 0-9**: The SDs of the lateral position and the headway from the lead car were calculated. The lead car repeatedly decelerated and accelerated (four speed variation events), but the distances travelled by it on a constant speed differed between the events. Therefore, the SDs of the lateral position and speed were compared between events 1 and 3 and separately between events 2 and 4. Moreover, when the lead car accelerated in session 1, it decelerated in session 2. This considered, it was decided not to analyse the data collected during the accelerations and decelerations of the lead car.

- **Minutes 10-19**: the SDs of the lateral position and speed were also calculated on the second half of the motorway drive. However, in contrast to minutes 0-9, no data were removed from the analysis because there were no speed variation events.

Two-way repeated measures ANOVAs were run for the driving performance data that were collected in the 20-minute drive. Factorial analyses with the ARTool package were performed for the CSS ratings just before, during, and directly after the motorway drive. A Wilcoxon signed-rank test was run for each session to compare the CSS ratings at the end of the flight with those just before the motorway drive. This comparison would reveal if the second break reduced participants' levels of fatigue and, thus, help to explain any differences in the expressions of fatigue found in the motorway drive.

6.4 Results

6.4.1 Age, driving experience, sleep-related factors of fatigue, and fatigue baselines

Fifteen males took part in the study. Similar, the percentage of female pilots worldwide is very small (International Society of Women Airline Pilots, 2019). The participants were asked about their age because older drivers and pilots are more likely to experience sleep-related fatigue (McCartt et al., 2009; Van Drongelen et al., 2017) due to the increased difficulty in adjusting their circadian rhythms to changes in the wake/sleep cycle (Harma et al., 1994), shorter sleep, and sleep of lower quality (Ohayon et al. 2004). The mean age was 20.4 years old (SD=1.72), so all the participants were very young. The participants also stated the year their license was issued because there is evidence that more experienced drivers may perform better in driving tasks (Hollopeter, 2011). The mean of the years they held a driver's license was 2.4 (SD=1.24), which suggests that all the participants had short driving experience. Reporting good or excellent health was an inclusion criterion, so 66.6% of the participants reported good and 33.4% excellent general overall health. None of the participants reported having a medical condition that could affect their ability to complete a short exercise of medium physical intensity or that a short physical exercise would harm their health. Moreover, all the participants stated that they had not received any medications or substances that could impact their alertness in the sessions. Therefore, none of the individuals was excluded due to health issues or medications/substances.

Before testing the hypotheses formed, it was decided to investigate if there were significant differences between the no-break and break sessions in sleep-related factors of fatigue and the fatigue baselines. These comparisons would show if any differences in the expressions of fatigue found in the sessions could be attributed to factors other than the effect of the driving and flying tasks completed and the breaks taken. To begin with sleep duration, all the participants reported at least 7 hours of sleep (which was an inclusion criterion). The mean was 7.8 hours (SD=0.61) in the no-break session and 7.6 hours (SD=0.57) in the break session. This suggests that all of them had slept enough before the sessions (Hirshkowitz et al., 2015). A paired-samples t-test showed that the difference in sleep duration between the no-break and the break session was not significant, $t(14) = 1.034$, $p = .319$. In contrast to the experiment in chapter 5, the participants were also asked how much time they had been awake since their last sleep. The data about the time awake were collected because the longer someone is awake, the higher the sleep pressure (and thus sleepiness) (Walker, 2017). The mean of the reported hours of being awake was 4.1 (SD=2.68) in the no-break session and 3.96 (SD=2.37) in the break session, which suggests that sleep pressure was low in both sessions (Walker, 2017). A paired-samples t-test showed that the between-session difference of the time awake was not significant, $t(14) = 0.281$, $p = .783$.

One item from the St. Mary's Hospital Sleep Questionnaire (see section 3.1.2.2) was used to identify if there was a significant difference in the quality of sleep between the no-break and the break session. The percentages of participants that selected each of the sleep quality categories are shown in Figure 6.16. As depicted, all the participants reported that they had slept '*fairly well*', '*well*' or '*very well*', so sleep quality the night before the trials should not have caused sleep-related fatigue. A Wilcoxon signed-rank test did not show a significant difference in the reported sleep quality between the two sessions, $z = 0.250$, $p = .803$.

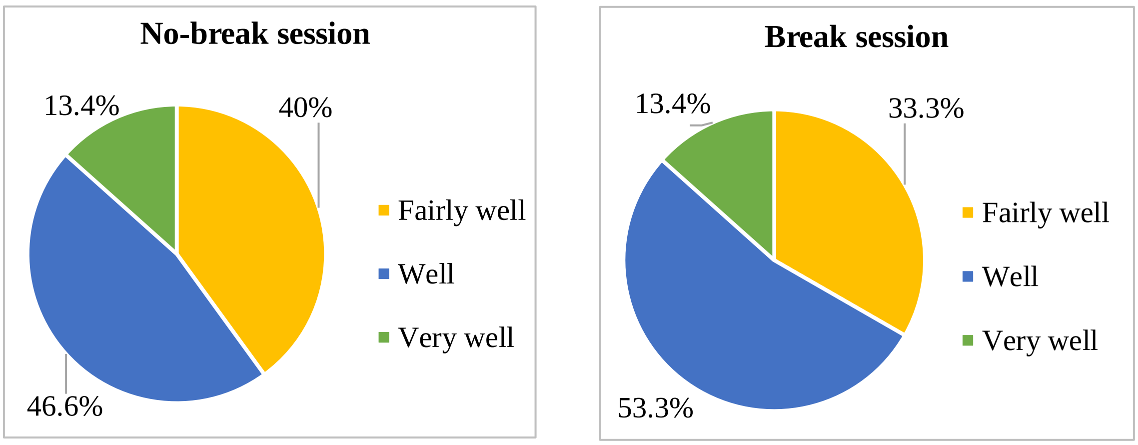


Figure 6.16 The sleep quality the night before the sessions in experiment 2.

The baselines of the PVT were compared between the sessions with paired-samples t-tests. The reaction times were not significantly different between the break (mean = 299.30 milliseconds, SE = 8.47) and the no-break session (mean = 293.31 milliseconds, SE = 5.93), $t(14) = 1.130$, $p = .277$. The difference in the number of lapses (i.e. reaction times > 500 milliseconds) between the break (mean = 0.67 lapses, SE = 0.60) and the no-break session (mean = 0.58 lapses, SE = 0.50) was not significant either, $t(14) = 0.564$, $p = .582$. A Wilcoxon signed-rank test showed that the baseline CSS in the no-break session (median = 2, 95% CI = 0.37) was not significantly different from the one on the break session (median = 2, 95% CI = 0.32), $z = 0.170$, $p = .865$.

To sum up, based on the results described in this section, the participants were young with short driving experience. That is, the effects of sleep-related contributors to fatigue may be intensified in the real-world, where many commuting pilots are older than in this study based on the survey described in chapter 4 (see Table 4.1). At the same time, more experienced drivers (as older pilots are expected to be) may cope better than the participants in this study with the effects of commuting induced fatigue on driving performance after duty. The results in this section also suggest that any differences in the expressions of fatigue found between the break and the no-break session are likely not attributable to sleep deprivation and poor quality of sleep the night before the trials. Based on the results described, the time awake since the last sleep and differences in the levels of fatigue at the beginning of the sessions should not have played a role either. However, fully controlling for the effect of sleep-related factors on fatigue is difficult. This limitation will be explained in the discussion section of this chapter and in chapter 7.

6.4.2 Fatigue in the 60-minute drive

According to hypothesis 1, the CSS ratings, the PVT reaction times, and the number of PVT lapses will be significantly higher at the end than at the beginning of the 60-minute drive in both the ‘break’ and ‘no-break’ sessions.

6.4.2.1 Self-reported fatigue

Figure 6.17 shows the medians of the baseline CSS ratings and the CSS ratings collected during and directly after the 60-minute drives. Some of the differences between the measurement time points are not depicted in this Figure because of the use of the medians (instead of the means). However, the significant differences of interest for the hypotheses tested are described in the next paragraph and shown as asterisks in Figure 6.17.

The analysis with the ARTool package showed that the interaction between the measurement time point and the session and the simple main effect of the session were not significant. In contrast, the simple main effect of the measurement time point was significant, $F(1, 11) = 56.765$, $p < .005$, $\eta_p^2 = .319$. This was further investigated with Friedman’s tests. In the no-break session, the CSS ratings were statistically significantly different between the measurement time points, $\chi^2(12) =$

145.571, $p < .005$, Kendall's coefficient of concordance = .77. The post-hoc analysis revealed a statistically significant increase in the ratings from the baseline (median = 2) to minute 39 (median = 3), $\chi^2(12) = 5.700$, $p = .005$, $r = .086$, onwards until minute 59 (median = 4), $\chi^2(12) = 8.233$, $p < .005$, $r = .442$. In the break session, the ratings were statistically significantly different between the measurement time points too, $\chi^2(11) = 130.438$, $p < .005$, Kendall's coefficient of concordance = .78. The post-hoc analysis revealed a statistically significant increase from the baseline (median = 2) to minute 39 (median = 4), $\chi^2(12) = 5.321$, $p = .023$, $r = .246$, onwards until minute 59 (median = 4), $\chi^2(12) = 8.107$, $p < .005$, $r = .387$.

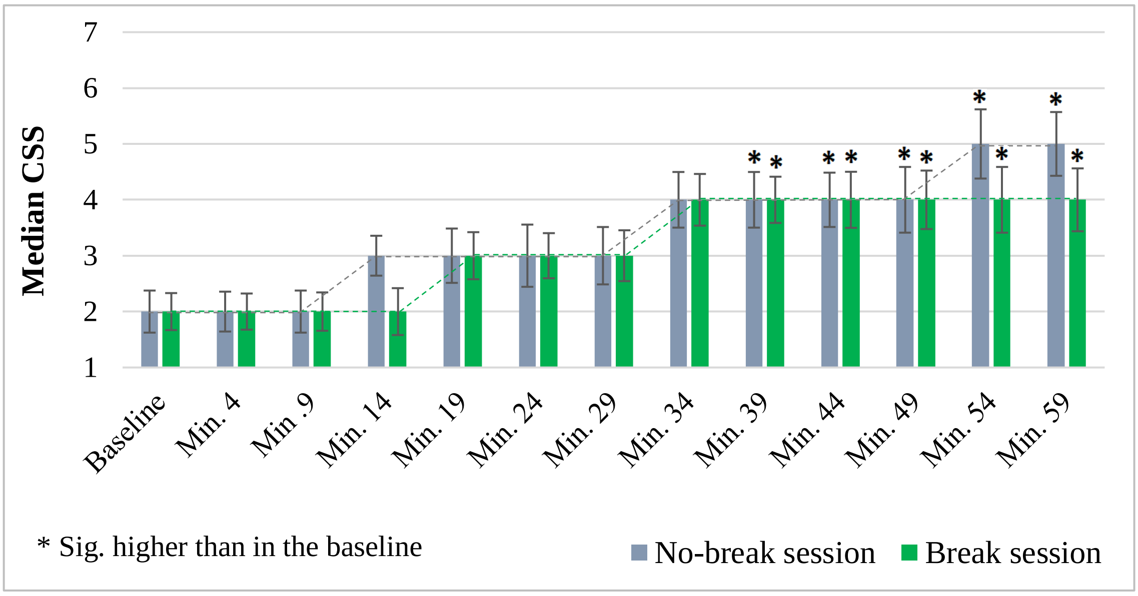


Figure 6.17 The CSS ratings before, during, and after the 60-minute drives (experiment 2).

6.4.2.2 Performance

The mean PVT reaction times before and after the 60-minute drives are shown in panel A of Figure 6.18 A two-way repeated measures ANOVA showed that the interaction between the measurement time point and the session and the simple main effect of the session were not significant. In contrast, the simple main effect of the measurement time point was significant, $F(1,14) = 12.008$, $p = .004$, $\eta^2 = .462$. This was further investigated with paired samples t-tests. In the no-break session, the PVT reaction times increased significantly from the baseline (mean = 299.34 milliseconds, SE = 8.47) to the measurement directly after the 60-minute drive (mean = 356.42, SE = 21.91), $t(14) = 2.927$, $p = .011$, $d = 0.75$. Similar, in the break session, the reaction times increased significantly from the baseline (mean = 293.36, SE = 5.93) to the measurement after the 60-minute drive (mean = 363.33, SE = 19.30), $t(14) = 3.901$, $p = .002$, $d = 1$.

The mean frequencies of lapses before and after the 60-minute drives are shown in panel B of Figure 6.18. The two-way mixed ANOVA showed that the interaction between the measurement time point and the session and the simple main effect of the session were not significant. In contrast, the simple main effect of the measurement time point was significant, $F(1, 38) = 44.564$, $p = .023$, $\eta^2 = .171$. This was further investigated with paired samples t-tests, which were not significant. However, there were trends for significantly more lapses after the 60-minute drive than before it in both the no-break ($p = .053$) and the break sessions ($p = .051$).

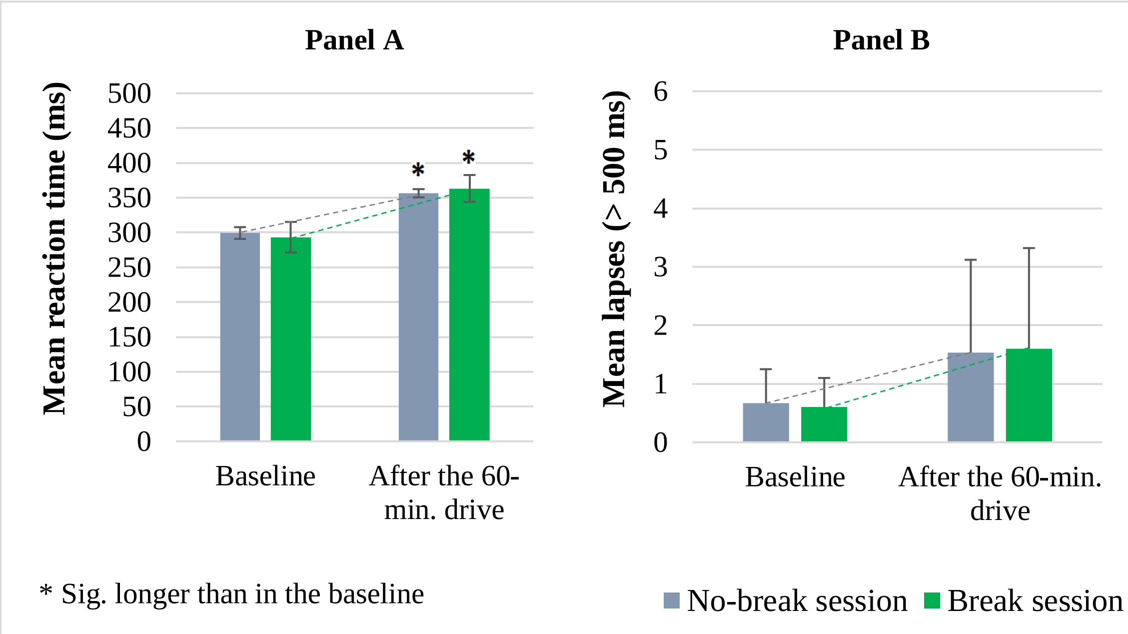


Figure 6.18 The PVT reaction times (panel A) and the frequency of lapses (panel B) before and after the 60-minute drives of experiment 2.

6.4.3 Fatigue in the X-Plane 10 flight

According to hypothesis 2, in the simulated flight, the participants would report significantly lower levels of fatigue and will perform better (i.e. lower RMSDs of heading, altitude, altitude change rate, rate of turn, and airspeed, and fewer errors in the communications task) in the 'break' than the 'no-break' session.

6.4.3.1 Self-reported fatigue

Figure 6.19 shows the medians of the CSS ratings just before and during the flying tasks. The medians of the CSS ratings at the end of the 60-minute drive are depicted too for comparison purposes. In both sessions, the participants completed a PVT between the 60-minute drive and the flight. Moreover, a break was taken after the PVT in the break session. The medians of the

CSS ratings at the end of the drive were compared to those just before the flight for both sessions to investigate if the first break reduced participants' levels of fatigue and if just completing the PVT had an alerting effect. Some of the differences between the measurement time points are not shown in this Figure because the use of the medians (instead of the means) resulted in some bars looking the same. However, the significant differences of interest for the hypotheses tested are described in the next paragraph and shown as asterisks, squares, and triangles in Figure 6.19.

Wilcoxon signed-rank tests revealed that the CSS ratings at the end of the 60-minute drive were not significantly different from those just before flying in the no-break session. In contrast, the participants reported significantly lower levels of fatigue just before the flight compared to the end of the drive in the break session, z = -2.961, p = .003, r = .611.

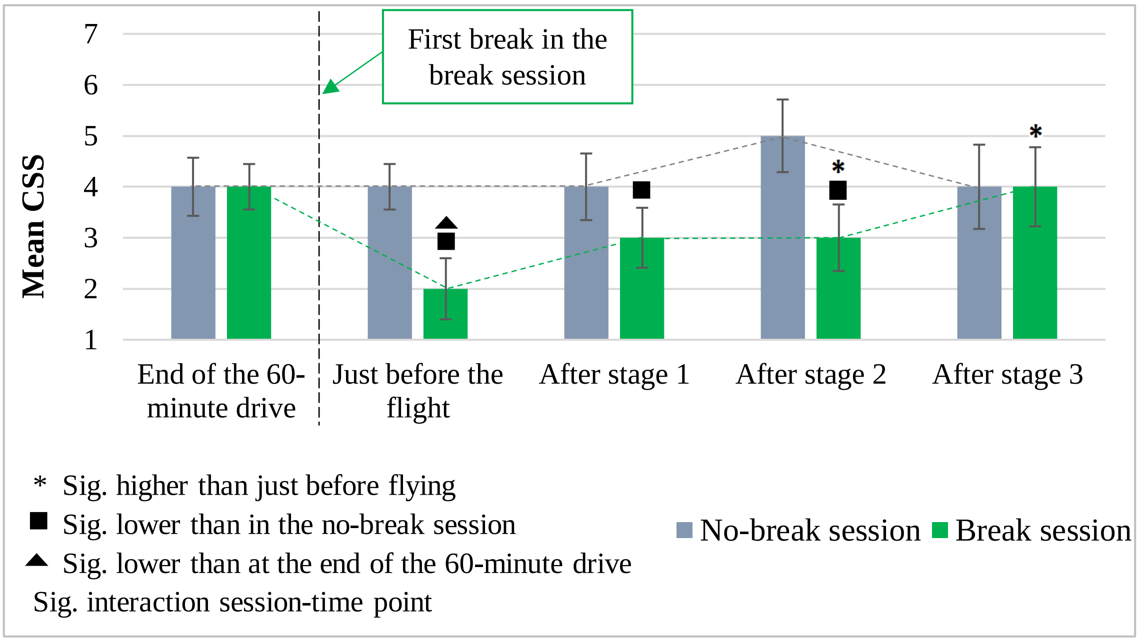


Figure 6.19 The CSS ratings before, during, and directly after the X-Plane 10 flight.

The CSS data collected in the flying tasks were analysed with factorial analysis with the ARTool package. The analysis showed a significant interaction between the session and the stage, F(3, 98) = 6.573, p < .005, η_p^2 = .212, which suggests that the ratings of fatigue changed differently over time between the sessions. The simple main effect of the session was significant, F(1) = 39.178, p < .005, η_p^2 = .335. Pairwise comparisons were performed with Wilcoxon signed-rank tests. The analyses showed that the CSS ratings were significantly higher in the no-break session (compared to the no-break one) just before the flight (z = 2.956, p = .003, r = .415), at the end of stage 1 (i.e. minute 14) (z = 2.714, p = .007, r = .389), and at the end of stage 2 of the flight (i.e. minute 38) (z = 2.714, p = .007, r = .548).

The simple main effect of the stage was significant too, $F(3, 98) = 8.985$, $p < .005$, $\eta_p^2 = .196$. This was further investigated with Friedman's tests. In the no-break session, the CSS ratings were statistically significantly different between the measurement time points, $\chi^2(3) = 12.267$, $p = .007$, Kendall's coefficient of concordance = .27. The post-hoc analysis did not reveal any significant differences between these time points in the no-break session, but a trend for an increase from the end of stage 1 to the end of stage 2 was found ($p = .065$). The reported levels of fatigue were statistically significantly different between the measurement time points in the break session too, $\chi^2(3) = 25.466$, $p < .005$, Kendall's coefficient of concordance = .60. The post-hoc analysis revealed a statistically significant increase in the reported levels of fatigue from just before the flight (median = 2) to the end of stages 2 (median = 5) ($p = .026$) and 3 (median = 4) ($p = .003$).

The number of participants that reported high levels of fatigue in the flights is shown in Table 6.2. A rating of six means '*extremely tired, very difficult to concentrate*', whereas a rating of seven refers to '*completely exhausted, unable to function properly*'. As can be seen, more participants reported being very fatigued after stages 2 and 3 in the no-break than in the break session.

Table 6.2 The frequency of high ratings of fatigue (CSS) in the X-Plane 10 flight.

	Rating	**Just before the flight**	**After stage 1**	**After stage 2**	**After stage 3**
No-break session	**6**	-	1	5	4
	7	-	-	-	1
Break session	**6**	1	1	1	3
	7	-	-	-	-

6.4.3.2 Performance

The means of the PVT reaction times before and after the simulated flight are shown in panel A of Figure 6.20. A two-way repeated measures ANOVA showed that the interaction between the session and the measurement time point and the simple main effect of the session were not significant. In contrast, the simple main effect of the measurement time point was significant, $F(1, 28) = 4.332$, $p = .047$, $\eta_p^2 = .134$. However, further analysis with paired-samples t-tests did not show a significant difference between the measurement time points in any of the sessions. The mean frequencies of the lapses are shown in panel B of Figure 6.20. The two-way mixed ANOVA did not show a significant interaction between the session and the measurement time point. The simple main effects of the measurement time point and the session were not significant either.

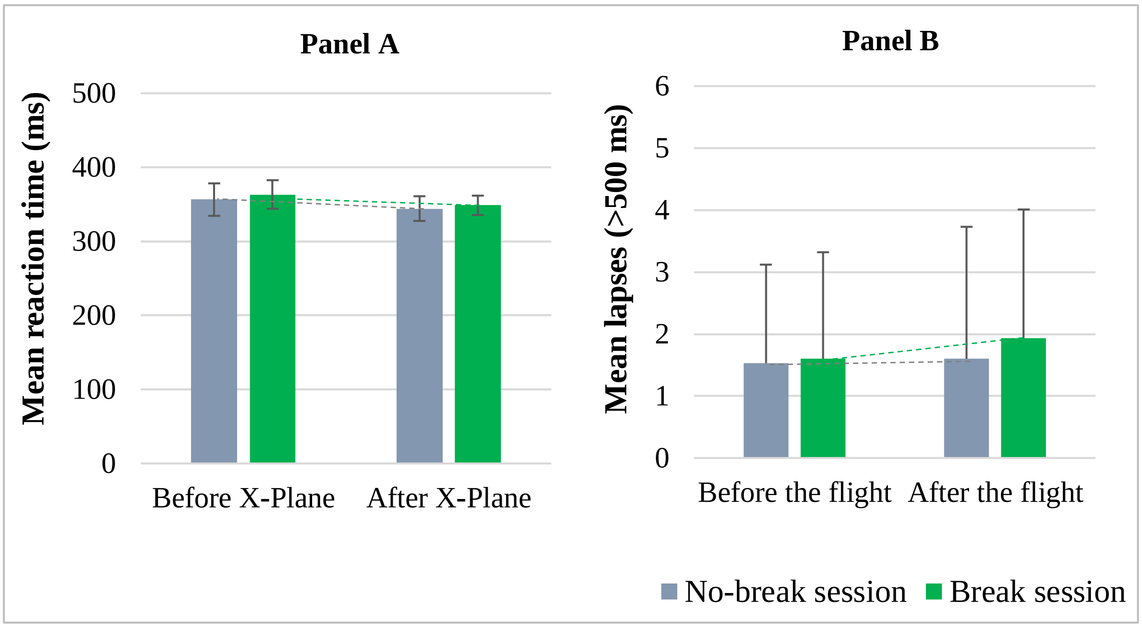


Figure 6.20 The reaction times (panel A) and the frequency of lapses (panel B) in the PVT before and after the X-Plane 10 flight.

Stages 1 and 3 of the flight included three straight flight sections and one section with a change of altitude each. The heading data for the turns were not analysed because of the nature of the manoeuvre. Stage 2 included only straight flying. Figure 6.21 shows the RMSDs of heading for those sections that required straight flying either with or without a change of altitude. A two-way 3x2 repeated measures ANOVA was performed to analyse the data collected in the sections, where the participants were asked to fly straight while maintaining a certain altitude. The analysis showed that the assumption of normality was violated. After a logarithmic transformation, the analysis showed that the interaction between the stage and the session and the simple main effect of the stage were not significant. In contrast, the simple main effect of the session was significant, $F(1, 14) = 4.677$, $p = .048$, $\eta_p^2 = .250$. Pairwise comparisons performed with paired samples t-tests showed that the RMSD of the heading in stage 1 was statistically lower in the break session than on the no-break one, $t(14) = 2.226$, $p = .043$, $d = 0.573$.

A 2x2 repeated measures ANOVA was run to analyse the heading data collected in the sections, where the participants were instructed to fly straight and change their altitude at the same time. The assumption of normality was not met, so a logarithmic transformation was performed. The new 2x2 repeated measures ANOVA showed that the interaction between the stage and the session and the simple main effect of the stage were not significant. In contrast, the simple main effect of the session was significant, $F(1, 14) = 5.718$, $p = .031$, $\eta_p^2 = .290$. Wilcoxon signed-rank tests revealed that the RMSD of the heading was statistically significantly lower in the sections with a change of altitude in the break session compared to the no-break one during stages 1 (i.e.

climb) (z = -2.385, p = .017, r = .633) and 3 (i.e. descent) (z = -2.442, p = .015, r = .565).

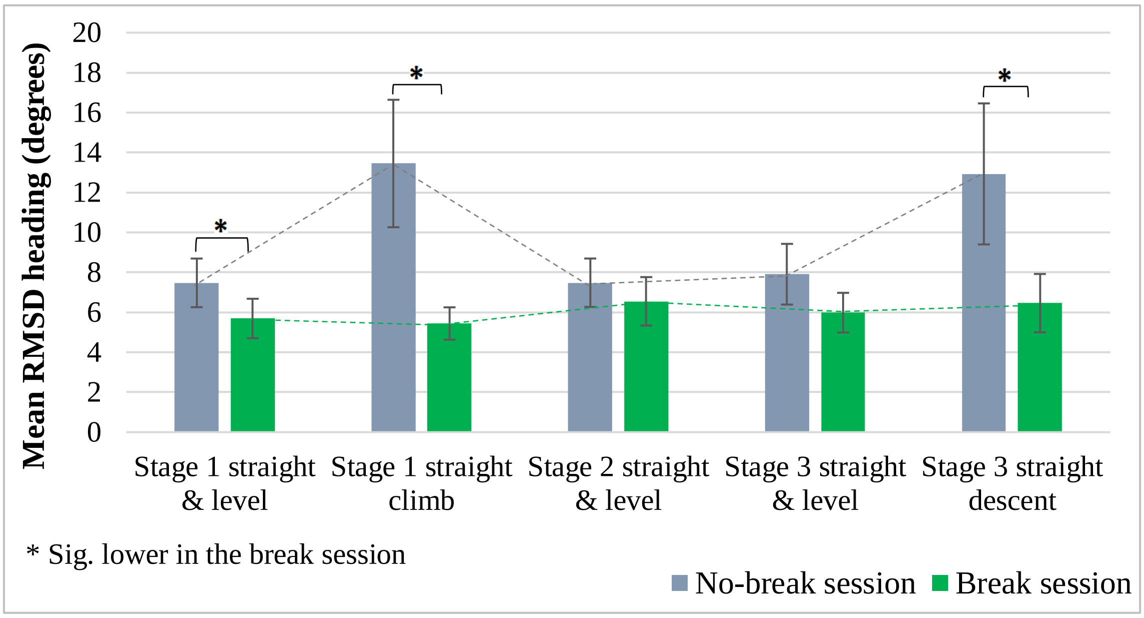


Figure 6.21 The RMSDs of the heading in the X-Plane 10 flight.

Each participant was instructed to complete two 360° turns on each of the stages 1 and 3 on a rate of turn of 3 degrees. The mean RMSDs of the rate of turn in both sessions are shown in Figure 6.22. A 2x2 repeated measures ANOVA showed that the assumption of normality was violated (logarithmic transformation). The new analysis revealed no significant interaction between the session and the stage of flight and no significant simple main effects. Although all the participants completed all the instructed turns, two participants completed one turn each to the opposite direction than instructed in stage 1 of the no- break session (Figure 6.23).

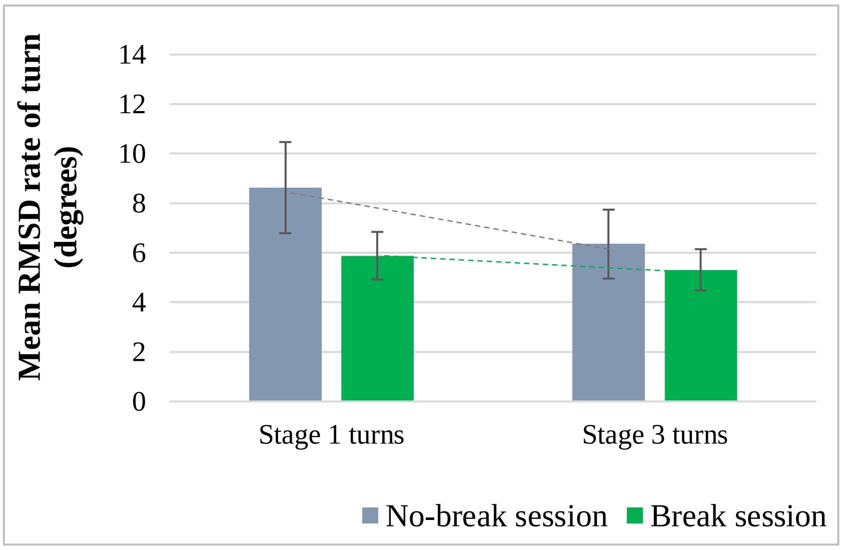


Figure 6.22 The RMSDs of the rate of turn in the X-Plane 10 flight.

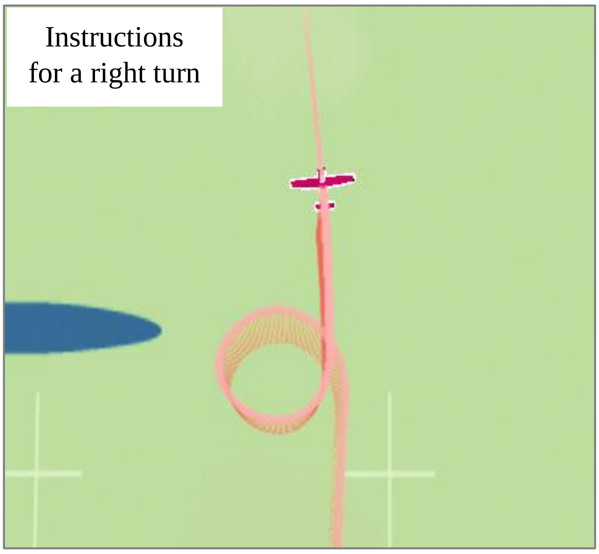


Figure 6.23 The flight path of a participant that turned to the wrong direction in the no-break session of the X-Plane 10 flight (Tacview, 2020).

Both the RMSDs of the altitude change rate and altitude were analysed because some manoeuvres required maintaining a target altitude and others climbing or descending at a target change rate. The mean RMSDs of the altitude change rate are shown in Figure 6.24. The target altitude change rate was 500ft. A 3x2 two-way repeated measures ANOVA showed that the assumption of normality was violated. The logarithmic transformation was not successful, so a non-parametric factorial analysis was performed with the ARTool package. The analysis showed that the interaction between the session and the stage and the simple main effect of the stage were not significant. In contrast, the simple main effect of the session was significant, $F(1, 42) = 4.798$, $p = .034$, $\eta_p^2 = .038$. However, the pairwise comparisons with Wilcoxon signed-rank tests did not show significant differences.

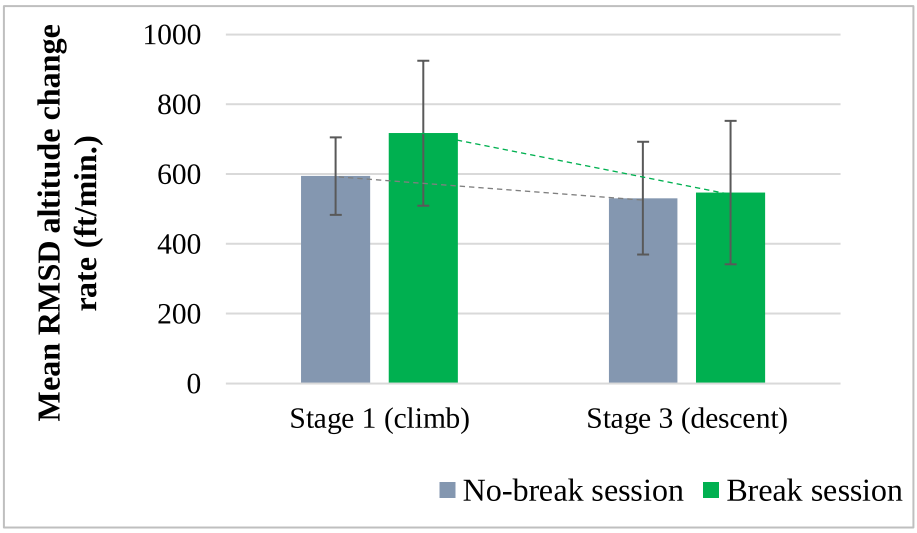


Figure 6.24 The RMSDs of the altitude change rate in the X-Plane 10 flight.

The altitude data were not analysed for the climb and the descent because of the nature of the manoeuvres. As mentioned in section 6.3.9, the RMSDs of the altitude were calculated separately for the sections that included straight flying and those with turns. The means of the RMSDs of altitude are shown in Figure 6.25. For the altitude data collected in the straight flying sections, a 3x2 two-way repeated measures ANOVA showed that the assumption of normality was violated. The analysis after a logarithmic transformation showed no significant interaction between the session and the stage of the flight and no significant simple main effects. The RMSDs of the altitude during the turns were analysed with a 2x2 two-way repeated measures ANOVA, but the assumption of normality was not met (square root transformation). A new two-way repeated measures ANOVA showed no significant interaction between the session and the stage of the flight and no significant simple main effects.

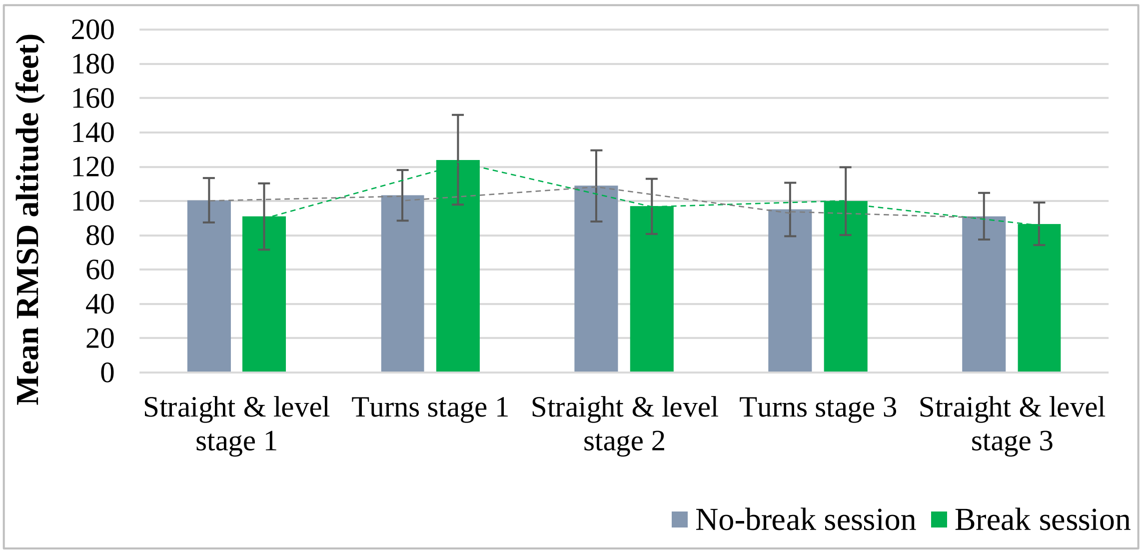


Figure 6.25 The RMSDs of the altitude in the X-Plane 10 flight.

The RMSDs of the airspeed were calculated separately for the sections with straight and level flying, the manoeuvres that required a change of altitude (i.e. climb or descent), and the turns. The means of the RMSDs of airspeed are shown in Figure 6.26. A 3x2 two-way repeated measures ANOVA of the airspeed data collected in the sections of flight that required straight and level flying showed that the assumption of normality was violated. After a square root transformation, a two-way repeated measures ANOVA revealed that the interaction between the session and the stage and the simple main effect of the stage were not significant. In contrast, the simple main effect of the session was significant, $F(1, 14) = 7.130$, $p = .018$, $\eta_p^2 = .337$. Pairwise comparisons with paired samples t-tests showed that the participants achieved higher accuracy in terms of airspeed during straight flying in stage 1 in the break session compared to the no-break one, $t(14) = 2.418$, $p = .03$, $d = 0.623$.

The RMSDs of the airspeed when changing altitude were analysed with a 2x2 two-way repeated measures ANOVA that showed that the assumption of normality was violated. The data were moderately, positively skewed, so a square root transformation was performed. The new two-way repeated measures ANOVA showed that the interaction between the session and the stage and the simple main effects were not significant. A 2x2 two-way repeated measures ANOVA was performed to analyse the RMSDs of the airspeed during the turns. The assumption of normality was not met, so a square root transformation was performed. The new two-way repeated measures ANOVA showed that the interaction between the session and the stage and the simple main effect of the stage were not significant. In contrast, the simple main effect of the session was significant, $F(1, 14) = 7.703$, $p = .015$, $\eta_p^2 = .355$. Pairwise comparisons with Wilcoxon signed-rank tests showed that the participants maintained the target airspeed in stage 1 more accurately in the break session than in the no-break one, $z = -2.215$, $p = .027$, $r = .220$.

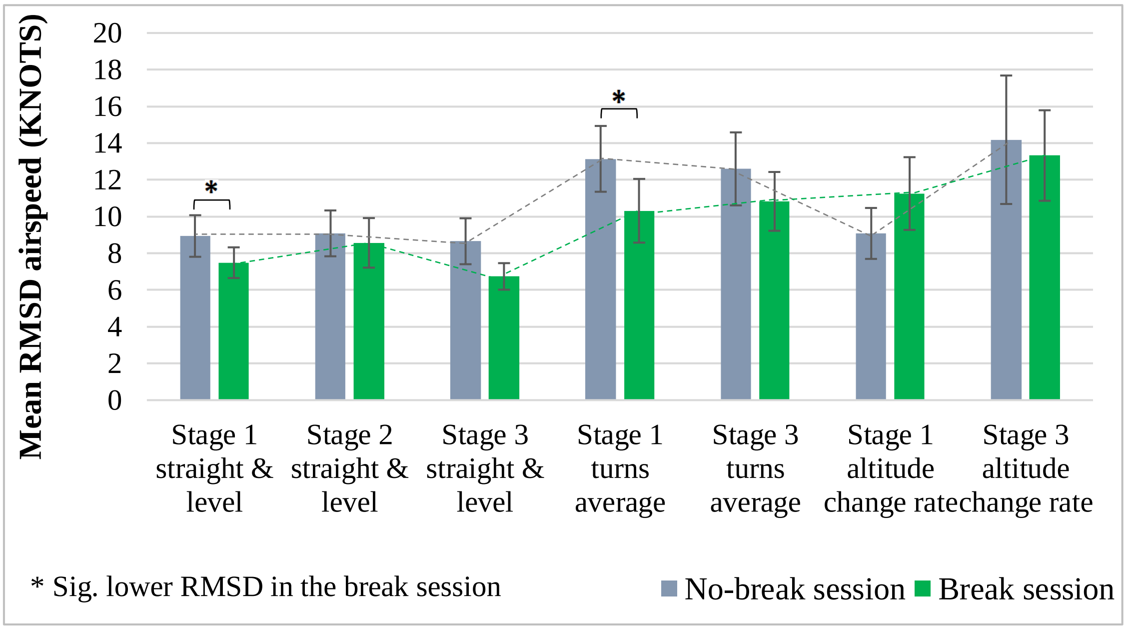


Figure 6.26 The RMSDs of airspeed in the X-Plane 10 flight.

The means of the frequencies of errors in the communications task are shown in Figure 6.27. A 2x2 two-way repeated measures ANOVA showed that the interaction between the session and the stage of the flight was not significant. In contrast, the simple main effect of the stage was significant, $F(1, 14) = 4.750$, $p = .047$, $\eta_p^2 = .253$. Pairwise comparisons with t-tests showed statistically significantly more errors during stage 3 compared to stage 1 only in the break session. The simple main effect of the session was significant too, $F(1, 14) = 23.500$, $p < .005$, $\eta_p^2 = .627$. Pairwise comparisons with paired samples t-tests showed significantly more errors in stage 1, $t(14) = 3.690$, $p = .002$, $d = 0.953$, and stage 3, $t(14) = 4.012$, $p = .001$, $d = 1.035$, when the participants did not have a break before the flight than when they had.

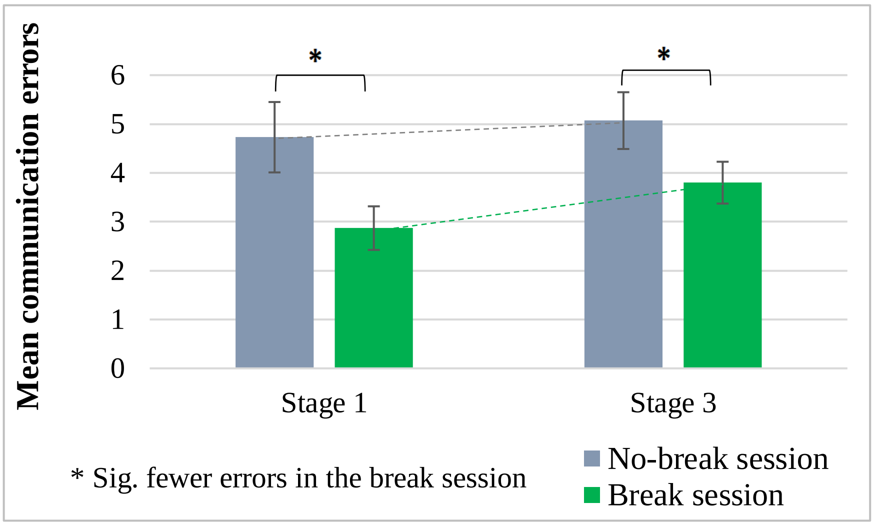


Figure 6.27 The frequencies of errors in the communications task of the X-Plane 10 flight.

6.4.4 Fatigue in the 20-minute drive

According to hypothesis 3, in the 20-minute simulated drive, the participants will report significantly lower levels of fatigue and will perform better (i.e. lower SDs of the lateral position, heading, and speed) in the 'break' than the 'no-break' session.

6.4.4.1 Self-reported fatigue

Figure 6.28 depicts the medians of the CSS ratings just before and during the motorway drive. The CSS ratings at the end of the flying tasks are shown too for comparison purposes. Some of the differences between the measurement time points are not shown in this Figure because of the use of the medians (instead of the means). Nevertheless, the significant differences of interest for the hypotheses tested are described in the next paragraphs and shown as asterisks and squares in Figure 6.28.

Wilcoxon signed-rank tests were performed to compare the ratings of fatigue at the end of the flight with those just before the motorway drive. This comparison would reveal if taking the second break or just completing the PVT reduced participants' levels of fatigue. The analysis showed a significant reduction of the reported levels of fatigue just before the drive compared to the end of the flight only in the break session, $z = -2.992$, $p = .003$, $r = .331$. The factorial analysis with the ARTool package of the CSS data collected during the motorway drive showed that the interaction between the measurement time point and the session was not significant. In contrast, the simple main effects of the session and the block were significant, $F(1) = 88.006$, $p < .005$, $\eta_p^2 = .112$ and $F(4) = 21.404$, $p < .005$, $\eta_p^2 = .224$ respectively.

The simple main effect of the session was explored with Wilcoxon signed-rank tests. These showed significantly higher ratings of fatigue in the no-break session compared to the break one on all the measurement time points. That is, just before the drive (p = .003, r = .421) and on minutes 4 (p = .018, r = .212), 9 (p = .005, r = .241), 14 (p = .004, r = .321), and 19 (p = .002, r = .455). The simple main effect of the block was further explored with Friedman's tests. For the no-break session, the analysis showed a significant difference in the CSS ratings between the blocks, $\chi^2(4) = 33.347$, $p < .005$, Kendall's coefficient of concordance = .55. The post-hoc analysis with Wilcoxon signed-rank tests revealed a statistically significant increase of the ratings of fatigue from just before the motorway drive (median = 3) to minutes 14 (median = 5) ($\chi^2(4)$ = 3.060, p = .022, r = .448) and 19 (median = 5), $\chi^2(4)$ = 3.349, p = .008, r = .565. As in the no-break session, the ratings of fatigue in the break session were statistically significantly different between the measurement time points, $\chi^2(4)$ = 35.789, $p < .005$, Kendall's coefficient of concordance = .63. The post-hoc analysis with Wilcoxon signed-rank tests showed a statistically significant increase of the ratings of fatigue from just before the motorway drive (median = 2.5) to minute 14 (median = 4.5), $\chi^2(4)$ = 3.645, p = .003, r = .441, and minute 19 (median = 4.5), $\chi^2(4)$ = 3.825, p = .001, r = .603.

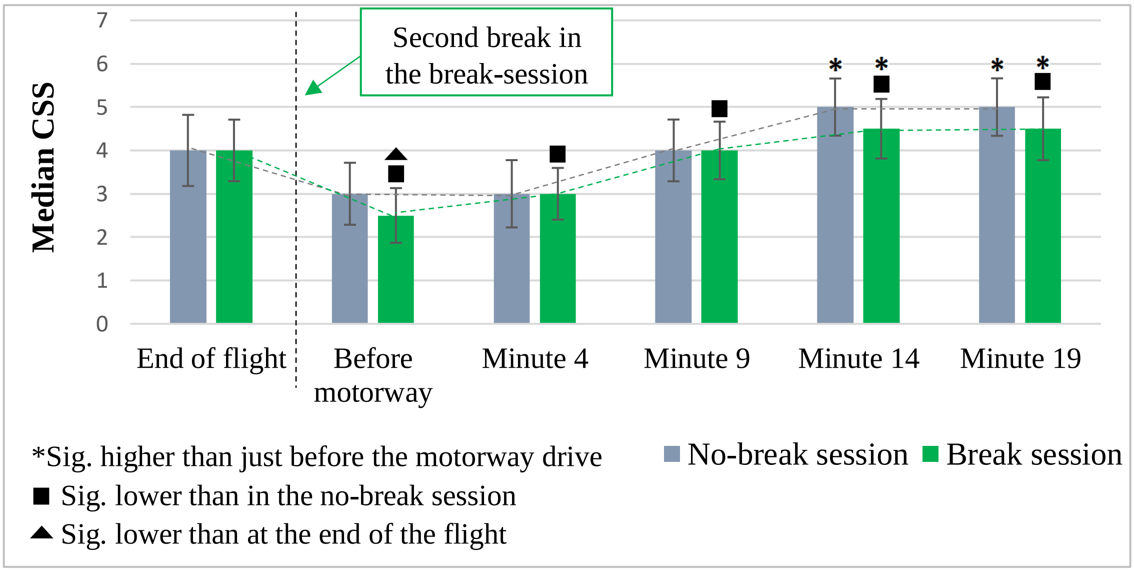


Figure 6.28 The CSS ratings in the motorway drive in experiment 2.

Table 6.3 shows the number of participants with high reported levels of fatigue on the CSS scale just before and in the motorway drive. Six means '*extremely tired, very difficult to concentrate*' and seven '*completely exhausted, unable to function properly*'. As depicted, more participants reported being very fatigued in all the blocks of the motorway drive in the no-break session.

Table 6.3 The frequency of high ratings of fatigue (CSS) in the motorway drive of experiment 2.

	Rating	Just before the 20-minute drive	Min. 4	Min. 9	Min. 14	Min. 19
No-break session	**6**	4	4	4	3	5
	7	-	-	1	2	2
Break session	**6**	1	1	1	2	3
	7	-	-	-	-	-

6.4.4.2 Performance

As mentioned in section 6.3.9, the driving performance data were analysed separately for the first (i.e. minutes 0-9) and the second half (i.e. minutes 10-19) of the 20-minute motorway drive. Moreover, there were four speed variation events in the first half of that drive, but the time the lead car drove on constant speed between consecutive events differed. Thus, the data of the first event were compared to those of the third, and the data of the second event to those of the fourth. The means of the SD of the lateral position per speed variation event are shown in Figure 6.29.

The 2x2 two-way repeated measures ANOVA showed that the assumption of normality was violated for the SD of the lateral position data collected in the first and third speed variation events, so a square root transformation was performed. The new analysis revealed that the interaction between the session and the speed variation event and the simple main effect of the session were not significant. In contrast, the simple main effect of the speed variation event was significant, $F(1,14) = 15.989$, $p = .001$, $\eta_p^2 = .533$. Pairwise comparisons (paired-samples t-tests) showed no significant difference in the SD of the lateral position between the first and third speed variation events in the no-break session. In contrast, in the break session, the SD of the lateral position was statistically significantly lower in the first than in the third speed variation event, $t(14) = 3.011$, $p = .009$, $d = 0.771$.

The SDs of the lateral position in the second and the fourth speed variation events were analysed with a 2x2 two-way repeated measures ANOVA, but normality was violated (square root transformation). The new analysis showed that the interaction between the session and the speed variation event and the simple main effect of the session were not significant. In contrast, the simple main effect of the speed variation event on the SD of the lateral position was significant, $F(1,14) = 6.992$, $p = .019$, $\eta_p^2 = .333$. Pairwise comparisons (paired-samples t-tests) showed no significant difference between the second and fourth speed variation events in the SD of the lateral position in the break session. In contrast, the SD of the lateral position was statistically significantly lower on the second than on the fourth speed variation event in the no-break session, $t(14) = 3.398$, $p = .004$, $d = 0.887$.

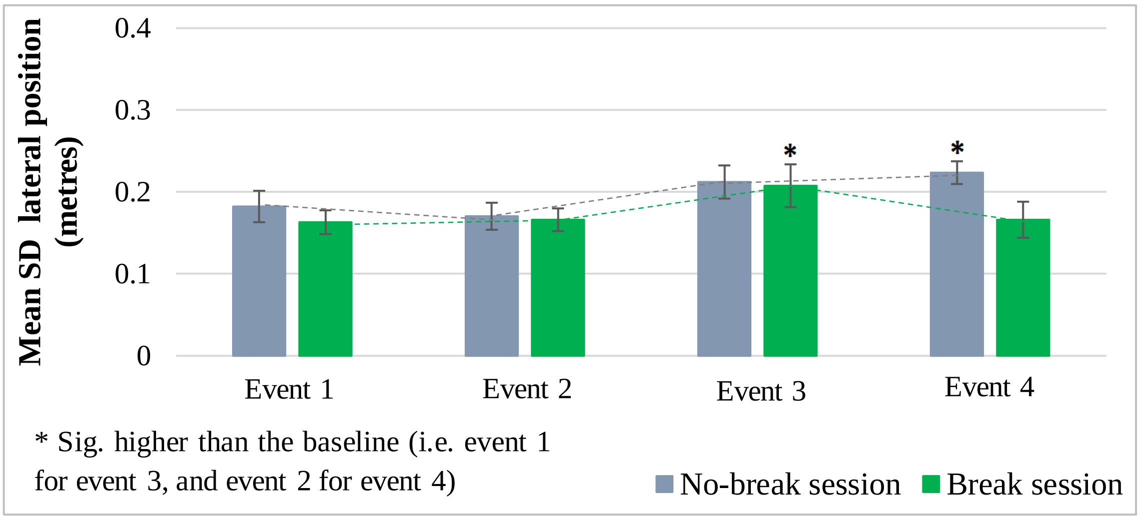


Figure 6.29 The SDs of the lateral position of the car per speed variation event (only when the lead car drove on a constant speed) in experiment 2.

The means of the SD of the headway from the lead car are shown in Figure 6.30.

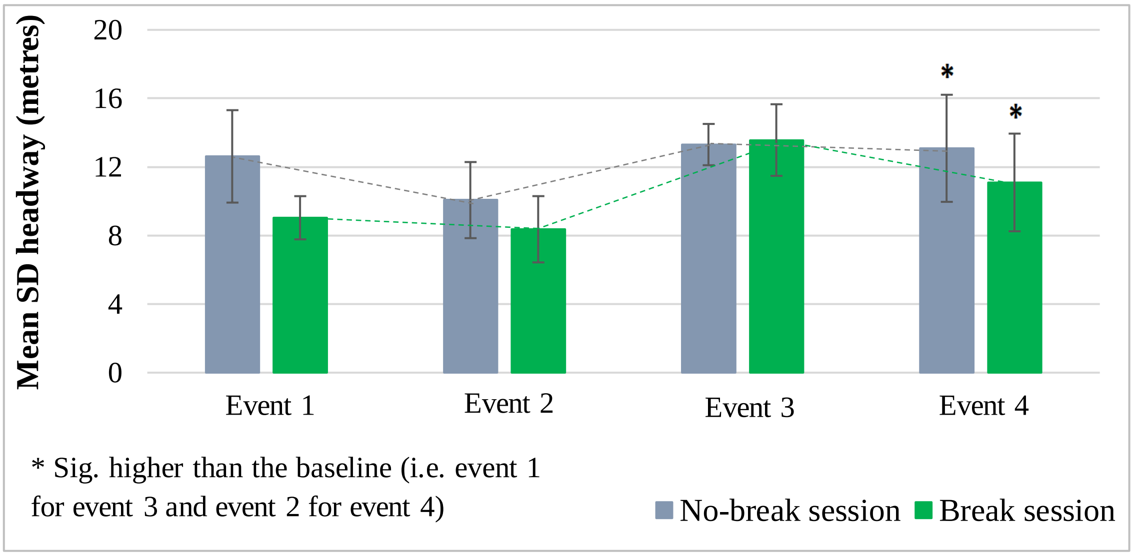


Figure 6.30 The SD of headway from the lead car per speed variation event (only when the lead car drove on a constant speed) in experiment 2.

A 2x2 two-way repeated-measures ANOVA for the first and third speed variation events showed that the assumption of normality was violated (logarithmic transformation). The new two-way repeated measures ANOVA showed that the interaction between the session and the speed variation event on the SD of headway and the simple main effects were not significant. The 2x2 two-way repeated measures ANOVA for the SDs of headway in the second and the fourth speed variation events showed that the assumption of normality was violated (logarithmic transformation). The new 2x2 two-way repeated measures ANOVA showed that the interaction between the session and the speed variation event and the simple main effect of the session were not significant. In contrast, the simple main effect of the speed variation event was significant, $F(1,14) = 15.168$, $p = .002$, $\eta_p^2 = .520$. Pairwise comparisons were run with paired-samples t-tests. For the no-break session, the SD of the headway was significantly higher in the fourth speed variation event than in the second, $t(14) = 2.966$, $p = .01$, $d = 6.29$. Similar, the SD of headway being higher on the fourth than on the second speed variation event of the break session, $t(14) = -2.288$, $p = .037$, $d = 5.903$.

The means of the SD of the lateral position in minutes 10-19 are depicted in Figure 6.31. A 2x2 two-way repeated measures ANOVA showed that the assumption of normality was violated. After a square root transformation, the 2x2 two-way repeated measures ANOVA on the transformed data showed no significant interaction between the session on the SD of the lateral position and the block and no significant simple main effects of the session and the block.

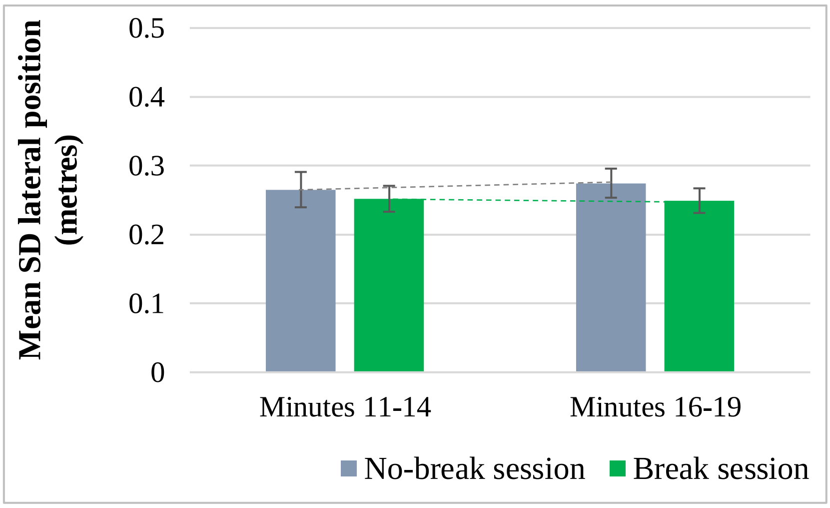


Figure 6.31 The SD of lateral position in minutes 10-19 of the motorway drive in experiment 2.

The means of the SD of speed in minutes 10-19 are shown in Figure 6.32. A 2x2 two-way repeated measures ANOVA showed that the assumption of normality was not met, so a logarithmic transformation was performed. A 2x2 two-way repeated measures ANOVA on the transformed data showed that the interaction between the stage and the session and the simple main effect of the stage were not significant. In contrast, the simple main effect of the session was significant,

$F(1,14) = 7.801$, $p = .014$, $\eta_p^2 = .358$. Pairwise comparisons with paired-samples t-tests showed that the SD of the speed was higher in the no-break session compared to the session with the breaks on minutes 10-14, $t(14) = 2.220$, $p = .043$, $d = 0.573$, and minutes 15-19, $t(14) = 2.181$, $p = .047$, $d = 0.561$.

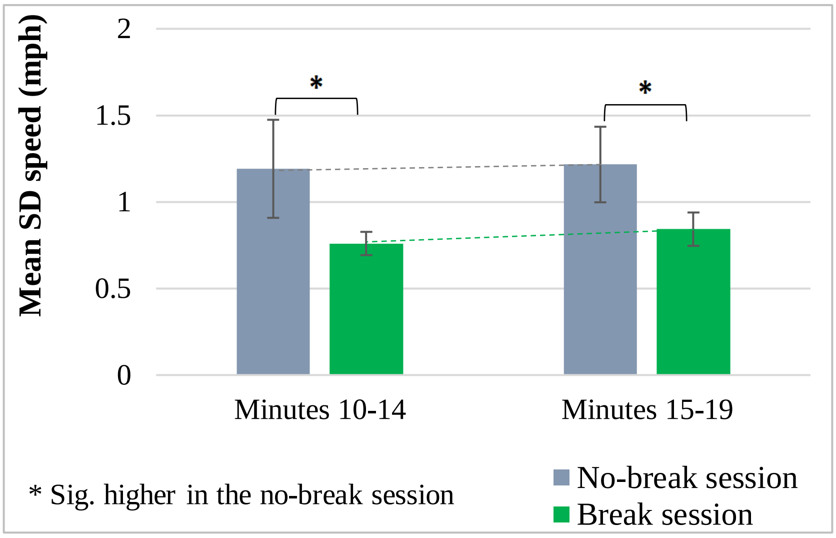


Figure 6.32 The SD of the speed of the car in minutes 10-19 of the motorway drive in experiment 2.

6.4.5 Summary of main results

60-minute drive:

- The CSS ratings increased significantly during the 60-minute drive in both sessions.
- The PVT reaction times after the 60-minute drive were significantly longer than the baseline in both sessions. Moreover, there were trends for more lapses in the PVT after the 60-minute drive compared to the baseline in both sessions.

Flying tasks:

- The CSS ratings just before the flying tasks were significantly lower than those at the end of the 60-minute drive only in the break session.
- Compared to the break session, the participants reported significantly higher levels of fatigue in the no-break session just before the flying tasks, at the end of stage 1, and at the end of stage 2.
- The CSS ratings changed differently over the flight between the sessions. The ratings increased significantly compared to just before the flying tasks only in the break session (after stages 2 and 3).

- The participants maintained the target heading more accurately in stage 1 and stage 3 in the break session compared to the no-break one.
- In the no-break session, two participants completed a turn to the wrong direction in stage 1.
- The participants maintained the target airspeed during the straight and level flying and the turns of stage 1 more accurately in the break session than in the no-break one.
- Compared to the no-break session, the participants made significantly fewer errors in the break session in the communications task in stages 1 and 3. Moreover, the number of errors in the communications task increased from stage 1 to stage 3 in the break session.

20-minute drive:

- The CSS ratings just before the motorway drive were significantly lower than those at the end of the flying tasks only in the break session.
- Compared to the no-break session, the participants reported significantly lower levels of fatigue just before the motorway drive and on minutes 4, 9, 14, and 19 of it in the break session.
- In the break session, the SD of the lateral position was statistically significantly lower on the first than on the third speed variation event.
- The SD of the lateral position was statistically significantly lower in the second than in the fourth speed variation event in the no-break session.
- In both sessions, the SD of the headway was significantly higher in the fourth than in the second speed variation event.
- The SD of the speed was higher in the no-break session compared to the break session on minutes 10-14 and 15-19.

6.5 Discussion

The research described in chapters 4 and 5 suggested that professional pilots can be at an increased risk of experiencing fatigue in flight and while driving after duty when they complete a long and demanding drive before flying. Drawing from Desmond and Hancock's (2001) theory of fatigue and the findings of the experiment of chapter 5, these adverse effects could be attributed to the fact that active fatigue induced by driving is accompanied by a depletion of attentional resources. Therefore, it is likely that there is a mismatch between the attentional resources needed for a task and those available in flight and while driving after it. If that was the case for commuting professional pilots, the use of appropriate intervention strategies should be considered to reduce the risk of accidents. The study described in this chapter explored the effectiveness of one of the self-initiated intervention strategies that commuting pilots could use to manage that risk. More specifically, the aim was to investigate if short bursts of physical activity can reduce the levels of fatigue experienced in the commuting cycle after a long and demanding drive.

Short bursts of physical activity were expected to be beneficial against fatigue because they combine postural change with physical exercise and mental disengagement (e.g. Abd-Elfattah et al., 2015). Fifteen individuals took part in two sessions each. During the sessions, they completed a long (60 minutes) and demanding simulated drive followed by simulated flying tasks (52 minutes) on the X-Plane 10 and a 20-minutes simulated motorway drive. Two 6-minute breaks that included physical exercise were completed before and after the flight in one of the sessions. Fatigue was measured by collecting subjective and performance data. The aim of the study was achieved because the long and demanding drives induced fatigue in both sessions, but the breaks helped the participants to reduce their levels of fatigue and perform better in the subsequent flying and driving tasks than in the no-break session.

6.5.1 The role of participant age, experience, substances, medications, the baselines of fatigue, and the investigated sleep-related factors

6.5.1.1 Age

All the participants in this study were males, but this is similar to the underrepresentation of female pilots in the population of pilots worldwide (Women in Aviation International, 2016). They were much younger than the average age of professional pilots in Europe and the USA (Civil Aviation Authority, 2018; Federal Aviation Administration, 2018). This was expected because all the participants were university students and none of them had a commercial pilot license. Older age can relate to higher levels of fatigue in driving and flying (e.g. McCartt et al., 2009). The reason is that older individuals may have more difficulties in adapting to changes in the wake/sleep cycle, may sleep less, and their sleep quality may be poorer (Harma et al., 1994; Ohayon et al., 2004). Hence, due to being older, many professional pilots who commute by driving may experience higher levels of fatigue in flight and while driving after it than those found in this study.

6.5.1.2 Experience

Airline pilots are more experienced than the participants in this study, which might allow them to mitigate any effects of commuting-induced active fatigue on flying performance by managing the workflow more effectively. However, workflow management was restricted in the simulated flight because of the instructions given to complete certain manoeuvres within pre-determined times. Thus, the effects of fatigue on flying performance found in this study may not differ significantly from those in professional pilots in situations in flight, where managing the workflow is not possible. For example, pilots may not be able to change the workflow in very demanding stages of flight or in emergency situations.

Based on the mean age of airline pilots worldwide, it is expected that many of them will have more years of driving experience than the participants in this study. Experienced drivers are better in paying attention to those parts of the road, where hazards are anticipated (Chapman et al., 2002; Crundall et al., 2012), and may perform better in simulated driving tasks (Hollopeter, 2011; Nabatilan et al., 2011). This considered, many professional pilots may be more likely to avoid any effects on driving performance after duty of fatigue induced by driving to the airport and flying than the participants in this study due to the expected longer driving experience. However, the findings of this study regarding driving performance after flight could apply more directly to young professional pilots and pilots who fail to use effective intervention strategies for fatigue before driving back home.

6.5.1.3 Medications, substances, and baselines of fatigue

Besides exploring the hypotheses of this study, it was deemed necessary to collect data about factors (other than the trials themselves) that could have contributed to any differences in the expressions of fatigue found between the sessions. The reason is that the effect of these factors could make the identification of the effect on fatigue of the 60-minute drives and the breaks difficult. To begin with, medications and substances that are known to increase or reduce alertness should not have played a role in the results of this study because not taking them was an eligibility criterion. Moreover, the differences in the expressions of fatigue found in the trials should not be attributed to between-session differences in the levels of fatigue experienced upon arrival for the study. This was supported by the absence of significant differences between the sessions in the baselines of CSS and PVT reaction times and lapses.

6.5.1.4 Sleep duration

Fully controlling for the effect of sleep-related factors on fatigue is difficult because of inter-individual differences in how these factors affect individuals. However, it is believed that between-session differences in the levels of sleepiness (as an expression of fatigue) were not caused by sleep deprivation the night before participation. This is based on the finding that all the participants reported at least 7 hours of sleep, which is considered adequate for adults (Hirshkowitz et al., 2015). Sleep duration was not measured by collecting physiological data (e.g. with actigraphy), but people are generally good at estimating their sleep duration (e.g. Kawada, 2008). Pilots are recommended to get eight hours of sleep every night (Federal Aviation Administration, 2009) but, as shown in the survey described in chapter 4, many of them sleep less than that. Therefore, professional pilots may be more likely to experience fatigue caused by sleep deprivation on top of any fatigue caused by driving to the airport and flying than the participants in this study. This could potentially contribute to more serious effects of fatigue on flying and driving performance than in this study. Future studies should also consider the potential effect of long-term sleep deprivation on commuting pilots' fatigue because pilots who suffer from long-term sleep deprivation might not recover fully with one night's sleep before commuting.

6.5.1.5 Sleep quality

Poor sleep quality is another factor that can contribute to fatigue in pilots and drivers (e.g. Reis et al., 2016). To explore its effect on participants' levels of fatigue, they rated their sleep quality the night before the sessions on an item taken from the St. Mary's Hospital Sleep Questionnaire. All the participants reported sleep of good quality (Figure 6.16) and the analysis suggested that any between-session differences in the expressions of fatigue should not be attributed to differences in sleep quality. It should be noted that only the most relevant item was used from that questionnaire and no physiological data were collected to infer sleep quality. However, single items (the one used in this study included) can give a relatively good indication of sleep quality (Clarke et al., 2013; Leigh et al., 1988; Snyder et al., 2018). In contrast to the participants in this study, sleep quality of professional pilots may be poor before some of their duties. On those occasions, the effects of active fatigue induced by driving to the airport on fatigue in flight and while driving after duty might be intensified. Future studies with more resources might benefit from collecting data about long-term sleep quality in commuting pilots because not sleeping well for several days might not be counteracted by one night of good quality sleep.

6.5.1.6 Circadian rhythms

The potential effects of the time of the day on fatigue in the flying and driving tasks were controlled by running the trials at 9 am for eight participants and 1 pm for the remaining seven. The preference for morning or evening tasks may affect the experience of sleep-related fatigue, but these data were not collected because exploring sleep-related fatigue in depth was out of the scope of this study. Moreover, the circadian low that occurs between 2 and 4 pm coincided with those sessions that started at 1 pm. However, any effects of the preference for mornings or evenings and the participation during the circadian low should have been controlled by running both sessions for each participant at the same time of the day. In real flights, experiencing a combination of task-related fatigue with fatigue caused by sleep-related factors is expected to be often. For example, many pilots drive to work and fly during the circadian low early in the morning or need to fly in the morning when they have a circadian preference for evenings. Thus, in professional pilots, the effects of fatigue induced by driving to the airport on fatigue in flying and subsequent driving tasks may be exacerbated compared to the findings of this study by the effects of the time of the day.

Sleep pressure increases with the time awake since the last sleep, but the participants in this study had neither been awake for too long (Walker, 2017) nor the time awake differed between the sessions. In real life, professional pilots may be awake for longer than the participants of this study when they start their commutes to the airport, for example, because they need to take care of an ill child. The effects of the time awake on fatigue in commuting pilots can be delayed by ensuring that the time awake is as short as possible when the commutes to the airport start. This means that feeling the need to sleep due to high sleep pressure may be postponed, hopefully, until pilots go to a rest facility after duty.

It is unclear if a disruption of the circadian rhythms due to time zone crossing and changes in the shift work patterns affected participants' levels of fatigue in the trials. In real life, such a disruption is possible, for example, in pilots that complete long-haul flights and may add to fatigue induced by driving to the airport. Thus, professional pilots may be at a higher risk of experiencing fatigue-related performance decrements after long and demanding drives to the airport than the participants in this study.

To sum up, based on the findings described above, any differences in the expressions of fatigue found between the sessions are likely not attributable to factors other than the breaks taken, although fully controlling for sleep-related fatigue and inter-individual differences is difficult. The findings of this study could apply to professional pilots, but age, flying experience, driving experience, and sleep-related factors might mediate fatigue in real life, either intensifying or suppressing its effects compared to this experiment.

6.5.2 Fatigue induction by the 60-minute drive

According to hypothesis 1, the CSS ratings, the PVT reaction times, and the number of PVT lapses will be significantly higher at the end than at the beginning of the 60-minute drive in both the 'break' and 'no-break' sessions. This expectation was based on the literature that suggests that fatigue can be induced by long and demanding driving (e.g. Cantin et al., 2009). The 60-minute drive was the same as in the study described in the manual group of chapter 5. Therefore, the induction of active fatigue was expected. Hypothesis 1 was confirmed by the results of the analysis of the CSS and PVT data.

6.5.2.1 CSS

The CSS ratings were collected every 4 minutes while driving. In both sessions, the participants started the drives on low levels of fatigue (i.e. a median of 2 = '*very lively, responsive, but not at peak*') (Figure 6.17). On average, the participants reported moderate levels of fatigue at the end of the 60-minute drives (a median of 4 '*a little tired, less than fresh*' in the break session and a median of 5 '*moderately tired, let down*' in the no-break one). However, driving induced fatigue in both sessions as evidenced by significant increases of the CSS ratings, so hypothesis 1 was confirmed. If the findings of this study applied to the population of professional pilots, this finding would indicate an increased risk of impaired performance in flight after a long and demanding drive to the airport.

The participants' CSS ratings in both sessions became significantly higher than the baseline after 39 minutes and remained significantly higher until the end of the 60-minute drive. The 39 minutes of driving that passed before finding a significant increase of the CSS ratings should be a relatively accurate value because people are generally good at detecting increased levels of fatigue (e.g. Horne and Baulk, 2004). Furthermore, the 39 minutes is very close to the 44 minutes of driving that had passed before finding the first significant increase of the CSS ratings compared to the baseline in the 60-minute drive that was completed by the manual group in the study of chapter 5 (Figure 5.24). The fact that the reported levels of fatigue did not differ significantly between the sessions and increased significantly after the same amount of time in both sessions indicates that the counterbalancing between the no-break and break sessions was effective.

6.5.2.2 PVT

The analysis of the PVT data supported hypothesis 1 because the reaction times after the 60-minute drive were significantly longer than the baselines in both sessions. There was also a trend in both sessions for more lapses after that drive than before it. In agreement with the study described in chapter 5 and Loh et al. (2004), these results suggest that the 5-minute PVT can show the effects of task-related fatigue on vigilance. It should be noted that PVT reaction times do not necessarily predict flying performance (e.g. Baulk et al., 2006) and this is why also measuring task-related performance is useful in fatigue studies. Nonetheless, if the findings of this study applied to professional pilots, the results of the analysis of the PVT data would indicate an increased risk of impaired performance in flight due to fatigue after long and demanding drives to the airport.

To sum up, the CSS and the PVT data suggested that the 60-minute drives induced fatigue in both sessions, confirming hypothesis 1. If the results of this study applied to professional pilots who drive to the airport, the risk of flying fatigued could be higher after long and demanding drives. Although the findings of this study suggest that even 40 minutes of demanding driving can be fatiguing, it is believed that the focus should be more on the increase of the risk of impaired flying performance associated with such a fatigue induction rather than on the absolute values found. That is, the same drive could induce more fatigue in some pilots than in others, for example, due to inter-individual differences in fatigue induction (Nilsson et al., 1997). On the other hand, some routes could induce fatigue more quickly than others due to differences in the levels of task load (Oron-Gilad and Ronen, 2007; Thiffault and Bergeron, 2003; Ting et al., 2008; Vogelpohl et al., 2019). Moreover, there is evidence that simulated driving induces fatigue more quickly than real driving (e.g. Davenne et al., 2012). Hence, drives to the airport of shorter or longer duration than the ones found in this study could increase the risk of flying fatigued in professional pilots.

6.5.3 The effect of the first break on the levels of fatigue experienced during the flight

Aeroplane incidents and accidents may be more likely to occur when pilots fly fatigued due to commuting to the airport. This expectation is based on accident investigation reports, documents published by aviation organisations, and the limited research on commuting pilots and other professionals (Azwar et al., 2018; Federal Aviation Administration, 2012[a,b]; Zakariassen et al., 2019). In addition, the studies described in chapters 4 and 5 suggested that the risk of impaired flying performance increases after long drives that induce active fatigue. All this considered, it was hypothesised that the participants would be less fatigued in the simulated flying tasks in the break session than in the no-break one (hypothesis 2). This hypothesis was confirmed by finding lower self-reported levels of fatigue and better flying performance when the participants had taken a break before the X-Plane 10 flight.

6.5.3.1 CSS

The literature suggests that disengaging mentally from a task can help individuals reduce their levels of fatigue (Neri et al., 2002). Nevertheless, just completing the PVT after the 60-minute drive did not help the participants with fatigue. This is evidenced by not finding a significant difference in the CSS ratings just before the flying tasks compared to those at the end of the 60-minute drive in the no-break session. The completion of a PVT after the 60-minute drive had not helped the participants in the manual group of the study described in chapter 5 to reduce their levels of fatigue either (Figure 5.33). The absence of a significant reduction in the no-break session of the CSS ratings directly after the PVT that followed the 60-minute drive might be attributed to the fact that the drive induced active fatigue (as suggested in the study of chapter 5). That is, the new task (i.e. the PVT) required continuing the investment of attentional resources, which is not helpful against active fatigue (Oron-Gilad et al., 2002). If the findings of this study applied to professional pilots, then completing pre-flight activities at the airport that require investing mental effort (e.g. prepare for flight) might not help them to reduce active fatigue induced by driving to the airport.

In contrast to completing the PVT, taking a break with physical activity helped the participants to reduce their levels of fatigue as evidenced by significantly lower CSS ratings after the first break compared to the end of the 60-minute drive in the break session. This reduction agrees with the literature that suggests that short breaks can reduce active fatigue when they disengage individuals mentally from a previous task and combine an upright body position with physical exercise of medium physical intensity (e.g. Abd-Elfattah et al., 2015). On the other hand, this result contradicts Horne and Reyner (1996), who had found no benefits in terms of self-reported sleepiness after 30-minute breaks. However, the participants in that study had remained seated during the break. The duration of the breaks used in this study was similar to that in Neri et al.'s (2002) study, but physical exercise was also used here in addition to mental disengagement. In contrast to studies that suggested the effectiveness of longer breaks (Hogervorst et al., 1996; Liang et al., 2009; Loy et al., 2013; Phipps-Nelson et al., 2010), the significant reduction of the CSS ratings after the first break in this study indicates that even 6-minute breaks can reduce the levels of active fatigue induced by driving.

As expected from hypothesis 2, the participants reported significantly lower levels of fatigue during the flying tasks in the break session than in the no-break one. This could be attributed to taking the first break since there was no significant difference between the sessions in the ratings of fatigue at the end of the 60-minute drive and the CSS ratings just before the flying tasks were significantly lower than at the end of the 60-minute drive only in the break session. Based on the analysis of the CSS data, taking the first break helped the participants to start the flight less fatigued than in the no-break session and this beneficial effect lasted until the end of stage 2 (i.e. 36 minutes of flying). The 36 minutes are longer than the 25 minutes found by Neri et al. (2002). This difference might be attributed to the fact that the participants in Neri et al.'s study stood and discussed with the experimenter during their breaks but did not do any physical exercise. This considered, the study described in this chapter suggests that including physical activity in breaks may increase the duration of their positive effects in terms of fatigue. It should be noted that there were more than 10 minutes without a subjective measurement of fatigue after the 36 and the 25 minutes in this and Neri et al.'s study respectively. Therefore, these durations might not reveal exactly how long the beneficial effects of the breaks lasted. However, what the results of both studies agree on is that a short break can help individuals with fatigue but not for long.

The absence of a significant between-session difference in the CSS ratings at the end of the flight (i.e. minute 52) when all the previous ratings in the flight were significantly different can be explained by the significant interaction between the session and the measurement time point in the CSS ratings. That is, the subjective ratings of fatigue changed differently over the flight in the two sessions with a significant within-session increase being found only in the break session (Figure 6.19). This might be attributed to the fact that the participants started the flight significantly less fatigued in the break session; thus, they were more sensitive in fatigue induction. In contrast, not having a break likely resulted in reaching a plateau of fatigue before the flight or early in it.

The absence of a significant increase of the CSS ratings during the flying tasks in the no-break session contradicts the findings of the experiment described in chapter 5. In that case, the CSS ratings increased significantly over the flying tasks for the manual group (Figure 5.33). The different patterns in the CSS ratings between this and the previous study might be explained by a difference in skills. That is, the participants in this study were highly skilled in completing the flying tasks and had used the X-Plane 10 software and equipment before. Therefore, their perceived task load might have been lower than that experienced by the manual group in the study of chapter 5. In turn, the lower perceived task load might have resulted in a slower depletion of their attentional resources; thus, in a higher resistance to fatigue induction in the X-Plane 10 flight.

In real flights, pilots spend some time in the aeroplane preparing for the flight, for example, by doing pre-flight checks. This means that, even if they took a break with physical activity just before entering the aeroplane, its benefits might dissipate before taking off. Future research could explore how any alerting effect of a break before entering the aeroplane changes by spending time in the aeroplane before taking off. In any case, it is important that pilots are alert during the pre-flight checks too in order to avoid mistakes, so a break before flight like the one tested in the study of this chapter could help.

If the alerting effect caused by a break like the one tested did not disappear before taking off, pilots could be helped to mitigate any effects of active fatigue induced by driving to the airport on their flying performance in short domestic flights. However, in longer flights, aeroplanes may still be on the cruise stage after 40 minutes of flying. In that case, short bursts of physical activity before flight could reduce the risk of performance decrements due to commuting-induced fatigue during takeoff, climb, and a part of the cruise stage. Therefore, pilots might be able to cope better with the usually higher requirement for the investment of attentional resources in the first stages of flight and avoid micro-sleeps during cruise (Cabon et al., 1993; Jen et al., 2009; Wright and McGown; 2001). However, as the beneficial effects of a break before flight started to dissipate and flying induced fatigue, the risk of experiencing impaired performance in flight would increase (e.g. Morris and Miller, 1996). If that was the case, short breaks of physical activity before flight should be accompanied by self-initiated intervention strategies for fatigue during flights to ensure pilots complete the cruise, descent, and landing stages as alert as possible.

6.5.3.2 PVT

The results of the analysis of the PVT data did not confirm hypothesis 2 because participants' reaction times and lapses after the flying tasks were not significantly different between the sessions (Figure 6.18). This result agrees with the absence of a significant difference between the sessions in the self-reported levels of fatigue at the end of the flights. No between-session differences were expected in the PVT before the flying tasks because the first break was taken after that PVT. The results of the analysis of the PVT data suggest that the flying tasks did not induce fatigue in the no-break and break sessions. The analysis of the CSS data also suggested that the flight did not induce fatigue in the no-break session. Nonetheless, the CSS data showed a significant increase in the levels of fatigue during the flying tasks in the break session. This disagreement between the results of the CSS and the PVT data in the break session might be attributed to the fact that the PVT that followed the 60-minute drive was completed before the first break, whereas the CSS rating was given just before the flying tasks. Since the first break reduced participants' levels of fatigue, their vigilance should have increased in flight. As a result, the comparison between the two PVTs could not show the direct effects of the flying tasks on fatigue.

The PVT reaction times after the MATB-II tasks in the manual group in the study of chapter 5 were significantly longer than those before them (Figure 5.34). This finding contradicts the PVT results in the no-break session in this study (the break session differed in including the break). An explanation might be the difference in motivation, since higher motivation may help to prevent a reduction of vigilance (Bonnefond et al., 2011). That is, the participants in the study described in this chapter might have been more motivated to perform well than the ones in the experiment of chapter 5 because of their personal and professional interest in the flying tasks.

6.5.3.3 X-Plane 10 performance

The analysis of the flying performance data confirmed hypothesis 2. More specifically, the participants maintained more accurately the target heading in stage 1 (during straight and level flying and in the climb) and stage 3 (during descent) in the break session than in the no-break one (Figure 6.21). Taking a break before flight also helped them to maintain more accurately the target airspeed during straight and level flying and in the turns completed in stage 1 (Figure 6.26). Moreover, the participants made fewer mistakes in the communications task in stages 1 and 3 in the break session (Figure 6.27). Finally, although no significant difference was found between the sessions in the accuracy of maintaining the target rates of turn, two participants completed turns to the wrong direction in stage 1 of the no-break session (Figure 6.23). Between-session differences in flying performance were expected because of the attentional deficits when individuals are fatigued (section 2.2.2). These deficits result in a difficulty in storing information in working memory, processing information to make arithmetic calculations, detecting unwanted deviations in system parameters, and responding accurately (Boksem et al., 2005; Dinges et al., 1997; Falleti et al., 2003; Hockey, 1997; Lorist et al., 2000; Sanders, 1998; Saxby et al., 2008). All these skills were required to perform the tasks included in the simulated flight.

In real flights, the reduced accuracy in controlling the joystick to maintain the target flight parameters (e.g. altitude) would probably reduce passenger comfort rather than contribute to accidents. However, the lower accuracy in the flight performance metrics found in this study suggests a higher risk of impaired performance, especially in high task load situations (e.g. due to a system failure). In those conditions, pilots' attentional resources might not suffice to cope with the tasks due to active fatigue induced by driving to the airport and flying. An interesting finding of this study is that two participants completed one 360° turn each to the opposite direction than instructed during stage 1 of the no-break session. Aeroplane accidents are rare, but it is those exceptional occasions of performance deterioration that relate to a particularly high risk of safety events.

Based on the findings of this study, a short burst of physical activity before flying can reduce the risk of missing critical information from the ATC and not acting upon it appropriately because of active fatigue induced by driving to the airport. The participants were told that all the tasks in flight had the same importance. Although data about task prioritisation were not collected, maybe the communications task was perceived as secondary by some of the participants. This may occur in real flights too. The reason is that the pilots are taught that the most important thing is to fly the aeroplane, followed in importance by understanding where they are going, and finally, communicating with the ATC and other aeroplanes ('aviate, navigate, communicate'). If that was also the case in this study, then performance in the communications task was more susceptible to the effects of fatigue than the primary tasks (Hockey et al., 1998).

The flight scenarios did not include takeoff, but task load in stage 1 was increased to simulate the high levels of task load often experienced during the takeoff stage of flights. The between-session differences in performance found in stage 1 suggest that the participants were more fatigued in the first minutes of the flight in the no-break session compared to the break one in line with the results of the analysis of the CSS data. If the findings of this study applied to the population of professional pilots, then active fatigue induced by driving to the airport would increase the risk of performance decrements in the takeoff and climb stages of flight. As suggested in this study, that risk could be reduced by taking a short break with physical activity before flight.

The literature suggests that micro-sleeps are more often in the cruise stage of flight (Cabon et al., 1993; Wright and McGown, 2001), but no significant differences were found in flying performance during stage 2 (which simulated cruise). This finding contradicts the significant difference between the sessions in the CSS ratings at the end of stage 2. An explanation for that might be that the between-session difference in the levels of fatigue during stage 2 was big enough to be captured on the subjective ratings but not to cause a between-sessions difference in performance. This explanation aligns with the finding that individuals can identify increased levels of fatigue before their performance deteriorates (Ingre et al., 2006[b]). It is also likely that the participants mainly experienced active fatigue when they started stage 2 because of the 60-minute drive and the high task load in stage 1. Therefore, the lower levels of task load in stage 2 could have helped them to avoid performance decrements because they did not have to invest many attentional resources. If active fatigue induced by driving to the airport had similar effects on professional pilots to the ones found in this study, then impaired performance would be likely in the cruise stage too, especially, in case of emergencies that required the investment of many attentional resources.

According to the CSS data, the beneficial effects of the first break disappeared as the flight progressed. As a result, the self-reported levels of fatigue did not differ between the sessions at the end of stage 3. Similar, in contrast to stage 1, no difference was found in stage 3 between the sessions in the accuracy of maintaining the target heading when flying straight and level and the airspeed during turns. However, in stage 3, the participants maintained the target heading during the decent more accurately and performed better in the communications task than in the no-break session. No CSS data were collected during stage 3, so it is unclear exactly how long the beneficial effect of the first break lasted. However, it appears that this effect continued after the 36 minutes of flight found by the analysis of the CSS data maybe because adding physical activity to the break resulted in longer-lasting effects than in Neri et al.'s (2002) study. In any case, the contradiction of the results of the analysis of the flying performance data in stage 3 suggests that the beneficial effect of the first break started to dissipate with time. In real flights, additional self-initiated intervention strategies might be needed after some time of flying to reduce the levels of fatigue.

The literature suggests that the risk of performance decrements due to fatigue is higher in the last stages of flight (e.g. Morris and Miller, 1996). This was supported in this experiment by finding significantly more errors in the communications task in stage 3 than in stage 1 in the break session. A similar increase was not found in the no-break session potentially because the participants had already started the flight closer to their plateau of fatigue and, thus, might have been less sensitive in further fatigue induction. The significant increase in the number of communication errors in the break session highlights the need for additional intervention strategies in flights, especially when these are long.

6.5.4 The effect of the second break on the levels of fatigue experienced in the 14-minute drive

As explained in the previous section, if the findings of this study applied to professional pilots, any benefits of taking a break with physical activity between driving to the airport and flying might dissipate before the end of the flights. Even if the alerting effect lasted until landing, pilots would then need to complete post-flight tasks while at the airport (e.g. file a post-flight report). As a result, they might start a drive back home fatigued potentially due to a combination of the effects of driving to the airport, pre-flight activities, flying, and post-flight tasks. The risk of experiencing fatigue-related performance decrements when driving after work has been associated in professionals other than pilots to extended shifts, sleep deprivation, consecutive and night shifts, and driving long distances after work (e.g. Barger et al., 2005). However, there is no literature on the risk of impaired driving performance after flight and on whether this relates to fatigue induced by driving before flying. To this end, it was explored if the participants would be less fatigued in the simulated motorway drive in the break session than in the no-break one. This hypothesis (hypothesis 3) was accepted based on the results of the analysis of the CSS and driving performance data.

6.5.4.1 CSS

Taking the break with the physical activity after the flying tasks helped the participants to reduce their levels of fatigue as evidenced by significantly lower CSS ratings just before the motorway drive compared to the end of the flight (Figure 6.28). This finding agrees with the literature that suggests that physical activity, an upright position, and mental disengagement from a previous task can have an alerting effect (e.g. Abd-Elfattah et al., 2015). Furthermore, this reduction of the CSS ratings suggests that even a 6-minute break can reduce the levels of task-related fatigue similar to what was found by Neri et al. (2002). At the same time, it contradicts the notion that breaks need to be long to be effective against task-induced fatigue (e.g. Hogervorst et al., 1996).

In contrast to the break session, completing only the PVT after the flight did not help with fatigue in the no-break session. Similar results were found for the PVT that was completed before the flight in this experiment and the PVTs of the manual group in the study described in chapter 5. These results might be explained by the fact that the participants in both studies should have mainly experienced active fatigue when the 20-minute drive started due to the demanding nature of the 60-minute drive and stages 1 and 3 of the flights. However, mental activation can only help against passive fatigue (Oron-Gilad et al., 2002).

The CSS ratings at the end of the flight were not significantly different between the sessions, which suggests that the benefits of the first break had dissipated by then. Nevertheless, as predicted by hypothesis 3, taking a break after flight resulted in significantly lower self-reported levels of fatigue during the motorway drive. The means of the CSS ratings were low to moderate throughout the motorway drive in both sessions, but half of the participants in the no-break session reported high or extremely high levels of fatigue at the end of that drive. That could indicate a higher risk of car crashes when no break with physical activity is taken after flight. Based on the CSS data, the motorway drive induced fatigue after 14 minutes of driving irrespective of whether a break was taken after flight or not. This finding is in line with the driver literature (e.g. Kecklund and Akerstedt, 1993) and suggests that the risk of fatigue-related crashes after duty may increase with the duration of the drives. This considered, using intervention strategies for fatigue during driving after duty should complement any breaks taken before those drives because the positive effects of breaks could disappear with time.

Despite the induction of fatigue in the motorway drive, the reported levels of fatigue remained significantly lower until the end of the drive (i.e. 19 minutes) in the break session compared to the no-break one. It is unclear how long the beneficial effects of the second break taken would last if the drive continued after the 20 minutes. Based on the analysis of the CSS and the performance data collected during the flight, the alerting effect of the break used might have lasted for 40 minutes or longer. If the duration of professional pilots' drives after duty were similar to those found for the drives to the airport in the survey described in chapter 4 (Figure 4.14), a 6-minute break after duty like the one tested could help many pilots prevent fatigue-related crashes for the whole duration of their drives. However, pilots might be benefitted by using intervention strategies during driving in addition to a pre-drive break to ensure they manage the risk of fatigue-related crashes in longer drives.

6.5.4.2 Driving performance

In regard to performance in the motorway drive, no significant between-sessions differences were found in the lateral position during the whole drive and the headway in the first 9 minutes of the drive. However, hypothesis 3 was supported by the finding that the participants maintained the target speed in the second half of the drive (i.e. minutes 10-19) more accurately in the break session than in the no-break one (Figure 6.32). This finding agrees with the significant differences between the sessions in the CSS ratings on minutes 14 and 19 of the motorway drive. Nevertheless, the CSS data suggested that taking the break after flight resulted in the participants being less fatigued in the first half of the drive too. Maybe this difference was not captured in performance in the first half of the drive because that part of the drives was more motivating than the second one due to the speed variation events. If that was the case, any between-session differences in fatigue might have been masked by the difference in motivation (Watling, 2016). In real drives after duty, the absence of significant between-session differences in driving performance in the first half of the simulated drive could indicate a reduction of the risk of crashing when the task load of the drives is high enough to keep the pilots engaged. Nonetheless, this benefit might not last until the end of the drives back home because the investment of effort would induce fatigue.

The induction of fatigue in the motorway drive was suggested by the analysis of the CSS. Moreover, the induction of fatigue was indicated by the reduction of the accuracy of maintaining the target lateral position and the headway from the lead car as the motorway drives progressed in both sessions. These findings agree with the literature that suggests that long driving increases individuals' levels of fatigue (e.g. Kecklund and Akerstedt, 1993). At the same time, the induction of fatigue in the 20-minute drive indicates that, if the findings of this study applied to professional pilots, the risk of car crashes after duty might increase with longer drives. As mentioned, using intervention self-initiated intervention strategies for fatigue while driving after duty might allow pilots to complete safely longer drives.

6.6 Conclusion

If the findings of this study applied to professional pilots, then:

- Completing a long and demanding drive could increase pilots' levels of fatigue, resulting in a higher risk of impaired performance in flight. This risk could reduce by taking a short break with physical activity of moderate intensity when pilots arrived at the airport and before flying.
- Pilots could be at an increased risk of car crashes after duty due to active fatigue induced by driving to the airport and flying. A short burst of physical activity after flight could help them to avoid driving fatigued back home or to a rest facility.
- The alerting effect of the break tested in this study could last longer than found in similar studies, but pilots would need to use additional self-initiated intervention strategies in flight and while driving after duty to extend their beneficial effects.

Chapter 7. Final discussion and conclusion

7.1 Overview

Pilot fatigue induced by commuting to the airport has recently drawn the attention of aviation authorities because it has been identified as a contributory factor in aeroplane incidents and accidents. Drawing from studies solely on drivers, the risk of flying fatigued may be higher when pilots travel to the airport by completing long and demanding drives. On the other hand, flying can increase pilots' levels of fatigue too, with cumulative fatigue potentially contributing to car accidents after duty. This considered, this thesis aimed to explore the relationship between pilot fatigue induced by long and demanding driving to the airport, fatigue in flight, and fatigue while driving after duty. This aim was achieved by:

- Exploring the extent to which fatigue induced by long and demanding drives to the airport affects negatively professional pilots' fatigue in flight and while driving after duty. This objective was achieved by conducting the survey described in chapter 4.
- Investigating if the type of fatigue induced by driving (i.e. active or passive) subsequently affects the levels of fatigue experienced in flying and driving. This was achieved with the experiment described in chapter 5.
- Exploring if short bursts of physical activity can reduce the levels of fatigue experienced in the commuting cycle after a long and demanding drive. This objective was achieved via the study described in chapter 6.

This chapter summarises the results of the research conducted in view of the existing literature. It also discusses its limitations, the implications for further study, and the development of interventions.

7.2 Summary of results

The objectives of the thesis were achieved by using a combination of quantitative and qualitative methods. The work described in this thesis lead to the following conclusions.

7.2.1 The extent to which active fatigue induced by driving to the airport affects professional pilots' fatigue in flight and while driving after duty (survey)

When this research began, there were no data regarding the extent to which fatigue induced by long and demanding driving to the airport is a problem among professional pilots worldwide. Before exploring that with a survey, an interview with an airline pilot was conducted to identify the key aspects of the issue. The interviewee highlighted the importance of researching active fatigue induced by driving to the airport because, in his opinion, this has resulted in many pilots flying fatigued and experiencing impaired flying performance. According to the interviewee, often pilots find it difficult to stay awake while driving after duty, probably because of the combined effect of fatigue caused by driving to the airport and flying. In addition, he suggested the provision of science-based training to pilots regarding commuting-related fatigue and changes to the fatigue policies and the rostering in order to manage this risk. These findings indicate that active fatigue induced by driving to the airport is an issue amongst professional pilots that needs further investigation in order to develop appropriate intervention strategies.

Based on the findings of the interview, an online survey followed to explore the extent of the issue among professional pilots worldwide. Four hundred ninenteen pilots flying from 52 countries were asked about the characteristics of their usual commutes and flights, their usual levels of fatigue during the commuting cycle (i.e. before and after flights), and impaired performance in flight and while driving after duty. The role of sleep duration and driving early in the morning was explored too to identify the potential interactive effect of task- and sleep-related factors in commuting pilots' fatigue.

The survey showed that two in three pilots usually commuted to and from the airport by driving. This suggests that the findings of the research described in this thesis could apply to many of the professional pilots worldwide. Driving has been found to be the most common commuting option in other professionals (e.g. Department for Transport, 2015) but not in pilots. That is, Kleinfehn (2016) and the National Research Council (2011) had suggested that most pilots commute by aeroplane. This difference could be attributed to the fact that the pilots in these studies had a home base in the US, which is a country with big geographical spread; thus, the distances most of the pilots there covered to get to the airport were longer than 150 miles. In contrast, most of the pilots in the survey of chapter 4 had a home base in Europe or Asia with more than 75% of them reporting a usual duration of commutes one-way of less than an hour. The durations of commutes found in the survey of chapter 4 are closer to those found by Friesacher (2015) in a study with pilots with a home base in Europe. This considered, pilots' commuting habits may differ by region. As a result, fewer pilots might be at an increased risk of being fatigued in flight in a region where, for example, most pilots commute by aeroplane because they can sleep during their commutes.

The participants rated their usual levels of fatigue during the commuting cycle on the CSS. Most of the pilots who commuted by driving usually avoided the long and demanding drives. Maybe because of that, their usual commutes to the airport were not reported as fatiguing. However, reporting longer and more mentally demanding drives to the airport related to usually experiencing higher levels of fatigue at the beginning and end of the flights, a more often experienced difficulty in staying awake in flight, and a higher agreement with the statement about the impaired flying performance after driving to the airport.

The preliminary analysis of the data in Brown and Whitehurst's (2011) study had not shown an association between longer commutes to the airport and fatigue-related performance decrements in flight. Similar, Roma et al. (2012) had not found a link between the duration of the commutes and the PVT reaction times of cabin crew before flying. However, Roma et al. suggested that commuting to the airport can induce fatigue in pilots when it is completed by driving because then they cannot sleep during the commute. Similar to Roma et al., aviation authorities have stated in investigation reports that long commutes to the airport can result in pilots flying fatigued and an increased risk of aeroplane accidents (e.g. Federal Aviation Administration, 2012[a]). The studies conducted by the National Research Council (2011) and Zakariassen et al. (2019) also suggested that the more time spent on commuting, the higher the risk of flying fatigued. A similar suggestion has been made in studies with other professionals (Azwar et al., 2018).

There is evidence from other studies that pilots often fly fatigued (Steptoe and Bostock, 2011) and fall asleep in flight when not allowed (Jen et al., 2009). However, in none of these studies was this risk associated with commuting-induced fatigue by asking them to rate their levels of fatigue during the commuting cycle as was the case with the participants in the study of this chapter. In addition, this survey adds to the existing literature by exploring fatigue specifically after driving and by collecting data from pilots from 52 countries. Furthermore, past knowledge is extended by suggesting that pilots fly fatigued not only because of long but also mentally demanding drives to the airport. If that was the case, pilots should avoid the adverse effect of such drives, for example, by staying at a hotel near the airport overnight, selecting the shortest route possible, or not driving in rush hour. Alternatively, self-initiated intervention strategies like the one tested in chapter 6 could be used while at the airport before flight.

Reporting driving more often between 2 and 6 am and sleeping less before the commutes to the airport related to higher usual self-reported levels of fatigue at the beginning and end of the flights and more frequently experienced difficulties in staying awake in them. The effects of the time of the day and sleep deprivation on pilots' fatigue in flight have been supported in other studies too (Caldwell et al., 2004; Le Duc et al., 1999). However, the importance of the finding of this survey lies in the identification of the potential co-existence of sleep- and task-related fatigue in pilots who commute by driving. This co-existence may intensify any adverse effects of commuting-induced fatigue on flying performance and make it difficult for pilots to identify the source of their fatigue in order to manage the risk properly. For example, if both sleep- and task-related fatigue was experienced while flying, pilots would need to use a different self-initiated intervention strategy for each (e.g. coffee and breaks respectively), unless sleeping was an option.

Based on pilots' self-reports, the time spent at the airport before duty usually helped them to start their flights less fatigued. This finding suggests that the risk associated with commuting-induced fatigue is usually managed effectively by pilots before taking off. It should be noted that the pilots were asked about their usual levels of fatigue, so it is possible that sometimes the time spent at the airport may not help some pilots to reduce any fatigue caused by driving to the airport. For example, pilots who usually take a nap at the airport after a long drive may not be able to do that on days with commutes longer than expected. This thesis focused mainly on those cases where pilots do not manage commuting-induced fatigue before flying because that is when the risk of impaired flying performance is expected to be higher.

Based on the findings of the survey, flying usually increases pilots' levels of fatigue. Therefore, the risk of fatigue-related performance decrements could be higher in the last stages of flights (i.e. descent and landing) in line with the previous literature (e.g. Morris and Miller, 1996). Active fatigue induced by driving to the airport may intensify the effects on flying performance of fatigue caused by flying. This interaction was not explored in this survey, but the finding that reporting longer and more mentally demanding drives to the airport related to higher levels of fatigue when entering and exiting the aeroplane suggests that fatigue induced by driving and flying may coexist. The potential magnification of the risk of impaired flying performance after long and demanding drives was supported in the studies described in chapters 5 and 6.

Many pilots reported that they usually did not manage to reduce their levels of fatigue when they finished their duties and before they drove back home, but it is unclear why. The participants reported that the time spent at the airport before flight usually reduced their levels of fatigue, but this was not the case for the time spent at the airport after duty. This difference might be attributed to spending less time at the airport after flight than before it because of the desire to go home the soonest. Moreover, pilots may focus less on managing fatigue after their duties compared to fatigue before them because only reporting fit-for-duty is regulated. Another explanation could be that any intervention strategies used before flight might not be equally effective after duty because of the accumulation of fatigue due to commuting to the airport, the completion of pre-flight tasks, flying, and post-flight tasks. Irrespective of why the pilots stated that they usually did not reduce their levels of fatigue before flying, this finding indicates an increased risk of car crashes and highlights the importance of using effective intervention strategies after duty.

Studies in professionals other than pilots suggest an increased risk of car crashes when people drive after work (Barger et al., 2005; Dorrian et al., 2008; Marcus and Loughlin, 1996; Nesthus et al., 2006; Scott et al., 2008). In those studies, the risk of driving accidents was associated with the extended work shifts, sleep deprivation before duty, and night shifts. The participants were asked about the frequency of drives to the airport between 2 and 6 am but not how often they drove after flight during the circadian low. In contrast, the link between duties and the risk of impaired driving performance was also indicated in this survey since more than 80% of the pilots agreed at least to a degree that their driving performance after duty had at some point been impaired by the flight time of that duty period. Similar to the literature in other professionals and the studies on the effects of sleep deprivation on driving performance (e.g. Connor et al., 2001; Stutts et al., 2003), the pilots who usually reported sleeping less also reported higher levels of fatigue when they started their drives after duty and a more frequently experienced difficulty in staying awake during those drives. Therefore, this survey extends the previous literature by suggesting that pilots may be at an increased risk of car crashes after work when the flights are fatiguing and/or they are sleep deprived.

In contrast to the literature in other professionals, the risk of decrements of driving performance after duty was also related to the drive to work. That is, the pilots who reported usually experiencing higher levels of mental demand while driving to the airport also reported higher levels of fatigue when starting their drives back home. Moreover, reporting more mentally demanding and longer drives to work related to experiencing more often a difficulty in staying awake while driving after duty and agreeing more with the statement about impaired driving performance after flight.

To sum up, the aim of the survey was achieved because it was suggested that most pilots worldwide commute by driving and that completing long and demanding drives to the airport played a role in their levels of fatigue in flight and while driving after duty. Furthermore, the adverse impact of commuting-induced fatigue may increase even further because of the potential additional effect of sleep-related fatigue. If the findings of this survey applied to the population of professional pilots, then appropriate intervention strategies should be developed to manage the risk of fatigue-related aeroplane and car accidents when the drives to the airport are long and mentally demanding.

7.2.2 The effects of the type of fatigue induced by driving on the levels of fatigue experienced in subsequent flying and driving tasks (experiment 1)

The experiment described in chapter 5 explored if the type of fatigue induced by driving (i.e. active or passive) subsequently affects the levels of fatigue experienced in flying and driving. Sixty non-pilots completed simulated driving and flying tasks divided into three groups (i.e. manual, automated, and control). All the groups completed 32 minutes of flying tasks on the MATB-II software followed by 14 minutes of simulated motorway driving. The manual and the automated group also completed a 60-minute simulated drive before the flying tasks. The only difference between the manual and the automated group was in the level of automation of the car driven. That is, level 0 of automation was used in the manual group and of automation in the automated. Fatigue during the trials was measured with the CSS scale the PVT, and performance in the flying and motorway driving tasks. Task load during the 60-minute drive was measured on the NASA-TLX scale.

The 60-minute drive induced fatigue in both the manual and the automated group as evidenced by increased CSS ratings and PVT reaction times. This finding is in line with previous literature on the effects of long drives on drivers' fatigue (e.g. Horne and Baulk, 2004). Although the levels of fatigue at the end of the 60-minute drive did not differ between the manual and the automated group, the total task load and the ratings of effort were significantly higher during the whole drive in the manual group. In addition, the levels of fatigue increased sooner during the 60-minute drive in the automated group similar to previous studies that compared driving of different levels of automation (e.g. Feldhütter et al., 2018). However, it is known that simulators may increase drivers' levels of fatigue more quickly than real driving (e.g. Davenne et al., 2012). Moreover, the levels of task load of a drive, inter-individual differences in fatigue induction, and the potential additional effect of sleep-related fatigue are expected to play a role in how quickly driving to the airport may induce fatigue to pilots. Therefore, a duration above which driving to the airport can induce fatigue cannot be safely determined based on the findings of this study.

The more quickly apparent induction of fatigue and the lower total task load and effort in the automated group suggest that the 60-minute drive induced active fatigue in the manual group and passive in the automated. If these findings applied to the population of pilots, a long drive to the airport might induce fatigue whether it was demanding or not. Nevertheless, fatigue after that drive would be expected to have different qualitative characteristics depending on whether it was caused by high or low task load (Desmond and Hancock, 2001). Aviation authorities, airlines, and pilots should consider these characteristics because, as explained in the next paragraphs, they could make a difference in the levels of fatigue experienced in flight and while driving after duty.

Based on the findings of the experiment described in chapter 5, professional pilots may be at an increased risk of fatigue-related aeroplane accidents after a long drive to the airport. This suggestion is based on the finding that the manual group reported higher levels of fatigue than the control just before and during the whole duration of the MATB-II tasks. Furthermore, the manual group performed worse than the control in the tracking task of the first multi-tasking block of the flight. The automated group also reported higher levels of fatigue than the control at the beginning of the flying tasks. These findings agree with the literature that suggests that pilots should not complete long commutes to the airport (e.g. Federal Aviation Administration, 2012[a]). However, in contrast to the existing literature, this study provided evidence of the risk of flying fatigued specifically after driving.

This thesis complimented the previous literature by suggesting that pilots' levels of fatigue in flight may be affected not only by the duration of the commutes to the airport but also by the type of fatigue induced. More specifically, the manual and the automated group completed a drive of the same duration (i.e. 60 minutes) before the MATB-II tasks and did not report significantly different levels of fatigue at the end of that drive or the beginning of the simulated flight. However, the manual group reported higher levels of fatigue than the automated at the end of the flying tasks, potentially because of a lower resilience in the induction of fatigue in the manual group. The lower resilience was evidenced by significant within-group increases of the CSS ratings until the end of the MATB-II tasks in the manual group but not in the automated one. The difference in this resilience can be explained by Desmond and Hancock's (2001) model of fatigue. Drawing from that, a long and demanding drive can deplete individuals' attentional resources with the need to continue to invest effort in subsequent flying tasks inducing even more fatigue. A long drive of low levels of task load can also cause fatigue, but individuals could have some attentional resources reserved because passive fatigue relates to a disengagement from the driving task. The participants did not disengage fully from the task because level 2 automation required monitoring the driving scene continuously. Therefore, some attentional resources were invested in the automated drive too. Nevertheless, the reservation of some attentional resources resulted in delaying the induction of fatigue in the flying tasks.

If the findings of this study applied to pilots, then commutes to the airport that induce active fatigue (e.g. drives in high traffic density) should be considered as potentially more hazardous than those that induce passive fatigue (e.g. when pilots commute by train). In addition, aviation authorities, airlines, and pilots might benefit by using intervention strategies that are appropriate for the type of fatigue caused by commuting. For example, the study of chapter 5 suggested that increasing task load can have an alerting effect when fatigue is passive but a negative impact when active.

The flying tasks induced fatigue in the manual group since the CSS ratings, the PVT reaction times, and the errors in the communications task increased with time. These results were expected based on the literature solely on pilots that suggests that flying can be fatiguing (e.g. Morris and Miller, 1996). If the findings of this study applied to professional pilots, they could experience a combination of commuting- and flying-induced fatigue at the end of their duties. Therefore, using an intervention strategy for commuting-induced fatigue before flight may help pilots, but additional intervention strategies in flight might be needed to compensate for the induction of fatigue in flight. The study described in chapter 6 provided evidence of that need. Similar, pilots may also need to consider using intervention strategies for fatigue after their duties to reduce the likelihood of car crashes caused by fatigue induced while driving to the airport and flying.

Of all the metrics used to infer fatigue, the CSS ratings were the most helpful in deducing differences in fatigue between the groups, followed by the measurement of flying and driving performance. The results of this study considered, using a combination of subjective scales and performance data appears to be an informative approach of measuring fatigue.

In general, the findings of the study described in chapter 5 suggest that aviation authorities, airlines, and pilots may need to approach the issue of commuting-related fatigue by looking at the whole commuting cycle rather than only driving or flying. This is not only supported by the between-group differences in the expressions of fatigue in flight mentioned but also to their differences in fatigue in the 14-minute drive. That is, fatigue while driving after flight did not relate only to the preceding flying tasks but to whether a long drive preceded the flight and what type of fatigue that drive induced. More specifically, the CSS ratings in the motorway drive were significantly higher in the manual than in the control and the automated group. Moreover, the manual group maintained the target driving speed less accurately than the automated one. The survey described in chapter 4 had also indicated that long and mentally demanding drives to the airport related to higher levels of fatigue while driving after duty.

The risk of fatigue-related crashes after flight has not been explored in pilots, but there is evidence of an increased risk of impaired driving performance after work from studies in other professionals (e.g. Barger et al., 2005). None of those studies had used simulated tasks as in the study in chapter 5. Furthermore, the risk of crashes after work had been linked there to sleep- and work-related factors. Similar to that literature, the study described in chapter 5 suggests that a fatiguing duty can increase the risk of car crashes after it but extends the existing knowledge by also linking that risk to fatigue induced by driving to work.

7.2.3 The effects of short bursts of physical activity on fatigue experienced in the commuting cycle after a long and demanding drive (experiment 2)

The study described in chapter 6 investigated if short bursts of physical activity can reduce the levels of fatigue experienced in the commuting cycle after a long and demanding drive. Fifteen participants with a Private Pilot's License completed two sessions that included the 60-minute simulated drive used in the experiment of chapter 5, 52 minutes of flying tasks on the X-Plane 10 software, and 20 minutes of simulated motorway driving. In contrast to the experiment of chapter 5, the motorway drive included both sections of low and moderate task load to explore if that would make a difference in the effects of fatigue on participants' driving performance after flight. The only difference between the two sessions was the use of two 6-minute breaks in only one of them. The first break was taken before the simulated flight and the second after it in the break session. The breaks included walking, jogging in place, star jumps, and discussion with the experimenter. Fatigue was measured with the CSS scale, the PVT, and performance in the flying and driving tasks.

The 60-minute drives induced fatigue in both sessions, which, according to the study described in chapter 5, should have been active. The induction of fatigue in the study of chapter 6 was evidenced by within-session increases of the CSS ratings and the PVT reaction times with the progression of the 60-minute drives. These findings agree with the literature solely on drivers that suggests that long and demanding driving can induce fatigue (e.g. Cantin et al., 2009). The medians of self-reported levels of fatigue at the end of these drives were moderate. Nonetheless, if the findings of this study applied to the population of professional pilots, they could still be at an increased risk of reporting fatigued for duty due to active fatigue induced by driving to the airport.

In the break session, the CSS ratings after the first and the second break were significantly lower than those at the end of the 60-minute drive and the flying tasks respectively. This finding suggests that the breaks used in this study had an alerting effect potentially due to the mental disengagement from flying and driving, the postural change, and the physical activity. As expected, the participants reported lower levels of fatigue in flight and performed better in the flying tasks in the break session than in the no-break one. The positive effects of the first break on flying performance were evidenced by a higher accuracy in maintaining the target heading and airspeed and fewer errors in the communications task. Moreover, two participants completed turns in the wrong direction in the flight of the no-break session. The breaks helped the participants to be less fatigued in the motorway drive too as suggested by the lower CSS ratings and the higher accuracy in maintaining the target speed while driving in the break session. If these findings applied to professional pilots, active fatigue induced by driving to the airport and fatigue caused by flying would increase the risk of aeroplane and car accidents. In that case, taking a break with physical activity before and after their duties could help them to reduce their levels of fatigue.

The between-session differences in flying and driving performance were expected because fatigue impairs attention, information processing, and response execution (e.g. Boksem et al., 2005). The alerting effects of the physical activity, the postural change, and the mental disengagement from the primary task have been shown in previous studies too (e.g. Abd-Elfattah et al., 2015). Some of these studies used longer breaks than the one tested in this study (e.g. Hogervorst et al., 1996). However, the findings of this experiment suggest that breaks as long as 6 minutes can also be effective against task-related fatigue similar to Neri et al. (2002). In addition, they suggest that adding physical activity to the postural change and the mental disengagement can extend the alerting effect of breaks to more than the 25 minutes found by Neri et al. (2002). This considered, if the findings of this study applied to professional pilots, taking breaks before and after flights that are similar to the one tested in this study could be a practical way of reducing the risk of fatigue-related aeroplane and car accidents when driving to the airport induces active fatigue.

Some pilots may find it impractical in real life to take breaks that include jog in place and star jumps before they enter the aeroplane and before they start their drives after duty. Moreover, it could be argued that pilots do not need to take breaks with physical activity before and after their flights because they already walk from the car park to the terminal and the aeroplane and walk from the aeroplane to their car after flight. However, as explained, any physical activity taken while at the airport would be more likely to be effective if it was not too long (to avoid inducing fatigue) and increased individuals' heart rate within an optimal range. This considered, quick walking to and from the aeroplane might be effective as well, but it needs investigation.

In real life, pilots spend time in the aeroplane preparing before flight. No research was located on how these tasks affect pilots' fatigue, but it is possible that any alerting effect of a break taken before the flights dissipates before takeoff. Irrespective of how pre-flight tasks affect pilots, it might be preferable to take a break with physical activity as much as possible nearer the start of the flights to increase the likelihood of maintaining alert while flying. Despite any beneficial effect of a break before takeoff on fatigue in flight, this effect could soon disappear because flying is fatiguing. In line with the literature on pilots (e.g. Morris and Miller, 1996), this was suggested in this study by finding an increase of the CSS ratings and the frequency of the errors in the communications task during the X-Plane 10 scenario in the break session. This considered, pilots might need to use additional intervention strategies during flights, especially when these are long.

Similar to flights, pilots might be benefitted from taking a break with physical activity before driving back home, but its alerting effect may not last for long because driving can be fatiguing (e.g. Thiffault and Bergeron, 2003). This was supported in the study of chapter 6 by finding that, despite taking a break before the motorway drive, the SD of the lateral position and the SD of the headway increased significantly during the drive. Hence, a break should be taken as much as possible before entering the car. At the same time, additional strategies for fatigue might be needed during the drives.

Aviation organisations and studies on commuting pilots and other professionals have suggested that completing long commutes to get to the airport may increase the risk of impaired flying performance in professional pilots (e.g. Azwar et al., 2018). The findings of the study described in chapter 6 are in line with this literature. Nevertheless, in contrast to the existing literature, these findings are based on experimental work and focused specifically on commuting to the airport by completing long and demanding drives.

Previous studies suggest an increased risk of car crashes after work in professionals other than pilots (e.g. Barger et al., 2005). This risk has neither been mentioned by aviation authorities nor researched. The findings of the study described in chapter 6 extend this literature by suggesting that pilots can also be at a similar risk while driving after duty. In addition, the risk of car crashes after work has been associated in the existing literature with sleep-related factors, a preceding work shift, or the drive after work itself. Besides the role of fatigue induced by working, this study suggests a link of the risk of car crashes after duty to fatigue induced by commuting to work as well. Therefore, researchers might benefit in the future by exploring the whole commuting cycle in order to identify the factors that contribute to fatigue while driving after duty.

To sum up, if the findings of this study applied to the population of professional pilots, that would mean that they could be at an increased risk of aeroplane accidents and car crashes after duty due to fatigue induced by completing long and demanding drives to the airport and flying. However, this risk could reduce by taking short breaks that combine physical activity, postural change, and mental disengagement from the primary task before and after flight. The potentially beneficial effects of these breaks may not last for long. Therefore, pilots should take them as close as possible to the flights and drives and consider using additional intervention strategies during these tasks. In addition, fatigue in commuting pilots cannot be only induced by commuting to the airport. For example, pilots may also be sleep deprived and awake for many hours before reporting for duty. Therefore, any interventions for active fatigue induced by driving to the airport (e.g. by taking a break similar to the one tested in this thesis) should be accompanied by interventions for sleep-related fatigue.

7.3 Limitations and further study

The studies described in this thesis improved the understanding of the adverse effects of active fatigue induced by driving to the airport on fatigue in flight and while driving after duty. However, they have some limitations. These limitations, along with recommendations about overcoming them in the future, are outlined in the next sections.

7.3.1 Survey

- The survey had an exploratory purpose due to the lack of previous studies on fatigue in commuting pilots. This means that it aimed to identify potential issues that could be further explored with experiments. As a result, there were questions about the task load and the duration of the drives and flights, fatigue, and performance impairment, but not questions that would link them directly. Future surveys could benefit from more direct questions on these links. For example, pilots could be directly asked about the effects of the task load and duration of driving to the airport on the levels of fatigue experienced in flight.

- The survey provided only a snapshot of pilots' usual commuting habits. These habits may vary in time, for example, because pilots decide to use public transport instead of their private vehicles. It is also likely that the duration of their drives differs between days, for example, because of the traffic density. Furthermore, pilots' commuting patterns may vary due to sleeping at hotels near the airport. Researchers with more resources could collect data about the effect of these variations of commuting habits on pilots' fatigue.

- The time spent at the airport before flight usually helped the pilots to reduce their levels of fatigue, but this was not the case with the time spent at the airport after flight. Future surveys could explore the role of the time spent at the airport on commuting pilots' fatigue by collecting data about the effect of pre- and post-flight tasks and the self-initiated intervention strategies used.

- Pilots working for a mainline airline and pilots with a home base in Europe were overrepresented in the survey. The commuting habits of pilots may vary depending on the type of employer because of potentially different scheduling patterns. Moreover, as mentioned in the survey of chapter 4, pilots' commuting habits in Europe may differ from those of pilots in other regions, such as in the US. Future studies could investigate if these factors affect pilots' commuting habits and fatigue during the commuting cycle.

- Fatigue in the survey was measured retrospectively. This approach was considered the most appropriate to collect data from as many pilots as possible worldwide. Nevertheless, people may not be able to rate their levels of fatigue accurately retrospectively due to the difficulty in recalling how fatigued they were. In addition, the CSS is used to rate fatigue on a specific moment in time. In contrast, in this survey, the pilots were asked to rate their usual levels of fatigue during the commuting cycle rather than fatigue during specific commutes and flights. This decision was made to explore the extent to which fatigue induced by long and demanding drives to the airport affects negatively professional pilots' fatigue in flight and while driving after duty. The comparisons between the ratings of fatigue on consecutive measurement time points also helped to reveal how these might have been related to driving, pre-flight tasks, flying, and post-flying activities. Nevertheless, asking the pilots about their usual levels of fatigue resulted in missing information about the variability of the ratings between days. This considered, the results of the analysis of the CSS data were treated as indications rather than definite links and differences because of these limitations. Researchers with more resources could conduct surveys to collect fatigue data about specific commutes and flights.

- Fatigue was not inferred from the analysis of physiological data due to limitations of the equipment used (i.e. eye tracker and wrist-worn device). Collecting this type of data when exploring fatigue in commuting pilots could help future researchers to detect variations of the levels of fatigue during driving and flying.

- The effect of sleep-related factors on fatigue could not be understood fully. A number of sleep-related factors can cause fatigue including operating during the circadian low, experiencing a disruption of the circadian rhythms due to time zone crossing, being sleep deprived and awake for a long time, and not sleeping well. Inter-individual differences in the preference for the time of the day and the response to sleep deprivation can play a role in fatigue too. Investigating in-depth the role of sleep-related fatigue was out of the scope of this thesis. Researchers with more resources could collect detailed sleep data (e.g. with morningness/eveningness questionnaires) when exploring fatigue in commuting pilots.

7.3.2 Behavioural experiments

- Older age may relate to higher levels of sleepiness (e.g. Harma et al., 1994) but the participants in the experiments were younger than the mean age of pilots worldwide (e.g. Civil Aviation Authority, 2018). Therefore, the adverse impact of fatigue induced by driving to the airport and flying may be exacerbated in the real world. Future studies with better access to pilots could investigate if this impact differs between younger and older pilots.

Using simulators is a safe and practical way of exploring driving and flying performance. Nonetheless, it also has disadvantages:

- The motivations for participation in simulator studies may differ from those when driving or flying in the real world. That is, participants in simulator studies are usually motivated by financial rewards and curiosity about the findings (Stunkel and Grady, 2011). In contrast, in real life, pilots drive to go to work and fly because this is their job. Hence, they are expected to be more motivated to perform well than the participants in the experiments of this thesis because of the safety risks associated with impaired performance in the real world.

- Simulated driving might induce fatigue more quickly than real driving (e.g. Davenne et al., 2012) and performance in simulated flying tasks may be more susceptible to the effects of fatigue than performance in real flights (Billings et al., 1975; Caldwell and Roberts, 2000). Moreover, low fidelity flying simulators may not predict flying performance on real aeroplanes (De Winter et al., 2012). Similar, driving might be more erratic in simulators than in real life (Blana, 2001; Blana and Golias, 2002; Reed and Green, 1999) especially when simulator fidelity is low (Jamson and Jamson, 2010). Therefore, the performance values in the studies described in chapters 5 and 6 should be interpreted as relative rather than absolute. Future studies might benefit from investigating commuting-induced fatigue in real-world operations.

- Pilots often use self-initiated intervention strategies when fatigued in-fight, such as reducing the workflow (Dawson et al., 2012). However, the participants in the experiment of chapter 6 did not have any control over the pace of the flying tasks. They were also not allowed to stand up to stretch or walk, which pilots can do when fatigued. Pilots may also use intervention strategies before and after their duties, such as sleeping in rest facilities. These strategies may help them to mitigate the effects of fatigue induced by driving to the airport and flying. Nonetheless, only one of the intervention strategies for fatigue was tested in the experiment described in chapter 6. Future studies could explore fatigue in commuting pilots when other intervention strategies are used while at the airport and in flight.

- Fatigue can hinder decision-making in pilots by increasing the difficulty in storing information in the working memory and planning and causing mental confusion (e.g. Falleti et al., 2003). The participants in the studies described in chapters 5 and 6 could have developed strategies to improve their multi-tasking performance. However, they were not asked to make decisions as pilots do in real life, for example, when they detect bad weather on their flight path. Researchers could explore in the future the effects of driving-induced active fatigue on pilots' decision-making in flight.

- The duration of the flying tasks in the experiments resembled that of some domestic flights but they could be longer to represent international flights as well. Similar, the motorway drives after the flying tasks could be longer as pilots may drive more than 20 minutes after duty.
- The participants in the experiment of chapter 6 completed the flying tasks alone, whereas commercial flights require at least two pilots in the cockpit. A second pilot can help to avoid errors by taking over additional tasks when the other pilot is fatigued. Moreover, pilots might reduce the likelihood of experiencing passive fatigue in the low task load stages of flight by discussing with their colleague. If that was the case, any adverse effects of fatigue induced by driving to the airport and flying might be mitigated. This considered, future studies could explore the effects of commuting-induced fatigue on flying performance when two pilots are in the cockpit.
- As in the survey, the role of the sleep-related factors in fatigue was not fully understood because this was out of the scope of the thesis. Researchers who explore fatigue in commuting pilots in the future might benefit from collecting detailed data about the factors that contribute to sleep-related fatigue. For example, the disruption of the circadian rhythms due to time zone crossing could be explored by asking pilots about their recent travels. In addition, future studies could explore how sleep deprivation, the time of the day, the time awake since the last sleep, circadian disruption, and sleep quality interact with task-related fatigue in pilots who commute by driving.

7.4 Contribution to existing knowledge

The research described in this thesis extends the understanding of the relationship between fatigue induced by long and demanding driving and pilots' fatigue in flight and while driving after duty in the following ways;

- It is the first to collect data about the commuting habits of pilots worldwide and suggest that driving is the most common means of commuting among professional pilots.
- This thesis is the first to suggest based on pilots' self-reports that many of them are at an increased risk of aeroplane accidents and car crashes after duty when they complete long and

demanding drives to the airport. It is also the first time that this risk was explored in experiments. Moreover, no other study that used both driving and flight simulation was sourced.

- A methodology that could be used in the future to explore commuting-related fatigue in pilots and other professionals was developed.
- The work described in this thesis suggests that using the literature solely on drivers and pilots to explore fatigue in commuting pilots may not capture all the aspects of the issue. That is, drawing from driver studies, the more fatigued the pilots finish their commutes to the airport, the more fatigued they would be in flight. However, the experiment in chapter 5 showed that the type of fatigue induced by driving plays a role too.
- This thesis provided evidence that short bursts of medium physical intensity can be an effective intervention strategy for fatigue induced by driving to the airport and flying in commuting pilots. The positive effects of using these short breaks may last longer than the breaks tested in other studies.

7.5 Practical implications

The research described in this thesis suggests an increased risk of impaired performance in flight and while driving after duty after long and demanding drives. The findings of this thesis should be validated in larger sample sizes before being used to;

- **Develop commuting policies.** Currently, only some airlines use policies regarding commuting. Drawing from the findings of the study described in chapter 5, more airlines and pilots might benefit from policies, such as using minibuses or taxis for pilot commuting. Travelling as a passenger might also induce fatigue, but the pilots would have the option to sleep during the commute or stay alert by discussing with other passengers and surfing the internet on their mobile phone. In addition, flying tasks could have an alerting effect when driving to the airport induced passive fatigue in contrast to active.

  Another policy currently in use by airlines is to ask the pilots to live within a certain range from their home base. This policy is mainly aimed to ensure that pilots on stand-by duty arrive quickly at the airport. Although reducing the time spent on commuting to the airport might help the pilots to avoid reporting fatigued for duty, this approach does not consider the effect of delays that may occur even in short commutes (e.g. due to high traffic density) and the means of transport used (e.g. car or aeroplane). As a result, the time on task may be longer than expected based on the distance from the airport. Airlines might benefit from updating their policies to also consider the usual duration of pilots' commutes and the means of transport used in addition to the distances covered.

- **Provide training**. Airlines provide fatigue training through Fatigue Risk Management Systems (FRMSs) to help pilots to avoid, detect, and manage high levels of fatigue before and during their flights. However, the FRMSs do not include information about the extent of the issue amongst pilots, how driving to the airport can affect flying and driving after duty, and what pilots could do to mitigate the associated safety risk. The findings of this thesis could be used to train pilots regarding the risk of fatigue-related impaired performance in flight and while driving after duty. Information about what can be done to manage that risk (e.g. breaks, use of rest facilities, and commuting by public means of transport) could be included too. Due to the high variability in pilots' commuting habits, training in commuting-induced fatigue might be one of the most effective interventions.

- **Improve scheduling.** The findings of this thesis could be used to improve the biomathematical models that predict pilots' fatigue in flight. These models are used by airline schedulers to provide quantitative and qualitative predictions of fatigue and performance in flight using equations. Based on the model outputs, airlines determine the start and finish times of duties, their total duration, the number of consecutive duties, and the days off work. The problem with these models is that their algorithms consider only sleep-related factors known to increase the levels of fatigue, such as the time of the day, the hours of sleep, and the time awake since last sleep. These models also consider the duration of commuting to and from the airport, but they do not take into consideration the levels of task load during commuting.

  The study described in chapter 5 suggested that the type of fatigue induced by driving can affect the levels of fatigue experienced in subsequent flying and driving tasks. Future studies could explore the role of various levels of task load (e.g. low, moderate, high) on fatigue in commuting pilots before continuing to any changes in these models. However, based on the findings of the study in chapter 5, adding a parameter of task load in the algorithms might increase the accuracy of the predictions. Airlines might be then able to adjust their scheduling practices to accommodate commuting habits. For example, a pilot might need the double time to cover the same distance compared to another pilot because of the need to commute by driving a very busy road. In that case, it might work better in terms of fatigue if the number of commutes in a month reduced for that pilot. That might be achieved, for example, by changing the schedules so that the duties were completed in blocks with sleep in-between at hotels near the airport.

- **Improve fatigue regulations.** The flight time limitation regulations do not count the time spent on commuting to or from the airport. More specifically, the flight time limitation that applies to pilots in Europe does not mention the risk of fatigue associated with commuting at all (European Union Aviation Safety Agency, 2017). On the other hand, the relevant regulation for the US mentions commuting as an important fatigue-related issue but does not regulate it (Federal Aviation Administration, 2011). Moreover, no regulations refer to the risk of car crashes after duty due to fatigue. Regulating pilots' commutes is difficult (as suggested in the interview of chapter 4) because this activity is part of their personal life. However, the Federal Aviation Administration (2011) stated that it is likely to conduct a rulemaking regarding pilot commuting to address the issue of fatigue in flight if more information was available about the risks posed by commuting and the potential countermeasures.

Regulators could use the findings of this thesis to explore further the effects of fatigue induced by commuting to the airport in large-scale studies that would involve professional pilots and complex commuting patterns. Depending on their findings, the regulations could be modified to include control mechanisms for commuting-related fatigue, such as systems to collect systematic data about pilots' commutes. Regulators could also raise awareness about the importance of managing the risk of accidents associated with commuting.

- **Help other commuting professionals.** The study described in chapter 5 explored the fundamental aspects of the effects of the type of fatigue induced by driving on fatigue experienced in subsequent flying and driving with non-pilots and low fidelity flying tasks. These tasks are similar to tasks completed by other professionals. Moreover, the benefits of the breaks used in the experiment of chapter 6 are expected to be effective for active fatigue induced by driving to work in any individual. Hence, other professionals who work in safety-critical roles could benefit from the findings of this research too.

7.6 Final conclusion

This thesis helped to identify the extent of a currently ill-explored safety risk for commuting professional pilots and improved understanding of the fundamental principles behind it. Moreover, it provided evidence of the effectiveness of a self-initiated intervention strategy that could be used by pilots who complete long and demanding drives to the airport to reduce their levels of fatigue in the commuting cycle. Thus, it contributed to flight and driving safety. It is hoped that this research will inspire other researchers to explore fatigue in commuting pilots and draw the attention of regulators, airlines, and pilots to this topic in order to develop control mechanisms.

Chapter 8. List of references

Abad, V.C. and Guilleminault, C. 2012. Polysomnographic evaluation of sleep disorders. In: Aminoff, M.J. and Saunders, W.B. ed. *Aminoff's electrodiagnosis in clinical neurology*: Elsevier, pp. 727-762.

Abd-Elfattah, H.M., Faten H. and Abdelazeim, S.E. 2015. Physical and cognitive consequences of fatigue: a review. *Journal of Advanced Research.* **6**(3), pp. 351-358. Doi: 10.1016/j.jare.20 15.01.011.

Aeschbach, D., Cajochen, C., Tobler, I., Dijk, D.J. and Borbely, A.A. 1994. Sleep in a sitting position: effect of triazolam on sleep stages and EEG power spectra. *Psychopharmacology*. **114**(2), 209-214. Doi: 10.1007/bf02244838.

Ahlstrom, U. and Friedman-Berg, F. 2006. Using eye movement activity as a correlate of cognitive workload. *International Journal of Industrial Ergonomics.* **36**, pp. 623-636. Doi: 10.1016/j.ergon.2006.04.002.

Airliners 2019. *Airbus A320-232 - JetBlue Airways*. [Online]. [Accessed 11 December 2019]. Available from: https://www.airliners.net/photo/JetBlue-Airways/Airbus-A320-232/16560.

Akerstedt, T., Ingre, M., Kecklund, G., Anund, A., Sandberg, D., Wahde, M., Philip, P. and Kronberg, P. 2010. Reaction of sleepiness indicators to partial sleep deprivation, time of day and time on task in a driving simulator – the DROWSI project. *Journal of Sleep Research*, **19**, pp. 298-309. Doi: 10.1111/j.1365-2869.2009.00796.x.

Akerstedt, T., Peters, B., Anund, A. and Kecklund, G. 2005. Impaired alertness and performance driving home from the night shift: a driving simulator study. *Journal of Sleep Research*. **14**(1), pp. 17-20. Doi: 10.1111/j.1365-2869.2004.00437.x.

Alessi, S.M. 1988. Fidelity in the design of instructional simulations. *Journal of Computer Based Instruction.* **15**(2), pp. 40-47.

Allen, M.T., Obrist, P.A., Sherwood, A. and Crowell, M.D. 1987. Evaluation of myocardial and peripheral vascular responses during reaction time, mental arithmetic, and cold pressor tasks. *Psychophysiology*. **24**(6), pp. 648-656. Doi:10.1111/j.1469-8986.1987.tb00345.x.

Anund, A., Kecklund, G., Peters, B. and Akerstedt, T. 2008. Driver sleepiness and individual differences in preferences for countermeasures. *Journal of Sleep Research*. **17**(1), pp.16-22. Doi: 10.1111/j.1365-2869.2008.00633.x.

Apparies, R., Riniolo, T. and Porges, S.A. 1998. Psychophysiological investigation of the effects of driving longer-combination vehicles. *Ergonomics.* **41**, pp. 581-592. Doi: 10.1080/001401

398186766.

Archer, J., Keno, H. and Kwon, Y. 2012. *Effects of Automation in the Aircraft Cockpit Environment: Skill Degradation, Situation Awareness, Workload.* Final project. West Lafayette: Purdue University.

Azwar, A. Susilowati, I., Dinar, A., Indriyani, K. and Wirawan, M. 2018. Impact of work-related and non-work-related factors on fatigue in production/shift workers. *KnE Life Sciences*. **4**, pp. 213-224. Doi: 10.18502/kls.v4i5.2554.

Azizan, A., Fard, M., Azari, M.F., Benediktsdóttir, B., Arnardóttir, E.S., Jazar, R. and Maeda, S. 2016. The influence of vibration on seated human drowsiness. Industrial Health. **54**(4), pp. 296-307. Doi: 10.2486/indhealth.2015-0095.

Badia, P., Myers, B.L., Boecker, M., Culpepper, J. and Harsh, J. 1991. Bright light effects on body temperature, alertness, EEG and behaviour. *Physiology and Behavior.* **50**, pp. 583-588. Doi: 10.1016/0031-9384(91)90549-4.

Balkin, T.J. and Wesensten, N.J. 2011. Differentiation of sleepiness and mental fatigue effects. In: Ackerman, L.P. ed. Cognitive Fatigue. *Multidisciplinary perspectives on current research and future applications*: American Psychological Association, pp. 47-66.

Barger, L.K., Cade, B.E., Ayas, N.T., Cronin, J.W., Rosner, B., Speizer, F.E. and Czeisler, C.A. 2005. Extended work shifts and the risk of motor vehicle crashes among interns. *The New England journal of medicine*. **352**(2), pp.125-134. Doi: 10.1056/NEJMoa041401.

Basner, M., David, F. and Dinges, D.F. 2011. Maximizing sensitivity of the Psychomotor Vigilance Test (PVT) to sleep loss. *Sleep.* **34**(5), pp. 581-591. Doi: 10.1093/sleep/34.5.581.

Basner, M., Hermosillo, E., Nasrini, J., McGuire, S., Saxena, S., Moore, T.M., Gur, R.C. and Dinges, D.F. 2017. Repeated administration effects on Psychomotor Vigilance Test Performance. *Sleep.* **41**(1), pp. 1-6. Doi: 10.1093/sleep/zsx187.

Baulk, S., Biggs, S., Reid, K., van den Heuvel, C. and Dawson, D. 2008. Chasing the silver bullet: measuring driver fatigue using simple and complex tasks. *Accident; analysis and prevention.* **40**, pp. 396-402. Doi: 10.1016/j.aap.2007.07.008.

Baulk, S.D., Biggs, S., van den Heuvel, C., Reid, K. and Dawson, D. 2006. *Managing driver fatigue: quantifying real world performance impairment*. 2006/01 final report. [Online]. Canberra city: Australian Transport Safety Bureau. [Accessed 11 December 2019]. Available from: https://www.infrastructure.gov.au/roads/safety/publications/2006/pdf/grant _200601.pdf.

Belenky, G., Wu, L.J., Zaslona, J.L. and Hodges, J. 2012. *Motorcoach driver fatigue study.*

FMCSA-RRR-12-042. [Online]. Washington: Department of Transportation. [Accessed 5 May 2019]. Available from: https://trid.trb.org/view/1261157.

Bent, B. and Dunn, J.P. 2020. Optimizing sampling rate of wrist-worn optical sensors for physiologic monitoring. Journal of Clinical and Transnational Science. Pp. 1-8. Doi: 10.1017/cts.2020.526.

Billings, C.E., Gerke, R.J. and Wick, R.L. 1975. Comparisons of pilot performance in simulated and actual flight. *Aviation, Space, and Environmental Medicine.* **46***, pp.* 304-308.

Blackwell, T., Redline, S., Ancoli-Israel, S., Schneider, J., Surovec, S., Johnson, N., Cauley, J. and Stone, K. 2008. Comparison of sleep parameters from actigraphy and polysomnography in older women: The SOF study. *Sleep*. **31**(2), pp. 283-91. Doi: 10.1093/sleep/31.2.283.

Blana, E. 2001. *The behavioural validation of driving simulators as research tools: a case study based on the Leeds driving simulator*. Ph.D. thesis. University of Leeds.

Blana, E. and Golias, J. 2002. Differences between vehicle lateral displacement on the road and in a fixed-base simulator. *Human Factors.* **44**(2), pp. 303-313. Doi: 10.1518/001872002449 7899.

Bleil, M.E., Gianaros, P.J., Jennings, J.R., Flory, J.D. and Manuck, S.B. 2008. Trait negative affect: toward an integrated model of understanding psychological risk for impairment in cardiac autonomic function. *Psychosomatic Medicine*. **70**(3), pp. 328-337. Doi: 10.1097/PSY.0b013e31816baefa.

Bodala, I., Ke, Y., Mir, H., Thakor, N. and Al-Nashash, H. 2014. Cognitive workload estimation due to vague visual stimuli using saccadic eye movements. In: *Proceedings of the 36th annual international conference of the IEEE Engineering in Medicine and Biology Society, 26-30 August 2014, Chicago* [Online]. Chicago: IEEE, pp. 2993-2996. [Accessed 11 December 2019]. Available from: https://ieeexplore.ieee.org/document/6944252.

Boksem, M.A., Meijman, T.F. and Lorist M.M. 2005. Effects of mental fatigue on attention: an ERP study. *Brain Research. Cognitive Brain Research.* **25**, pp. 107–116. Doi: 10.1016/j.co gbrainres.2005.04.011.

Bonnefond, A., Doignon-Camus, N., Hoeft, A. and Dufour, A. 2011. Impact of motivation on cognitive control in the context of vigilance lowering: An ERP study. *Brain and Cognition.* **77**(3), pp. 464-471. Doi: 10.1016/j.bandc.2011.08.010.

Bonnet, M.H. 2000. Sleep deprivation. In: Kryger, M.H., Roth, T. and Dement, W.C. ed. *Principles and practice of sleep medicine*. 3rd Edn. Philadelphia: Saunders, pp. 53–71.

Bourgeois-Bougrine, S., Cabon, P., Mollard, R., Coblentz, A., Speyer, J.J., 2003. Fatigue in aircrew from short-haul flights in civil aviation: the effects of work schedules. *Human Factors and Aerospace Safety*. **3**, 177-187.

Bristol Online Survey. 2017. Jisc. [Software]. [Accessed 1 February 2017].

Brookhuis, K., De Waard, D. and Mulder, B. 1994. Measuring driving performance by car-following in traffic. *Ergonomics*. **37**(3), pp. 427-434. Doi: 10.1080/00140139408963661.

Brooks, J.O., Goodenough, R.R., Crisler, M.C., Klein, N.D., Alley, R.L., Koon, B.L., Logan, W.C., Ogle, J.H., Tyrrell, R.A. and Wills, R.F. 2010. Simulator sickness during driving simulation studies. *Accident Analysis and Prevention.* **42**(3), pp. 788-796. Doi: 10.1016/j.aap.2009.04.013.

Brown, I. 2001. Coping with driver fatigue: is the long journey nearly over?. In: Hancock, P.A. and Desmond, P.A. eds. *Stress, Workload, and Fatigue*. Mahwah: Lawrence Erlbaum, pp. 596-606.

Brown, L.J. and Whitehurst, G. 2011. The effects of commuting on crewmember fatigue. In: *Proceedings of the international symposium on aviation psychology, 2-5 May 2011, Dayton.* [Online]. Dayton: Wright State University, pp. 422. [Accessed 11 December 2019]. Available from: https://corescholar.libraries.wright.edu/cgi/viewcontent.cgi?article=1072& context=isap_2011.

Burton, D.A., Stokes, K. and Hall, G.M. 2004. Physiological effects of exercise. *Continuing Education in Anaesthesia Critical Care and Pain.* **4**(6), pp. 185-188. Doi: 10.1093/bjaceaccp /mkh056.

Buysse, D.J., Reynolds, C.F., Monk, T.H., Berman, S.R. and Kupfer, D.J. 1989. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. *Journal of Psychiatric Research.* **28**(2), pp. 193-213. Doi: 10.1016/0165-1781(89)90047-4.

Cabon, P., Coblentz, A. Mollard, R. and Fouillot, J. 1993. Human vigilance in railway and long-haul flight operation. *Ergonomics.* **36**, pp. 1019-1033. Doi: 10.1080/00140139308967974.

Caffier, P.P., Erdmann, U. and Ullsperger, P. 2003. Experimental evaluation of eye-blink parameters as a drowsiness measure. *European Journal of Applied Physiology*. 89(3-4), pp. 319-325. Doi: 10.1007/s00421-003-0807-5.

Cajochen, C., Zeitzer, J.M., Czeisler, C.A. and Dijk, D.J. 2000. Dose-response relationship for light intensity and ocular and electroencephalographic correlates of human alertness. *Behavioural Brain Research.* **115**, pp. 75–83. Doi: 10.1016/s0166-4328(00)00236 -9.

Caldwell, J.A., Caldwell, J.L., Brown, D.L. and Smith, J.K. 2004. The effects of 37 hours of continuous wakefulness on the physiological arousal, cognitive performance, self-Reported mood, and simulator flight performance of F-117A pilots. *Military Psychology*. **16**(3), pp. 163-181. Doi: 10.1207/s15327876mp1603_2.

Caldwell, J. and Ramspott, S. 1998. Effects of task duration on sensitivity to sleep deprivation using the Multi-Attribute Task Battery. *Behavior Research Methods, Instruments, and Computers*. **30**, pp. 651-660. Doi: 10.3758/BF03209483.

Caldwell, J.A., and Roberts, K.A. 2000. Differential sensitivity of using simulators versus actual aircraft to evaluate the effects of a stimulant medication on aviator performance. *Military Psychology* 12(4), 277-291. Doi: 10.1207/S15327876MP1204_04.

Caldwell, J.A., Mallis, M., Caldwell, J.L., Paul, M., Miller, J. and Neri, D. 2009. Fatigue countermeasures in aviation. *Aviation Space and Environmental Medicine*. **80**(1), pp. 29-59. Doi: 10.3357/asem.2435.2009.

Caldwell, J.A., Prazinko, B. and Caldwell, J.L. 2003. Body posture affects electroencephalographic activity and Psychomotor Vigilance Task performance in sleep-deprived subjects. *Clinical Neurophysiology*. **114**(1), pp. 23-31. Doi: 10.1016/s1388-2457(0 2)00283-3.

Campbell, R.D. and Bagshaw, M. 2002. *Human Performance and Limitations in Aviation*. New Jersey: John Wiley & Sons.

Cantin, V., Lavalliere, M., Simoneau, M. and Teasdale, N. 2009. Mental workload when driving in a simulator: effects of age and driving complexity. *Accident Analysis & Prevention*. **41**(4), pp. 763-771. Doi: 10.1016/j.aap.2009.03.019.

Carskadon, M.A. 1989. Measuring daytime sleepiness. In: Kryger, M.H., Roth, T., Dement, W.C. eds. *Principles and Practice of Sleep Medicine*. 3rd Edn. Philadelphia: Saunders, pp. 961-962.

Caska, T.J. and Molesworth, B. 2007 The effects of low dose caffeine on pilot performance. *International Journal of Applied Aviation Studies*. **7**, pp. 244-255.

Chang, C., Leopold, D.A., Schölvinck, M.L., Mandelkow, H., Picchioni, D., Liu, X., Frank, Q.Y., Turchi, J.N. and Duyn, J.H., 2016. Tracking brain arousal fluctuations with fMRI. In: *Proceedings of the National Academy of Sciences*. **113**(16), pp. 4518–4523. Doi: 10.1073/p nas.1520613113.

Chapman, P., Underwood, G. and Roberts, K. 2002. Visual search patterns in trained and untrained novice drivers. *Transportation Research Part F*. **5**, pp. 157-167. Doi: 10.1016/S1 369-8478(02)00014-1.

Civil Aviation Authority 2018. Licence Holders by Age and Sex. [Online]. [Accessed 15 March 2020]. Available from: https://www.caa.co.uk/Data-and-analysis/Approved-persons-and-organisations/Datasets/Licence-holders-by-age-and-sex/.

Clarke, L.L., Wilson, S. and Kirwan, J.R. 2013. Using actigraphy to measure sleep patterns in rheumatoid arthritis: A pilot study in patients taking night-time prednisone. Musculoskeletal Care. **11**(3), pp. 179-185. Doi: 10.1002/msc.1052.

Co, E.L., Gregory, K.B., Johnson, J.M. and Rosekind, M.R. 1999. *Crew Factors in Flight Operations XI: A Survey of Fatigue Factors in Regional Airline Operations*. NASA/TM-1999-208799. [Online]. California: National Aeronautics and Space Administration. [Accessed 11 December 2019]. Available from: https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20000032967.pdf.

Commission Regulation (EU) No. 83/2014 of 29 January 2014. [Online]. [Accessed 11 December 2019]. Available from: https://www.easa.europa.eu/document-library/regulations/commission-regulation-eu-no-832014.

Connor, J., Whitlock, G., Norton, R. and Jackson, R. 2001. The role of driver sleepiness in car crashes: a systematic review of epidemiological studies. *Accident Analysis and Prevention.* **33**(1), pp. 31-41. Doi: 10.1016/s0001-4575(00)00013-0.

Craye, C., Rashwan, A., Kamel, M. and Karray, F. 2015. A multi-modal driver fatigue and distraction assessment system. *International Journal of Intelligent Transportation Systems Research.* **14**, pp. 1-17.

Crundall, D., Chapman, P., Trawley, S., Collins, L., van Loon, E., Andrews, B. and Underwood, G. 2012. Some hazards are more attractive than others: drivers of varying experience respond differently to different types of hazard. *Accident Analysis & Prevention.* **45**, pp. 600-609. Doi: 10.1016/j.aap.2011.09.049.

Cummings, P. Koepsell, T., Moffat, J. and Rivara, F. 2001. Drowsiness, counter-measures to drowsiness, and the risk of a motor vehicle crash. *Injury Prevention Journal of the International Society for Child and Adolescent Injury Prevention.* **7**. pp. 194-199. Doi: 10.1136/ip.7.3.194.

Da Silva, F.P. 2014. Mental workload, task demand and driving performance: what relation?. *Procedia - Social and Behavioral Sciences.* **162**, pp. 310-319. Doi: 10.1016/ j.sbspro.2014.12.212.

Davenne, D., Lericollais, R., Sagaspe, P., Taillard, J., Gauthier, A., Espie, S. and Philip, P. 2012. Reliability of simulator driving tool for evaluation of sleepiness, fatigue and driving performance. *Accident Analysis and Prevention.* **45**, pp. 677-682. Doi: 10.1016/j.aap.2011.

09.046.

Dawson, D., Chapman, J. and Matthew, J.W. 2012. Fatigue-proofing: a new approach to reducing fatigue-related risk using the principles of error management. *Sleep Medicine Reviews.* **16**(2), pp. 167-175. Doi: 10.1016/j.smrv.2011.05.004.

De Valck, E. and Cluydts, R. 2001 Slow-release caffeine as a countermeasure to driver sleepiness induced by partial sleep deprivation. *Journal of Sleep Research.* **10**, pp. 203-209.

De Weerd, A., de Haas, S., Otte, A., Trenité, D.K., van Erp, G., Cohen, A., de Kam, M. and van Gerven, J. 2004. Subjective sleep disturbance in patients with partial epilepsy: a questionnaire-based study on prevalence and impact on quality of life. *Epilepsia*. **45**(11), pp. 1397-404. Doi: 10.1111/j.0013-9580.2004.46703.x.

De Winter, J.C.F., Dodou, D. and Mulder, M. 2012. Training effectiveness of whole-body flight simulator motion: a comprehensive meta-analysis. *The International Journal of Aviation Psychology*. **22**(2), pp. 164-183. Doi: 10.1080/10508414.2012.663247.

Department for Transport 2015. *Transport Statistics Great Britain 2015*. [Online]. London: Department for Transport. [Accessed 11 December 2019]. Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/495910/tsgb-2015-print-ready-version.pdf.

De Souza, L., Benedito-Silva, A.A., Pires, M.L.N., Poyares, D., Tufik, S. and Calil, H.M. 2003. Further validation of actigraphy for sleep studies. *Sleep*. **26**(1), pp. 81–85. Doi: 10.1093/sleep/26.1.81.

Desmond. P. and Hancock, P. 2001. Active and passive fatigue states. In: Hancock, P. and Desmond, P. eds. *Stress, workload, and fatigue*. Mahwah: Lawrence Erlbaum Associates, pp. 455-465.

Di Stasi, L.L., Catena, A., Canas, J.J., Macknik, S.L. and Martinez-Conde, S. 2013. Saccadic velocity as an arousal index in naturalistic tasks. *Neuroscience & Biobehavioral Reviews.* **37**(5), pp. 968-975. Doi: 10.1016/j.neubiorev.2013.03.011.

Dinges, D., Pack, F., Williams, K,, Gillen, K., Powell, J.W., Ott, G.E., Aptowisz, C. and Pack, A.I. 1997. Cumulative sleepiness, mood disturbance, and psychomotor vigilance performance decrements during a week of sleep restricted to 4–5 hours per night. *Sleep.* **20**, pp. 267-277.

Dorrian, J., Tolley, C., Lamond, N., van den Heuvel, C., Pincombe, J., Rogers, A.E. and Drew, D. 2008. Sleep and errors in a group of Australian hospital nurses at work and during the commute. *Applied Ergonomics.* **39**(5), pp. 605-613. Doi: 10.1016/j.apergo.2008.01.012.

Deary, I. and Der, G. 2005. Reaction time, age, and cognitive ability: Longitudinal findings from age 16 to 63 years in representative population samples. *Aging*. **12**, pp. 187-215. Doi: 10.1080/13825580590969235.

Du, H., Zhao, X., Zhang, X., Zhang, Y. and Rong, J. 2015. Effects of fatigue on driving performance under different roadway geometries: A simulator study. *Traffic Injury Prevention*. **16**(5), pp. 468-473. Doi: 10.1080/15389588.2014.971155.

Dunn, O.J. 1964. Multiple comparisons using rank sums. Technometric., **6**(3), pp. 241-252.

Ellis, B.W., Johns, M.W., Lancaster, R., Raptopoulos, P., Angelopoulos, N. and Priest, R.G. 1981. The St. Mary's hospital sleep questionnaire: a study of reliability. *Sleep*. **4**(1), pp. 93-97. Doi: 10.1093/sleep/4.1.93.

Empatica, 2019. *E4 wristband*. [Online]. [Accessed 14 December 2019]. Available from: https://www.empatica.com/research/e4/.

Eriksen, C.A. and Akerstedt, T. 2006. Aircrew fatigue in trans-atlantic morning and evening flights. *Chronobiology International.* **23**(4), pp. 843-858. Doi: 10.1080/0742052060086557 4.

Eriksen, C., Akerstedt, T., and Nilsson, J. 2006. Fatigue in trans-atlantic airline operations: Diaries and actigraphy for two- vs. three-pilot crew. Aviation, space, and environmental medicine. **77**(6), pp. 605-612. PMID: 16780238.

European Commission 2020. *Mobility and Transport. Road Safety*. [Online]. [Accessed 31 May 2020]. Available from: https://ec.europa.eu/transport/road_safety/specialist/knowledge/fatiq ue/fatigue_and_road_crashes/conclusions_en.

European Union Aviation Safety Agency FTL 2017. Flight and duty time limitations and rest requirements. [Online]. [Accessed 11 December 2019]. Available from: https://understandin geasa2016ftl.files.wordpress.com/2017/04/easa-combined-ftl-2017.pdf.

Eurostat 2017. *Road accident fatalities – statistics by type of vehicle.* [Online]. [Accessed 2 June 2020]. Available from: https://ec.europa.eu/eurostat/statistics-explained/index.php/Road_ac cident_fatalities_-_statistics_by_type_of_vehicle.

Fairclough, S. and Venables, L. 2006. Prediction of subjective states from psychophysiology: A multivariate approach. *Biological psychology*. **71**(1), pp. 100-110. Doi: 10.1016/j.bi opsycho.2005.03.007.

Falleti, M.G., Maruff, P., Collie, A., Darby, D.G. and McStephen, M. 2003. Qualitative similarities in cognitive impairment associated with 24 h of sustained wakefulness and a blood alcohol concentration of 0.05%. *Journal of Sleep Research.* **12**(4), pp. 265-274.

Doi: 10.1111/j.1365-2869.2003.00363.x.

Federal Aviation Administration 2018. U.S. Civil Airmen Statistics. [Online]. [Accessed 15 March 2020]. Available from: https://www.faa.gov/data_research/aviation_data_statistics/civil_airmen_statistics/.

Federal Aviation Administration 2012a. *Fitness for duty*. Report No.: AC 117-3. [Online]. Washington: Federal Aviation Administration. [Accessed 11 December 2019]. Available from: https://www.faa.gov/regulations_policies/advisory_circulars/index.cfm/go/document.information/documentID/1020389.

Federal Aviation Administration 2012b. *Fatigue education and awareness training program*. Report No.: AC 117-2. [Online]. Washington: Federal Aviation Administration. [Accessed 11 December 2019]. Available from: https://www.faa.gov/regulations_policies/advisory_circulars/index.cfm/go/document.information/documentid/1020388.

Federal Aviation Administration 2011. Final rule FAA-2009-1093. Flightcrew member duty and rest requirements. [Online]. [Accessed 11 December 2019]. Available from: https://www.faa.gov/regulations_policies/rulemaking/recently_published/media/2120-aj58-finalrule.pdf.

Federal Aviation Administration 2010. *Fitness for duty*. Report No.: AC-120 FIT. [Online]. Washington: Federal Aviation Administration. [Accessed 11 December 2019]. Available from: https://beta.regulations.gov/document/FAA-2009-1093-0022.

Federal Aviation Administration 2009. *Flightcrew member duty and rest requirements*. Report No.: FAA-2009-1093; Amdt. Nos. 117-1, 119-16, 121-357 RIN 2120–AJ58. [Online]. Washington: Federal Aviation Administration. [Accessed 11 December 2019]. Available from: https://www.faa.gov/regulations_policies/rulemaking/recently_published/media/2120-aj58-finalrule.pdf.

Feldhutter, A., Hecht, T., Kalb, L. and Bengler, K. 2018. Effect of long periods of conditionally automated driving on the development of fatigue: With and without non-driving-related activities. *Cognition, Technology and Work*, **12**(1), pp. 309-318. Doi: 10.1007/s10111-018-0524-9.

Fernandez-Mendoza, J., Calhoun, S.L., Bixler, E.O., Karataraki, M., Liao, D., Vela-Bueno, A., Ramos-Platon, M.J., Sauder, K.A., Basta, M. and Vgontzas, A.N. 2011. Sleep misperception and chronic insomnia in the general population: The role of objective sleep duration and psychological profiles. *Psychosomatic Medicine*. **73**(1), pp. 88-97. Doi: 10.1097/PSY.0b013e3181fe365a.

Ferreira, C., Deslandes, A., Moraes, H., Cagy, M., Pompeu, F., Basile, L., Piedade, R. and

Ribeiro, P. 2006. Electroencephalographic changes after one night of sleep deprivation. *Arquivos de Neuro-Psiquiatria*. **64**(2-B), pp. 388-393. Doi: 10.1590/s0004-282x200600030 0007.

Flaps2Approach 2019. *Boeing 737 Master caution system*. [Online]. [Accessed 12 December 2019]. Available from: http://www.flaps2approach.com/journal/2014/3/14/boeing-737-ng-master-caution-system-six-packs-installed-and.html.

Fishman, M.G., Christina, R.W. and Anson, G.J. 2008. Memory drum theory's C movement. *Research Quarterly for Exercise and Sport*. **79**(3), pp. 312-318. Doi: 10.1080/ 02701367.2008.10599494.

Freard, D., Jamet, E., Le Bohec, O., Poulain, G. and Botherel, V. 2007. Subjective measurement of workload related to a multimodal interaction task: NASA-TLX vs. Workload Profile. In: *Proceedings of the 12th international conference on human-computer interaction: intelligent multimodal interaction environments*, 22-27 July 2007, Beijing. [Online]. Beijing: HCI International, pp. 60-69. [Accessed 28 January 2020]. Available from: https://link.springer.com/chapter/10.1007/978-3-540-73110-8_7.

Freeman, E.W., Sammel, M.D., Gross, S.A. and Pien, G.W. 2015. Poor sleep in relation to natural menopause: a population-based 14-year follow-up of midlife women. *Menopause*. **22**(7), pp. 719-726. Doi: 10.1097/GME.0000000000000392.

Friesacher, T. and Greaves, M. 2016. The effect of commuting on pilot self-assessment of stress and performance. In: *Proceedings of the International Society of Air Safety Investigators annual conference*, 17-20 October 2016, Reykjavik. [Online]. Reykjavik: ISASI, pp. 1-18. [Accessed 11 December 2019]. Available from: https://www.isasi.org/Documents/library/technical-papers/2016/Thurs/Friesacher_The%20effect%20of%20commuting%20on%20pilot%20self-assessment%20of%20stress%20and%20performance%20Amend.pdf.

Friesacher, T. 2015. *Aspects of a European airline pilot commute*. [Online]. Master thesis, Cranfield University. [Accessed 13 December 2019]. Available from: http://www.aeroexpert.eu/wp-content/uploads/2016/02/Aspects-of-a-European-Airline-Pilot-Commute-Friesacher-2015.pdf.

Fukuda, K., Stern, J.A., Brown, T.B. and Russo, M.B. 2005. Cognition, blinks, eye-movements, and pupillary movements during performance of a running memory task. *Aviation, Space and Environmental Medicine.* **76**(7), pp. 75–85.

Galy, E., Paxion, J., and Berthelon, C. 2017. Measuring mental workload with the NASA-TLX needs to examine each dimension rather than relying on the global score: an example with driving. *Ergonomics*. **61**(4), pp. 517-527. Doi: 10.1080/00140139.2017.1369583.

Galley, N. 1993. The evaluation of the electrooculogram as a psychophysiological measuring instrument in the driver study of driver behaviour. *Ergonomics*. **36**(9), pp. 1063–1070. Doi: 10.1080/00140139308967978.

Gander, P.H., Mulrine, H.M., Van den Berg, M.J., Smith, A.T., Signal, T.L., Wu, L.J. and Belenky, G. 2015. Effects of sleep/wake history and circadian phase on proposed pilot fatigue safety performance indicators. *Journal of Sleep Research.* **24**(1), pp. 110-119. Doi: 10.1111/jsr.12197.

Gershon, P., Ronen, A., Oron-Gilad, T. and Shinar, D. 2009. The effects of an Interactive Cognitive Task (ICT) in suppressing fatigue symptoms in driving. *Transportation Research Part F.* **12**, pp. 21-28. Doi: 10.1016/j.trf.2008.06.004.

Gimeno, P., Pastor, G. and Choliz, M. 2006. On the concept and measurement of driver drowsiness, fatigue and inattention: implications for countermeasures. *International Journal of Vehicle Design.* **42**, pp. 67 - 86. Doi: 10.1504/IJVD.2006.010178.

Gokhale, B.N. 2010. *Report on Accident to Air India Express Boeing 737-800 aircraft VT-AXV on 22nd May 2010 at Mangalore.* [Online]. New Delhi: Government of India Ministry of Civil Aviation. [Accessed 11 December 2019]. Available from: http://dgca.nic.in/accident/reports/VT-AXV.pdf.

Goldstein, B., Woolf, P.D., DeKing, D., DeLong, D.J., Cox, C. and Kempski, M.H., 1994. Heart rate power spectrum and plasma catecholamine levels after postural change and cold pressor test. *Pediatric Research.* **36**(3), pp. 358-363. Doi: 10.1203/00006450-199409000-00016.

Golosheykin, S., Grant, J.D., Novak, O.V., Heath, A.C. and Anokhin, A.P. 2017. Genetic influences on heart rate variability. *International Journal of Psychophysiology*. **115**, pp. 65-73. Doi: 10.1016/j.ijpsycho.2016.04.008.

Graaumans, J.M., Jackson, M., Belenky, G., Vila, B. and Dongen, H. (2010). Time on task effect in reaction times during a simulated driving task. *Sleep-Wake Research in the Netherlands*. **21**, pp. 73-76.

Green, M. 2000. How long does it take to stop? Methodological analysis of driver perception-brake times. *Transportation Human Factors* **2**(3), pp. 195-216. Doi: 10.1207/STHF0203_1.

Grier, R.A., Warm, J.S., Dember, W.N., Matthews, G., Galinsky, T.L. and Parasuraman, R. 2003. The vigilance decrement reflects limitations in effortful attention, not mindlessness. *Human Factors*. **45**(3), pp. 349-359. Doi:10.1518/hfes.45.3.349.27253.

Guo, Z., Chen, R., Zhang, K., Pan, Y. and Wu, J. 2016. The impairing effect of mental fatigue on visual sustained attention under monotonous multi-object visual attention task in long durations: An event-related potential based study. *PLOS ONE.* **11**(9), pp. 1-13. Doi: 10.1371/journal.pone.0163360.

Hakkanen, H., Summala, H., Partinen, M., Tiihonen, M. and Silvo, J. 1999. Blink duration as an indicator of driver sleepiness in professional bus drivers. *Sleep*. **22**(6), pp. 798–802. Doi: 10.1093/sleep/22.6.798.

Hamada, T., Ito, T., Adachi, K., Nakano, T. and Yamamoto, S. 2003. Detecting method for drivers' drowsiness applicable to individual features. *IEEE Xplore*. **2**, pp. 1405-1410. Doi: 10.1109/ITSC.2003.1252715.

Hammerton, G., Heron, J., Mahedy, L., Maughan, B., Hickman, M. and Murray, J. 2018. Low resting heart rate, sensation seeking and the course of antisocial behaviour across adolescence and young adulthood. *Psychological Medicine*. **48**, pp. 2194–2201. Doi: 10.1017/S0033291717003683.

Harma, M., Hakola, T., Akerstedt, T. and Laitinen, J. 1994. Age and adjustment to night work. *Occupational and Environmental Medicine.* **51**(8), pp. 568-573. Doi: 10.1136/ oem.51.8.56 8.

Harrison, Y. and Horne, J.A. 2000. The impact of sleep deprivation on decision making: A review. *Journal of Experimental Psychology: Applied.* **6**(3), pp. 236-249. Doi: 10.1037//1076-898x .6.3.236.

Hart, S.G. and Staveland, L.E. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In: Hancock, P.A. and Meshkati, N. eds. *Advances in psychology*, pp. 139–183. Doi: 10.1016/S0166-4115(08)62386-9.

Hayashi, M., Chikazawa, Y., and Hori, T. 2004. Short nap versus short rest: Recuperative effects during VDT work. *Ergonomics*. **47**(14), pp. 1549-1560. Doi: 10.1080/00140130412331293 346.

Health and safety executive 2019. *Human factors: fatigue*. [Online]. [Accessed 11 December 2019]. Available from: https://www.hse.gov.uk/humanfactors/topics/fatigue.htm.

Heinrich Heine Universität Dusseldorf. 2017. G*Power 3.1.9.3. [Software]. [Accessed 5 August 2017].

Helton, W.S., Hollander, T.D., Warm, J.S., Matthews, G., Dember, W.N., Wallaart, M., Beauchamp, G., Parasuraman, R. and Hancock, P.A. 2005. Signal regularity and the mindlessness model of vigilance. *British Journal of Psychology*. **96**(2), pp. 249-261. Doi: 10.1348/000712605X38369.

Hirshkowitz, M., Whiton, K., Albert, S.M., Alessi, C., Bruni, O., Don Carlos, L. et al. 2015.

National sleep foundation's sleep time duration recommendations: methodology and results summary. *Sleep Health.* **1**(1), pp. 40-43. Doi: 10.1016/j.sleh.2014.12.010.

Hockey, R. 2011. A motivational control theory of cognitive fatigue. In Ackerman, P.L. ed. *Cognitive fatigue: multidisciplinary perspectives on current research and future applications.* Washington: American Psychological Association, pp. 167-188.

Hockey, R. 1997. Compensatory control in the regulation of human performance under stress and high workload: A cognitive-energetical framework. *Biological Psychology*. **45**(1), pp. 73-93. Doi: 10.1016/s0301-0511(96)05223-4.

Hockey, R., Wastell, D.G. and Sauer, J. 1998. Effects of sleep deprivation and user interface on complex performance: a multilevel analysis of compensatory control. *Human Factors*. **40** (2), pp. 233-253. Doi: 10.1518/001872098779480479.

Hoermann, H., Gontar, P. and Haslbeck, A. 2015. Effects of workload on measures of sustained attention during a flight simulator night mission. In: *Proceedings of the international symposium on aviation psychology, 4-7 May 2015, Dayton.* [Online]. Dayton: ISAP, pp. 602-607 [Accessed 11 December 2019]. Available from: https://corescholar.libraries.wright.edu /isap_2015/5/.

Hogervorst, E., Riedel, W., Jeukendrup, A. and Jolles, J. 1996. Cognitive performance after strenuous physical exercise. *Perceptual and Motor Skills.* **83**, pp. 479-488. Doi: 10.2466/pms.1996.83.2.479.

Hollopeter, J.N. 2011. *Response of novice and experienced drivers to lateral control intervention to prevent lane departures*. Master's thesis. University of Iowa.

Holtzer, R., Stern, Y., Rakitin, B. C. 2005. Predicting age-related dual-task effects with individual differences on neuropsychological tests. *Neuropsychology*, **19**, pp. 18–27. D oi:10.1037/089 4-4105.19.1.18.

Horne, J. and Baulk, S. 2004. Awareness of sleepiness when driving. *Psychophysiology*. **41**, pp. 161-165. Doi: 10.1046/j.1469-8986.2003.00130.x.

Horne, J.A. and Ostberg, O. 1976. A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms. *International Journal of Chronobiology*. **4**(2), pp. 97–110.

Horne, J.A. and Reyner, L.A. 1996. Counteracting driver sleepiness: effects of napping, caffeine, and placebo. *Psychophysiology*. **33**(3), pp. 306-309. Doi: 10.1111/j.1469-8986.1996.tb004 28.x.

Horne, J.A. and Reyner, L.A. 1995. Driver sleepiness. *Journal of Sleep Research*. **4**, pp. 23-29.

Doi: 10.1111/j.1365-2869.1995.tb00222.x.

Hwang, Y., Yoon, D., Kim, H.S. and Kim, K.H. 2014. A validation study on a subjective driving workload prediction tool. *IEEE Transactions on Intelligent Transportation System*. **15**, pp. 1835-1843. Doi: 10.1109/TITS.2014.2334664.

IBM. 2016. IBM SPSS Statistics (24.0). [Software]. [Accessed 20 February 2017].

Ingre, M, Akerstedt, T., Peters, B., Anund, A. and Kecklund, G. 2006a. Subjective sleepiness, simulated driving performance and blink duration: examining individual differences. *Journal of Sleep Research*. **15**, pp. 47–53. Doi: 10.1111/j.1365-2869.2006.00504.x.

Ingre, M, Akerstedt, T., Peters, B., Anund, A., Kecklund, G. and Pickles, A. 2006b. Subjective sleepiness and accident risk avoiding the ecological fallacy. *Journal of Sleep Research*. 15(2), pp. 142–148. Doi: 10.1111/j.1365-2869.2006.00517.x.

Injury Facts 2020. *Historical fatality trends*. [Online]. [Accessed 31 May 2020]. Available from: https://injuryfacts.nsc.org/motor-vehicle/historical-fatality-trends/deaths-and-rates/.

International Civil Aviation Organization (ICAO) 2019. *API data service*. [Online]. [Accessed 11 December 2019]. Available from: https://www.icao.int/safety/iStars/Pages/API-Data-Service.aspx.

International Civil Aviation Organization (ICAO) 2016. *Manual for the oversight of fatigue management approaches*. Doc 9966. [Online]. Montreal: International Civil Aviation Organization. [Accessed 11 December 2019]. Available from: https://www.icao.int/safety/fatiguemanagement/FRMS%20Tools/9966_cons_en.pdf.

International Society of Women Airline Pilots, 2019. *Worldwide airline female pilot numbers by region*. [Online]. [Accessed 11 December 2019]. Available from: https://www.iswap.org/content.aspx?page_id=86&club_id=658242&sl=1504342051.

Jafarian, S., Gorouhi, F., Taghva, A. Lotfi, J. 2008. High-altitude sleep disturbance: results of the Groningen Sleep Quality Questionnaire survey. *Sleep Medicine*. **9**(4), pp. 446-9. Doi: 10.1016/j.sleep.2007.06.017.

Jahn, G., Oehme, A., Krems, J.F. and Gelau, C. 2005. Peripheral detection as a workload measure in driving: effects of traffic complexity and route guidance system use in a driving study. *Transportation Research Part F: Traffic Psychology and Behaviour.* **8**(3), pp. 255-275. Doi: 10.1016/j.trf.2005.04.009.

Jamson, S.L. and Jamson, H. 2010. The validity of a low-cost simulator for the assessment of the effects of in-vehicle information. *Safety Science.* **48**(10), pp. 1477-1483. Doi: 10.1016/j.ssci.2010.07.008.

Jarosch, O., Bellem, H. and Bengler, K. 2019. Effects of task-induced fatigue in long conditional automated driving. *Human Factors*. **61**(7), pp. 1186-1189. Doi: 10.1177/001872081881622 6.

Jen, J.R., Hsu, C.C., Yang, H. and Ho, H. 2009. An investigation of fatigue issues on different flight operations. *Journal of Air Transport Management.* **15**(5), pp. 236-240. Doi: 10.1016/ j.jairtraman.2009.01.001.

Johns, M. 1992. A new method for measuring daytime sleepiness: the Epworth Sleepiness scale. *Sleep.* **14**, pp. 540-545. Doi: 10.1093/sleep/14.6.540.

Johns, M. 1998. Rethinking the assessment of sleepiness. *Sleep Medicine Reviews.* **2**(1), pp. 3-15. Doi: 10.1016/s1087-0792(98)90050-8.

Kahneman, D 1973. *Attention and effort*. New Jersey: Prentice-Hall.

Kaida, K., Takahashi, M., Akerstedt, T., Nakata, A., Otsuka, Y., Haratani, T. and Fukasawa, K. 2006. Validation of the Karolinska sleepiness scale against performance and EEG variables. *Clinical Neurophysiology.* **117**, pp. 1574-1581. Doi: 10.1016/j.clinph.2006.03.011.

Kaplan, K.A., Itoi, A. and Dement, W.C. 2007. Awareness of sleepiness and ability to predict sleep onset: can drivers avoid falling asleep at the wheel?. *Sleep Medicine.* **9**, pp. 71-79.

Karanikas, N. 2016. Human factors science and safety engineering. Can the STAMP model serve in establishing a common language?. In: *Proceedings of the EAAP 32 conference, 26-30 September 2016, Cascais*. [Online]. Cascais: EAAP, pp.132-149. [Accessed 11 December 2019]. Available from: https://www.researchgate.net/publication/313737062_Human_Factors_Science_and_Safety_Engineering_Can_the_STAMP_Model_Serve_in_Establishing_a_Common_Language.

Karlen, W., Cardin, S., Thalmann, D. and Floreano, D. 2010. Enhancing pilot performance with a symbodic system. In: *Proceedings of the annual international conference of the IEEE engineering in medicine and biology society, 31 August – 4 September 2010, Buenos Aires*. [Online]. Buenos Aires: IEEE Engineering in Medicine and Biology Society, pp. 6599-6602. [Accessed 11 December 2019]. Available from: https://ieeexplore.ieee.org/document/5627127.

Karrer, K., Briest, S., Vohringer-Kuhn, T., Baumgarten, T. and Schleicher, R. 2005. Driving without awareness. In: G. Underwood. ed. *Traffic and transport psychology*. Amsterdam: Elsevier, pp. 455–470.

Kawada T. 2008. Agreement rates for sleep/wake judgments obtained via accelerometer and sleep diary: a comparison. *Behavior Research Methods.* **40**(4), pp. 1026–1029. Doi: 10.3758/BRM

.40.4.1026.

Kazemi, Z., Mazloumi, A., Nasl, Saraji, G. and Barideh, S. 2016. Fatigue and workload in short and long-haul train driving. *Work*. **54**(2), pp. 425-433. Doi: 10.3233/WOR-162328.

Kecklund, G. and Akerstedt, T. 1993. Sleepiness in long distance truck driving: an ambulatory EEG study of night driving. *Ergonomics*. **36**(9), pp. 1007-1017. Doi: 10.1080/00140139308 967973.

Klein, D.C., Moore, R.Y. and Reppert, S.M. 1991. *Suprachiasmatic nucleus: the mind's clock*. New York: Oxford UP.

Kleinfehn, A.D. 2016. *Regional airline pilot commute: how commuting by air affects pilots' satisfaction with life*. Master thesis, University of North Dakota.

Ko, L.W., Komarov, O., Hairston, W.D., Jung, T.P. and Lin, C.T. 2017. Sustained attention in real classroom settings: an EEG study. *Frontiers in human neuroscience.* **11**, pp. 388. Doi: 10.3389/fnhum.2017.00388.

Konietschke, F., Bathke, A.C., Hothorn, L.A. and Brunner, E. 2010. Testing and estimation of purely non-parametric effects in repeated measures designs. *Computational Statistics and Data Analysis*. **54**(8), pp. 1895-1905. Doi: 10.1016/j.csda.2010.02.019.

Korber, M., Cingel, A., Zimmermann, M. and Bengler, K. 2015. Vigilance decrement and passive fatigue caused by monotony in automated driving. *Procedia Manufacturing*. **3**, pp. 2403-2409. Doi: 10.1016/j.promfg.2015.07.499.

Kosmadopoulos, A., Charli, S., Xuan, Z., Darwent, D., Matthews, R., Dawson, D. and Roach, G. 2017. The efficacy of objective and subjective predictors of driving performance during sleep restriction and circadian misalignment. *Accident Analysis & Prevention.* **99**, pp. 445-451. Doi: 10.1016/j.aap.2015.10.014.

Krauchi K, Cajochen C and Wirz-Justice A. 1997. A relationship between heat loss and sleepiness: effects of postural change and melatonin administration. *Journal of Applied Physiology.* **83**(1), pp. 134–139. Doi: 10.1152/jappl.1997.83.1.134.

Kuo, J., Lenne, M. G., Myers, R., Collard-Scruby, A., Jaeger, C., & Birmingham, C. 2017. Real-time assessment of operator state in air traffic controllers using ocular metrics. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 9-13 October, Austin. [Online]. Austin: Human Factors and Ergonomics Society, pp. 257–261. [Accessed 12 May 2017]. Available from: https://journals.sagepub.com/doi/abs/10.1177/15419312136 01547?journalCode=proe.

Kuo, T.B., Lin, T., Yang, C.C., Li, C.L., Chen. C.F. and Chou, P. (1999) Effect of aging on gender differences in neural control of heart rate. *The American Journal of Physiology*. **277**(6), pp. H2233-H2239. Doi: 10.1152/ajpheart.1999.277.6.H2233.

Kushida, C.A., Chang, A., Gadkary, C., Guilleminault, C., Carrillo, O. and Dement, W.C. 2001. Comparison of actigraphic, polysomnographic, and subjective assessment of sleep parameters in sleep-disordered patients. *Sleep Medicine.* **2**(5), pp. 389-396. Doi: 10.1016/s1389-9457(00)00098-8.

Lal, S. and Craig, A. 2001. A critical review of the psychophysiology of driver fatigue. *Biological Psychology*. **55**, pp. 173-194. Doi: 10.1016/s0301-0511(00)00085-5.

Lal, S. and Craig, A. 2002. Driver fatigue: electroencephalography and psychological assessment. *Psychophysiology.* **39**, pp. 313-321. Doi: 10.1017/s0048577201393095.

Laminar Research 2017. X-Plane 11 Cessna 172. Pilot's Operating Manual. [Online]. [Accessed 15 May 2019]. Available from: https://www.x-plane.com/manuals/C172_Pilot_Operating_Manual.pdf.

Laminar Research. 2014. X-Plane 10. [Software]. [Accessed 10 October 2018].

Lamond, N., Jay, S., Dorrian, J., Ferguson, S., Roach, G. and Dawson, D. 2008. The sensitivity of a palm-based Psychomotor Vigilance Task to severe sleep loss. *Behavior Research Methods.* **40**, pp. 347-352. Doi: 10.3758/BRM.40.1.347.

Landstrom, U., Englund, K., Nordstrom, B. and Stenudd, A. 2000. Laboratory studies on the effects of temperature variations on drowsiness. *Perceptual and Motor Skills*. **89**, pp. 1217-1229. Doi: 10.2466/pms.1999.89.3f.1217.

Lauderdale, D.S., Chen, J.H., Kurina, L.M., Waite, L.J. and Thisted, R.A. 2016. Sleep duration and health among older adults: associations vary by how sleep is measured. *Journal of Epidemiology and Community Health*. **70**(4), pp. 36-366. Doi: 10.1136/jech-2015-206109.

Lechin, F., van der Dijs, B., Orozco, B., Lechin, M.E., Báez, S., Lechin, A.E., Rada, I., Acosta, E., Arocha, L., Jiménez, V., León, G. and García, Z. 1995. Plasma neurotransmitters, blood pressure, and heart rate during supine-resting, orthostasis, and moderate exercise conditions in major depressed patients. Biological Psychiatry. **38**(3), pp. 166-173. Doi: 10.1016/0006-3223(94)00258-5.

Le Duc, P., Riley, D., Hoffman, S., Brock, M., Norman, D., Johnson, P., Williamson, R. and Estrada, A. 1999. *The Effects of Sleep Deprivation on Spatial Disorientation.* Report No. 2000-09. [Online]. Fort Rucker: U.S. Army Aeromedical Research Laboratory. [Accessed 11 December 2019]. Available from: https://www.usaarl.army.mil/TechReports/2000-09.PDF.

Ledger, H. 2013. The effect cognitive load has on eye blinking. *The Plymouth Student Scientist.* **6**(1), pp. 206-223.

Lee, K.A., Hicks, G. and Nino-Murcia, G. 1991. Validity and reliability of a scale to assess fatigue. *Psychiatry Research*. **36**(3), pp. 291-298. Doi: 10.1016/0165-1781(91)90027-m.

Lee, Y.C. and Liu, B.S. 2003. Inflight workload assessment: comparison of subjective and physiological measurements. *Aviation, space, and environmental medicine*. **74**(10), pp. 1078-1084.

Leger, D., Philip, P., Jarriault, P., Metlaine, A. and Choudat, D. 2009. Effects of a combination of napping and bright light pulses on shift workers' sleepiness at the wheel: A pilot study. Journal of Sleep Research. 18, pp. 472–479. Doi: 10.1111/j.1365-2869.2008.00676.x.

Leigh, T.J., Bird, H.A., Hindmarch, I., Constable, P.D. and Wright, V. 1988. Factor analysis of the St. Mary's Hospital Sleep Questionnaire. *Sleep*. **11**(5), pp. 448-453. Doi: 10.1093/sleep/ 11.5.448.

Lenne, M.G., Thomas, J., Triggs, J. and Redman, R. 1997. Time of day variations in driving performance. *Accident Analysis and Prevention.* **29**(4), pp. 431-437. Doi: 10.1016/s0001-4575(97)00022-5.

Lenne, M.G., Triggs, T.J., Redman, J.R. 1998. Interactive effects of sleep deprivation, time of day, and driving experience on a driving task. *Sleep.* **21**(1), pp. 38-44. Doi: 10.1093/sleep/21.1.38.

Liang, W.C., Yuan, J., Sun, D.C. and Lin, M.H. 2009. Changes in physiological parameters induced by indoor simulated driving: effect of lower body exercise at mid-term break. *Sensors.* **9**(9), pp. 6913–6933. Doi: 10.3390/s90906913.

Lim, J. and Dinges, D.F. 2008. Sleep deprivation and vigilant attention. *Annals of the New York Academy of Sciences.* **1129**, pp. 305-322. Doi: 10.1196/annals.1417.002.

Loh, S., Lamond, N., Dorrian, J., Roach, G. and Dawson, D. 2004. The validity of psychomotor vigilance tasks of less than 10-minute duration. *Behavior Research Methods, Instruments, and Computers.* **36** (2), pp. 339-346. Doi: 10.3758/bf03195580.

Lopez, N., Previc, F.H., Fischer, J., Heitz, R.P. and Engle, R.W. 2012. Effects of sleep deprivation on cognitive performance by United States Air Force pilots. *Journal of Applied Research in Memory and Cognition.* **1**(1), pp. 27-33. Doi: 10.1016/j.jarmac.2011.10.002.

Lorist, M.M., Boksem, M.A.S., Ridderinkhof, K.R. 2005. Impaired cognitive control and reduced cingulate activity during mental fatigue. *Cognitive Brain Research*. **24**, pp. 199-205. Doi: 10.1016/j.cogbrainres.2005.01.018.

Lorist, M.M. and Faber, L.G. 2011. In: Ackerman, P.L. ed. *Cognitive fatigue. Multidisciplinary Perspectives on current research and future applications.* Washington: American Psychological Association, pp. 105-126.

Lorist, M.M., Klein, M., Nieuwenhuis, S., de Jong, R., Mulder, G., and Theo, F. 2000. Mental fatigue and task control: planning and preparation. *Psychophysiology*. **37**(5), pp. 614-625. Doi: 10.1111/1469-8986.3750614.

Louw, T. and Merat, N. 2017. Are you in the loop? Using gaze dispersion to understand driver visual attention during vehicle automation. *Transportation Research Part C Emerging Technologies.* **76**, pp. 35-50. Doi: 10.1016/j.trc.2017.01.001.

Loy, B. 2013. The effect of a single bout of exercise on energy and fatigue states: a systematic review and meta-analysis. *Fatigue: Biomedicine, Health & Behavior.* **1**, pp. 223-242. Doi: 10.1080/21641846.2013.843266.

Magnusson, S. 2002. Similarities and differences in psychophysiological reactions between simulated and real air-to-ground missions. *The International Journal of Aviation Psychology*. **12**(1), pp. 49-61. Doi: 10.1207/S15327108IJAP1201_5.

Mansikka, H., Virtanen, K. and Harris, D. 2019. Comparison of NASA-TLX scale, modified Cooper–Harper scale and mean inter-beat interval as measures of pilot mental workload during simulated flight tasks. *Ergonomics*. **62**(2), pp. 246-254. Doi: 10.1080/00140139.201 8.1471159.

Marcus, C.L. and Loughlin, G.M. 1996. Effect of sleep deprivation on driving safety in housestaff. *Sleep.* **19**, pp. 763-766. Doi: 10.1093/sleep/19.10.763.

Marcus, J.H. and Rosekind, M.R. 2017. Fatigue in transportation: NTSB investigations and safety recommendations. *Injury Prevention.* **23, pp.** 232-238. Doi: 10.1136/injuryprev-2015-041791.

Marino, M., Li, Y., Rueschman, M.N., Winkelman, J.W., Ellenbogen, J.M., Solet, J.M., Dulin, H., Berkman, L.F. and Buxton, O.M. 2013. Measuring sleep: Accuracy, sensitivity, and specificity of wrist actigraphy compared to polysomnography. *Sleep*. **36**(11), pp. 1747–1755. Doi: 10.5665/sleep.3142.

Matthews, G. and Desmond, P.A. 2002. Task-induced fatigue states and simulated driving performance. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology.* **55**, pp. 659-686. Doi: 10.1080/02724980143000505.

Matthews, R.W., Ferguson, S.A., Sargent, C., Zhou, X., Kosmadopoulos, A. and Roach, G. 2017. Using interstimulus interval to maximise sensitivity of the Psychomotor Vigilance Test to fatigue. *Accident Analysis & Prevention*. **99**, pp. 406-410. Doi: 10.1016/j.aap.2015.10.013.

Matthews, G., Joyner, L., Gilliland, K., Campbell, S. E., Falconer, S. and Huggins, J. 1999. Validation of a comprehensive stress state questionnaire: Towards a state 'Big Three'?. In: Mervielde, I., Deary, I.J., De Fruyt, F. and Ostendorf, F. eds. *Personality psychology in Europe.* Tilburg: Tilburg University Press, pp. 315-340.

May, J. and Baldwin, C. 2009. Driver fatigue: the importance of identifying causal factors of fatigue when considering detection and countermeasure technologies. *Transportation Research Part F: Traffic Psychology and Behaviour.* **12**, pp. 218-224. Doi: 10.1016/j.trf.20 08.11.005.

McCall, C. and McCall, W.V. 2012. Comparison of actigraphy with polysomnography and sleep logs in depressed insomniacs. *Journal of Sleep Research.* **21**, pp. 122-127. Doi: 10.1111/j.1 365-2869.2011.00917.x.

McCartt, A., Mayhew, D., Braitman, K., Ferguson, S. and Simpson, H. 2009. Effects of age and experience on young driver crashes: review of recent literature. *Traffic injury prevention*. **10**(3), 209-219. Doi: 10.1080/15389580802677807.

McKie, R and O'Hanlon, J. 1978. A study of the combined effects of extended driving and heat stress on driver arousal and performance. In: *Proceedings of a symposium on vigilance II: relationships among theory, physiological correlates, and operational performance, 3-6 August 1976, St Vincent.* New York: Plenum Press, pp. 537-558. [Accessed 5 December 2017].

McMorris, T. and Keen, P. 1994. Effect of exercise on simple reaction times of recreational athletes. *Perceptual and Motor Skills*. **78**(1), pp. 123-130. Doi: 10.2466/pms.1994.78.1.123.

Mehler, B., Reimer, B., and Wang, Y.A. 2011. Comparison of heart rate and heart rate variability indices in distinguishing single-task driving and driving under secondary cognitive workload. In: *Proceedings of the sixth international driving symposium on human factors in driver assessment, training and vehicle design, 27-30 June 2011, Olympic Valley*. Iowa city: The University of Iowa, pp. 590-597. [Accessed 11 December 2019]. Available from: https:/ /core.ac.uk/download/pdf/194407777.pdf.

Menghini, L., Gianfranchi, E., Cellini, N., Patron, E., Tagliabue, M. and Sarlo, M. 2019. Stressing the accuracy: wrist-worn wearable sensor validation over different conditions. *Psychophysiology.* **56**(7), pp. 1-15. Doi: 10.1111/psyp.13441.

Mets, M., Baas, D., Boven, I., Olivier, B. and Verster, J. 2012. Effects of coffee on driving performance during long simulated highway driving. *Psychopharmacology*. **222**, pp. 337-342. Doi: 10.1007/s00213-012-2647-7.

Miller, C.B., Gordon, C.J., Delwyn, L.T., Bartlett, J., Grunstein, R.R., D'Rozario, A.L. and Marshall, N.S. 2015. Agreement between simple questions about sleep duration and sleep diaries in a large online survey. *Sleep Health*. **1**(2), pp. 133-137. Doi: 10.1016/j.sleh.2015.0 2.007.

Miller, J.C. and Narvaez, A. 1986. A comparison of two subjective fatigue checklists. In: *Proceedings of the 10th Psychology in the Department of Defense symposium*. [Online]. Colorado Springs: U.S. Air Force Academy, pp. 1-6. [Accessed 11 December 2019]. Available from: https://www.researchgate.net/publication/315735301_A_comparison_of_two_subjective_fatigue_checklists.

Milstein, N. and Gordon, I. 2020. Validating measures of electrodermal activity and heart rate variability derived from the Empatica E4 utilized in research settings that involve interactive dyadic states. *Frontiers in Behavioral Neuroscience.* **14**, pp. 1-13. Doi: 10.3389/fnbeh.2020.00148.

Miro, E., Cano-Lozano, M.C. and Buela-Casal, G. 2002. Electrodermal activity during total sleep deprivation and its relationship with other activation and performance measures. *Journal of Sleep Research.* **11**, pp. 105-112. Doi: 10.1046/j.1365-2869.2002.00286.x.

Morad, Y., Lemberg, H., Yofe, N. and Dagan, Y. 2000. Pupillography as an objective indicator of fatigue. *Current eye research.* **21**, pp. 535-542. Doi: 10.1076/0271-3683(200007)2111-ZFT535.

Moroney, B.W., Warm, J.S. and Dember, W.N. 1995. Effects of demand transitions on vigilance performance and perceived workload. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 9-13 October 1995, San Diego.* [Online]. San Diego: Human Factors and Ergonomics Society, pp. 1375-1379. [Accessed 11 December 2019]. Available from: https://journals.sagepub.com/Doi/10.1177/154193129503902104.

Morris, T.L. and Miller, J. 1996. Electrooculographic and performance indices of fatigue during simulated flight. *Biological Psychology*. **42**, pp. 343-360. Doi: 10.1016/0301-0511(95)0516 6-x.

Mueller, S.T. and Piper, B.J. 2014. The Psychology Experiment Building Language (PEBL) and PEBL test battery. *Journal of Neuroscience Methods.* **222**, pp. 250-259. Doi: 10.1016/j.jneu meth.2013.10.024.

Mulder, M., Abbink, D.A., van Paassen, M.M. and Mulder, M. 2010. Haptic gas pedal support during visually distracted car following. *IFAC Proceedings Volumes*. **43**(3), pp. 322-327. Doi: 10.3182/20100831-4-FR-2021.00057.

Mulder-Hajonides van der Meulen, W., Wijnberg., J.R., Hollander, J.J., De Diana, I.P.F. and van den Hoofdakker, R.H 1980. Measurement of subjective sleep quality. *European Sleep Research Society Abstracts*. **5**, pp. 98.

Nabatilan, L., Aghazadeh, F., Nimbarte, A., Harvey, C. and Chowdhury, S. 2011. Effect of driving experience on visual behavior and driving performance under different driving conditions. *Cognition, Technology and Work*. **14**(4), pp. 355-363. Doi: 10.1007/s10111-011-0184-5.

National Highway Traffic Safety Administration 2015. *Drowsy driving 2015*. Report No. DOT HS 812446. [Online]. Washington: National Highway Traffic Safety Administration. [Accessed 31 May 2020]. Available from: https://crashstats.nhtsa.dot.gov/Api/Public/View Publication/812446.

National Highway Traffic Safety Administration 2002. *National survey of distracted and drowsy driving. Attitudes and behaviour: 2002*. Report No. DOT HS 809 566. [Online]. Washington: National Highway Traffic Safety Administration. [Accessed 3 January 2018]. Available from: https://www.nhtsa.gov/sites/nhtsa.dot.gov/files/hs809566v1.pdf.

National Research Council 2011. *The effects of commuting on pilot fatigue*. [Online]. Washington: National Academy of Sciences. [Accessed 11 December 2019]. Available from: https://www.nap.edu/catalog/13201/the-effects-of-commuting-on-pilot-fatigue.

National Transportation Safety Board 2019[a]. *NTSB 2019-2020 most wanted list of transportation safety improvements*. [Online]. [Accessed 11 December 2019]. Available from: https://www.ntsb.gov/safety/mwl/Pages/default.aspx.

National Transportation Safety Board 2019[b]. *Reduce fatigue-related accidents*. [Online]. [Accessed 11 December 2019]. Available from: https://www.ntsb.gov/safety/mwl/Pages/mwlfs-19-20/mwl2.aspx.

National Transportation Safety Board 2018. *Taxiway overflight Air Canada flight 759 Airbus A320-211, C-FKCK, San Francisco, California, July 7, 2017*. Report No. report NTSB/AIR-18/01 PB2018-101561. [Online]. Washington: National Transportation Safety Board. [Accessed 11 December 2019]. Available from: https://www.ntsb.gov/investigations/AccidentReports/Reports/AIR1801.pdf.

National Transportation Safety Board 2010. *Loss of control on approach Colgan Air, Inc. operating as continental connection flight 3407 Bombardier DHC-8-400, N200WQ, Clarence center, New York, February 12, 2009.* Report No. NTSB/AAR-10/01. [Online]. Washington: National Transportation Safety Board. [Accessed 11 December 2019]. Available from: https://www.ntsb.gov/investigations/AccidentReports/Reports/AAR1001.p df.

Neri, D.F., Oyung, R.L., Colletti, L.M., Mallis, M.M., Tam, P.Y. and Dinges, D.F. 2002. Controlled activity as a fatigue countermeasure on the flight deck. *Aviation Space and Environmental Medicine.* **73**(7), pp. 654-664.

Nesthus, T., Cruz, C., Hackworth, C. Boquet, A. 2006. *An assessment of commuting risk factors for air traffic control specialists*. Report No. DOT/FAA/AM-06/13. [Online]. Washington: Federal Aviation Administration. [Accessed 11 December 2019]. Available from: https://www.semanticscholar.org/paper/An-Assessment-of-Commuting-Risk-Factors-for-Air-Nesthus-Cruz/71447d51cda5976517eeeb5cddb44533e9215341.

Netherlands Aerospace Centre and Netherlands Institute of Neurosciences 2019. *Independent research on Pilot Fatigue Measurement by the Netherlands Aerospace Centre*. Report No. CAP1756. [Online]. The Netherlands: Netherlands Aerospace Centre. [Accessed 11 December 2019]. Available from: http://publicapps.caa.co.uk/docs/33/Pilot%20Fatigue%20Measurement%20Research_Final%20Report_v0.9.1clean.pdf.

Nilsson, T, Nelson, T.M. and Carlson, D. 1997. Development of fatigue symptoms during simulated driving. *Accident Analysis and Prevention*. **29**(4), pp. 479-488. Doi: 10.1016/s0001-4575(97)00027-4.

Noguchi, K., Gel, Y.R., Brunner, E. and Konietschke, F. 2012. nparLD: an R software package for the non-parametric analysis of longitudinal data in factorial experiments. *Journal of Statistical Software*. **50**(12), pp. 1–23. Doi: 10.18637/jss.v050.i12.

Nordbakke, S. and Sagberg, F. 2007. Sleepy at the wheel: Knowledge, symptoms and behaviour among car drivers. *Transportation Research Part F: Traffic Psychology and Behaviour*. **10**(1), pp. 1–10. Doi: 10.1016/j.trf.2006.03.003.

O'Donnell, R.D. and Eggemeier, F.T. 1986. Workload assessment methodology. In Boff, K.R., Kaufman, L. and Thomas, J.P. eds. *Handbook of perception and human performance.* New Jersey: John Wiley & Sons, pp. 1-49.

O'Hagan, A., Issartel, J., Wall, A., Dunne, F., Boylan, P., Groeneweg, J., Herring, M., Campbell, M. and Warrington, G. 2019. 'Flying on empty' – effects of sleep deprivation on pilot performance. *Biological Rhythm Research*., pp. 1-22. Doi: 10.1080/09291016.2019.1581481.

Ohayon, M.M., Carskadon, M.A., Guilleminault, C. and Vitiello, M.V. 2004. Meta-analysis of quantitative sleep parameters from childhood to old age in healthy individuals: developing normative sleep values across the human lifespan. *Sleep.* **27**, pp. 1255–1273. Doi: 10.1093/sleep/27.7.1255.

Ohayon, M., Wickwire, M., Hirshkowitz, M.M., Avidan, A., Daly, J., Dauvilliers, Y., Ferri, R., Fung, D., Hazen, A., Lichstein, K., Mallampalli, M. G., Rawding, R., Scheer, F.A., Somers, V. and Vitiello, M.V. 2017. National Sleep Foundation's sleep quality recommendations: first report. *Sleep Health*. **3**(1), pp. 6-19. Doi: 10.1016/j.sleh.2016.11.006.

Ollander, S., Godin, C., Campagne, A. and Charbonnier, S. 2016. A comparison of wearable and stationary sensors for stress detection. In: *IEEE International Conference on Systems, Man, and Cybernetics (SMC), 9-12 October 2016, Budapest*. [Online]. Budapest: IEEE, pp. 004362-004366. [Accessed 13 October 2019]. Available from: https://ieeexplore.ieee.org/document/7844917.

Oron-Gilad, T. and Ronen, A. 2007. Road characteristics and driver fatigue: A simulator study. *Traffic Injury prevention*. **8**, pp. 281-289. Doi: 10.1080/15389580701354318.

Oron-Gilad, T., Ronen, A., Cassuto, Y., Shinar, D. 2002. Alertness maintaining tasks while driving. In: *Proceedings of the Human Factors and Ergonomics Society annual meeting*, 29 September – 4 October 2002, Baltimore. [Online]. Baltimore: Human Factors and Ergonomics Society, pp. 46. [Accessed 11 December 2019]. Available from: https://www.sciencedirect.com/science/article/abs/pii/S0001457507001753?via%3Dihub.

Ottati, W. L., Hickox, J. C. and Richter, J. 1999. Eye scan patterns of experienced and novice pilots during visual flight rules (VFR) navigation. *Proceedings of the Human Factors and Ergonomics Society annual meeting, 27 September – 1 October 1999, Houston.* [Online]. Houston: Human Factors and Ergonomics Society, pp. 66–70. [Accessed 11 December 2019]. Available from: https://journals.sagepub.com/Doi/abs/10.1177/154193129904300114.

Pack, A.I., Pack, A.M., Rodgman, E., Cucchiara, A., Dinges, D.F. and Schwab, C.W. 1995. Characteristics of crashes attributed to the driver having fallen asleep. *Accident Analysis and Prevention*. **27**(6), pp. 769-775.

Parkes, A., Ashby, M. and Fairclough, S. 1991. The effects of different in-vehicle route information displays on driver behaviour. In: *Proceedings of the Vehicle Navigation and Information Systems Conference, 20-23 October 1991, Warrendale*. [Online]. Warrendale: IEEE, pp. 61-70. [Accessed 13 December 2019]. Available from: https://ieeexplore.ieee.org/document/1623612.

Paul, M.A., Pigeau, R.A. and Weinberg, H. 2001. CC130 pilot fatigue during re-supply missions to former Yugoslavia. *Aviation Space and Environmental Medicine.* **72**(11), pp. 965-973.

PEBL. 2018. PEBL: The Psychology Experiment Building Language. [Software]. [Accessed 15 February 2018].

Petrie, K.J., Powell, D., Broadbent, E. 2004. Fatigue self-management strategies and reported fatigue in international pilots. *Ergonomics*. **47**(5), pp. 461–468. Doi: 10.1080/00140130310 00085653.

Petrilli, R., Roach, G., Dawson, D. and Lamond, N. 2006. The sleep, subjective fatigue, and sustained attention of commercial airline pilots during an international pattern. *Chronobiology International.* **23**, pp. 1357-1362. Doi: 10.1080/07420520601085925.

Petrilli, R.M., Roach, G.D., Lamond, N., Dawson, D. and Thomas, M.J.W. 2007. Effects of flight duty and sleep on the decision-making of commercial airline pilots. In: Anca, J. ed. *Multimodal safety management and human factors: crossing the borders of medical, aviation, road and rail industries*. Aldershot: Ashgate, pp. 259–270.

Philip, P., Sagaspe, P., Taillard, J., Valtat, C., Moore, N., Akerstedt, T., Charles, A. and Bioulac, B. 2005a. Fatigue, sleepiness, and performance in simulated versus real driving conditions. *Sleep.* **28**(12), pp. 1511–1516. Doi: 10.1093/sleep/28.12.1511.

Philip, P., Sagaspe, P., Moore, N., Taillard, J., Charles, A., Guilleminault, C. and Bioulac, B. 2005b. Fatigue, sleep restriction and driving performance. *Accident Analysis & Prevention*, **37**(3), pp. 473-478. Doi: 10.1016/j.aap.2004.07.007.

Phipps-Nelson, J., Redman, J. and Rajaratnam, S. 2010. Temporal profile of long, night-time driving performance: breaks from driving temporarily reduce time-on-task fatigue but not sleepiness. *Journal of Sleep Research.* **20**, pp. 404-415. Doi: 10.1111/j.1365-2869.2010.009 00.x.

Phipps-Nelson, J., Redman, J.R., Dijk, D.J. and Rajaratnam, S.W. 2003. Daytime exposure to bright light, as compared to dim light, decreases sleepiness and improves psychomotor vigilance performance. *Sleep.* **26**, pp. 695–700. Doi: 10.1093/sleep/26.6.695.

Picard, R.W. 2009. Future affective technology for autism and emotion communication. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences.* **364**(1535), pp. 3575–3584. Doi: 10.1098/rstb.2009.0143.

Pien, G.W., Sammel, M.D., Freeman, E.W., Lin, H. and DeBlasis, T.L. 2008. Predictors of sleep quality in women in the menopausal transition. Sleep. **31**(7), pp. 991-999. Doi: 10.5665/sleep/31.7.991.

Pilcher, J.J., Band, D., Odie-Dusseau, H.N. and Muth, E.R. 2007. Human performance under sustained operations and acute sleep deprivation conditions: toward a model of controlled attention. *Aviation Space and Environmental Medicine*. **78**(5), pp. B15-B24. PMID: 175473 01.

Pilcher, J.J. and Huffcutt, A.I. 1996. Effects of sleep deprivation on performance: a meta-analysis. *Sleep: Journal of Sleep Research & Sleep Medicine*. **19**(4), pp. 318-326. Doi: 10.1093/sleep/19.4.318.

Portnoy, J. and Farrington, D.P. 2015. Resting heart rate and antisocial behavior: An updated systematic review and meta-analysis. *Aggression and Violent Behavior*. **22**, pp. 33-45. Doi: 10.1016/j.avb.2015.02.004.

Posner, M.I. 1980. Orienting of attention. *Quarterly Journal of Experimental Psychology*. **32**, pp. 3-25. Doi: 10.1080/00335558008248231.

Powell, D., Spencer, M., Holland, D., Broadbent, E. and Petrie, K. 2007. Pilot fatigue in short-haul operations: effects of number of sectors, duty length, and time of day. *Aviation, space, and Environmental Medicine*. **78**, pp. 698-701. Doi: 10.22237/jmasm/1493597220.

QualtricsXM 2016. *Determining sample size: how to make sure you get the correct sample size*. [Accessed 11 November 2016]. Available from: https://www.qualtrics.com/experience-management/research/determine-sample-size/.

Quer, G., Gouda, P., Galarnyk, M., Topol, E.J. and Steinhubl, S.R. 2020. Inter- and intraindividual variability in daily resting heart rate and its associations with age, sex, sleep, BMI, and time of year: Retrospective, longitudinal cohort study of 92,457 adults. *PLoS ONE*. **15**(2), pp. e0227709. Doi: 10.1371/journal.pone.0227709.

R-Project. 2018. The R Project for Statistical Computing (2018). [Software]. [Accessed 1 August 2018].

Rapino, M.A. and Fields, A.K. 2013. Mega commuters in the U.S.. In: *Proceedings of the association for public policy analysis and management conference, 7-9 November 2013, Washington*. [Online]. Washington: Association for public policy and analysis, pp.1-8. [Accessed 11 December 2019]. Available from: https://www.census.gov/content/dam/Census/library/working-papers/2013/demo/SEHSD-WP2013-03.pdf.

Recarte, M.A., Perez, E., Conchillo, A. and Nunes, L.M. 2008. Mental workload and visual impairment: differences between pupil, blink, and subjective rating. *The Spanish Journal of Psychology*. **11**(2), pp. 374–385.

Reed, M.P. and Green, P.A. 1999. Comparison of driving performance on-road and in a low-cost simulator using a concurrent telephone dialing task. *Ergonomics*. **42**, pp. 1015–1037. Doi: 10.1080/001401399185117.

Reid, G.B. and Nygren, T.E. 1988. The subjective workload assessment technique: a scale procedure for measuring mental workload. In: Hancock, P.A. and Meshkati, N. eds. *Human* *mental workload*. New York: Elsevier Science, pp. 185-218.

Reis, C., Mestre, C., Canhão, H., Gradwell, D. and Paiva, T. 2016. Sleep complaints and fatigue of airline pilots. *Sleep Science*. **9**(2), pp. 1-5. Doi: 10.1016/j.slsci.2016.05.003.

Reyner, L.A. and Horne, J.A. 2002. Efficacy of a 'functional energy drink' in counteracting driver sleepiness. *Physiology and Behavior*. **75**, pp. 331–335. Doi: 10.1016/S0031-9384(01) 00669 -2.

Reyner, L.A. and Horne, J.A. 2000. Early morning driver sleepiness: effectiveness of 200 mg caffeine. *Psychophysiology*. **37**, pp. 251–256.

Reyner, L.A. and Horne, J.A. 1998. Falling asleep while driving: are drivers aware of prior sleepiness? *International Journal of Legal Medicine*. **111**(3), pp. 120-123. Doi: 10.1007/s00 4140050131.

Reyner, L.A. and Horne, J.A. 1995. Driver sleepiness. *Journal of Sleep Research*. **4**, pp. 23-29. Doi: 10.1111/j.1365-2869.1995.tb00222.x.

Richter, S., Marsalek, K., Glatz, C. and Gundel, A. 2005. Task-dependent differences in subjective fatigue scores. *Journal of Sleep Research*. **14**(4), p. 393-400. Doi: 10.1111/j.1365 -2869.2005.00473.x.

Richter, P., Wagner, T. and Heger G. 1998. Psychophysiological analysis of mental load during driving on rural roads - a quasi-experimental study. *Ergonomics*. **41**(5), pp. 593-609. Doi: 10.1080/001401398186775.

Roach, G., Petrilli, R., Dawson, D. and Thomas, M. 2006. The effects of fatigue on the operational performance of flight crew in a B747-400 simulator. In: *Proceedings of the Seventh International AAvPA Symposium, 9-12 November, Manly*. [Online]. Manly: AAvPA, pp. 1-7. [Accessed 4 January 2017]. Available from: https://pdfs.semanticscholar.org/e911/30146 7e243b37935cbae80545d261506d5c0.pdf.

Rognin, L., Grimaud, I., Hoffman, E. and Zeghal, K. 2004. Assessing the impact of a new instruction on air traffic controller monitoring tasks. In: *Proceedings of the International Conference on Human–Computer Interaction in Aeronautics, 29 September – 1 October 2004, Toulouse.* [Online]. Toulouse: AAAI Press, pp. 1-6. [Accessed 11 December 2019]. Available from: https://www.eurocontrol.int/eec/public/standard_page/DOC_Conf_2004_0 07.html.

Roma, P.G., Hursh, S.R., Mead, A.M. and Nesthus, T.E. 2012. *Analysis of commute times and neurobehavioural performance capacity in aviation cabin crew*. Report No. DOT/FAA/AM-12/14. [Online]. Washington: Federal Aviation Administration. [Accessed 11 December

2019]. Available from: https://www.faa.gov/data_research/research/med_humanfacs/oamtechreports/2010s/media/201214.pdf.

Roscoe, A.H. 1993. Heart rate as a psychophysiological measure for in-flight workload assessment. *Ergonomics*. **36**(9), pp. 1055-1062. Doi: 10.1080/00140139308967977.

Rosekind, M., Co, E., Gregory, K. and Miller, D. 2000. *Crew factors in flight operations XIII: a survey of fatigue factors in corporate/executive aviation operations*. Report No. NASA/TM-2000-209610. [Online]. California: National Aeronautics and Space Administration. [Accessed 11 December 2019]. Available from: https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20010039028.pdf.

Rosekind, M.R., Gander, P.H., Miller, D.L., Gregory, K.B., Smith, R.M., Weldon, K.J., Lebacqz, J.V. 1994. Fatigue in operational settings: examples from the aviation environment. *Human Factors*. **36**(2), pp. 327–338. Doi: 10.1177/001872089403600212.

Roskam, A.J., Brookhuis, K.A., De Waard, D., Carsten, O.M.J., Read, L., Jamson, S., Ostlund, J., Bolling, A., Nilsson, L., Anttila, V., Hoedemaeker, M., Janssen, W.H., Harbluk, E., Johansson, M., Tevell, J., Santos, M., Fowkes, J., Engstrom, V. 2002. HASTE Deliverable 1: Development of Experimental Protocol [Online]. Leeds: Institute for Transport Studies. [Accessed 5 December 2017]. Available from: http://www.its.leeds.ac.uk/projects/haste/HASTE%20Deliverable%201%20v1.1.pdf.

Rubio, S., Diaz, E., Martin, J. and Puente, J.M. 2004. Evaluation of subjective mental workload: A comparison of SWAT, NASA-TLX, and Workload Profile methods. *Applied Psychology*. **53**(1), pp. 61-86. Doi: 10.1111/j.1464-0597.2004.00161.x.

Rupp, T.L., Wesensten, N.J. and Balkin, T.J. 2012. Trait-like vulnerability to total and partial sleep loss. *Sleep*. **35**(8), pp. 1163-1172. Doi: 10.5665/sleep.2010.

SAE International 2019. *Levels of driving automation*. [Online]. [Accessed 11 December 2019]. Available from: https://www.sae.org/news/2019/01/sae-updates-j3016-automated-driving-graphic.

Saffarian, M., de Winter, J. and Senders, J. 2015. Measuring drivers' visual information needs during braking: a simulator study using a screen-occlusion method. *Transportation Research Part F: Traffic Psychology and Behaviour*. **33**, pp. 48-65. Doi: 10.1016/j.trf.2015. 07.001.

Sagberg, F., Jackson, P., Kruger, H.P., Muzet, A.G. and Williams A. 2004. *Fatigue, sleepiness and reduced alertness as risk factors in driving* 2004. Report No.739/2004. [Online]. Oslo: Institute for Transport Economics. [Accessed 11 December 2019]. Available from: https://www.toi.no/getfile.php/131016/Publikasjoner/T%C3%98I%20rapporter/2004/739-2004/739-2004-2.pdf.

Sahayadhas, A., Sundaraj, K. and Murugappan, M. 2012. Detecting driver drowsiness based on sensors: a review. *Sensors.* **12**(12), pp. 16937-16953. Doi: 10.3390/s121216937.

Salvucci, D.D. and Taatgen, N.A. 2008. Threaded cognition: an integrated theory of concurrent multitasking. *Psychological Review*. **115**(1), pp. 101–130. Doi.org/10.1037/0033-295X.115.1.101.

Samn, S. and Perelli, L. 1982. *Estimating aircrew fatigue: a technique with application to airlift operations.* Report No. SAM-TR-82-21. [Online]. Texas: USAF School of Aerospace Medicine. [Accessed 11 December 2019]. Available from: https://www.researchgate.net/publication/235150215_Estimating_Aircrew_Fatigue_A_Technique_with_Application_to_Airlift_Operations.

Sanders, A.F. 1998. *Elements of Human Performance: Reaction Processes and Attention in Human Skill.* Malwah: Lawrence Erlbaum Associates.

Santiago-Espada, Y., Myer, R.R., Latorella, K.A. and Comstock, J.R. 2011. The Multi-Attribute Task Battery II (MATB-II) software for human performance and workload research: a user's guide. NASA/TM–2011-217164. Hampton: Langley research Center.

Sarter, M., Givens, B. and Bruno, J.P. 2001. The cognitive neuroscience of sustained attention: where top-down meets bottom-up. *Brain Research Reviews.* **35**(2) pp. 146-160. Doi: 10.1016/S0165-0173(01)00044-3.

Satou, Y., Ando, H., Nakiri, M., Nagatomi, K., Yamaguchi, Y., Hoshino, M., Tsuji, Y., Muramoto, J., Mori, M., Hara, K. and Ishitake, T. 2007. Effects of short-term exposure to whole-body vibration on wakefulness level. Industrial Health. **45**(2), pp. 217-23. Doi: 10.2486/indhealth.45.217.

Satou, Y., Ishitake, T., Ando, H., Nagatomi, K., Hoshiko, M., Tsuji, Y., Tamaki, H., Shigemoto, A., Kusano, M., Mori, M. and Hara, K. 2009. Effect of short-term exposure to whole body vibration in humans: relationship between wakefulness level and vibration frequencies. Kurume Medical Journal. **56**(1-2), pp. 17-23. Doi: 10.2739/kurumemedj.56.17.

Saxby, D., Matthews, G., Hitchcock, E., Warm, J., Funke, G. and Gantzer, T. 2008. Effect of active and passive fatigue on performance using a driving simulator. In: *Proceedings of the Human Factors and Ergonomics Society 52nd annual meeting, 22-26 September 2008, New York*. [Online]. New York: Human Factors and Ergonomics Society, pp. 1751-1755 [Accessed 11 December 2019]. Available from: https://journals.sagepub.com/doi/ abs/10.1177/154193120805202113?journalCode=proe.

Saxby, D.J.G., Matthews, J.S., Warm, E.M., Hitchcock, C. and Neubauer, C. 2013. Active and passive fatigue in simulated driving: discriminating styles of workload regulation and their safety impacts. *Journal of Experimental Psychology: Applied.* **19**, pp. 287-300. Doi: 10.1037 /a0034386.

Schleicher, R., Galley, N., Briest, S. and Galley, L. 2008. Blinks and saccades as indicators of fatigue in sleepiness warnings: looking tired?. *Ergonomics.* **51**, pp. 982-1010. Doi: 10.1080/ 00140130701817062.

Schmeisser, E.T., Fuller, D.R. and Van de Pol, C. 2008. Predictability of pilot performance from simulated to real flight in the UH-60 (black hawk) helicopter. Report No. 2008-06. Fort Rucker: U.S. Army Aeromedical Research Laboratory. [Accessed 10 December 2019]. Available from: https://www.usaarl.army.mil/TechReports/2008-06.PDF.

Schmidt, E. and Bullinger-Hoffmann, A. 2017. Mitigating passive fatigue during monotonous drives with thermal stimuli: insights into the effect of different stimulation durations. *Accident Analysis and Prevention.* **126**, pp. 115-121. Doi: 10.1016/j.aap.2017.12.005.

Schömig, N., Hargutt, V., Neukum, A., Petermann-Stock, I. and Othersen, I. 2015. The interaction between highly automated driving and the development of drowsiness. *Procedia Manufacturing.* **3**, pp. 6652-6659. Doi: 10.1016/j.promfg.2015.11.005.

Schubert, C., Lambertz, M., Nelesen, R.A., Bardwell, W., Choi, J.B. and Dimsdale, J. E. 2009. Effects of stress on heart rate complexity - a comparison between short-term and chronic stress. *Biological psychology.* **80**(3), pp. 325–332. Doi: 10.1016/j.biopsycho.2008.11.005.

Schuurmans, A., de Looff, P., Nijhof, K., Rosada, C., Scholte, R., Popma, A. and Otten, R. 2020. Validity of the Empatica E4 wristband to measure Heart Rate Variability (HRV) parameters: a comparison to Electrocardiography (ECG). *Journal of medical systems*. **44**, pp. 190-201. Doi: 10.1007/s10916-020-01648-w.

Schwarz, J.F.A., Ingre, M., Fors, C., Anund, A., Kecklund, G., Taillard, J., Philip, P. and Akerstedt, T. 2012. In-car countermeasures open window and music revisited on the real road: popular but hardly effective against driver sleepiness. *Journal of Sleep Research.* **21**, pp. 595–599. Doi: 10.1111/j.1365-2869.2012.01009.x.

Scott, L., Hwang, W., Rogers, A., Nysse, T., Dean, G. and Dinges, D. 2008. The relationship between nurse work schedules, sleep duration, and drowsy driving. *Sleep.* **30**, pp. 1801-1807. Doi: 10.1093/sleep/30.12.1801.

Shattuck, N.L. and Matsangas, P. 2015. Psychomotor vigilance performance predicted by Epworth Sleepiness Scale scores in an operational setting with the United States Navy. *Journal of Sleep Research*. **24**(2), pp. 174-180. Doi:10.1111/jsr.12243.

Shen, J., Barbera, J. and Shapiro, C. 2006. Distinguishing sleepiness and fatigue: focus on definition and measurement. *Sleep Medicine Reviews.* **10**, pp. 63-76. Doi: 10.1016/j.smrv.2005.05.004.

Shiferaw, B., Downey, L., Westlake, J., Stevens, B., Rajaratnam, S., Berlowitz, D., Swann, P. Howard, M. 2018. Stationary gaze entropy predicts lane departure events in sleep-deprived drivers. *Scientific Reports*. **8**(1), pp. 1-10. Doi: 10.1038/s41598-018-20588-7.

Signal, T.L., Gale, G. and Gander, P.H. 2005. Sleep measurement in flight crew: Comparing actigraphic and subjective estimates to polysomnography. *Aviation, Space, and Environmental Medicine*. **76**(11), pp. 1058-1063.

Silva, G.E., Goodwin, J.L., Sherrill, D.L., Arnold, J.L., Bootzin, R.R., Smith, T., Walsleben, J.A., Baldwin, C.M. and Quan, S.F. 2007. Relationship between reported and measured sleep times: the sleep heart health study (SHHS). *Journal of Clinical Sleep Medicine*. **3**(6), pp. 622-630.

Sirevaag, E.J. and Stern, J.A. 2000. Ocular measures of fatigue and cognitive factors. In: R.W. Backs and W. Boucsein. eds. *Engineering psychophysiology*. Mahwah: Lawrence Erlbaum Associates, pp. 269–288.

Skype. n.d. Microsoft (7.30.0.105). [Software]. [Accessed 15 January 2017].

Snyder, E., Cai, B., DeMuro, C., Morrison, M.F., Ball, W. 2018. A new single-item sleep quality scale: Results of psychometric evaluation in patients with chronic primary insomnia and depression. *Journal of Clinical Sleep Medicine*. **14**(11), pp. 1849-1857. Doi: 10.5664/jcsm.7478.

Spencer, M.B. and Robertson, K.A. 2002. *Aircrew alertness during short-haul operations, including the impact of early starts*. Report No. CRO10406/1.0. [Online]. Farnborough: QinetiQ Centre for Human Sciences. [Accessed 11 December 2019]. Available from: https://www.eurocockpit.be/sites/default/files/Spencer-Robertson-2002.pdf.

StackExchange 2019 *Aviation*. [Online]. [Accessed 11 December 2019]. Available from: https://aviation.stackexchange.com/questions/19229/how-would-the-airbus-a320-pfd-respond-to-failure-of-all-pitot-and-aoa-sensors.

Statistics Canada 2011. *Commuting to work*. Report No. 99-012-2011003. [Online]. Ottawa: Statistics Canada. [Accessed 11 December 2019]. Available from: https://www12.statcan.gc.ca/nhs-enm/2011/as-sa/99-012-x/99-012-x2011003_1-eng.pdf.

Steptoe, A. and Bostock, S. 2011. *A survey of fatigue and well-being among commercial airline pilots*. Final Report. London: UCL psychology group department of epidemiology and public health.

Sterman, B., Schummer, G., Dushenko, T. and Smith, J. 1987. Electroencephalographic correlates of pilot performance: simulation and in-flight studies. In Jessen, K. ed. *Electrical and magnetic activity of the central nervous system: research and clinical applications in aerospace medicine*. Neuilly-sur-Seine: NATO Advisory Group for Aerospace Research and Development, pp. 1-16.

Stern, J.A., Walrath, L.C. and Goldstein, R. 1984 The endogenous eyeblink. *Psychophysiology*. **21**(1), pp. 22–33. Doi: 10.1111/j.1469-8986.1984.tb02312.x.

Stunkel, L. and Grady, C. 2011. More than the money: A review of the literature examining healthy volunteer motivations. *Contemporary Clinical Trials*. **32**, pp. 342-352. Doi: 10.1016/j.cct.2010.12.003.

Stutts, J.C., Wilkins, J.W., Osberg, J.S. and Vaughn, B.V. 2003. Driver risk factors for sleep-related crashes. *Accident Analysis and Prevention.* **35**(3), pp. 321-331. Doi: 10.1016/s0001-4575(02)00007-6.

Summala, H., Hakkanen-Nyholm, H., Mikkola, T. and Sinkkonen, J. 1999. Task effects on fatigue symptoms in overnight driving. *Ergonomics*. **42**(6), pp. 798–806. Doi: 10.1080/0014013991 85298.

Tabachnick, B. G. and Fidell, L. S. 2007. *Using multivariate statistics*. Boston: Pearson.

Tacview. 2020. Tacview. The universal flight analysis tool (1.7.4). [Software]. Accessed 07 March 2020].

Teasdale, N., Cantin, V., Blouin, J. and Simoneau, M. 2004. Attentional demands while driving in a simulator: effects of driving straights on open roads, approaching intersections and doubling maneuvers. *Advances in Transportation Studies.* **41**, pp. 75-84.

Teh, E., Jamson, S., Carsten, O. and Jamson, H. 2014. Temporal fluctuations in driving demand: the effect of traffic complexity on subjective measures of workload and driving performance. *Transportation Research Part F: Traffic Psychology and Behaviour.* **22**, pp. 207-217. Doi: 10.1016/j.trf.2013.12.005.

Terry, P., Lane, A. and Fogarty, G. 2003. Construct validity of the profile of mood states - adolescents for use with adults. *Psychology of Sport and Exercise.* **4**, pp. 125-139. Doi: 10.1016/S1469-0292(01)00035-8.

Thiffault, P. and Bergeron, J. 2003. Monotony of road environment and driver fatigue: a simulator study. *Accident Analysis and Prevention.* **35**, pp. 381–391. Doi: 10.1016/s0001-4575(02)00014-3.

Thomas, M.J. and Ferguson, S.A. 2010. Prior sleep, prior wake, and crew performance during normal flight operations. Aviation Space and Environmental Medicine. **81**(7), pp. 665-670. Doi: 10.3357/ASEM.2711.2010.

Tietzel, A.J. and Lack, L.C. 2002 The recuperative value of brief and ultra-brief naps on alertness and cognitive performance. *Journal of Sleep Research*. **11**, pp. 213 – 218. Doi: 10.1046/j.1365-2869.2002.00299.x.

Ting, P.H., Hwang, J.R., Doong, J.L. and Jeng, M.C. 2008. Driver fatigue and highway driving: a simulator study. *Physiology and Behavior.* **94**(3), pp. 448-453. Doi: 10.1016/j.physbeh.2008.02.015.

Tonetti, L., Pasquini, F., Fabbri, M., Belluzzi, M. and Natale, V. 2008. Comparison of two different actigraphs with polysomnography in healthy young subjects. *Chronobiology International*. **25**(1), pp. 145-153. Doi: 10.1080/07420520801897228.

Tops, M., Lorist, M.M., Wijers, A.A. and Meijman, T.F. 2004. To stress or relax: Neurochemical aspects of activity and rest. *Gedrag Organisatie*. **17**, pp. 32-42.

Tornros, J. 1998. Driving behaviour in a real and a simulated road tunnel - a validation study. *Accident Analysis and Prevention,* **30**, pp. 497-503. Doi: 10.1016/S0001-4575(97)00099-7.

Tsang, P.S. and Velazquez, V.L. 1996. Diagnosticity and multidimensional subjective workload ratings. *Ergonomics.* **39**(3), pp. 358–381. Doi: 10.1080/00140139608964470.

Tucker, A.M., Dinges, D.F. and Van Dongen, H.P. 2007. Trait interindividual differences in the sleep physiology of healthy young adults. *Journal of Sleep Research.* **16**(2), pp. 170-180. Doi: https://doi.org/10.1111/j.1365-2869.2007.00594.x.

U.S. Department of Transportation 2018. *FAA approval of aviation training devices and their use for training and experience*. Advisory Circular 61-136B. [Online]. Washington: U.S. Department of Transportation. [Accessed 11 December 2019]. Available from: https://www.faa.gov/documentLibrary/media/Advisory_Circular/AC_61-136B.pdf.

Van de Water, A.T., Holmes, A. and Hurley, D.A. 2011. Objective measurements of sleep for non-laboratory settings as alternatives to polysomnography - a systematic review. *Journal of Sleep Research.* **20**, pp. 183-200. Doi: 10.1111/j.1365-2869.2009.00814.x.

Van der Hulst, M., Meijman, T. And Rothengatter, T. 2001. Maintaining task set under fatigue: A study of time-on-task effects in simulated driving. *Transportation Research Part F: Traffic Psychology and Behaviour*. **4**(2), pp. 103-118. Doi: 10.1016/S1369-8478(01)00017-1.

Van der Linden, D. and Eling, P. 2006. Mental fatigue disturbs local processing more than global processing. *Psychological Research.* **70**, pp. 395-402.

Van der Linden, D., Frese, M. and Meijman, T.F. 2003. Mental fatigue and the control of cognitive processes: effects on perseveration and planning. *Acta Psychologica.* **113**, pp. 45-65. Doi: 10.1016/S0001-6918(02)00150-6.

Van Dongen, H., Baynard, M.D., Maislin, G. and Dinges, D.F. (2004). Systematic interindividual differences in neurobehavioral impairment from sleep loss: Evidence of trait-like differential vulnerability. *Sleep*. **27**(3), pp. 423-433.

Van Dongen, H.P. and Belenky, G. 2009. Individual differences in vulnerability to sleep loss in the work environment. Industrial Health. **47**(5), pp. 518-526. Doi: 10.2486/indhealth.47.518.

Van Drongelen, A., Boot, C.R., Hlobil, H., Smid, T. and van der Beek, D. 2017. Risk factors for fatigue among airline pilots. *International Archives of Occupational and Environmental Health.* **90**, pp. 39. Doi: 10.1007/s00420-016-1170-2.

Vakulin, A., Baulk, S.D., Catcheside, P.G., Anderson, R., van den Heuvel, C.J., Banks, S. and McEvoy, R.D. 2007. Effects of moderate sleep deprivation and low-dose alcohol on driving simulator performance and perception in young men. *Sleep*. **30**(10), pp. 1327–1333. Doi: 10.1093/sleep/30.10.1327.

Van Orden, K.F., Jung, T.P. and Makeig, S. 2000. Combined eye activity measures accurately estimate changes in sustained visual task performance. *Biological Psychology*. **52**(3), pp. 221–240. Doi: 10.1016/S0301-0511(99)00043-5.

Van Veen, S. 2016. *Driver vitalization. Investigating sensory stimulation to achieve a positive driving experience*. Ph.D. thesis. TU Delft.

Veltman, J.A. and Gaillard, A.W.K. 1998. Physiological workload reactions to increasing levels of task difficulty. *Ergonomics.* **41**(5), pp. 656–669. Doi: 10.1080/001401398186829.

Verwey, W. and Veltman, H. 1996. Detecting short periods of elevated workload: a comparison of nine workload assessment techniques. *Journal of Experimental Psychology: Applied.* **2**, pp. 270-285.

Verwey, W.B. and Zaidel, D.M. 1999. Preventing drowsiness accidents by an alertness maintenance device. *Accident Analysis and Prevention.* **31**, pp. 199-211. Doi: 10.1016/s000 1-4575(98)00062-1.

Vincent, G., Aisbett, B., Hall, S. and Ferguson, S. 2015. Fighting fire and fatigue: Sleep quantity and quality during multi-day wildfire suppression. *Ergonomics*. **59**(7), pp. 1-24. Doi: 10.1080/00140139.2015.1105389.

Vink, J.M., Jacqueline M., Vink, A.S., Groot, G., Kerkhof, A. and Boomsma, D.I. 2001. Genetic analysis of morningness and eveningness. *Chronobiology International.* **18**(5), pp. 809-822. Doi: 10.1081/cbi-100107516.

Vogelpohl, T., Kuhn, M., Hummel, T. and Vollrath, M. 2019. Asleep at the automated wheel-sleepiness and fatigue during highly automated driving, *Accident Analysis and Prevention.* **126**, pp. 70-84. Doi: 10.1016/j.aap.2018.03.013.

Walker, M. 2017. *Why we Sleep. Unlocking the Power of Sleep and Dreams.* New York: Scribner.

Walsleben, J., Norman, R., Novak, R., O' Malley, E., Rapoport, D. and Strohl, K. 1999. Sleep habits of Long Island rail road commuters. *Sleep.* **22**, pp. 728-734. Doi: 10.1093/sleep/22.6.728.

Wang, Y., Zhu, H., Zhou, F. and Zhao, Y. 2018. How driving duration influences drivers' visual behaviors and fatigue awareness: A naturalistic truck driving test Study. *Ethiopian Journal of Health Development*. **32**(1), pp. 1-10.

Warm, J., Parasuraman, R. and Matthews, G. 2008. Vigilance requires hard mental work and is stressful. *Human factors.* **50**, pp. 433-441. Doi: 10.1518/001872008X312152.

Waterhouse, J., Drust, B., Weinert, D., Edwards, B., Gregson, W., Atkinson, G., Kao, S., Aizawa, S. and Reilly, T. 2005. The circadian rhythm of core temperature: origin and some implications for exercise performance. *Chronobiology International.* **22**(2), pp. 207-225. Doi: 10.1081/CBI-200053477.

Watling, C.N. 2016. *The sleep and wake drives: Exploring the genetic and psychophysiological aspects of sleepiness, motivation, and performance*. Ph.D. thesis. Queensland University.

Watling, C.N., Smith, S.S. and Horswill, M.S. 2014. Stop and revive? The effectiveness of nap and active rest breaks for reducing driver sleepiness. *Psychophysiology*. **51**, pp. 1131-1138. Doi: 10.1111/psyp.12256.

Wehrens, S., Christou, S., Isherwood, C., Middleton, B., Gibbs, M., Archer, S., Skene, D. and Johnston, J. 2017. Meal timing regulates the human circadian system. *Current Biology*. **27**(12), pp. 1768-1775. Doi: 10.1016/j.cub.2017.04.059.

Wickens, C.D. 1984. Processing resources in attention. In Parasuraman, R. and Davies, D.R. eds. *Varieties of attention.* London: Academic Press, pp. 63-102.

Wierwille, W.W. and Casali, J.G.A. 1983. Validated rating scale for global mental workload measurement applications. In: Proceedings of the 27th Annual Meeting of the Human Factors Society, 10-14 October, Santa Monica. [Online]. Santa Monica: Human Factors Society, pp. 129-133. [Accessed 11 January 2018]. Available from: https://journals.sagepub.

Wierwille, W. and Eggemeier, F.T. 1993. Recommendations for mental workload measurement in a test and evaluation environment. *Human Factors.* **35**(2), pp. 263–281. Doi: 10.1177/001872089303500205.

Wierwille, W. and Ellsworth, A. 1994. Evaluation of driver drowsiness by trained raters. *Accident Analysis and Prevention.* **26**(5), pp. 571-581. Doi: 10.1016/0001-4575(94)90019-1.

Wikipedia 2020. *Primary flight display*. [Online]. [Accessed 14 May 2020]. Available from: https://en.wikipedia.org/wiki/Primary_flight_display.

Wikipedia 2019. *Cessna 172*. [Online]. [Accessed 11 December 2019]. Available from: https:// en.wikipedia.org/wiki/Cessna_172.

Wilhelm, B., Wilhelm, H., Luedtke, H.D., Streicher, P. and Adler, M. 1998. Pupillographic assessment of sleepiness in sleep-deprived healthy subjects. *Sleep.* **21**(3), pp. 258-265.

Williamson, A.M., Feyer, A.M., Mattick, R.P., Friswell, R. and Finlay-Brown, S. 2001. Developing measures of fatigue using an alcohol comparison to validate the effects of fatigue on performance. *Accident Analysis and Prevention*. **33**, pp. 313-326. Doi: 10.1016/j.aap.201 4.04.007.

Williamson, A., Friswell, R., Olivier, J. and Grzebieta, R. 2014. Are drivers aware of sleepiness and increasing crash risk while driving? *Accident Analysis & Prevention*. **70**, pp. 225-234. Doi: 10.1016/j.aap.2014.04.007.

Williamson, A., Lombardi, D., Folkard, S. and Stutts, J. 2011. The link between fatigue and safety. *Accident Analysis and Prevention.* **43**(2), pp. 498-515. Doi: 10.1016/ j.aap.2009.11.0 11.

Wilson, G., Russell, C. and Caldwell, J. 2006. Performance and psychophysiological measures of fatigue effects on aviation related tasks on varying difficulty. *International Journal of Aviation Psychology*. **17**(2): pp. 51. Doi: 10.1080/10508410701328839.

Winter, J., Happee, R., Martens, M. and Stanton, N. 2014. Effects of adaptive cruise control and highly automated driving on workload and situation awareness: a review of the empirical evidence. *Transportation Research Part F.* **27**, pp. 196-217. Doi: 10.1016/j.trf.2014.06.016.

Wobbrock, J.O., Findlater, L., Gergle, D. and Higgins, J.J. 2011. The aligned rank transform for non-parametric factorial analyses using only ANOVA procedures. In: *Proceedings of the CHI, 7–12 May 2011, Vancouver*. [Online]. Vancouver: CHI, pp. 143-146. [Accessed 06 January 2020]. Available from: http://faculty.washington.edu/wobbrock/pubs/chi-11.06.pdf.

Woods, D.L., Wyma, J.M., Yund, E.W., Herron, T.J., and Reed, B. 2015. Factors influencing the latency of simple reaction time. *Frontiers in human neuroscience*. **9**, pp. 131-143. Doi: 10.3389/fnhum.2015.00131.

Women in Aviation 2012. *Women in Aviation. The facts*. [Accessed 10 November 2016]. Available from: https://www.wai.org/resources/waistats.

Wright, N. and McGown, A. 2001. Vigilance on the civil flight deck: incidence of sleepiness and sleep during long-haul flights and associated changes in physiological parameters. *Ergonomics*. **44**(1), pp. 82-106. Doi: 10.1080/00140130150203893.

Wu, Y.H., Aksan, N., Rizzo, M., Stangl, E., Zhang, X. and Bentler, R. 2014 Measuring listening effort: driving simulator versus simple dual-task paradigm. *Ear Hear*. **35**(6), pp. 623-32. Doi: 10.1097/AUD.0000000000000079.

Wyon, D.P., Wyon, I. and Norin, F. 1996. Effects of moderate heat stress on driver vigilance in a moving vehicle. *Ergonomics*. **39**(1), pp. 61-75. Doi: 10.1080/00140139608964434.

Xiao, Y.M., Wang, Z.M., Wang, M.Z. and Lan, Y.J. 2005. The appraisal of reliability and validity of subjective workload assessment technique and NASA-task load index. *Zhonghua Lao Dong Wei Sheng Zhi Ye Bing Za Zhi*. **23**(3), pp. 178-81PMID: 16124892.

Young, M.S and Stanton, N.A. 2004. Taking the load off: investigations of how adaptive cruise control affects mental workload. *Ergonomics*. **47**(9), pp. 1014-1035. Doi: 10.1080/0014013 0410001686348.

Zakariassen, E., Waage, S., Harris, A., Gatterbauer-Trischler, P., Lang, B., Voelckel, W., Pallesen, S. and Bjorvatn, B. 2019. Causes and management of sleepiness among pilots in a Norwegian and an Austrian air ambulance service - a comparative study. *Air Medical Journal*. **38**(1), pp. 25-29. Doi: 10.1016/j.amj.2018.11.002.

Zhang, H., Wu, C., Yan, X. and Qiu, T.Z. 2016. The effect of fatigue driving on car following behavior. Transportation Research Part F: Traffic Psychology and Behaviour. 43(), pp. 80-89. Doi: 10.1016/j.trf.2016.06.017.

Zhao, C., Zhao, M., Liu, J. and Zheng, C. 2012. Electroencephalogram and electrocardiograph assessment of mental fatigue in a driving simulator. *Accident Analysis and Prevention*. **45**, pp. 83-90. Doi: 10.1016/j.aap.2011.11.019.

Zhu, M., Liu, W. and Wargocki, P. 2019. Changes in EEG signals during the cognitive activity at varying air temperature and relative humidity. *Journal of Exposure Science and Environmental Epidemiology*. **30**, pp. 285-298. Doi: 10.1038/s41370-019-0154-1.

Zijlstra, F. 1993. *Efficiency in work behaviour: a design approach for modern tools*. Ph.D. thesis, Delft University.

Zils, E., Sprenger, A., Heide, W., Born, J. and Gais, S. 2005. Differential effects of sleep deprivation on saccadic eye movements. *Sleep*. **28**(9), pp. 1109-1115. Doi:10.1093/sleep/28 .9.1109.

Chapter 9. Appendices

Appendix A. The methods used in this thesis to collect data about fatigue, control for the effect of contributors to it, and measure its expressions.

Contributors controlled by study design	
Task load	• Increased task load in the 60-minute drives (both experiments) • Changes of task load between stages of flight (both experiments) • Reduced task load in the short drives (both experiments) • Increased task load in the first half of the 20-minute drive (experiment 2)
Time on task	60-minute drives (both experiments)
Time awake since last sleep	Trials run on certain times of the day (both experiments), asked the participants in experiment 2
Time of day	Trials run on certain times of the day (both experiments)
Ambient temperature	Between 22 and 25°C (both experiments)
Vibration	No vibration of the seats (both experiments)
Body posture	Participants sat upright in both experiments but stood up during the breaks in experiment 2
Light	Ambient light adjusted to room light (both experiments)
Contributors controlled by collecting data	
Task load	Item developed for this thesis (survey) and NASA TLX (experiment 1)
Age	Age data (all studies)

Sleep duration	Usual sleep duration (survey) and self-reported sleep duration the night before the trials (both experiments)
Sleep quality	Item from St Mary's hospital sleep questionnaire (both experiments)
Data collected to measure the expressions of fatigue	
Subjective fatigue	CSS (survey and experiments)
Behaviour	• Vigilance: PVT (both experiments) • Task-related performance: diving and flying simulation (both experiments)

Appendix B. The MATB-II main scenario

MATB-II main scenario

Block	Active tasks	Events	Start time	End time
1 (00:00 – 04:00)	Tracking	Tracking	00:00	04:00
CSS				
2 (04:01 – 08:00)	Tracking, monitoring, and communications	Tracking	04:01	08:00
		Red light	04:10	04:20
		Scale 2 down	04:10	04:20
		Scale 3 down	04:30	04:40
		Green light	04:30	04:40
		COM 2, 124.450	04:30	04:55
		Scale 4 down	05:00	05:10
		Green light	05:00	05:10
		Red light	05:20	05:30
		Red light	05:40	05:50
		NAV 2, 114.600	05:40	06:05
		Green light	05:55	06:05
		Red light	06:20	06:30
		Green light	06:20	06:30
		Scale 2 up	06:40	06:50
		Scale 4 up	06:40	06:50
		NAV 1, 115.400	06:40	07:05
		Green light	07:05	07:15
		Scale 1 down	07:05	07:15
		Red light	07:15	07:25
		Red light	07:30	07:40
		Com 1, 128.475	07:30	07:55
		Green light	07:45	07:55
CSS				
3 (08:01 – 24:00)	Tracking	Tracking	08:01	24:00

CSS				
4 (24:01 – 28:00)	Tracking, monitoring, and communications	Tracking	24:01	28:00
		Red light	24:10	24:20
		Scale 2 down	24:10	24:20
		Scale 1 up	24:30	24:40
		Green light	24:30	24:40
		NAV 2, 112.450	24:30	24:55
		Scale 3 down	25:00	25;10
		Green light	25:00	25:10
		Red light	25:20	25:30
		Red light	25:40	25:50
		COM 2, 127.525	25:40	26:05
		Green light	25:55	26:05
		Red light	26:20	26:30
		Green light	26:20	26:30
		Scale 3 up	26:40	26:50
		Scale 1 up	26:40	26:50
		NAV 1, 113.600	26:40	27:05
		Green light	27:05	27:15
		Scale 4 up	27:05	27:15
		Red light	27:15	27:25
		Red light	27:30	27:40
		COM 1, 125.550	27:30	27:55
		Green light	27:45	27:55
CSS				
5 (28:01 – 32:00)	Tracking	Tracking	28:01	32:00
CSS				

Appendix C. The X-Plane 10 tasks

Stage	Instructions for manoeuvres	Timing (min.)	Number of communications tasks	Timing (min.) of arithmetic tasks
1	360° straight and level, 80 knots, 6,000 feet MSL	00:00 – 03:00	6	00:15 – 00:23.5
				00:50 – 00:58.5
				01:15 – 01:23.5
				01:35 – 01:43.5
				02:10 –02:18.5
				02:40 –02:48.5
	Right 360° turn (rate 1), 70 knots, 6,000 feet MSL	03:00.01 – 05:30	4	03:20 – 03:28.5
				03:40 – 03:48.5
				04:10 – 04:18.5
				04:45 – 04:53.5
	360° straight and level, 80 knots, 6,000 feet MSL	05:30.01 – 07:30	4	05:35 – 05:43.5
				06:15 – 06:23.5
				06:45 – 06:53.5
				07:10 – 07:18.5
	Straight climb from 6,000 to 6,500 MSL, 70 knots, climb rate 500 feet/minute	07:30.01 – 09:30	2	07:50 – 07:58.5
				08:30 – 08:38.5
	360° straight and level, 80 knots, 6,500 feet MSL	09:30.01 – 11:30	4	09:45 – 09:53.5
				10:10 – 10:18.5
				10:35 – 10:43.5
				10:00 -11:08.5
	Left 360° turn (rate 1), 70 knots, 6,500 feet MSL	11:30.01 – 14:00	4	11:50 – 11:58.5
				12:15 – 12:23.5
				12:40 – 12:48.5
				13:00 – 13:08.5
CSS				

2	360° straight and level, 80 knots, 6,500 feet MSL	15:01.01 – 37.00	0	-
CSS				
3	Left 360° turn (rate 1), 70 knots, 6,500 feet MSL	38:00.01 –40:30	4	38:20 – 38:28.5
				38:45 – 38:53.5
				39:20 – 39:28.5
				39:40 – 39:48.5
	360° straight and level, 80 knots, 6,500 feet MSL	40:30.01 – 42:30	4	40:45 – 40:53.5
				41:10 – 41:18.5
				41:35 – 41:43.5
				42:00 – 42:08.5
	Straight descent from 6,500 to 6,000 feet MSL, 70 knots, descent rate of 500 feet/minute	42:30.01 – 44:30	2	42:50 – 42:58.5
				43:30 – 43:38.5
	360° straight and level, 80 knots, 6,000 feet MSL	44:30.01 – 46:30	4	44:35 – 44:43.5
				45:05 – 45:13.5
				45:35 – 45:43.5
				46:00 – 46:08.5
	Right 360° turn (rate 1), 70 knots, 6,000 feet MSL	46:30.01 – 49:00	4	46:50 – 46:58.5
				47:10 – 47:18.5
				47:40 – 47:48.5
				48:15 – 48:23.5
	360° straight and level, 80 knots, 6,000 feet MSL	49:00.01 – 52:00	6	49:15 – 49:23.5
				49:50 – 49:58.5
				50:15 – 50:23.5
				50:35 – 50:43.5
				51:10 – 51:18.5
				51:40 – 51:48.5
CSS				