Czech Technical University in Prague
Faculty of Transportation sciences
Department of Air transportation

Metody hodnocení únavy u pilotů
Methods for evaluation of fatigue in pilots

Bakalářská práce

Bachelor thesis

Author: **Marina Zhuzhgova**

Bachelor thesis supervisor: **Ing. Lenka Hanáková, doc. Ing. Bc. Vladimír Socha, Ph.D.,**

Akademic year: **2020/2021**

ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE
Fakulta dopravní
děkan
Konviktská 20, 110 00 Praha 1

K621 .... **Ústav letecké dopravy**

ZADÁNÍ BAKALÁŘSKÉ PRÁCE
(PROJEKTU, UMĚLECKÉHO DÍLA, UMĚLECKÉHO VÝKONU)

Jméno a příjmení studenta (včetně titulů):

Marina Zhuzhgova

Kód studijního programu a studijní obor studenta:

B 3710 – LED – Letecká doprava

Název tématu (česky): **Metody hodnocení únavy u pilotů**

Název tématu (anglicky): Methods for Evaluation of Fatigue in Pilots

Zásady pro vypracování

Při zpracování bakalářské práce se řiďte následujícími pokyny:

- Cílem práce je vytvořit přehled stávajících metod hodnocení únavy u pilotů na základě publikovaných odborných výstupů, a s ohledem na analýzu jejich výsledků selektovat takové metody, které vykazují nejvyšší potenciál jako kvantifikátory únavy u pilotů.
- Vypracujte rešerši současného stavu v oblasti subjektivního a objektivního hodnocení únavy u leteckých specialistů s primárním zaměřením na piloty.
- Proveďte selekci článků vhodných pro následnou analýzu.
- Proveďte statistickou analýzu dat získaných z vybraných článků a dále určete metody s nejvyšší importancí (v případě dostatečného množství dat), případně formulujte doporučené postupy pro hodnocení únavy vyplývající z vykonané analýzy (v případě nedostatečného množství dat pro statistickou analýzu).
- Proveďte diskuzi a formulujte závěry práce.

ČVUT
FD

Rozsah grafických prací:	dle pokynů vedoucího bakalářské práce
Rozsah průvodní zprávy:	minimálně 35 stran textu (včetně obrázků, grafů a tabulek, které jsou součástí průvodní zprávy)
Seznam odborné literatury:	BORENSTEIN, M. et al. Introduction to Meta-Analysis. 2011. John Wiley & Sons. CALDWELL, John A. Fatigue in aviation. Travel Medicine and Infectious Disease. 2005, 3(2), 85-96.

Vedoucí bakalářské práce: **Ing. Lenka Hanáková**

doc. Ing. Bc. Vladimír Socha, Ph.D.

Datum zadání bakalářské práce: **9. října 2019**
(datum prvního zadání této práce, které musí být nejpozději 10 měsíců před datem prvního předpokládaného odevzdání této práce vyplývajícího ze standardní doby studia)

Datum odevzdání bakalářské práce: **1. prosince 2020**

a) datum prvního předpokládaného odevzdání práce vyplývající ze standardní doby studia a z doporučeného časového plánu studia
b) v případě odkladu odevzdání práce následující datum odevzdání práce vyplývající z doporučeného časového plánu studia

ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE Fakulta dopravní

L. S.

Kraus

doc. Ing. Jakub Kraus, Ph.D.
vedoucí
Ústavu letecké dopravy

doc. Ing. Pavel Hrubeš, Ph.D.
děkan fakulty

Potvrzuji převzetí zadání bakalářské práce.

Marina Zhuzhgova
jméno a podpis studenta

V Praze dne 16. července 2020

CZECH TECHNICAL UNIVERSITY IN PRAGUE
Faculty of Transportation Sciences
Dean's office
Konviktská 20, 110 00 Prague 1, Czech Republic

K621 .... **Department of Air Transport**

BACHELOR'S THESIS ASSIGNMENT

(PROJECT, WORK OF ART)

Student's name and surname (including degrees):

Marina Zhuzhgova

Code of study programme code and study field of the student:

B 3710 – LED – Air Transport

Theme title (in Czech): **Metody hodnocení únavy u pilotů**

Theme title (in English): Methods for Evaluation of Fatigue in Pilots

Guides for elaboration

During the elaboration of the bachelor's thesis follow the outline below:

- The aim of the thesis is to create an overview of existing methods of fatigue evaluation in pilots on the basis of scientific publications and with regard to the analysis of their results select those methods that show the highest potential as pilots fatigue quantifiers.
- Prepare a state-of-the-art review of subjective and objective fatigue evaluation in aviation specialists with a primary focus on pilots.
- Select suitable articles for subsequent analysis.
- Perform a statistical analysis of the data obtained from the selected articles and determine the methods with the highest importance (in case of sufficient data), or formulate recommended procedures for fatigue evaluation resulting from the performed analysis (in case of insufficient data for statistical analysis).
- Discuss outcomes of the thesis and formulate conclusions.

Graphical work range: According to the instructions of the supervisor

Accompanying report length: At least 35 pages of text (including pictures, graphs and tables included in the accompanying report)

Bibliography: BORENSTEIN, M. et al. Introduction to Meta-Analysis. 2011. John Wiley & Sons.

CALDWELL, John A. Fatigue in aviation. Travel Medicine and Infectious Disease. 2005, 3(2), 85-96.

Bachelor's thesis supervisor: **Ing. Lenka Hanáková**
doc. Ing. Bc. Vladimír Socha, Ph.D.

Date of bachelor's thesis assignment: **October 9, 2019**
(date of the first assignment of this work, that has be minimum of 10 months before the deadline of the theses submission based on the standard duration of the study)

Date of bachelor's thesis submission: **December 1, 2020**
a) date of first anticipated submission of the thesis based on the standard study duration and the recommended study time schedule
b) in case of postponing the submission of the thesis, next submission date results from the recommended time schedule

L. S.

Kraus

doc. Ing. Jakub Kraus, Ph.D.
head of the Department
of Air Transport

doc. Ing. Pavel Hrubeš, Ph.D.
dean of the faculty

I confirm assumption of thesis assignment.

Marina Zhuzhgova
Student's name and signature

Prague .... July 16, 2020

Poděkovaní

V průběhu celého procesu vytváření bakalářské práce byla jsem velice podporovaná skupinou lidí, kterým bych chtěla poděkovat v této části práce. Chtěla bych zaprvé poděkovat doc. Ing. Bc. Vladimíru Sochovi,Ph.D. a Ing. Lence Hanákové za profesionální dohled nad tímto projektem, neocenitelná doporučení a nápady k opravám a k vylepšení práce. Navíc bych chtěla ocenit obrovskou pomoc moje rodiny, která mě nadšeně podporovala na moje celé cestě ke vzdělávání.

Acknowledgements

During the entire process of the thesis creation I was receiving a lot of support from a particular group of people, whom I would like to thank in this part of the paper. I would like to thank both doc. Ing. Bc. Vladimír Socha,Ph.D. and Ing. Lenka Hanáková for their professional supervision of this project, invaluable recommendations and ideas for corrections and improvements of the study. Moreover, I would like to acknowledge an inestimable support of my family, who enthusiastically encouraged me on my whole path of education.

Čestné Prohlášení

Přikládám tímto k posouzení a obhajobě bakalářskou práci, zpracovanou na závěr studia na ČVUT v Praze Fakultě dopravní.

Prohlašuji, že jsem předloženou práci vypracovala samostatně a že jsem uvedla veškeré použité informační zdroje v souladu s Metodickým pokynem o dodržování etických principů při přípravě vysokoškolských závěrečných prací.

Nemám závažný důvod proti užívání tohoto školního díla ve smyslu § 60 Zákona č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon).

Statutory Declaration

I hand in for assessment and defense a bachelor thesis, prepared at the end of my studies at the Czech Technical University in Prague, Faculty of Transport.

I herewith formally declare that I have written the submitted bachelor thesis independently. I did not use any outside support except for the quoted literature and other sources mentioned in the paper, accordingly to Methodical instructions on compliance with ethical principles in the preparation of university theses. I am aware that the violation of this regulation will lead to failure of the thesis.

I have no objection against using this thesis in the sense of § 60 Zákona č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon).

In Prague on ……………………………

Marina Zhuzhgova ……………………………

Signature

Název práce: **Metody hodnocení únavy u pilotů**

Autor: Marina Zhuzhgova

Obor: Letecká doprava

Druh práce: Bakalářská práce

Počet stran: 96

Vedoucí práce: doc. Ing. Bc. Vladimír Socha, Ph.D., Ing. Lenka Hanáková

Abstrakt:

S rychlým rozvojem letectví se stává únava pilotů, která je jedním z klíčových parametrů bezpečnosti letectví, žhavým problémem. Hodně leteckých společností se snaží implementovat strategie zmírnění únavy, ale jen málo studií se zabývá zkoumáním existujících metod hodnocení únavy. Hlavním cílem této práce je vytvořit přehled stávajících nástrojů pro hodnocení únavy shromážděním a analýzou článků o měření únavy pomocí metaanalýzy. Praktická část této práce provádí statistickou analýzu výsledků zkoumaných článků a vyčleňuje metody s největším významem. Později je udělán závěr, že je nezbytné zlepšovat způsoby detekce únavy u pilotů zavedením novějších přesnějších metod hodnocení ospalosti a uznáváním omezení aktuálně používaných metod.

Klíčová slova: únava, letectví, ospalost, fyziologické signály

Title: **Methods for evaluation of fatigue in pilots**

Author: Marina Zhuzhgova

Study field: Air Transport

Type of study: Bachelor thesis

Amount of pages: 96

Thesis supervisor: doc. Ing. Bc. Vladimír Socha, Ph.D., Ing. Lenka Hanáková

Key words: fatigue, aviation, sleepiness, physiological signals

Abstract:

With the aviation sector growing rapidly, fatigue in pilots, as one on the crucial components for aviation safety, is becoming a pressing issue. Many airlines are trying to implement fatigue mitigation strategies, however not many studies are dedicated to discussion of already used methods of evaluating states of low vigilance in detail. The main goal of the current study is to make an overview of existing fatigue assessment tools by means of collecting and analyzing articles on fatigue measurement via the meta-analysis. The practical part of this thesis performs a statistical analysis on results of the papers looked into and outlines methods with the greatest significance. Later in the study a conclusion is made that it is necessary to improve the way fatigue is detected in pilots nowadays with an introduction of newer more accurate methods of drowsiness evaluation and by acknowledging the limitations of the currently used methods.

List of Tables

List of Figures

List of Abbreviations

KSS	Karolinska Sleepiness Scale
SP	Samn-Perelly Scale
NTLx	NASA Task Load Index
PVT	Psychomotor Vigilance Task
FRMS	Fatigue Risk Management System
FSS	Fatigue Severity Scale
ESS	Epworth Sleepiness Scale
RPE	Rating of Perceived Exertion
GVS	Global Vigor Scale
SP	Samn-Perelli
VAS	Vigor Analogue Scale
ULR	ultra long range
TOC	top of climb
TOD	top of descent
SPI	safety performance indicator
LR	long range
TST	total sleep time
SAFE	System for Aircrew Fatigue Evaluation
FDP	flight duty period
BAM	Boeing Alertness Model
EEG	electroencephalogram
GCC	Gulf Cooperation Council
TMD	total mood disturbance
OSA	obstructive sleep apnea
FAI	fatigue assessment instrument
SD	sleep deprivation
FL	flight level
ASR	automatic speech recognition
CM	correlation metric
SOL	sleep onset latency
MFCC	Mel-frequency cepstral coefficient
ROC	receiver operating characteristic
DAL	digital amplitude length
CVR	cockpit voice recorder
RT	reaction time

RS	reaction speed
CFF	Critical Flicker Frequency
MATB	Multi-Attribute Task Battery
OSPAN	Operation Span Task
RMSE	root mean square error
OSCORE	absolute OSPAN score
TSCORE	total number correct score
MFRRT	mean of the reciprocal of the fastest 10% of reaction times
MRRT	mean reciprocal reaction time
MEDRT	median reaction time
ECG	electrocardiography
RespR	respiration rate
EDA	electrodermal activity
HR	heart rate
fNIRS	functional near infrared spectroscopy
PPG	photoplethysmogram
HRV	heart rate variation
POMS	profile of mood states
SBP	systolic blood pressure
BR	breathing rate
MFI	Mental Fatigue Index
PFI	Physical Fatigue Index
LF	low frequency
HF	high frequency
RMSSD	square root of the mean differences of successive NN intervals
SDNN	SD of normal to normal beats
TP	total power of the HRV spectrum
EOG	electrooculogram
EO	eye opening
PD	pupil diameter
PERCLOS	change in percentage of eyelid closure over the pupil time
BF	blinking frequency

Contents

Introduction

The goal of the presented bachelor thesis is to create a review of the currently used methods for evaluation of fatigue in pilots in regard to subdivision of those methods into objective and subjective type.

With an airline industry being one of the most rapidly growing ones and the number of daily operations steadily increasing over the years the problem of pilot fatigue is becoming a pressing issue, as numerous statistics, reports from pilots and operational flight studies reveal. Despite pilot fatigue factor being a crucial component of an aviation safety, fatigue is often not being detected or reported in time, which may undoubtedly lead to catastrophic causalities.

Due to the nature of the human being and the pilots' profession being generally associated with considerable workload, as long duty hours, sleep deprivation and the disruption of circadian rhythm, it is not possible, at least nowadays, to entirely avoid the presence of fatigue in aviators. However, a possible solution might be an invention of more advanced methods for fatigue prevention and timely detection and subsequently their correct implementation into practice.

Among the most commonly used methods for fatigue evaluation belong Karolinska Sleepiness Scale (KSS), Samn-Perelly Scale (SP), NASA Task Load Index (NTLx), various Psychomotor Vigilance Tasks (PVT) and measurement of physiological signals. In some studies, multiple methods are used in combination to produce a mixed method technique or questionnaire for fatigue detection and qualification in order to achieve the highest level of accuracy.

Though this study does not comprehensively examine all existing methods for evaluation of fatigue in pilots, it provides a detailed insight into commonly used methods, aiming to measure their effectiveness and to outline the most reliable techniques for fatigue assessment.

The study has been conducted as a systematic recherché with commentaries provided with an ultimate goal of the research to perform a meta-analysis of chosen articles and to summarize the results or experiments they present. The articles used in this study have been obtained from the electronic databases SCOPUS and Google Scholar. Chapter 1 discusses the choice of meta-analysis and the algorithm of collecting articles for this study. Chapters 2 and 3 review the chosen articles and outline their significant results, for subjective and objective fatigue assessment tools, respectively. A statistical analysis across groups of studies and cross-studies is performed in Chapter 4. Finally, a conclusion is drawn on a basis of the analysis and recommendations are given.

Chapter 1

Meta-analysis

1.1 Meta-analysis

Meta-analysis is an acknowledged type of a statistical analysis, a method implying a systematic collection of qualitative data from multiple studies created in the past to identify their common grounds. In doing so meta-analysis essentially uses a quantitative effect, as results, being consistent across studies, have a greater power statistically than each study result taking separately out of a context.

Undoubtedly, conclusions about the validity of the hypothesis in the modern-day study cannot be based on a single study, as presented results might me misleading or insignificant. On the other hand, while looking into multiple studies, one might notice the extreme heterogeneity of the results or results being inconsistent, varying from study to study. The key advantage of meta-analysis is that it overcomes this issue by allowing to pool information from unlimited amount of studies dedicated to the same area of research and then evaluate the evidence in results of included studies to draw a statistically significant conclusion. Additionally, very often studies are conducted with a limited number of participants without any diversity, making the achieved result not significant enough, meta-analysis overcomes this issue by looking at summarized results collected from many more subjects across the studies.

One of the benefits of the meta-analysis is that it helps to see the overall picture of a studied topic more visually and identify questions left without answers or problematic areas that might require further work. In case of continuing inconsistency of the results the method also might point out reasons for abnormality. Moreover, the meta-analysis is beneficial in outlining common points and crisscrossing areas in investigated studies, correlations and effects of studied subjects on each other. The most common or with the highest importance subjects or methods could be also be outlined from the whole pool by means of the meta-analysis.

Meta-analysis has been wildly used for researches in different fields starting from the beginning of the 20th century. Nevertheless, the most common example of a meta-analysis application in the modern world is probably its usage by pharmaceutical companies, when it is necessary to present statistically proven results of a research in order for a newly created product to be approved to be introduced to the general public.

As for the choice of the meta-analysis for this particular study on methods for evaluation fatigue in pilots, it is due to its ability to collect information on various fatigue methods being examined in previously conducted studies by authors that naturally achieved various results not only in terms of results' inconsistency from study to study, but also in regards to different fatigue assessment tools used in each experiment. In present times already a huge amount of means for fatigue evaluation have been created and proposed, both from subjective and objective spectrum, and thus a high heterogeneity might be already expected among the results of studies that will be analyzed further in this thesis.

1.2 Selection of articles

The main criteria for the selection of articles to be used in the meta-analysis were their availability, quality, accessibility of the content and the primary focus on measurement of fatigue among pilots. For the purpose of the analysis were chosen studies that described an experiment performed on a test group that required a fatigue assessment by either an objective or a subjective method.

Studies were obtained from electronic databases SCOPUS and Google Scolar. For the sake of filtering the articles in those databases to outline ones connected to fatigue a search query "pilot fatigue" was used. However, a further adjustment for the query changing it into "pilot fatigue, aviation" was necessary, as the original pool of results contained too many articles from fields not related to aviation where fatigue was examined. A total number of articles for the 1st step of analysis was 694. Later 8 duplicating articles were removed. In the next phase articles were selected from the pool of results judging primarily on their abstract, this way 547 articles were deemed not to be according to the requirements of this thesis for further meta-analysis or their field of investigation was not aviation. Additionally, articles without full text or text in a foreign language were removed. During closer analysis and studying of the text of the remaining 139 articles, 96 articles had to be excluded due to various reasons. Despite the best effort to filter out non-aviation related articles, some 13 more articles were not focused primarily on pilots or just mentioned a possibility of the method reviewed to be applied to pilot's fatigue. Multiple articles turned out to be reviews, conference papers or reports, with no actual experiment being set, and thus with no results to evaluate. Quite a significant number of articles were dedicated to development and testing of a fatigue risk management system (FRMS) or evaluation of effectiveness of an automatic fatigue recognition system, such articles we also not included into the meta-analysis. Finally, the bigger part of articles being removed on the last step while having a full-text accessible were ones that did not reach any significant results, were in some way reduced or were redirecting to other articles for some parts of research or results.

Even though the majority of articles without a clear presentation of results in terms of precise measured values, and thus not applicable to statistical analysis, were excluded, some still remained in the study either in order for this paper to be more inclusive discussing fatigue evaluation tools, due to some methods being represented by just few studies or due to the character of those studies and the conclusions drawn in them.

The final number of studies used in meta-analysis is 43. A flow diagram of the article selection process done for meta-analysis in this paper is presented below in figure 1.

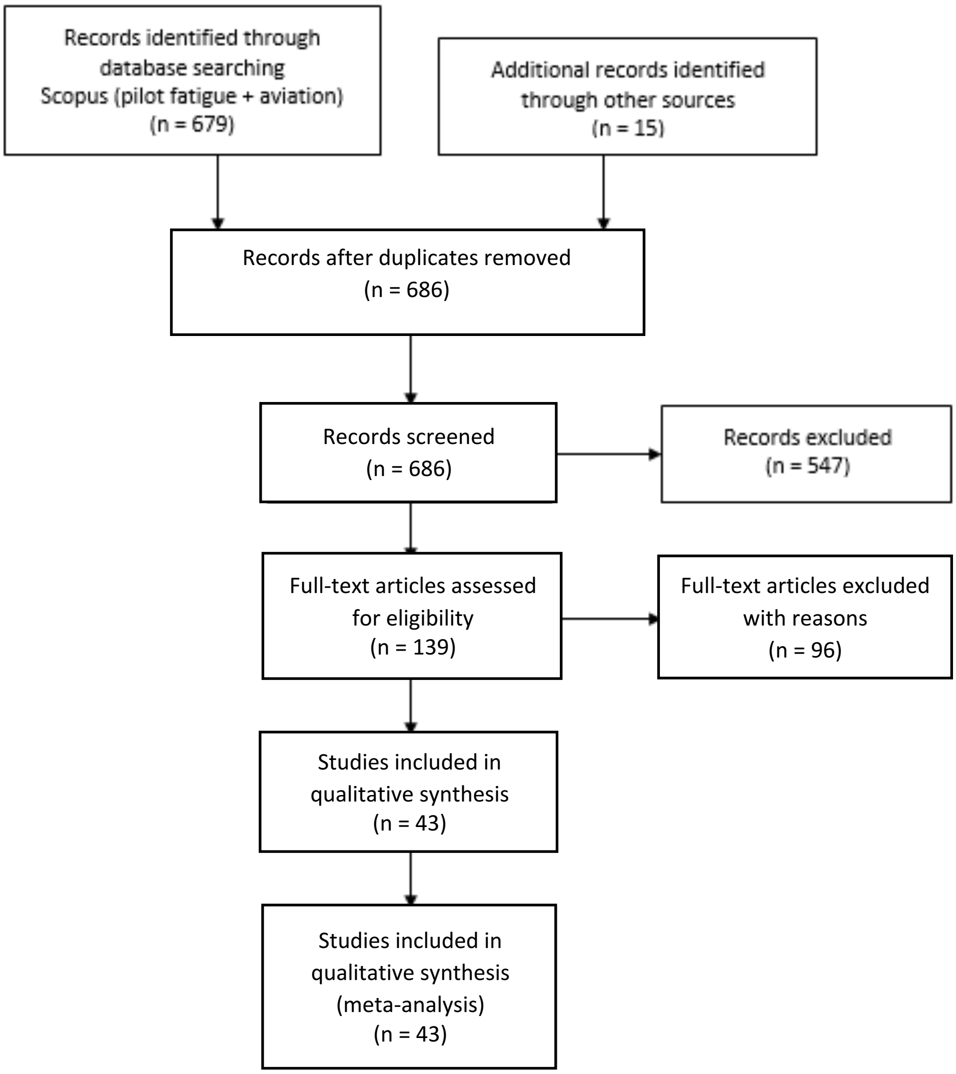

Figure 1: Flow diagram of studies selection for the meta-analysis of the methods for evaluation of fatigue in pilots

Chapter 2

Subjective methods for evaluation of fatigue in pilots

Subjective are the tools for fatigue evaluation that are based on pilot’s personal judgement, feelings and interpretations. They are often influenced by emotions and are not based on facts. Measurements from this group are often presented in a form of questionnaire, checklist or scale and are frequently intended for self-administration. Although not being supported by measurable facts, the subjective fatigue assessment tools are important for understanding pilot’s mental perception of fatigue. In this chapter the following subjective fatigue assessment methods will be discussed: KSS, Fatigue Severity Scale (FSS), Epworth Sleepiness Scale (ESS), Stanford Sleepiness Scale (SSS), Rating of Perceived Exertion (RPE), Global Vigor Scale (GVS), SP, Visual Analogue Scale (VAS) and NTLx.

2.1 Karolinska Sleepiness scale (KSS)

The following part of the text is dedicated to studies where KSS was used as a method for evaluation of fatigue in pilots during an experiment session. KSS is one of the most often used in aviation methods for assessing levels of sleepiness in pilots. KSS is a 9-point scale on which a subject is asked to choose a level of sleepiness that describes subject’s mental and cognitive state in the last 10 minutes best. Points are from 1 - extremely alert to 9 - extremely sleepy, fighting sleep.

Information presented in the reviewed articles is included in table 2.1. Each article chosen describes clearly a methodology of an experiment, achieved results and their interpretation and analysis.

Table 2.1: Meta-analysis of studies examining Karolinska Sleepiness Scale method

Year	Method used	Number of participants	Conclusion	Result
2012	Sleepiness and sleep on ULR flights were assessed. Sleepiness was self-rated by KSS for each flight on every sleep opportunity, pre-flight, at TOC, before the start and at the end of each operating period, at TOD and post-flight.	44	Operating and relief crews were equally sleepy (1) on the outbound flight. Sleepiness elevated at TOD (2) and post-flight (3) when compared to pre-flight sleepiness (4). Pre-sleep sleepiness on return flight was higher (5) than during the outbound flight, post-sleep sleepiness was equal (6). The relief crew were sleepier than the operating crew at TOC (7) and TOD (8) on return flight. Sleepiness was reduced on	**Wilcoxon signed rank/ t-test** (1) KSS < 5; P > .05 (2) KSS = 3.6 ± 1.3; P < .01 (3) KSS = 5.9 ± 1.4; P < .0001 (4) KSS = 2.9 ± 0.7 (5) KSS = 6.5 ± 1.1; Z = −4.1; P < .001 (6) KSS = 4.3 ± 1.5; P > .05 (7) KSS = 4.8 ± 1.8 and 3.4 ± 1.5; Z = −2.8; P = .005 (8) KSS = 4.2 ± 1.8 and 3.2 ± 1.2; Z = −2.05; P = .04 (9) KSS = 5.2 ± 1.4 (pre-sleep); KSS = 3.4 ± 1.1 (post-sleep); Z = −4.3; P < .001

			outbound (9) and inbound (10) flights after sleeping. Post-sleep sleepiness (11) remained stable across the layover and recovery days.	(10) Z = 6.6; P < .001 (11) KSS = 4.0 ± 1.5
2015	Sleepiness of a crew was monitored on ULR flight. Sleepiness was rated pre-flight, at TOC, at TOD, and after landing and during flight leg (outbound and inbound).	55	A great change in sleepiness throughout the flight (1) was noticed. Higher sleepiness rating was measured on the inbound flight compared to the outbound flight (2). Pilots were sleepier at TOD with growing number of hours awake (3). If compared to baseline, crew members were equally sleepy while waking up on any of the post-flight days (4).	**ANOVA** (1) $F_{3,\ 311}$ = 94.40; P < .001 (2) $F_{1,\ 313}$ = 14.32; P < .001 (3) $F_{1,75}$ = 5.06; P = .03 (4) $F_{5,\ 237}$ = 0.57; P =.72
2014	Sleepiness was measured by KSS pre-flight and at TOD as a part of many SPIs measurement on LR and ULR flights in order to assess its accuracy in hopes of using KSS in FRMS.	133	Pre-flight sleepiness (1) and TOD sleepiness (2) were compared among flights. The proportion of crew members rating 7 on the KSS showed no significant: difference for pre-flight sleepiness (3) or TOD sleepiness (4).	**ANOVA** (1) P = .039 (2) P = .255 **The Chi-squared statistic** (3) P = .372; (4) P = .096;
2014	Sleepiness was measured as one of fatigue indicators on various flights in order to compare predictions of SAFE and real measurements.	324	Predicted by SAFE overall fatigue values were closely correlated with measured subjective fatigue.	**Pearson** (1) r = 0.85
2014	Fatigue levels in pilots were compared when finishing a duty late or starting a duty early. Sleepiness as a fatigue indicator.	40	Sleepiness increased in the late evening hours with an increase of time awake. Sleepiness was higher in pilots with FDP late at night.	**NA**
2018	Factors influencing pilots' layover sleep on out-and-back trips as well as a correlation between layover sleep and an inbound pre-flight fatigue. Sleepiness was rated with KSS as one of the SPIs at duty start for the inbound flight.	299	KSS scores showed correlation with amounts of TST (1) and time awake of the pilot (2) and decreased with greater amount of TST. The lowest KSS score was for pilots starting their shift 12:00–15:59 (3), the highest was for those on duty between 00:00–03:59 (4).	**ANOVA** (1) F = 29.15; P < .0001 (2) F = 0.02; P = .8981 **KSS score** (3) KSS = 2.54 (4) KSS = 3.07
2016	Was investigated if subjective measurements of in-flight sleep are as effective as actigraphic measurements. For each outbound and inbound flight sleepiness was measured on KSS at top of descent.	298	Every 1-h increase in sleep duration, decreased sleepiness ratings (1). The sleepiest at TOD were the pilots taking the first break than those taking the second (2) or third break (3). Sleepiness was higher at TOD on flights arriving between 02:00–05:59 domicile time than on flights arriving between 10:00–13:59 (4), 14:00–17:59 (5), and 22:00–01:59 (6). Similarly, higher sleepiness at TOD was measured on flights arriving between 6:00–09:59 than on flights arriving between 10:00–13:59 (7), 14:00–17:59 (8), or 22:00–01:59 (9).	**KSS points** (1) by 0.6 points **ANOVA** (2) t_{352} = 7.31; P < .0001 (3) t_{406} = 4.69; P < .0001 (4) t_{462} = 4.95; P < .0001 (5) t_{427} = 5.92; P < .0001 (6) t_{506} = 4.48; P < .0001 (7) t_{452} = 6.21; P < .0001 (8) t_{470} = 6.89; P < .0001 (9) t_{482} = 6.04; P < .0001

2016	A study took place in order to investigate whether fatigue is affected by a number of flight segments, especially landings and take offs. Subjective sleepiness was rated by each pilot at predetermined times via the KSS.	24	A major effect of condition (1) and a test number (2) was discovered on KSS. Interaction of condition by test number (3) was proven insignificant. The duration of prior sleep proved to be greatly affecting sleepiness (4), when sleepiness was lower even after a 1 hour longer sleep (5). KSS score was affected by participant individuality and a test number (6). Correlation between KSS and PVT lapse were moderate (7).	**ANOVA** (1) $F_{1,437}$ = 17.0; P < .001 (2) $F_{9,437}$ = 7.9; P < .001 (3) $F_{9,437}$ = 1.3; P = .21 (4) $F_{1,436}$ = 28.5; P < .001 **KSS points** (5) by 0.4 ± 0.1 units **Pearson** (6) r = 0.63; P < .001 (7) r = 0.28; P < .001
2014	Sleepiness of pilots on transmeridian flights was examined, study aimed to measure the accuracy of KSS method and if it reflects circadian phase during flight operations. Pilots completed KSS ratings pre-flight and at TOD.	237	Every 1 hour increase in TST decreased, KSS ratings (1). Every 1 hour increase in time awake at the TOD increased KSS ratings (2). Every 1 hour increase in total in-flight sleep decreased KSS ratings (3). There was a difference in pre-flight sleepiness between flights departing between 14:00 -17:59 (4), 22:00 - 01:59 (5) or 02:00 - 05:59 (6). Landing crew scored as sleepier (7). Pre westward outbound flights sleepiness differed from the pre westward inbound (8) or pre eastward outbound flights (9). KSS at the TOD was various for arriving between 06:00 -09:59 (10), 10:00 - 13:59 (11) or 14:00 - 17:59 (12) flights.	**ANOVA/ KSS points** (1) by KSS =0.95 (2) by KSS = 0.2 (3) by KSS = 0.3 (4) KSS = 2.1 (5) KSS = 2.7; P = .0030 (6) KSS = 3.0; P = .0008 (7) by KSS = 0.1 (8) by KSS = 0.4; P = .0224 (9) by KSS = 0.4; P = .0043 (10) KSS = 4.2 (11) KSS = 3.7; P = .0433 (12) KSS = 3.4; P = .0022
2015	During each experimental session participants were instructed to collect data via the 9-point KSS. The comparison was later made between relief and landing and tired and rested crews.	32	Rested and fatigued pilots showed different KSS ratings (1). Sleepiness between short and long-hauls was similar. Sleepiness grew towards the end of each session (2), more rapidly per hour in fatigued pilots (3).	**ANOVA/ KSS points** (1) KSS = 0.85 (2) KSS = 0.22; P < .0001 (3) KSS = 0.25, P < .0001
2014	Measurement of sleepiness via KSS took place as a part of a bigger experiment aimed at evaluating a performance of an algorithm for an automatic detection of fatigue.	14	Participants were checked to not be sleepy before the test (1). Sleepiness was detected only in one pilot after the flight (2). KSS scores before (3) and after the flight (4) showed to be similar.	**KSS points** (1) KSS < 5 (2) KSS > 7 (3) KSS = 2.6 ± 0.9 (4) KSS = 3.2 ± 1.7
2014	Sleepiness was measured before and after each sleep period via KSS in a study aimed to investigate the effectiveness of a fatigue battling guidance material on a westward ULR flight and return LR flight.	52	First post-flight day showed an increased sleepiness (1). The analyses took of the effect of time (2), crew position (3), and flight leg (4) and the interactions of these factors (time **x** position (5), time **x** leg (6), leg **x** position (7) ,time **x** position **x** leg (8)) with sleepiness took place.	**ANOVA** (1) $F_{5, 266}$ = 7.22; P < .001 (2) $F_{3, 282}$ = 35.48; P < .001 (3) $F_{1, 45}$ = 10.55; P = .002 (4) $F_{1, 289}$ = 1.89; P = .171 (5) $F_{3, 282}$ = 7.56; P < .001 (6) $F_{3, 281}$ = 3.20; P = .024 (7) $F_{1, 289}$ = 0.04; P = .841 (8) $F_{3, 281}$ = 5.27; P = .002

2010	KSS was a chosen method for measuring subjective sleepiness in the current study with an aim to assess the influence of an additional day's layover on pilots' fatigue.	125	Sleepiness increased gradually on Brisbane route (1) and Los Angeles Route (2). Time by schedule interaction, being significant, was detected in sleepiness on the Brisbane route (3). Sleepiness was bigger (4) on a return than on an outward flight, with no effect of an additional layover day noticed.	**ANOVA** (1) $F_{1,134}$) = 113.9; P < .001 (2) $F_{1,137}$) = 259.0; P < .001 (3) $F_{4,116}$) = 4.65; P < .01 (4) $F_{1,42}$) = 6.17; P < .05
2006	The study set a goal of investigating differences in sleepiness and sleep on westward morning and evening flights. KSS rating was taken every 2–3 hours from the day before to the day after the experimental flights.	14	The outwardbound flight data indicated an effect of time (1) and interaction (2) on sleepiness. KSS stayed the same on a morning and on an evening flight (3) during homeward-bound flight and the day after the homeward-bound flight (4), however timing has a major effect on sleepiness on homeward-bound (5) and the day after the homeward-bound flight (6). No effect of the morning vs. evening flight on sleepiness on outward-bound and homeward-bound flight (7) was found, however, the direction of flight was significant (8), not the interaction (9).	**ANOVA** (1) $F_{7,12}$ = 19.1; P < .0001 (2) $F_{7,84}$ = 2.6; P < .0163 (3) $F_{1,12}$ = 0.7 (4) $F_{1,12}$ = 0.3 (5) $F_{7,12}$ = 29.7; P = .0001 (6) $F_{5,12}$ = 21.1; P = .0001 (7) $F_{1,12}$ =1.2 (8) $F_{1,12}$ =6.3; P = .0276 (9) $F_{1,12}$ =3.0

ULR - ultra long range, KSS - Karolinska Sleepiness Scale, TOC - top of climb, TOD - top of descent, FRMS - Fatigue Risk Management System, SPI - safety performance indicator, LR - long range, SAFE - System for Aircrew Fatigue Evaluation, FDP - flight duty period, TST - total sleep time, PVT - Psychomotor Vigilance Task, BAM - Boeing Alertness Model

2.1.1 Assessment of sleep and sleepiness on an ultra long range flights

The study [1] set a goal of sleep and sleepiness evaluation of a flight crew during an ultra long range (ULR) operation on routes between Doha, Qatar, and Houston, United States. Data were collected from 44 participating pilots on 11 trips across the same 9-hrs roster. A guidance was provided to the crew prior to the outbound flight, both ULR flights, layover and the first recovery night on sleep and nap schedule, diet, exercise and light regime. For sleep evaluation Actiwatch devices were worn, and a sleep diary with notes, sleep qualification and further sleep need was filled in, sleepiness in the experiment was self-rated by the pilots on a 9-point KSS. Crew members were instructed to rate their sleepiness pre-flight, at top of climb (TOC), at start and end of each operating period, at top of descent (TOD) and post-flight. Critical phases of landing and take off were operated by 2 members - "operating crew" on the outbound flight. Then during cruise operating crew changes with "relief crew" in work / rest periods. The comparison in sleepiness ratings was made between those two crews on both outbound and return flights.

Crew sleepiness on the outbound flight remained below a KSS of 5 with both crews showing the same rating. Relief and operating crew were indicating increased sleepiness at TOD and post-flight in comparison to pre-flight sleepiness.

Measured pre-sleep sleepiness on the return flight was significantly higher than on the outbound flight, possibly because it was a night flight and occurred after a layover in a different time-zone, but post-sleep sleepiness did not differ significantly stayed almost the same. Greater sleepiness was measured in the relief crew in comparison to the operating crew at TOC and TOD.

In-flight sleep seemed to greatly reduce sleepiness on both outbound and flights, post-sleep sleepiness on both flights remained equal. Post-sleep sleepiness, if compared to the first rest day before the trip, showed no change during the layover and recovery days.

The study showed that KSS shows expected results of fatigue level elevation at moments of sleep deprivation, accumulated tiredness, night shifts and challenging flight parts. It also revealed that provided guidance helped pilots keep the KSS score under control.

2.1.2 Monitoring pilots' sleepiness during an ultra long range trip

The current study [2] aimed not only to prove the effectiveness of a FRMS for 55 pilots on a westward outbound Johannesburg-New York ULR, but also to assess whether to measure the suitability of chosen data collection measurements. In addition to actigraph and a sleep/duty diary, KSS was used to assess sleepiness during pre-flight, at TOC and TOD, after landing and during flight both outbound and inbound leg.

Subjective sleepiness showed variation in ratings across the flight, the crew reported smaller sleepiness before flight, with ratings increasing during the flight. With longer time awake at TOD, crew members felt sleepier. Fatigue, sleepiness, PVT response speed and total in-flight sleep or total sleep in 24-hrs prior to TOD showed no correlation. On any of the following after the flight days sleepiness stayed on the same level after waking up. Sleepiness was higher on a crucial part of the flight - at TOD, similarly, the inbound flight sleepiness rating was much higher than the outbound flight one.

2.1.3 Crew sleepiness for FRMS implementation

The paper (3) describes an experiment set in hopes of measuring the effectiveness of some proposed safety performance indicators (SPIs) and outlining the most accurate ones to be possibly implemented into FRMS. As a part of an experiment, data were measured in multiple separate experiment sessions during long range (LR) and ULR outbound and inbound flights. Sleepiness was rated pre-flight and at TOD on a 9-point KSS scale among other SPIs. Assessment showed that an inbound flight crew scored lower pre-flight sleepiness in comparison to the outbound flights crew, that could be explained with the fact, that the crew on the inbound flight obtained a large amount of total sleep time (TST) in 24-hrs prior to duty, the flight also took place after a 1 day layover. The amount of obtained sleep proved to have a big influence on the in-flight sleepiness. Evaluating whether the proportion of crew members rating more than 7 on the KSS, which is considered to be a high level of sleepiness, showed no differences between flights, pre-trip or at TOD.

2.1.4 Sleepiness assessment as a fatigue indicator for further comparison to the predicted fatigue values of the System for Aircrew Fatigue Evaluation (SAFE)

The main goal of the study [4] by Air New Zealand was to make a comparison between predictions of a bio-mathematical model SAFE, and measured fatigue in a total of 324 pilots during multiple flights conducted over a 10 years period. Sleepiness was measured on a 9-point KSS as one of fatigue indicators.

Overall predicted by SAFE values were closely correlated with subjective fatigue. Due to the fact, that the comparison and all the presented measurements' results were demonstrated for the general "fatigue" indicator, consisting of SP, KSS, fatigue and drowsiness scales, it is impossible to use present study results for statistical analysis or to draw any significant conclusions on sleepiness.

2.1.5 Sleepiness in short-haul pilots during early and late shifts

Presented study [5] measured fatigue in 40 short-haul pilots, comparing fatigue and sleepiness in pilots with early and late (afternoon and evening) shifts. Fatigue indicators were assessed via a duty and sleep log, actigraphy, SP, NASA and KSS at the end of each flight and each flight duty period (FDP).

Overall fatigue was proven to be higher at duty end if FDP for a pilot started in the afternoon and evening, than in the morning with fatigue levels depending more on time awake at duty start, and not showing correlation with prior sleep period time. Increase of sleepiness in participants showed correlation with an increase of time awake, reaching maximum in late evening hours, with higher ratings for pilots finishing FDPs late at night.

Measured sleepiness data were presented on a graph with no precise reached values given, which made it impossible to include the study into statistical analysis.

2.1.6 Sleep and fatigue on long-haul layovers

Next paper [6] presented results of measurements of six experiments with total of 299 pilots participating, that were set in order to look closely at factors influencing layover sleep on out-and-back trips for pilots. It also assessed the correlation between layover sleep and inbound SPIs, including KSS ratings taken at duty start for the inbound flight.

There was a decrease in KSS scores with TST increasing in 24 hours before the duty. Time of waking up highly affected levels of sleepiness at top of descent, with pilots, whose duty started between 12:00–15:59, showing the lowest KSS ratings and pilots on duty between 00:00–03:59 scoring the highest. Sleepiness was shown to grow across the day similarly to all the previous studies.

2.1.7. An influence of circadian variation on sleepiness

The comparison in effectiveness and accuracy was made between subjective measurements of in-flight sleep, such as KSS, and actigraphic measurements for monitoring fatigue in pilots in the current paper [7]. Sleepiness was measured on KSS for each flight at top of descent on inbound and outbound flights.

Results uncovered variations in sleepiness influenced by sleep duration: with increase in sleep duration sleepiness ratings decreased. Rest break number affected the sleepiness in a way, that pilots who took the first break demonstrated increased sleepiness at TOD compared to those taking second or third break. Correlation was found between sleepiness levels and domicile arrival times, as sleepiness was rated the highest at TOD on flights arriving between 02:00–05:59 or between 6:00–09:59 domicile time than on any other flights. In some way the influence of timing of the flight in regard to the domicile time is resembling one shown in previous studies.

2.1.8 An effect of take offs and landings on sleepiness

The aim of the study [8] was to investigate the influence of number of flight segments, especially take offs and landings, on fatigue. Subjective fatigue assessments were done by each of 24 participating pilots via KSS at the predetermined times during a simulated flight. A comparison was made between a 9-hrs duty day with 5 short segments and a 9-hrs, single-segment duty day. Each flight crew participated in the study for two days with duty starting and finishing always at the same time.

Sleepiness was mostly affected by condition (multi-segment or single-segment duty day) and by test number (1 - 10). Sleepiness turned out to be more prominent in the multi-segment duty day than in the single-segment duty day, moreover, sleepiness was accumulating across the test bouts not affected by the type of a duty day. Prior sleep duration had an obvious great impact on KSS ratings. Inter-individual differences in participants and test numbers were discovered to significantly influence the fatigue as well. The objective measures for fatigue assessment showed moderate correlation with the objective measures of fatigue, as comparison of PVT lapses and KSS ratings revealed.

2.1.9 Sleepiness during transmeridian flight operations

Described study [9] investigated if KSS, among other methods for fatigue measurement, accurately reflects history of sleep and wakefulness and circadian phase. 237 pilots rated their sleepiness on inbound and outbound transmeridian flights pre-flight and at TOD. Data were compared between relief and landing crews and between directions of a flight.

TST in the 24 h prior to duty start was discovered to have a significant effect on KSS rating, with more sleep leading to lower KSS rating. In the same way, in-flight sleep reduced sleepiness in pilots.

Sleepiness was to a great extent affected by the departure time being the lowest before flights departing between 14:00 and 17:59 hours, as for sleepiness at TOD, it was the highest for flights arriving between 06:00 and 09:59 hours, an effect of time of arrival was noticeable. Direction of the flight showed a big influence on fatigue, with westward outbound flights being started with lower sleepiness than westward inbound flights or eastward outbound flights.
KSS ratings increased parallel to hours awake at TOD. Relief crew reported on overall higher sleepiness.

2.1.10 An impact of fatigue on commercial flight operations

The goal of the present research [10] was to investigate if varying levels of fatigue and workload somehow affect performance and physiological responses of pilots and to compare measurements' results to fatigue predictions of Boeing Alertness Model (BAM). 32 pilots rated their sleepiness on a 9-point KSS during each experimental session, data were recorded both under rested and fatigued conditions and long and short haul flights.

With time into the session sleepiness grew, the increase was more rapid for fatigued participants. Predictably, KSS ratings were lower for rested pilots, however, ratings for sleepiness on long and short haul flights were almost similar.

2.1.11 Sleepiness detected by a system of an in-flight automatic fatigue detection

The main focus of the study [11] with 14 participants was to rate the accuracy of a proposed in-flight automatic detection system of fatigue, that works based on electroencephalogram (EEG) measurements, sleepiness here was measured after each real 8-hrs flight via KSS.

Mainly the purpose of KSS rating in the study was to make sure participants are not sleepy at the beginning of the physiological measurements. Surprisingly, mean assessed sleepiness before and after the flight was not much different with the score more than 7 according to KSS, considered as sleepiness, being detected only in one pilot.

Due to this study providing too insufficient results with little emphasis put on sleepiness' measurement, the data in the study will not be used for statistical analysis.

2.1.12 Sleepiness evaluation after a fatigue mitigation strategy application

The study [12] strived to measure the effectiveness of a fatigue battling advisory material (as recommendations on napping and breaks) on a westward ULR flight and return LR flight. 52 pilots, divided into primary, so operating take off and landing, and relief crews, in addition to wearing actigraph and completing a sleep/duty diary, rated their sleepiness via KSS.

This study showed a pattern of sleepiness increasing across the flight, then it drops slightly after an in-flights sleep. Sleepiness increased for all crews after waking on the first post-flight day. The second rest break was the point of time with the highest number of KSS ratings more than 7, additionally, two primary crew members scored more than 7 according to KSS at TOD on the inbound flight.

Relief crew showed greater sleepiness than primary crew at TOC, as well as sleepiness at TOD and after landing on the inbound flight. For the relief crew the inbound flight was met with higher sleepiness at TOC than on the outbound flight, they were also sleepier at TOD and after landing in comparison to TOC or pre-flight scores.

The study concluded, that proposed fatigue mitigation techniques proved to be effective, moreover continuing, that, as recommended, 4 recovery days off duty were enough to recover after an ULR flight.

2.1.13 An effect of an additional layover day on sleepiness

An experiment [13] was conducted with data being collected from 125 pilots on both outward and return flights on a route to Brisbane with a 1 day layover or on a route to Los Angeles with a 1 or 2 day layover in order to investigate the effect of an additional day's layover on reducing sleepiness. KSS was used as a method of measuring subjective sleepiness.

Sleepiness was found to be greater on the return than on the outward flight, regardless of layover duration. Time by schedule interacted with sleepiness greatly on the Brisbane route. Order within flights showed significant effect on variables on both routes, sleepiness increased throughout the flights in sleepiness on both routes. The additional night of layover on Los Angeles route shower zero effect on sleepiness on the return flight.

2.1.14 A comparison of sleep and sleepiness on morning and evening flights

The aim of the investigation [14] was to compare sleepiness and sleep on westward morning and evening flights. Participants were instructed to rate their sleepiness on a 9-point KSS every 2–3 hours from the day before to the day after the experimental flights took place flights. Sleepiness reached the highest score late in the evening crew and after awakening on the morning of the outward-bound flight, being also the biggest at the destination. Interestingly, homeward-bound flight showed no increase or decrease of sleepiness during morning and evening, however in those flights timing had a crucial effect on sleepiness. The day after the homeward-bound flight showed no difference in morning and evening flight sleepiness, however, timing showed to be influential to sleepiness.

Generally, sleepiness increased towards the later part of the flight, was stable during almost all the evening flight and achieved maximum score at bedtime. The direction of flight indicated its importance on sleepiness occurrence in pilots.

2.2 Fatigue Severity Scale (FSS)

In this part of the text are discussed articles dedicated to FSS method. FSS, much like KSS discussed in the previous subchapter, is a 9-item scale focused on assessing the severity of fatigue. Subjects are required to choose an option from 1 – strongly disagree to 7 – strongly agree to each of the 9 statements about fatigue the subject is experiencing. The higher the total score, the more severe fatigue is, a maximum score thus being 63.

The table 2.2 represents a summary and statistically proven results of experiments based on fatigue severity evaluation via FSS.

Table 2.2: Meta-analysis of studies examining Fatigue Severity Scale method

Year	Method used	Number of participants	Conclusion	Result
2018	FSS was used to assess fatigue as a part of a bigger questionnaire offered to pilots in GCC countries. The main aim of the study was to investigate fatigue, sleepiness and depression on commercial flights.	328	224 participants had severe fatigue (1). Correlation was found between severe fatigue and amount of duty hours between midnight and 6:00 in the 7 days before the study (2), daytime sleepiness (3), depression (4) and a risk of OSA (5).	**Student's t tests** (1) FSS ≥ 36 **Chi-squared** (2) 14.6 ± 29.4 or 10.8 ± 13.7 hrs; P = .046 (3) 39.3 vs 23.1%; P = .004 (4) 41.5 vs 20 19.2%; P < .001 (5) 33.5 vs 21 20.2%; P = .014

2018	The study's goal was to measure how does a 24-hrs sleep deprivation affect pilots' fatigue. Fatigue was measured via FSS in a testing period with no sleep opportunity and the second period with 8-hrs sleeping opportunity.	7	Fatigue grew throughout the testing, changing in 0-hrs (1), 8-hrs (2) and 16-hrs (3) in the 'no sleep' condition, it reached maximum at 24-hrs (4) for a 'no sleep' group. An interaction was found for fatigue and TMD (5). Compared to 0-hrs, for 'no sleep' conditions fatigue scores were high (6). Sleepiness was higher between the 'sleep' and 'no sleep' conditions at the 24-hrs point (7).	**ANOVA** (1) $F_{3,18} = 19.819$; P = .009 (2) $F_{3,18} = 19.819$; P = 013 (3) $F_{3,18} = 19.819$; P = .016 (5) $F_{3,18} = 7.895$; P = .001 (6) $F_{3,18} = 4.132$; P = .026 **Bonferroni test** (4) $t_6 = - 4.885$; P = .003 (7) $t_6 = 24.074$; P = .007
2014	FAI test that included a FSS was administered to participants before and after each simulated flight session aiming to assess their fatigue under workload.	21	Comparing ratings of fatigue severity before (1) and after (2) the simulated flight, no significant differences were found.	**T-test** (1) FSS = 3.7261.45; P = .702 (2) FSS = 3.8061.03; P = .702

FSS - Fatigue Severity Scale, GCC - Gulf Cooperation Council, OSA - obstructive sleep apnea, TMD - Total Mood Disturbance, FAI - fatigue assessment instrument

2.2.1 Severe fatigue in commercial pilots in Gulf Cooperation Council (GCC) countries

328 commercial pilots in the GCC countries participated in the current study [15] focused on investigating and measuring fatigue, sleepiness, depression and obstructive sleep apnea (OSA) in pilots. Data on fatigue were recorded when participants were asked to fill in an electronic questionnaire, part of which was a FSS, with a score of ≥ 36 being considered a severe fatigue.

Questionnaire results indicated that the majority of participating pilots, more precisely, 68,3%, scored enough to be considered severely fatigued. More detailed analysis revealed that pilots with ratings of fatigue indicating "severe fatigue" in general had more hours on duty between 24:00 and 06:00 in a week prior to the experiment. A significant correlation was also found between severe fatigue levels, depression, sleepiness and tendency to have OSA.

2.2.2 An effect of a 24-hrs sleep deprivation on pilots' fatigue severity

Fatigue was assessed in this article [16] via FSS among other measurements of mood and pilots' personal attributes in order to study the effect on them of a 24-hrs sleep deprivation. Testing for 7 participants was divided into two periods: with and without an 8-hrs sleep opportunity.

Fatigue grew gradually throughout the testing in no sleep period parallel to the amount hours on duty and reached maximum at a 24-hrs with no sleep point. Fatigue at this point was much higher in a group without an additional sleep opportunity, and higher in this group compared to the 0-hrs into the testing measurements. Subjective fatigue scores interacted with Total Mood Disturbance scores.

2.2.3 Fatigue severity before and after a simulated flight

The current study [17] provides a description of an experiment that aimed to create a standard real-time test to measure both physiological and psychological fatigue in pilots. As a part of that Fatigue Assessment Instrument (FAI) test, FSS was used to self-rate fatigue for 21 participants, each of them rated fatigue before and after a simulated flight.

Post-trial analysis indicated no changes in fatigue severity before and after the simulated flight. The reasons for this lack of change could be the small sample size in the experiment or individual perception of fatigue of the participants.

2.3 Epworth Sleepiness Scale (ESS)

Discussed in table 2.3 below is ESS. ESS is a questionnaire that was developed to assess daytime sleepiness. Each subject self-rates his chances of falling asleep on a 4-point scale in regard to one of the 8 questions, thus the maximum score of 24 being attributed to the highest sleepiness.

Articles describing in detail a usage of ESS are presented in table 2.3 with results of test sessions.

Table 2.3: Meta-analysis of studies examining Epworth Sleepiness questionnaire method

Year	Method used	Number of participants	Conclusion	Result
2014	The study aimed to test an algorithm for an automatic detection of fatigue based on EEG recordings during 15 LR flights.	14	Before the experiment subjects were administered ESS and were allowed to participate only their sleepiness was normal (1)	**ESS points** (1) ESS < 10
2018	Sleepiness was self-rated via ESS by participants as a part of a questionnaire aiming to evaluate sleepiness, fatigue, depression, and OSA in commercial airline pilots in the GCC countries.	328	112 pilots had excessive daytime sleepiness (1). Severe fatigued pilots often showed excessive daytime sleepiness (2), as well as a correlation between a higher percentage of long-haul flights and sleepiness (3) was discovered.	**T-tests** (1) ESS ≥ 10 **Chi-squared** (2) 39.3 vs 23.1%; P = .004 (3) ESS > 10; r = 0.112; P = .043

EEG - electroencephalogram, LR - long range , ESS - Epworth sleepiness Scale, OSA - obstructive sleep apnea

2.3.1 A system of an in-flight automatic fatigue detection

The current study [11] set a goal to rate the accuracy of a proposed in-flight automatic detection system of fatigue, that works based on EEG measurements, on 15 LR flights. Participating 14 subjects were administered ESS before the experiment, later only subjects with normal sleepiness were allowed to participate.

Study provides no statistics for ESS scores and does not directly use ESS for any further analysis, thus this article cannot be later used in any statistical analysis of the researched chosen articles.

2.3.2 Sleepiness in commercial pilots in Gulf Cooperation Council (GCC) countries

Next article [15] described an experiment aimed to investigate and measure fatigue, sleepiness, depression and OSA in GCC countries' pilots. Data from participants were recorded via an electronic questionnaire, part of which was ESS to assess subjective sleepiness.

The correlation was discovered between severe fatigue and excessive sleepiness, with more fatigued pilots being more prone to being too sleepy. Another correlation indicated that higher sleepiness might be connected to the greater amount of long-haul flights. Out of 328 participants 112 scored enough to be qualified as having excessive daytime sleepiness.

2.4 Stanford Sleepiness Scale (SSS)

SSS is a scale for evaluating sleepiness subjectively at specific moments in time. The scale consists of only one task - subjects are asked to choose 1 of 7 statements representing their subjective feeling of sleepiness the best. The statements are rated 1-7 points and, thus, the highest level of sleepiness corresponds to 7 points.

Articles with comprehensive results are listed in table 2.4 below and then described in detail.

Table 2.4: Meta-analysis of studies examining Stanford Sleepiness Scale method

Year	Method used	Number of participants	Conclusion	Result
2007	An experiment was conducted under conditions of a 37-hrs SD in order to investigate the effect of caffeine on sleepiness. Two groups, caffeine and placebo one, rated their sleepiness after each flight in simulator via SSS.	13	Subjective sleepiness increased in total from (1) to (2) throughout 4 simulator trials under the SD. The caffeine (3) and placebo groups (4) scored almost the same on sleepiness.	**Wilcoxon-signed ranks test** (1) $Z = 2.71$; $P = .01$ (2) $Z = 3.0$; $P = .05$ **Kruskall Wallis test** (3) $X^2 = 0.006$; $P = .68$ (4) $X^2 = 0.169$; $P = .94$

2013	Drowsiness was rated via SSS by each participant before each of the tests on various FL in an experiment aimed to assess a correlation between sleepiness and a 32-hrs continuous SD.	4	Subjective feeling of sleepiness via SSS score was significantly increasing along with the SD time. SSS score reached a significant level (1) at 04:00 .	**T-test** (1) $P < .05$
2011	Article describes an investigation into the effects of pre-duty sleep, pre-duty travel time, and workload on fatigue of helicopter pilots. Sleepiness was rated on SSS in each test session across duty and relief days.	24	The lowest sleepiness was always recorded pre-duty (3), and the highest sleepiness was recorded at bedtime for both duty days (1) and days off (2). Time of departure affected sleepiness, with before noon departure resulting in higher level of pilot sleepiness (6). Less TST before the duty resulted in greater sleepiness (4). Reaction time and sleepiness showed a correlation (5), as well as a higher number of flight hours and higher sleepiness at the end of the duty (7). Moreover, the effect of the demanding work on sleepiness halfway duty (8) and at the end of duty (9) was discovered, as well as a correlation between workload and sleepiness levels at the end of duty (10).	**ANOVA** (3) $P < .001$ (1) $F_{4,904} = 148.27$; $P < .001$, (2) $F_{1,268} = 66.43$; $P < .001$ **T-test** (6) $t_{222} = -2.49$; $P < .05$ **Pearson coefficient** (4) $r = -0.27$; $P < .05$ (5) $r = 0.41$; $P < .05$ (7) $r = 0.26$; (8) $r = 0.19$; $P < .05$ (9) $r = 0.30$; $P < .05$ (10) $r = 0.22$; $P < .05$

SD - sleep deprivation, SSS - Stanford Sleepiness Scale, FL - flight level, TST - total sleep time

2.4.1 An effect of caffeine on sleepiness of pilots under a sleep deprivation

The effect of caffeine on performance of pilots was examined [18] under the condition of a 37-hrs sleep deprivation (SD) during 4 simulated flight trials. The study aimed to investigate whether caffeine could be considered an effective countermeasure against fatigue, thus participants were divided into placebo and caffeine intake groups. Data on sleepiness from 13 pilots were recorded 30 minutes after each simulated flight via SSS.

Recordings of sleepiness indicate a gradual increase in SSS scores after each flight under continuous SD with exception of results after second and third trials. However, caffeine intake showed no effect on sleepiness, as subjective sleepiness between two groups remained almost similar.

2.4.2 An effect of a 32-hrs continuous sleep deprivation on sleepiness

The goal of the study [19] was to assess if continuous 32-hrs SD affects pilots' cognitive behavior. Participants were asked to rate their sleepiness on SSS before each simulated flight trial. Flights were simulated on different flights levels (FL).

Subjective sleepiness was gradually increasing along with sleep deprivation, also, similar to sleepiness measured via KSS, it was significantly increasing along with the SD time. Sleepiness reached a score, consistent with significant sleepiness, on at 04:00.

2.4.3 Sleepiness during helicopter operations

The paper [20] studies how are connected alertness and fatigue of pilots and their pre-duty sleep, pre-duty travel time, and on-duty workload. 24 helicopter pilots were asked to rate their sleepiness on SSS from 1 to 7 during test sessions that were built in pilots' normal duty schedule after waking up, pre-duty, halfway-duty, end-duty, and at bedtime.

Sleepiness was found to be always the lowest at pre-duty session and the highest at bedtime both for duty and rest days. Sleepiness measured before and halfway the duty time was almost on a same level. Additionally, results indicated that "morning" and "evening" types on overall scored the same on SSS. Length of pre-duty TST was, as in KSS method experiments, affecting sleepiness on duty.

Further analysis revealed a correlation between reaction time in test sessions and sleepiness, a correlation between workload and sleepiness at the end of duty, and a correlation between number of flights and sleepiness at the end of duty: sleepier pilots had longer reaction time and greater workload at duty or more flights. Higher demand on duty was found to negatively affect sleepiness halfway and at the duty end.

Sleepiness varied between pilots starting their duty at different time, being especially high for those on duty starting before 12:00.

2.5 Rating of Perceived Exertion (RPE)

RPE is a subjective physical activity intensity level evaluation tool, a scale from 6 to 20 designed to give an approximate estimate of the subject's actual heart rate when the chosen rating is multiplied by 10.

The rating was met only in 1 article as a method for fatigue level assessment in pilots. The results of this experiment are presented below in table 2.5.

Table 2.5: Meta-analysis of studies examining Rating of Perceived Exertion method

Year	Method used	Number of participants	Conclusion	Result
2013	The effects of a 32-hrs SD on pilots' cognitive behavior was investigated. Participants were asked to rate their fatigue via RPE before each simulated flight test session on a different flight level.	4	The feeling of fatigue and workload reported by subjects both increased during the 32-hrs SD. RPE score reached a significant level (1) at 04:00.	**T-test** (1) P < .05

SD - sleep deprivation, RPE - Rating of Perceived Exertion

2.5.1 An effect of a 32-hrs continuous sleep deprivation on exertion

The study [19] aimed to investigate whether continuous 32-hrs SD affects pilots' cognitive behavior. Before each test session in a flight simulator with varying FLs data were collected on a subjective fatigue from each pilot via the RPE.

The significant level of fatigue by participants was reached at 04:00 on the 2nd day of an experiment, fatigue, as in experiments using FSS method increased gradually across time on duty and test sessions. Nevertheless, the study does not provide a precise number value of the recorded fatigue level, measurement results are presented with rough values on a graph, thus the study will not be included into the statistical analysis.

2.6 Global Vigor Scale (GVS)

Presented below table 2.6 includes a description of an article that set an experiment with pilot's vigor measured via GVS. GVS consists of 4 VASs regarding answers about alertness, sleepiness, motivational loss and weariness. Each VAS is scored from 0 to 100, standardly higher scores meaning higher levels of vigor.

Table 2.6: Meta-analysis of studies examining Global Vigor Scale

Year	Method used	Number of participants	Conclusion	Result
2011	The study reviewed factors that could possibly affect alertness and vigilance of helicopter pilots. During each test across normal duty and relief days levels of vigor and alertness were assessed via GVS after wake up, pre-duty, halfway-duty, end-duty, and at bedtime.	24	Both on duty days (1) and days off (2), bedtime vigor was lower, with reaching maximum at pre-duty sessions (3). No significant differences were found in vigor levels between pre-duty and halfway-duty sessions. Shorter TST significantly affected lower vigor levels (4). A correlation was found between vigor scores and sleepiness (5). Vigor at bedtime reached higher scores in summer than in winter, both on relief (6) and working days (7).	**ANOVA** (1) $F_{4,904} = 45.61$; $P < .001$ (2) $F_{1,268} = 14.79$; $P < .001$ (3) $P < .001$ **Pearson coefficient** (4) $r = 0.24$; $P < .05$ (6) $r = -0.55$; $P < .05$ **T-test** (6) $t_{132} = 6.13$; $P < .0001$ (7) $t_{210} = 7.35$; $P < .0001$

GVS - Global Vigor Scale, TST - total sleep time

2.6.1 Vigor during helicopter operations

The reviewed paper [20] describes an experiment set for 24 helicopter pilots in order to follow changes in their cognitive behavior under continuous 32-hrs SD. Participants were instructed to rate their levels of vigor and alertness using the GVS, consisting of multiple visual analogue scales concerning alertness, sleepiness, effort, and weariness, during each test sessions after wake up, pre-duty, halfway-duty, end-duty, and at bedtime during normal duty rosters when pilots.

Vigor reached the lowest level always at bedtime, being similar to levels at wake-up sessions, while being at its highest at pre-duty. No difference was found in vigor scores measured on duty or relief day or recorded by morning or evening types. However, the season was found to have an impact on bedtime vigor levels in a sense of bedtime vigor being lower in winter on both day types. Vigor levels stayed the same during pre-duty and halfway-duty tests. Significant correlation was discovered between sleepiness and vigor levels.
The confirmation of pre-duty TST length to greatly affect in-duty levels of vigor could be compared with the same influence of pre-duty TST on sleepiness via KSS and SSS.

2.7 Samn-Perelli fatigue scale (SP)

SP scale is another extremely widely spread self-administered subjective fatigue assessment scale. SP scale is a 7-point for with scores from 1 - fully alert, wide awake to 7 - completely exhausted, unable to function effectively. Scores over 5-6 are attributed to presence of severe fatigue.

Multiple articles covering usage of SP in experiments were collected. Their description of a methodology in articles and results of the experiments set are presented in table 2.7 and below.

Table 2.7: Meta-analysis of studies examining Samn-Pirelly scale method

Year	Method used	Number of participants	Conclusion	Result
2006	In order to establish the influence of prior sleep on fatigue ratings of pilots on international flights and to collect overall data on amount of sleep, subjective fatigue, and pilot's sustained attention an experiment was set with participants being asked to rate their fatigue via the 7-point SP checklist before and after each flight.	19	Flight sector (1) and stage of flight (2) were discovered to have an effect on fatigue. Fatigue was at its lowest after the third sector with the biggest layover opportunity in comparison to all other flight sectors (3).	**ANOVA** (1) $F_{3,24} = 4.95$; $P < .01$ (2) $F_{1,24} = 40.04$; $P < .001$ (3) $SP < 4$; $P < .05$
2015	The effectiveness of a FRMS and the chosen method of collecting data on subjective fatigue- SP checklist was tested. Pilots' fatigue was monitored on ULR flights with data being recorded before and after each sleep episode and at different times in flight.	55	Pilots rated their fatigue as the lowest pre-flight, fatigue was noticed to grow throughout the flight (1). Measurements at TOD showed fatigue increasing parallel to time on duty (2). Inbound flight fatigue reached higher scores in comparison to the outbound flight one (3). Fatigue, sleepiness, PVT response speed, TST in-flight or TST in the 24 hours prior to TOD showed no correlation. If compared to baseline (4), fatigue remained the same (5) on all post-flight days.	**ANOVA** (1) $F_{3, 315} = 89.23$; $P < .001$ (2) $F_{1,56} = 6.58$; $P = .01$ (3) $F_{1, 317} = 20.83$; $P < .001$ (4) $F_{5, 237} = 0.57$; $P = .72$ (5) $F_{5,237} = 1.07$; $P = .38$
2014	SP was one of the proposed SPIs that could possibly be used in future FRMS on LR and ULR flights. SP was used to evaluate pre-flight and TOD fatigue.	133	The proportion of crew members rating 5 on SP on different flight segments was tested to show pre-flight fatigue (1) not being much different to TOD fatigue (2).	**Chi-squared test** (1) $P = 0.103$; $P < .05$ (2) $P = 0.349$; $P < .05$
2014	The study compared predictions of SAFE to the measured fatigue indicator value, consisting of multiple methods, as well as a 7-point SP fatigue scale.	324	Values predicted with SAFE were close to ones derived from pilots' measurements of subjective fatigue (1).	**Pearson coefficient** (1) $r = 0.85$
2014	The study's goal was to compare overall fatigue between pilots starting their shift early or late. For that purpose, a 7- point SP scale was used to rate subjective fatigue at the end of each flight and each FDP.	40	Fatigue at the end of the FDP was proven to be significantly higher for pilots starting their shift in the afternoon or evening hours (1).	**ANOVA** (1) $t_{51} = 3.92$; $P < 0.001$

2018	The relationship between a layover sleep and pre-flight SPIs was examined. Pilots rated their subjective fatigue on SP fatigue scale at duty start for the inbound flight.	299	It was discovered that every additional hour in TST in 24-hrs prior to duty start decreased fatigue ratings (1). Time awake and fatigue ratings at TOD showed correlation, with lowest fatigue (2) measured on shifts starting 12:00–15:59 and the highest (3) on shifts starting 00:00–03:59.	**SP points** (1) by SP = 0,13 (2) SP = 1.96 (3) SP = 2,59
2016	SP crew status check was used in a study aiming to compare effectiveness of subjective measurements of an in-flight sleep and actigraphy. Fatigue was self-rated by each pilot at TOD on an outbound and inbound flight.	298	Plots taking the first break reached a higher score at TOD than pilots taking the second break (1) or the third break (2). Those taking the second break were less fatigued at TOD than those taking the third break (3). Further effects of arrival time (4) and sleep duration (5) were found.	**ANOVA** (1) $t_{356} = 6.92$; $P < .0001$ (2) $t_{411} = 3.99$; $P = .0002$ (3) $t_{370} = 22.82$; $P = .0051$ (4) $F_{16.97}= 4, 423$; $P < .0001$ (5) $F_{41.15}= 1, 551$; $P < .0001$
2016	The study took place to assess whether there is a connection between a number of flight segments: landings and take offs, and subjective fatigue. Each pilot self-rated their fatigue via SP at particular times.	24	SP subjective fatigue was mostly affected by the time of the duty day (1) and a number of a test session (2). Previous TST duration revealed to have a huge influence on SP ratings (3), each additional hour of sleep reducing fatigue (4). SP subjective fatigue was affected not only by a test session number, but by subjects' individuality (5). Subjective fatigue and a number of PVT lapse showed a correlation (6).	**ANOVA** (1) $F_{1,418} = 20.5$; $P < .001$ (2) $F_{9,418} = 9.3$; $P < .001$ (3) $F_{1,417} = 15.4$; $P < 001$ **SP points** (4) by SP = 0.5 ± 0.1 **Pearson** (5) $r = 0.63$; $P < .001$ (6) $r = 0.26$; $P < .001$
2014	The article is dedicated to examining the reliability of multiple subjective and objective fatigue assessment tools, among which is a SP scale. During transmeridian flight operations pilots completed SP ratings, that were used for cross and in flight comparisons.	237	Relief crew were found to be more fatigued (1) than the landing crew. Each 1 hour increase in TST in the 24-hrs prior to duty start, SP rating decreased (2). Pre-flight SP ratings of fatigue were lower for flights departing at 14:00 - 17:59 (3) or at 18:00 - 21:59 (4) than those at 22:00 - 01:59 (5) or 02:00–05:59 hours (6). Flights arriving at 06:00 - 09:59 (7) showed higher SP ratings than flights arriving at 10:00 - 13:59 (8). Before westward outbound lights, pilots rated themselves as less fatigued than before westward inbound flights (9) or eastward outbound flights (10). SP fatigue at TOD gradually increased throughout the time awake each hour (11).	**SP points** (1) by SP = 0.3 (2) by SP = 0.12 (3) SP = 1.9 (4) SP = 2.1 (5) SP = 2.5; $P = .0003$ (6) SP = 2.7; $P = .0004$ (7) SP = 3.4 (8) SP = 2.8; $P = .0159$ (9) by SP = 0.4; $P = .0099$ (10) by SP = 0.4; $P = .0043$ (11) by SP = 0.1
2015	The influence of fatigue and workload on pilots' performance was established. Predictions of the BAM tool and in-test measurements of fatigue via a 7-point SP scale were compared.	32	The data from this study indicate that the KSS and SP ratings are strongly related.	**NA**

2014	Fatigue battling guidance material was offered to participating pilots, the effectiveness of proposed guidelines was evaluated on ULR and LR flights. Fatigue was rated before and after each sleep period on the SP Status Check among other measurements taken during the experiment.	52	Compared to baseline pilots scored higher on SP on a first day after the flight (1). Fatigue ratings show a pattern of gradually increasing fatigue across the flight, with slight decreases in fatigue following each in-flight sleep period. The relationship between fatigue and: time (2), crew position (3), flight leg (4) and the interactions of these factors (time **x** position(5), time **x** leg (6), leg **x** position (7) ,time **x** position **x** leg (8)) was analyzed.	**ANOVA** (1) $F_{5,\,290} = 3.62$; $P = .003$ (2) $F_{3,\,285} = 65.33$; $P < .001$ (3) $F_{1,\,45} = 11.61$; $P = .001$ (4) $F_{1,\,292} = 3.80$; $P = .052$ (5) $F_{3,\,285} = 8.35$; $P < .001$ (6) $F_{3,\,284} = 1.20$; $P =.031$ (7) $F_{1,\,292} = 0.77$; $P = .382$ (8) $F_{3,\,284} = 5.19$; $P = .002$
2012	The aim of this study was to examine the impact of layover length on sleep, subjective fatigue levels, and capacity to sustain attention of long-haul pilots. Participants rated their subjective fatigue level in their duty diary using a 7-point SP Fatigue Checklist	19	Layover length greatly affected levels of fatigue (1), with longer layover resulting info the lower fatigue. Stage of trip also showed an effect on subjective fatigue (2), end of the flight was usually connected to higher, than at the start, levels of fatigue. Layover duration and stage of trip did not interact in their influence on SP ratings (3).	**Mixed-model regression** (1) $F_{1,123} = 11.3$; $P = .001$ (2) $F_{1,22} = 17.5$; $P = .001$ (3) $F_{9,25} = 3.8$; $P = .55$
2010	The study investigated if an additional layover day could be beneficial for subjective fatigue reduction on Brisbane and Los Angeles route. SP scale was used as a mean of assessing a subjective fatigue.	125	Order within flight increased fatigue throughout the Brisbane (1) and Los Angeles (2) flights. An additional night of layover on Los Angeles route resulted in no effect in fatigue, however, sleepiness increased between the first leg and the second without a layover for both routes (3). A relationship was discovered between fatigue and schedule (4) on the Brisbane route.	**ANOVA** (1) $F_{1,134} = 258.7$; $P < .001$ (2) $F_{1,137} = 274.8$; $P < .001$ (3) $F_{1,64} = 12.8$; $P < .001$ (4) $F_{4,116} = 4.22$; $P < .01$
2008	The goal of a study was to receive accurate information of a relationship between fatigue, duty length and time of day. Pilots on multiple flights filled in a questionnaire with a SP scale at TOD prior to the end of each flight.	3181	Was showed a correlation between VAS fatigue and SP fatigue (1).	**Pearson coefficient** (1) $r = 0.89$

SP - Samn-Perelli, FRMS - Fatigue Risk Management System, ULR - ultra long range, TOD - top of descent, PVT - Psychomotor Vigilance Task, TST- total sleep time, SPI - safety performance indicators, LR - long range, SAFE - System for Aircrew Fatigue Evaluation, FDP - flight duty period, BAM - Boeing Alertness Model

2.7.1 Sleep and fatigue on international flights

The current study [21] had two goals: to collect data on subjective fatigue, sustained attention and amount of sleep of 19 pilots before and after international flights, and to investigate if and how precisely prior to the duty sleep influences subjective fatigue levels. Data were collected on return Australia - Europe flights on four sectors. Pilots were asked to self-rate their fatigue via SP checklist before and after each flight.

Fatigue was discovered to be different in various flight sectors and also stages of flight, being, for example, significantly lower in the beginning of a flight compared to the end of it. The lowest fatigue score was detected in a Europe - Asia sector, possibly due to pilots having a bigger layover in London, than on any other sectors, an opportunity for a longer sleep.

2.7.2 Monitoring pilots' fatigue during an ultra long range trip

Described experiment with 55 pilots [2] was set to test the effectiveness of a FRMS on a westward outbound Johannesburg-New York ULR and tried to measure its chosen method for assessing subjective fatigue – SP checklist was reliable. Additionally, actigraph, sleep/duty diary and KSS were used for data collection. SP ratings were recorded pre-flight, at TOC and TOD, after landing and during flight both outbound and inbound leg.

Fatigue, sleepiness, PVT response speed and total in-flight sleep or total sleep in 24-hrs prior to TOD showed no correlation. Fatigue was discovered to change dramatically across the flight, predictably reaching the lowest scores pre-flight. Pilots scored as more fatigued on the inbound flight than the outbound flight. The time on duty directly affected levels of fatigue at TOD, likewise to sleepiness via KSS. On post-flight days fatigue remained on a same level.

2.7.3 Crew fatigue assessment for FRMS implementation

The study [3] aimed to assess the effectiveness of SP and other SPIs to choose the most reliable ones for a possible future implementation of those SPIs into a FRMS. A total of 133 pilots in a four-person flight deck crews participated in collecting the data on fatigue via SP during multiple separate experiment sessions during LR and ULR outbound and inbound flights.

Main focus was on the proportion of pilots who rated their fatigue more than 5 – severe fatigue and their proportion in different flight segments. The analysis discovered that their proportion in fatigue measurements of pre-flight and TOD fatigue was almost the same.

2.7.4 Fatigue assessment as a fatigue indicator for further comparison to the predicted fatigue values of the System for Aircrew Fatigue Evaluation (SAFE)

Air New Zealand created a study [4] dedicated to testing a proposed SAFE model and to make a comparison between predicted values and values derived from subjective ratings of 324 pilots during multiple flights conducted over a 10-year period. Fatigue was measured on a 7-point SP as one of fatigue indicators.

The analysis showed predicted by SAFE values to be consistent with data on subjective fatigue collected from participants. Even though the study uses SP method of evaluating subjective fatigue, it is not entirely possible to use recorded results for future statistical analysis, as in the experiment a general “fatigue” indicator, that is counted on of SP, KSS, fatigue and drowsiness scales.

2.7.5 Subjective sleepiness in short-haul pilots during early and late shifts

The reviewed study [5] aimed to investigate how does the awake time affect fatigue in 40 short-haul pilots. A comparison was drawn between fatigue in pilots with early and late (afternoon and evening) shifts. In addition to a duty and sleep log, actigraphy, NASA and KSS methods, SP method was used to assess subjective fatigue with data collected at the end of each flight and each FDP.

If looking at an overall fatigue that includes all the indicators of it as both subjective sleepiness and fatigue, a connection was found between time of a FDP beginning and levels of fatigue that pilots reported at duty end, with higher fatigue for those beginning shifts in the afternoon or evening. Fatigue measured at duty end also showed no correlation with prior sleep period time.

Fatigue at the end of FDP was discovered to be much higher, reaching moderate and even severe level, for those pilots, who had a late finishing shift, in comparison to pilots having an early starting one. Interestingly, fatigue in pilots on evening shifts was higher despite them having a shorter duty period than those, who started shift in the morning. Critical values for fatigue were reached around 22:00 hours.

Unfortunately, most of the data were presented in rough values on a histogram, which made it impossible, to use the results of this particular study in future statistical analysis.

2.7.6 Sleep and fatigue on long-haul layovers

The present study [6] combined data from six studies to look into the factors affecting pilots’ layover sleep on inbound and outbound flights and the relationship between layover sleep and inbound SPIs. 299 participating pilots rated their fatigue via SP at duty start for the inbound flight.

As in multiple previous studies increase in TST in the 24-hrs prior to duty start was found to have a direct effect on fatigue at TOD. Moreover, time awake and fatigue at TOD showed a correlation: here the lowest fatigue rating was recorded on shifts starting 12:00–15:59 and the highest - on shifts starting 00:00–03:59.

2.7.7. An influence of circadian rhythms on fatigue and its subjective measurements

An experiment aiming at investigating if subjective measurements of in-flight sleep and actigraphic measurements could be equally reliable took place in a next study [7]. Pilots were asked to rate their fatigue subjectively via SP at TOD for a LR outbound and then inbound flight.

Data were obtained from 298 pilots. A noticeable connection was discovered between fatigue at TOD and time of arrival of the flight: arrivals between 02:00–05:59 and 06:00–09:59 showed especially high subjective fatigue ratings. Increase in TST decreased fatigue ratings. The relationship between a number of break taken by a pilot and his level of fatigue at TOD was also revealed, subjects taking the first break had higher levels of fatigue at TOD than subjects taking the second or the third one, results indicated, that the duration of time awake at TOD was affecting subjective fatigue greatly.

2.7.8 A fatiguing effect of take offs and landings

The current article [8] describes an experiment with a goal of assessing the influence of number of flight segments on fatigue, particular attention was given to effects of multiple take offs and landings. Each of 24 participating pilots self-rated their fatigue via SP at specific times during a simulated flight session. Two simulated duty days were compared: a 9-hrs duty day with 5 short segments and a 9-hrs, single-segment duty day, participants were taking part in the experiment two days in a row with duty starting and finishing always at the same time.

A type of the flight, multi-segment or single-segment and a number of test bout, 1 – 10, were discovered to have a slight effect on fatigue, thus multiple take offs and landings did have a fatiguing effect.

Prior TST showed a great influence on subjective ratings of fatigue, an additional hour of sleep led to reduction of subjective fatigue. SP subjective rating were somehow influenced by a number of PVT lapse, they also were influenced by characteristics of each participant.

2.7.9 Fatigue during transmeridian flight operations

The study [9] investigated if some methods for fatigue measurement, including SP scale, are accurate when reflecting a history of sleep and wakefulness and circadian phase. Pilots were asked to rate their pre-flight and at TOD fatigue via SP scale on both inbound and outbound transmeridian flights. A comparison was made between measurement results of relief and landing crew, and between directions and timings of flights.

Surprisingly, relief crew rated themselves as more fatigued. The reduction of sleepiness was noticed with increase of prior 24-hrs TST. In general, at TOD fatigue was increasing with an increase of time awake. Moreover, pre-flight fatigue was lower before westward outbound lights, than before westward inbound flights or eastward outbound flights, direction of a flight was also found to have an effect on fatigue.

Analysis of measured data from 237 participants indicated, that there was a relationship between flight departure time and SP ratings, the lowest ratings being recorded for flights departing between 22:00 and 01:59 or between 18:00 and 21:59, compared to other flights. Time of arrival influenced fatigue levels at TOD as well, fatigue reached the highest among flights score for a flight arriving between 06:00 and 09:59.

2.7.10 An impact of fatigue on commercial flight operations

The purpose of this research [10] was to compare predictions of BAM tool and measured levels of fatigue, it also aimed to assess the effects of fatigue and varying workload levels on pilot's. A 7-point SP scale was used in each testing session to determine 32 pilots' subjective fatigue.

The study results were not suitable to be included into the future statistical analysis, as no measured values were included into the paper and only a correlation between KSS and SP scale ratings was noted.

2.7.11 Fatigue and sleepiness evaluation after a fatigue mitigation strategy application

Fatigue mitigation strategy was offered to 52 pilots on ULR and LR westward and return flights. The current article [12] describes an experiment set to measure the effectiveness of a proposed strategy by collecting sleepiness and fatigue measurements' samples. As a part of a study, pilots were required to self-rate their fatigue on a SP checklist before and after each sleep period. A comparison was made between landing and relief crews' ratings.

Fatigue showed an increasing pattern throughout the flight; however, each in-flight sleep period decreases it slightly. The first post-flight day brought in higher fatigue results than at the baseline.

The outbound flight results for relief crew indicated a higher level of fatigue at TOC, TOD, and after landing compared to pre-flight, fatigue was also higher after landing than at TOC. Same crew was more fatigued at TOD and after landing compared to pre-flight or TOC on the inbound flight. As for the landing crew, they showed greater level of fatigue before the inbound flight than before outbound at TOD. On the outbound flight their SP ratings were greater both after landing and at TOD, lower at TOC and pre-flight. During an inbound flight landing crew reached lower levels of fatigue than the relief crew at both TOC and after landing, during an outbound flight lower levels of fatigue at TOC and TOD were also detected among landing crew members.

The rating of SP > 5, that is considered a significant fatigue, occurred most often during the outbound flight before the second rest break, same timing for occurrence of severe fatigue was noticed during the inbound flight. In general, relief crew reported severe fatigue after the landing more often, than the landing crew.

Due to the study design, when measured data are presented not in precise values, but via histograms and box plot diagrams with rough value, it is impossible to use the study results in statistical analysis.

2.7.12 An effect of layover duration on subjective fatigue

The current study [22] with 19 participants focuses on investigation what exactly is a relationship between the duration on layover, attentiveness and subjective fatigue in pilots during LR flights. A SP rating was made at the start and the end of each flight sector with participants flying on international routes with short or long layover and at least 4 days off at home prior to the duty.

A strong effect of the layover duration on subjective fatigue levels was discovered, pilots with longer layovers reported lower fatigue. Greater scores in SP checklist were achieved at the end of flights compared to their start and to any type of a day off.

2.7.13 An effect of an additional layover day on fatigue

Presented article [13] discusses the outcome of adding an additional layover day on possible reduction of fatigue. 125 participants were either on outward or return flight on a route to either Brisbane with a 1 day layover or to Los Angeles with a 1 or 2 day layover. Subjective fatigue was assessed by means of a SP checklist.

The additional night of layover had zero effect on the return flight on the Los Angeles route. Nevertheless, on both routes without a layover an increase was detected in subjective fatigue. The order demonstrated to have a particular effect on fatigue on both routes, which lead to the increase of fatigue on both routes. Moreover, a relationship between schedule and fatigue was found on the Brisbane route.

2.7.14 A relationship between fatigue on duty, duration of duty and the flight timing

The investigation [23] into the effect of flight timing and duration of being on duty on subjective fatigue took place with the main goal being establishing duty time limitations to improve aviation safety. Data were collected from 3181 pilots in a form of a questionnaire, which participants were asked to fill in at TOD during the finishing sector of the flight. SP scale was among other subjective fatigue assessment methods included into the questionnaire.

Fatigue on flights occurring at 03:00 was not reported to reach severe levels even after 12 hours into the duty, nevertheless, fatigue on flights occurring between 21:00 - 00:00 reached a level, considered to be severe fatigue much faster - only after 6-7 hours on duty. Circadian rhythms might have influenced fatigue, as it reached its peak between 02:00 – 06:00, time considered to be a circadian low. Moreover, fatigue at TOD was affected by a mixture of factors as the number of sectors, time of day, the type and the length of duty. VAS on fatigue and SP scale showed common results.

Data on fatigue are presented as a histogram in the current article, no precise values, suitable for future statistical analysis were provided.

2.8 Visual Analogue Scale (VAS)

VAS is a measurement tool that is used for measurements of various parameters, including fatigue, as in studies described in articles in table 2.8. By VAS are measured characteristics that range across a continuum of values and thus cannot be measured precisely. VAS usually comes in a form of a 100 mm line on which a subject is asked to put a mark responding to his current level of fatigue. Results are then determined as a distance between 0 - no fatigue and subject's mark.

Table 2.8: Meta-analysis of studies examining Visual Analogue Scale method

Year	Method used	Number of participants	Conclusion	Result
2014	Predictions of SAFE on fatigue in pilots were compared to measurements of fatigue via VAS scales.	324	No information on VAS results provided.	**NA**

2010	An experiment was set on two flight routes to test the benefits of an additional layover day on reducing fatigue. A 100-mm VAS was used as one of the methods of measuring fatigue.	125	An order during each flight affected the increase of VAS fatigue on both Brisbane (1) and Los Angeles routes (2). Fatigue on both outbound and inbound flight was showing a pattern of increase across the flight. Both times the fatigue grew across the first leg to the second in case of flight without a layover (3). The return flight results demonstrated a higher level of fatigue (4).	**ANOVA** (1) $F_{1,132} = 49.7$; $P < .001$ (2) $F_{1,137} = 265.1$; $P < .001$ (3) $F_{1,63} = 7.3$; $P < .01$ (4) $F_{1,42} = 6.53$; $P < .05$
2008	The study's aim was to collect information of a relationship between fatigue, duty length and time of day. Pilots on multiple flights filled in a questionnaire with fatigue being rated on a VAS scale at TOD prior to the end of each flight.	3181	Was showed a correlation between VAS fatigue and SP fatigue (1).	**Pearson** (1) $r = 0.89$

SAFE - System for Aircrew Fatigue Evaluation, VAS - Visual Analogue Scale, TOD - top of descent, SP - Samn-Perelli

2.8.1 Fatigue assessment for further comparison to the predicted fatigue values of the System for Aircrew Fatigue Evaluation (SAFE)

Air New Zealand study [4] had a goal to make test the accuracy of a bio-mathematical model SAFE predictions. 324 pilots participated and rated their fatigue on VAS during flights that took place over a 10 years period. 2 100 mm VAS were filled in during each flight.

The study's primary focus was the comparison of SAFE model predictions to measured subjective fatigue, thus no detailed information and no actual measured values on VAS fatigue measurements were provided, being deemed to be able to provide little additional information.

2.8.2 An effect of an additional layover day on fatigue

The study [13] was conducted to test if adding an additional layover day is in some way affecting fatigue of 125 pilots on outward and return flights on two flight routes: to Brisbane with a 1-day layover and to Los Angeles with a 1 or 2 day layover. 100 mm VAS was a chosen method for subjecting fatigue assessment.

The return flight results demonstrated a higher level of fatigue. Fatigue on both outbound and inbound flight was gradually increasing across the flight. For both routes without a layover fatigue was discovered to grow across the first leg to the second in case of flight without a layover. A correlation between order during each flight and the increase of VAS fatigue on both routes was found.

2.8.3 A relationship between fatigue on duty, duration of duty and the flight timing

The main goal of the study [23] was to establish the relationship between fatigue, flight timing and duration of being on duty. Data, collected in a form of filled in questionnaires, that consisted, among other methods for fatigue evaluation, of 100-mm VAS, was collected from 3181 pilots. Questionnaires were filled in at TOD during the finishing sector of the flight.

The results on subjective fatigue measurement with VAS showed a close correlation to SP scale, the results were so similar, the results from VAS were not included into the discussion.

2.9 NASA Task Load Index (NTLx)

According to its name, NTLx was developed by the Human Performance Group at NASA with a goal of subjectively assessing workload perceived by subject. NTLx total rating consists of ratings of further 6 subjective subscales as mental demand, physical demand, temporal demand, performance, effort and frustration with scores from 0 to 20. The overall score is later comprised from ratings of subscales when each of them is given some weight. Outcome of the measurement is a workload index rated from 0 to 100.

The table 2.9 combines results from 5 chosen articles describing experiments with the usage of NTLx.

Table 2.9: Meta-analysis of studies examining NASA Task Load Index method

Year	Method used	Number of participants	Conclusion	Result
2019	A 24-hrs experiment took place with a main goal of investigating the effect of pilot's fatigue on results of multiple performance and psychological tests. Measurements included workload evaluation, using NTLx.	8	The mental demand parameter (1), the physical demand parameter (2), the temporal demand parameter (3), the performance parameter (4), the effort parameter (5) and the frustration parameter (6) all did not reveal any significant change between individual pairs of measurements.	**Friedman test** (1) $X^2 = 11.72$; $P = .11$. (2) $X^2 = 3.61$; $P = .82$ (3) $X^2 = 8.61$; $P = .28$ (4) $X^2 = 4.86$; $P = .67$ (5) $X^2 = 4.86$; $P = .67$ (6) $X^2 = 2.87$; $P = .89$
2014	Fatigue levels of pilots finishing their duty late at night were analyzed and compared with pilots starting their duty early. Subjects were asked to rate their workload via NTLx at the end of each flight and each FDP.	40	Workload showed little influence on fatigue at duty end.	**NA**

2019	The study aimed to investigate if static upright balance index and brain-blood oxygen levels could be used as workload indicators. Subjects filled in the NTLx form to evaluate pilot workload.	15 -1st phase 50 – 2nd phase	The following indicators were measured: mental demand for real (1) and simulated flight (2), physical demand for real (3) and simulated flight (4), temporal demand for real (5) and simulated flight (6), performance for real (7) and simulated flight (8), effort for real (9) and simulated flight (10), frustration for real (11) and simulated flight (12). Workload after real flight and simulated flight tasks was highly correlated (13). Total scores in the experimental group (14) were higher than those in control group (15) after the simulated task.	**NTLx points** (1) 3.42 ± 0.96 (2) 3.71 ± 1.34 (3) 2.03 ± 1.35 (4) 0.90 ± 0.97 (5) 1.25 ± 0.93 (6) 2.24 ± 1.12 (7) 1.09 ± 0.62 (8) 1.22 ± 0.84 (9) 3.02 ± 1.16 (10) 3.03 ± 1.25 (11) 0.71 ± 0.40 (12) 1.32 ± 0.70. **Pearson coefficient** (13) $P < .01$ **t-test** (14) $P < .001$ (15) $P = .002$).
2013	Subjects rated their workload via NTLx after each simulated flight test session in an experiment set to examine the effects of 32 hours of a continuous SD on pilot's cognitive behavior in a simulated flight.	4	Six NTLx parameter reached significant level at 04:00 on the 2nd day.(1)	**T-test** (1) $P < 0.05$
2015	Monitoring of multiple parameters, including workload, took place on ULR flights, to measure the accuracy of those methods. Pilots were asked to rate their workload via NTLx at the each of each flight.	81	Results not included in the study.	**NA**

NTLx - NASA Task Load Index, FDP - flight duty period, SD - sleep deprivation

2.9.1 A relationship between fatigue, workload and the results of psychological and performance testing

In this study [24] 8 measurements of the data (from 21:00 to 18:00) from 8 pilots took place during a flight in a flight simulator, among other methods of fatigue, workload and performance evaluation, NTLx questionnaire was used to self-rate the performance, evaluating mental, temporal and physical demand, performance, effort and frustration.

Results indicated that participants were the least affected by time pressure, while frustration affected them the most.

No significant change between individual pairs of measurements of mental demand parameter was found. Subjective ratings of stress grew along 1-3 test sessions to then decrease and finally reach the maximum value during the 6th measurement. The level of frustration only showed a decrease during the 7th measurement.

The individual pairs of measurements of the physical demand and then temporal demand parameter also showed no significant variation. The temporal demand, however, kept increasing from the 2nd to the 4th measurement to then decrease and increase one more time. The performance parameter showed no great variation as well. The effort parameter demonstrated an increase between the 3rd and the 5th measurements.

2.9.2 Workload in short-haul pilots during early and late shifts

Presented study [5] tried to evaluate the relationship between the duty start and pilot's fatigue. Workload and fatigue measurements of 40 short-haul pilots, that were either on morning or evening duty were compared. Subjects filled in a NTLx questionnaire at the end of each flight and each FDP.

Workload was found to have a minor effect of pilot's fatigue and thus measured values of workload were not included int the study. Since no values were obtained from the article, it will not be used in statistical analysis.

2.9.3 Workload assessment via static upright balance index and brain blood oxygen levels

The current paper [25] aimed to investigate whether it is possible to measure objectively pilot's workload by the means of recording and analyzing data on static upright balance index and brain blood oxygen levels. Subjects were asked to rate their workload before and after the task, pilots were divided into 2 groups: control (resting) and flight simulator one, the NTLx was used to compare workload on a real flight with one on a simulated flight in pilots.

A significant correlation was found between real and simulated flight fatigue, proving, that an experiment set to measure workload on a simulator could be reliable. For the simulated flight task experiments, experimental group workload was higher than the control group results.

2.9.4 An effect of a 32-hrs continuous sleep deprivation on workload

A study was conducted [19] to follow the change in pilots' cognitive behavior throughout performing performance tasks under continuous 32-hrs SD. Participants were asked to rate their workload via NTLx before each test in a flying simulator. A total of 5 flying segments were simulated on different FLs.

NTLx questionnaire provided results on 6 measured parameters: mental, physical and temporal demand, performance, effort and frustration level. The parameters did not increase throughout the tests at the same degree, thought all reached significant level at 04:00 on the

2nd day of the experiment. The greatest change was noticed in mental demand, then effort and performance. The smallest change was noticed in frustration level.

The change of each NTLx parameter was presented on a graph with no values of workload measurements' results given. Due to this obstacle, this study cannot be used for statistical analysis.

2.9.5 Monitoring pilots' workload during an ultra long range trip

The article [2] describes monitoring of pilots' fatigue, sleep, and performance on ULR. The goal was to test some of the subjective fatigue collection methods on reliability to possibly later implement them into FRMS. Crew members were instructed to rate their workload via NTLx at the end of each flight.

The results of NTLx measurements were not included in the study, thus the current study will be excluded from the list of chosen articles to be used in meta-analysis.

Chapter 3

Objective methods for evaluation of fatigue in pilots

Methods for objective evaluation of fatigue are supported by fact-based information, that could be measured. The biggest benefit of an objective fatigue assessment tools is them being

Not dependent on mood and perceived feelings of fatigue, results are less fuzzy in interpretation. Objective fatigue measurements are often based either on monitoring of actual parameters of human's body functioning as heart rate of eye movements or on evaluating the performance of a subject by asking him to repeatedly go through the same task over and over again to detect changes in performance parameters. In this chapter the following objective fatigue detecting tools are reviewed: voice analysis, PVT, physiological signals and eye movements.

Detailed descriptions of those methods and articles chosen for the meta-analysis with clearly established results are given in the following subchapters.

3.1 Voice Analysis

Voice analysis is, essentially a graphical representation of an electronic speech recording of the subject's voice. Each human is believed to have unique characteristics of speech, moreover, thus a potential automatic speech recognition of speech is possible, when phrases, recorded as a baseline are compared to the phrases recording under workload – on duty. There are numerous techniques of voice analysis and their enhances variations, but the most common detection of fatigue through voice is based on measuring deviations in voice, often on particular chosen vocals.

Some articles, covering the topic of an accuracy and potential of speech analysis as a tool for fatigue detection are presented below together with measured values in table 3.1.

Table 3.1: Meta-analysis of studies examining Voice Analysis method

Year	Method used	Number of participants	Conclusion	Result
2006	A voice analysis was proposed as a method for evaluating fatigue, the study measured the accuracy of the method. Speech recordings were compared at test and pre-test with the change of voice being tracked in changes of sounds 'p' and 't'.	6 in the 1st test 2 in the 2nd test	The correlation coefficient between time awake and changes in sounds 'p' (1) and 't' (2) was discovered. A correlation between test sessions at 12-hrs and 78-hrs awake (3) was lower in comparison to tests sessions at 12-hrs and 39-hrs awake (4), thus indicating, that voice does change under the effect of fatigue during sleep deprivation.	**Pearson coefficient** (1) r = -0.89 (2) r = 0.67 (3) r = 0.19 (4) r = 0.82

2010	The study examined effects of sleep deprivation on cognitive performance of pilots. An analysis was performed on voice recordings at the laboratory after waking up and during the flight, on departure, at one stopover and on arrival. Maximal Lyapunov exponent λ, DAL and F_0 were measured.	2	Maximal Lyapunov exponent λ showed an increase for the 2nd recording, which could be caused by fatigue. Comparison of CVR showed that F_0 parameter was lower during the recording on the ground (1) that during the flight recording (2). DAL is bigger during the flight voice recording (3), than during the ground voice recording (4). The parameters were growing under the workload.	**F_0** (1) 117.03 (16.10) (2) 130.61 (15.44). **DAL** (3) 9.32 (σ = 2.64), (4) 5.73 (σ = 1.22)
2013	The reviewed article presents results of an experiment conducted with an aim to prove effectiveness of a speech analysis with a Cepstrum Coefficient Speech Analysis for detecting fatigue in pilots.	195	A positive correlation was found between speech deviation scores and PVT RTs and lapses (1). Speech deviation was not detected for an international crew, however domestic operation crew revealed a decrease from pre-trip (3) to post-trip (4). In regional crew results indicated that speech deviation scores decreased from pre-trip (5) to post-trip (6), same as decrease between pre-trip (7) to post-trip (8) for the junior-level crew results.	**Spearmann test** (1) R^2s > .013, ps < .001 **T-test** (3) 14.30 + 0.03; t_{1331} = 2.81; P < .01 (4) 14.18 + 0.03; t_{1331} = 2.81; P < .01 (5) 14.30 + 0.05; t_{419} = 2.33; P < .05 (6) 14.13 + 0.05; t_{419} = 2.33; P < .05 (6) 14.41 + 0.06; t_{457} = 2.60; P < .01 (7) 14.20 + 0.06; t_{457} = 2.60; P < .01

DAL - digital amplitude length, F_0 - fundamental frequency, CVR - cockpit voice recorder, PVT - Psychomotor Vigilance Task, RT - reaction time

3.2.1 An automatic speech recognition system (ASR) with correlation metric (CM) for detecting states of low vigilance in pilots

The current paper [26] discusses first a reliability of a fatigue-detection via a speech recognition system, that has a big potential to be used for improving safety in aviation. Data from subjects were collected after a night of sleep deprivation, participants were asked to pronounce sentences with words from a special set of words. Later an analysis was performed to study the relationship of changes in pronunciation and fatigue and the accuracy of the proposed system.

In the first part of the experiment, a correlation between changes in time, that participants needed to fall asleep, the sleep onset latency (SOL), and changes in sounds ‘p’ and ‘t’ was discovered. The time awake was a predicting factor for pronunciation of 79%, of sounds ‘p’ and 45% of sounds ‘t’. The circadian peak at 16:00 was discovered to affect fatigue in voice.

Formant frequencies, that are connected to one’s level of alertness were evaluated and 19 of them showed great correlation with measured reaction time. Mel-frequency cepstral coefficients (MFCCs) were used to measure fatigue with analysis performed on sound ‘t’. MFCC vector was found to change with an increase of fatigue, as a correlation between test sessions at 12-hrs and 78-hrs awake was lower in comparison to tests sessions at 12-hrs and 39-hrs awake.

The system with CM usage proved to be an effective tool for fatigue detection, as a receiver operating characteristic (ROC) curve was plotted and the results indicated the ROC curve to have areas of 0.80-0.85 for the two subjects – indicators of a high prediction ability.

3.2.2 Voice analysis to detect fatigue in pilots

The main goal of the the current study [27] was to investigate the relationship between changes in pilot's voice and fatigue. Two experiments were set: voice was collected in a laboratory after waking up to detect sleep inertia in LR flights, later voice recordings were collected from participating pilots on departure, at one stopover and on arrival during short-haul flights. A comparison was made between vowel pronunciation at cockpit and in laboratory.

The digital amplitude length (DAL) parameter, that could potentially indicate drowsiness was detected to jump, increasing, in the beginning of the 2nd recording. During the 2nd experiment an observation was made, that vowels of recording 2 were having higher Maximal Lyapunov exponent λ values, but only for the 2nd pilot. The 1st pilot's speech analysis showed the same result, but more subtle. Results might indicate fatigue, as the recording took place at 08:00 after waking up.

Comparison of cockpit voice recorder (CVR) during the flight and on the ground revealed some variations in fundamental frequency F_0 parameter, as it was lower during the recording on the ground than during the flight recording. In the same part of an experiment DAL is bigger during the flight voice recording, than during the ground voice recording. Increasing workload was forcing those two parameters to grow.

3.2.3 Fatigue detection with a Cepstrum Coefficient Speech Analysis

The paper [28] presents results of a testing of a voice analysis system based on Cepstrum Coefficient modeling. Baseline speech models were created during test sessions with PVT tasks, those were later used for comparison to phrases recorded during actual flights before and after each workday and sleep from 195 pilots.

Significant positive correlations were detected between speech deviation scores and PVT RTs and lapses, as the worse was the PVT performance, the higher the speech scores were.

Speech deviation was not detected in international crew results, however domestic operation crew samples revealed a decrease from pre-trip to post-trip. Speech deviation scores decreased dramatically post-trip in regional crew results, the same decrease was recorded for the junior-level crew results.

The article confirms suitability of the chosen speech analysis method using a Cepstrum Coefficient for detecting fatigue in pilots on duty.

3.2 Psychomotor Vigilance Task (PVT)

Another frequently used method for assessing levels of fatigue objectively is PVT. PVT is a sustained-attention, reaction-timed task that measures the speed with which subjects respond to a visual stimulus – response speed (RS). The classic version of the PVT test is a simple task requiring participants to press the button as soon as a light turns on aimed at sustained attention, number of lapses in attention, response time or speed evaluation. Further versions of PVT will be discussed in the papers shortly described in table 3.2: PalmPilot PVT, OR test, Critical Flicker Frequency (CFF), Reaction time test, Sternberg dual-task test, the Multi-Attribute Task Battery (MATB) and Operation Span Task (OSPAN).

Table 3.2: Meta-analysis of studies examining Psychomotor Vigilance Task methods

Year	Method used	Number of participants	Conclusion	Result
2019	The article presents results of a 24-hrs experiment conducted to examine the relationship between fatigue in pilots and their results of psychological and performance tests. Measurements included performance testing, using the so-called OR-test.	8	Statistically significant differences between distributions in the observed measurements were found for the 1st (1), 2nd (2), 3rd (3) and 4th (4) tasks. The longest response time was found for task 1 (5), for task 2 (6), for task 3 (7) and for task 4 (8). The shortest times being (9), (10), (11), (12) for tasks 1,2,3 and 4 respective.	**Friedmann's test** (1) $X^2 = 33.66$; $P = 1.99\times10^{-5}$ (2) $X^2 = 30.07$; $P = 9.20\times10^{-5}$ (3) $X^2 = 12.95$; $P = 7.33\times10^{-2}$ (4) $X^2 = 67.28$; $P = 5.22\times10^{-12}$ **Reaction time** (5) 3382 ms; $P < .01$ (6) 3583 ms; $P < .01$ (7) 7922 ms; $P < .01$ (8) 9324 ms; $P < .01$ (9) 2673 ms; $P < .01$ (10) 2877 ms; $P < .01$ (11) 6915 ms; $P < .01$ (12) 7088 ms; $P < .01$
2006	The study examined the relationship between international pilots' amount of sleep, fatigue, and sustained attention before and after flights, it was also researched weather prior duty and sleep could be a marker for prediction of subjective fatigue and attention. A Palm Pilot based PVT was completed before and after each flight.	19	A flight sector showed no effect on RS (1) but was revealed a significant that a stage of flight affected pilots' mean RS the most (2). RS was increasing towards the end of flights in comparison to the start of duty, sleep in prior 24-hrs turned out to be significant predictor of sustained attention (3). No interaction was discovered between flight sector and stage of flight (4) on RS.	**ANOVA** (1) $F_{3,21} = 1.06$; $P = .39$ (2) $F_{1,21} = 7.97$; $P < .05$ (3) $F = 0.1$; $P = .005$ (4) $F_{3,21} = 1.53$; $P = .24$

2008	The current study investigated the effect of unscheduled naps on duty on alertness and vigilance. Coastal pilots participated in recording data during a single 28 day period or a two 14 day periods, when they completed a Palm Pilot based PVT every 4 hours, at the end and at the beginning of flight.	17	No significant changes of reaction time were detected during the test sessions.	**NA**
2015	The study's goal was to track fatigue, sleep, and performance of pilots on ULR flights. It was also an aim to evaluate whether chosen methods for fatigue evaluation are accurate in order to implement them into FRMS. Performance was measured using a Palm PVT.	81	PVT RS showed a decline from TOC to TOD for both flights (1) and was slower on the inbound flight than the outbound flight (2). The fastest 10% of responses were recorded on the outbound leg (3). Similarly, RS declined for the fastest (4) and slowest 10% of responses (5), and lapses (6).	**ANOVA** (1) $F_{1, 96} = 11.97$; $P < .001$ (2) $F_{1, 95} = 5.97$; $P = .02$ (3) $F_{1, 89} = 10.60$; $P = .002$ (4) $F_{1,91} = 5.35$; $P = .02$ (5) $F_{1,97} = 12.44$; $P < .001$ (6) $F_{1,96} = 15.79$; $P < .001$
2014	The study offered subjective and objective SPIs for future FRMS and evaluated their accuracy. The performance on duty days was measured with a Palm PVT pre-trip and at TOD for subjects from airlines 1 and 3 and with a longer PVT at TOD only for subjects from airline 2.	133	PVT response speed did not show a lot of variation between flights, as it was indicated from measurements of pre-flight RS (1), TOD RS (2), pre-flight slowest 10% of responses (3) and TOD slowest 10% of responses (4) among different flights.	**ANOVA** (1) $P = .757$ (2) $P = .461$ (3) $P = .048$ (4) $P = .497$
2014	Air New Zealand set an experiment to test the accuracy of SAFE predictions on states of low vigilance in pilots. Data on performance were collected in a series of flights with the Pilot Alertness Test.	324	The predicted by SAFE values were more correlated with subjective fatigue than with reaction time (1).	**Pearson coefficient** (1) $r = 0.57$
2019	Experimentally was investigated whether changes in static upright balance function and brain-blood oxygen parameters could be used as reliable indicators of pilot workload. Reaction time test and Sternberg dual-task test were used to measure performance.	50	No significant differences in the answer and trace scores between the experimental (1) and control group (2) were discovered for the Sternberg dual-test task. Variation of response times, errors and correct answers were more prominent in the experimental (3) than in the control group (4). For the experimental group amount of correct answers dropped after the test on simulator.	**T- test** (1) $P = .142$ (2) $P = .462$ (3) $P < .001$ (4) $P = .002$

2018	The objective of the experiment was to investigate effects of a 24-hrs SD on fatigue, mood and performance of pilots. Performance was measured using the PVT. Reaction time and lapses in attention were assessed.	7	The 24:00 measurements revealed more lapses in attention (1) and slower RTs (2) for the period with no sleep than for a period with a sleep opportunity. For a period with no sleep opportunity during the 24:00 measurement more lapses in attention, compared to the 00:00 (3) or 16:00 (4) measurements, and slower reactions, compared to the 16:00 measurement were recorded. Lapses in attention (5) and reaction time (6) showed an interaction effect.	**ANOVA** (1) $t_6 = 23.029$; P = .023 (2) $t_6 = 22.627$; P = .039 (3) $F_{1.083,18} = 13.001$, P = .049 (4) $F_{1.083,18} = 13.001$; P = .039 (5) $F_{3,18} = 10.299$; P = 0.030 (6) $F_{3,18} = 8.599$; P = .001 (7) $F_{3,18} = 7.511$; P = .002
2016	A study was conducted to investigate the relationship between number of flight segments, especially landings and take offs and pilot's fatigue and performance. A comparison was made between performance on a 9-hrs duty day with many take offs and landings and a 9-hrs day with a single take off and landing, PVT test was used.	24	Flying pilots had slower reaction time than monitoring (1). Prior sleep duration was discovered to be crucial for performance prediction (2). To be insignificant was found an interaction of condition and the number of test session (3), as well as an influence of the test number (4) on performance. Response speed was mainly affected by condition (5), the number of a test session (6), and an interaction condition **x** number of a test session (7). Errors in PVT start time mostly were influenced by the type of the day (8). A modest correlation was revealed between SP subjective fatigue and PVT lapses (9) and a KSS subjective sleepiness and PVT lapses (10).	**ANOVA** (1) $F_{1,436} = 6.8$; P = .010 (2) $F_{1,436} = 20.5$; P < .001 (3) $F_{9,437} = 0.9$; P = .52 (4) $F_{9,437} = 1.1$; P = .38 (5) $F_{1,437} = 15.4$; P < .001 (6) $F_{9,437} = 2.1$; P = .031 (7) $F_{9,437} = 2.6$; P = .007 (8) $F_{1,437} = 12.8$; P < .001 **Pearson coefficient** (9) r = 0.26; P < .001 (10) r = 0.28; P < .001
2014	Presented paper discusses an accuracy and effectiveness of a PVT, along with methods for evaluating fatigue and sleepiness, investigated is also if PVT results could reflect sleep history during transmeridian flight operations. Pilots from multiple airlines were asked to undertake PVT.	237	RS decreased with every additional hour of TST in 24 hours prior to duty (1) and increased with every additional hour of duty prior to TOD (2). RS was slower before westward outbound than before westward inbound (3), eastward outbound (4) or eastward inbound flights (5). RS was higher for flights arriving between 14:00–17:59 (6) than between 02:00–05:59 (7) or 06:00–09:59 (8).	**ANOVA** (1) by RS = 0.004 resp./s (2) by RS = 0.03 resp./s (3) by RS = 0.3 resp./s (4) by RS = 0.3 resp./s; P = .0204 (5) by RS = 0.3 resp./s; P = .0058 (6) RS = 4.1 resp./s (7) RS = 3.9 resp./s; P = .0041 (8) RS = 3.9 resp./s; P = .0050
2015	The purpose of this research was to evaluate effects of varying levels of fatigue and workload on pilot performance and physiological responses, and to determine the ability of the BAM tool to accurately predict the risk of fatigue. Subjects were asked to do PVT during every test session.	32	Rested and fatigued conditions showed different response time results, with fatigued pilots being faster (1). A correlation between change of the RT and rested vs. fatigued conditions was significant. Pilots became increasingly slower with every hour into the experiment (2) during the fatigued sessions. Lapses' results differed between long and short haul flights, being higher for short hauls (3), but not over time and between rested and fatigued conditions.	**RT** (1) 6.7 ms; P = .0255 (2) 2.30 ms; P < .001 **Lapses** (3) 1.09 lapses; P < .01

2014	A CFF test was used to measure performance in the current study that examined the efficiency of a series of test approaches aimed to provide physiological and psychological pilot fatigue assessments.	21	Results indicated that CFF became significantly smaller after the experiment, with gradual change from before the test (5), after the 1st (4) and 2nd (3) time out of simulation cabin and after the trial (4).	**T-test** (1) 39.666 +/- 3.19 Hz (2) 38.116 +/- 3.92 Hz (3) 38.34 +/- 62.49 Hz (4) 37.50 +/- 62.69 Hz
2014	This study investigated whether a fatigue battling strategy offered to participants was effective for pilots on a westward ULR flight and return LR flight. Performance of subjects was assessed with PVT at crucial times.	52	An analysis of results was made to draw correlations between response speed and time (1), crew position (2), and flight leg (3) and interactions of these different factors.	**ANOVA** (1) $F_{2,165} = 14.27$; P < .001 (2) $F_{1,44} = 1.75$; P = .193 (3) $F_{1,166} = 5.64$; P = .019
2012	The study investigated an influence of the layover length on sustained attention of long-haul pilots. Sustained attention was assessed using PVT pre-trip, during the layover and post-trip	19	An effect of layover length on sustained attention (1) was proven, with longer layover resulting into better sustained attention. No correlation between time on duty and sustained attention (2), as well as between the interaction layover length x time on duty and sustained attention (3) was significant.	**Mixed-model regression** (1) $F_{1,123} = 12.5$, P = .001 (2) $F_{10,28} = 2.1$; P = .0 (3) $F_{9,28} = .8$; P = .66
2012	The effects of a 35-hrs SD on performance in a simulated flight were examined. MATB, OSPAN and a 10-minute PVT were used as a method of performance assessment. Tracking RMSE, OSCORE, TSCORE, lapses and other parameters were measured.	10	RMSE changed dramatically across the SD period (1), getting drastically different from baseline on the 5th measurement (2). OSCORE (3) and TSCORE (4) were to a great extent influenced by fatigue. TSCORE reached a statistically significant low level at the 6th session measurement, while both TSCORE (5) and OSCORE (6) reached statistically significant low level at the 7th measurement. The number of PVT lapses grew with time (7), reaching a significantly different from baseline level at the 7th measurement (8), as well as MEDRT (9), MFRRT (10) and MRRT (11). The between-subject correlation was high (12). Low was a between-subject correlation of the RMSE and the change in flight performance (13).	**ANOVA** (1) $F_{5.43, 29.67} = 5.50$; P < .05 (2) P = .017 (3) $F_{4.04, 36.38} = 2.84$; P = .038 (4) $F_{3.43, 30.90} = 2.73$; P = .054 (5) P = .017 (6) P = .028 (7) $F_{2.42, 21.80} = 5.93$; P = .006 (8) P = .034 (9) $F_{2.28, 20.54} = 5.71$; P = .009 (10) $F_{7.51, 67.56} = 31.29$; P < .001 (11) $F_{3.35, 30.11} = 10.44$; P < .001 **Pearson coefficient** (12) r = −0.69 (13) r = 0.22
2011	The influence of a pre-duty sleep, pre-duty travel time, and workload on the alertness of pilots was measured. Helicopter pilots performed PVT sessions after waking up, pre-duty, halfway-duty, end-duty, and at bedtime during normal duty rosters.	24	A correlation was established between sleepiness and RT (1) and between RT and tracking error (2). Higher age correlated with bigger tracking error (3) and longer RT (4). Shorter pre-duty TST tended to result in longer RT (5). Another correlation was found between longer duty and bigger tracking errors and longer RT during pre-duty (7), halfway duty (8), end-duty (9), and at bedtime (10).	**Pearson coefficient** (1) r = 0.41; P < .05 (2) r = 0.61, P < .05 (3) r = 0.37; P < .05 (4) r =0.61; P < .05 (5) r = - 0.27; P < .05 (6) $r_{er} = 0.52$ and $r_{RT} = 0.70$; P < .05 (7) $r_{er} = 0.53$ and $r_{RT} = 0.55$; P < .05 (8) $r_{er} = 0.50$ and $r_{RT} = 0.37$; P < .05 (9) $r_{er} = 0.54$ and $r_{RT} = 0.50$; P < .05

2010	The relationship between an additional layover day and fatigue was studied in an experiment with pilots on two different duties: a two-pilot crew and a three-pilot crew. Pilot Alertness Test was used as a method for measuring objective fatigue and performance.	125	On Brisbane (1) and Los Angeles flights (2) an increase in RT was found, caused by an effect of order, both for outward and return routes. An interaction for leg **x** order was found for RT (3).	**ANOVA** (1) $F_{1,134} = 58.6$; $P < .001$ (2) $F_{1,137} = 89.4$; $P < .001$ (3) $F_{1,124} = 26.1$; $P < .001$
2006	The goal of the article was to examine opportunities to recover before an inbound flight for pilots on short international flights. Subjects undertook a PVT before and after each flight.	19	A layover length **x** time of the testing interaction was found (1). A variation of RS across the flights was found (2), it varied for both the short (3) and longer (4) layover. The difference between the RS at the end of the outbound flight and pre-trip for the longer layover was not drastic (5).	**ANOVA** (1) $F_{3,51} = 4.3$; $P = .009$ (2) $F_{3,51} = 6.9$; $P = .001$ (3) $F_{3,24} = 5.8$; $P = .004$ (4) $F_{3,27} = 5.4$; $P = .005$ (5) $P = .073$

PVT - Psychmotor Vigilance Task, RS - reaction speed, ULR - ultra long range, FRMS - Fatigue Risk Management System, TOC - top of climb, TOD - top of descend, SPI - Safety Performance Indicator, SAFE - System for Aircrew Fatigue Evaluation, SD - sleep deprivation, RT - reaction speed, KSS - Karolinska Sleepiness Scale, TST - total sleep time, BAM - Boeing Alertness Model, CFF - Critical Flicker Frequency, LR - long range, MATB - Multi-Attribute Task Battery, OSPAN - Operation Span Task, RMSE - root mean square error, OSCORE - absolute OSPAN score, TSCORE - total number correct score, MEDRT - median reaction time, MFRRT - mean of the reciprocal of the fastest 10% of reaction times, MRRT - mean reciprocal reaction time

3.2.1 A relationship between fatigue, workload and the results of psychological and performance testing

OR-test, a variation of PVT test was used in this study [24] to assess a workload of subjects who participated in the test that included a simulated flight. The test was performed by 8 participants in approximately 3-hrs intervals and was repeated eight times for each subject.

Results indicated that subjects experienced the highest level of fatigue in early morning hours. The longest response time for task 1 was found in the 4th measurement and the shortest response time was detected at the 7th measurement, possible explanation might be that the subjects were too fatigued at the time of the 4th measurement, as they were awake for more than 20 hours during the morning hours, however the 7th test session took place in the daylight, the statistical difference has been found between the 4th and the 5th measurements.

As for task 2, the longest response time was found in the 5th measurement, the shortest – in the 6th measurement, results are supported by a significant statistical difference between the 5th and the 6th measurement, the 5th measurement took place when morning changed into the day.

As for the task 3, the longest response time was detected again in the 4th measurement, the shortest – in the 8th measurement, it could be proven by a statistically significant difference between the 4th and 7th test sessions' results. Measurements in task 4 indicated the longest response time being found in the 1st measurement, the shortest – for the 8th measurement, possibly it being connected to the fact that subjects got accustomed to the test with every try.

3.2.2 Sustained attention on international flights

The current study [21] strived to investigate how does sustained attention, among other factors as fatigue and amount of sleep, change in 19 pilots before and after international flights, and a later assessment of the influence of prior and duty on sustained attention was made. Data were collected on return Australia - Europe flights on four sectors: Australia - Asia, Asia - Europe, Europe - Asia, and Asia - Australia. Subjects completed a 5-min PVT task before and after each flight.

A stage of flight was shown to affect subject's RT the most, as at the end of the duty response speed much slower than in the end of flights. The flight sector measurements were recorded showed no effect on response speed. For stage of flight and flight sector no interaction effect was discovered.

3.2.3 A relationship between irregular naps and an alertness of pilots

The paper [29] describes an experiment conducted with participation of 17 coastal pilots in order to assess the effectiveness of short irregular naps on duty on fatigue management and to investigate its influence of alertness. A PVT was used to measure performance and vigilance of pilots every 4 hours into the duty, at the beginning and at the end of duty.

The study revealed no change of RT with increase of time on duty, it also detected no effect of circadian rhythms on RT.

The current study published all the RT measurements plotted on a graph, with no precise values presented. Due to this and due to fact, that no change in RT was detected, the study will not be used in statistical analysis.

3.2.4 Monitoring pilots' sleepiness during an ultra long range trip

The goal of analyzed study [2] was to assess an accuracy and effectiveness of chosen methods for detecting pilots' cognitive states and their later implementation into FRMS. Measurements were made for 55 pilots on a Johannesburg-New York ULR flight. Subjective methods for evaluating fatigue and sleepiness were used in addition to Palm PVT that was used to assess performance during pre-flight, at TOC and at TOD, after landing and during a flight on both outbound and inbound leg.

RS was discovered to be, on average, higher on an outbound flight than on the inbound flight. RS was found to decline from TOC to TOD and be slower on the inbound flight compared to the outbound flight for both flight legs. The performance was found to decline across the flight similarly for all measurements of the fastest 10% of responses, the slowest 10% of responses and lapses.

Analysis revealed no connection between PVT results and fatigue, sleepiness, total in-flight sleep or total sleep in the 24-hrs before to TOD.

3.2.5 A crew performance evaluation method for FRMS implementation

The study [3] offered various SPIs to be used for subjective and objective evaluation of pilots' sleepiness, sleep, fatigue and performance and measured their effectiveness to possibly use then in FRMS. Data were recorded from 133 landing pilots during multiple test sessions during LR and ULR outbound and inbound flights. Palm PVT was used to evaluate the performance of pilots from airlines 1 and 3 pre-trip, at TOD and at TOD only for subjects from airline 2.

Variations in PVT response speed were not as significant as variations of total sleep in 24-hrs before duty or TOD, or time awake at duty start or TOD. PVT performance pre-flight and at TOD was found to be affected by departure time and the flight duration, RS was found to be faster for pilots with more sleep in the last 24-hrs and less time awake.

The study found no statistically significant variations of RS using PVT, thus it will not be used in statistical analysis.

3.2.6 Performance assessment as a fatigue indicator for further comparison to the predicted fatigue values of the System for Aircrew Fatigue Evaluation (SAFE)

The study [4] by Air New Zealand aimed to test the effectiveness of SAFE bio-mathematical predictions of fatigue in pilots and its actual measurements recorded for 324 pilots during multiple flights conducted over a 10-year period. Performance was assessed using a Pilot Alertness Test developed by Air New Zealand, which is a sustained choice reaction time task.

Predictions of SAFE tended to underestimate reaction time towards the start of a duty period and to overestimate it at the end and be more correlated with subjective fatigue than with RT.

The study shows a possibility of PVT tasks to be used not only to assess pilots' performance, but to also validate the accuracy of safety models, however, it provided little information about recorded reaction time values and their change throughout the flight, thus the study will not be used in statistical analysis.

3.2.7 Workload and performance assessment via static upright balance index and brain blood oxygen levels

The current paper [25] set an experiment with an objective to test if it is possible to measure and predict pilot's workload and performance by means of following changes in static upright balance index and brain blood oxygen levels. Sternberg dual-task test and RT tests were used to objectively measure subjects' performance. Participants were divided into 2 groups: control (resting) and flight simulator one, later results of performance tasks were compared between those two groups.

The experimental and control groups showed the same answer and trace scores before and after the simulator test results of the Sternberg dual-task test. As for the RT test, for the experimental group there was a noticeable difference in performance before and after simulated tasks, the amount of errors was bigger and the amount of correct answers was smaller, compared to the control group.

3.2.8 The effect of a 24-hrs sleep deprivation on pilots‘ performance

The current article [16] assessed various measurements of mood, fatigue and pilots' performance to establish the effect of 24-hrs continuous SD on those parameters. 7 participants were tested in two periods: with and without a 8-hrs sleep opportunity with testing period consisting of four testing sessions performed at 00:00, 08:00, 16:00 24:00. Objective fatigue in a sense of performance was determined using the PVT.

Compared to the period with sleep opportunity, the no sleep opportunity period showed a lot more lapses in attention and slower reactions at 24:00 measurement. The 24:00 measurement in general revealed a higher number of lapses in attention than 00:00 and 16:00 measurement sessions in period with no sleep opportunity. Moreover, the 24:00 measurement was connected to the significantly slower reactions than a 16:00 measurement for a period with no sleep. A major interaction effect was discovered for lapses in attention and RTs.

3.2.9 A fatiguing effect of take offs and landings

The study [8] presents results of an experiment that took place during two simulated 9-hrs duty days: one with a single take off and landing, and another one with multiple take off and landings. The goal was to investigate how does a number of flight segments affect the pilots' performance, among other factors. 24 participating were administered a PVT task 10 times on each simulator day.

Differences in PVT lapses results appeared to be increasing with a number of a test session for both types of days. RT was mostly affected by condition, by a number of a test session and by an interaction of condition **x** test session number, errors in a start time of PVT sessions were mostly affected by condition, the greater number of mistakes was made in a multi-segment duty day. Sessions' results on a multi-segment day revealed greater fatigue compared to the results from a single-segment day.

Monitoring pilots showed longer RT than flying pilots. Sleep prior to the duty was discovered to greatly influence performance on duty. Results indicated, that results of subjective and objective measurements of fatigue were moderately correlated.

3.2.10 Performance during transmeridian flight operations

The reviewed study [9] investigated if PVT and methods of assessing subjective fatigue could reflect sleep history and circadian phase. A total of 237 pilots of 4 airlines were asked to take part in PVT to measure their performance. Performance was assessed for relief and landing crews on inbound and outbound transmeridian flights pre-flight and at TOD, except for Delta Air relief crews, who were not doing PVT at TOD and Singapore Airlines pilots, who did not do a pre-flight PVT.

Measured PVT RS showed an increase along to increasing time on duty. Sleep in 24-hrs prior to duty was shown to largely affect RS, increasing of prior sleep resulted in shortening or a RS, however no correlation between TST on duty and RS was found.

RS pre-flight was significantly influenced by a domicile departure time, as well as the arrival time: RS was significantly higher for flights arriving at 14:00–17:59 hours than for flights arriving at any other time. Moreover, a direction and type of a flight affected RS too, as RS pre-flight was slower before westward inbound, eastward outbound or eastward inbound flights than before westward outbound flight.

3.2.11 An impact of fatigue on performance of pilots during commercial flight operations

The goal of the present research [10] was to investigate if varying levels of fatigue and workload have an impact on performance and physiological responses of pilots, moreover, it aimed to compare experiment measurements' results to fatigue predictions of BAM. Participating 32 pilots were instructed to do a PVT with mean RTs and a number of lapses measured during each experimental session, data were recorded both under rested and fatigued conditions and long and short haul flights.

PVT lapses were discovered to be drastically differ between short and long-haul flights, being longer for long-haul flights. Nevertheless, lapses did not seem to be affected nor by time nor by the level of fatigue. There was a big difference between results of RT of rested and fatigued pilots, surprisingly, fatigued pilots had faster response. The change of RT and accumulation of fatigue were revealed to be closely correlated. An increase in duration of RT was detected for fatigued pilots, as pilots were getting slower and slower each hour into the experiment. Not to be significant was found the change of length of RT with time on duty for both long and short haul flights.

3.2.12 Performance assessment before and after a simulated flight

The article [17] describes an experiment that took place for 21 pilots in order to prepare a real-time test to measure physiological and psychological fatigue in pilots. As a part of that FAI test, CFF test was administered to pilots 3 times, including before and after the simulated flight to objectively evaluate their performance.

The results of CFF indicated that the overall CFF became slower with time, which was statistically proven by a comparison of before and after the flight measurements. Pilots became increasingly more fatigued after the simulation experiment.

3.2.13 Performance evaluation after an application of a fatigue mitigation strategy

The study [12] examined the effect of fatigue mitigation strategy offered to pilots on ULR outbound and return LR flights. The mitigation strategy recommended, among other measures, a schedule of napping and breaks. 52 pilots have been divided into primary, operating take off and landing, and relief crews, and instructed to participate in PVT to measure their performance.

RS was recorded to be faster on the outbound flight than on the inbound. Only the primary crew on the outbound flight results did not display a slowing down of the RS across the flight, pre-flight RS was faster than the RS at TOC and TOD, while, when comparing the RS at TOC and TOD, at TOC it was faster. Mean RS was similar for both groups of participants.

3.2.14 An effect of layover duration on sustained attention

The experiment [22] with 19 participants was set to assess the influence of a layover duration on sustained attention of pilots, as well as on fatigue during LR. Sustained attention was measured on a variation of PVT at a particular point during the flight.

Pilots with a short layover were revealed to have higher levels of sustained attention disruption, than ones with a long layover, thus a cruciality of a layover length was proven by a PVT. Time on duty or its interaction with a layover length were not found to have any correlation with sustained attention.

3.2.15 Effects of sleep deprivation on cognitive performance

The study [30] focused on an evaluation of the effect of a 35-hrs continuous SD on pilots‘ performance. 10 pilots participated in an experiment that included a 3-hrs flight in a flight simulator and three performance evaluating tests: MATB, PVT, and OSPAN. Root mean square error (RMSE), an absolute OSPAN score (OSCORE), a total number correct score (TSCORE), lapses, a mean of the reciprocal of the fastest 10% of reaction times (MFRRT), a mean reciprocal reaction time (MRRT) and a median reaction time (MEDRT) were measured.

OSCORE and TSCORE showed signs of being strongly affected by accumulating fatigue, scores getting lower and lower with time into SD, as well as PVT. A dramatic change, compared to baseline, was detected in TSCORE starting from the 04:00 6th measurement, as well as in TSCORE and in OSCORE starting from 07:00 7th measurement. RMSE variated dramatically over the whole SD period, constantly increasing, while being different from the baseline measurement already at 01:00 5th measurement. No significant correlation between the change of MATB RMSE over the experiment and the change of in-flight performance of subjects was discovered. However, the correlation was high for the change in OSCORE with the change in flight period RMSE between subjects. The number of lapses increased parallel to time into the test and growing fatigue, similar to MEDRT, MFRRT and MRRT, reaching a significantly different from baseline value at 07:00 7th measurement. Here the between-subject correlation was high.

3.2.16 Performance of pilots during helicopter operations

The article [20] describes an experiment set with an aim to assess changes in performance, among other parameters, under a continuous 32-hrs SD. 24 participants were instructed to undertake a PVT test in each testing session, for each duty day a maximum amount of test sessions being five, to assess their RT.

Pre-duty TST showed a significant effect on performance, the shorter TST being connected to longer RT, moreover, longer time on duty resulted in more tracking errors in performance and also in longer RT during pre-, halfway, and end-duty, and at bedtime. Halfway-duty measurements that had the biggest tracking errors and also longest RT were observed at halfway-duty. Morning and evening duties showed approximately the same level of performance. As for personal pilots' characteristics, elder pilots were reported to have a bigger tracking error and a longer RT. A correlation between RT and tracking error, as well as between sleepiness and RT was discovered.

3.2.17 An effect of an additional layover day on performance of pilots

The study [13] tries to evaluate an effectiveness of an additional layover day on reduction of fatigue. A total number of pilots of 125 participated in the study while being on an outward or a return flight on a route to Brisbane with a 1-day layover or to Los Angeles with a 1 or 2 day layover. Performance was assessed via Pilot Alertness Test 3 times on flights to and from Brisbane and 4 times on flights to and from Los Angeles.

On both routes an effect of order within flight was discovered to affect the performance, following the reveal that a RT was getting longer on both destinations' flights, the results were equal for the inbound and outbound flights.

3.2.18 A profitability of short layovers on recovery of pilots

The experiment described in the article [31] sets a goal of examining if short layovers are beneficial to some extent in helping international pilots to recover. In order to answer that question, 19 pilots were asked to participate in the study. Data from both the captain and the first officer operating were collected on the return flight from Australia to Los Angeles after pilots had either short of long layover.

Great timing of the measurement **x** layover length interaction was found. A RS was changing for each type of flight after both long and short layovers, indicating, however, no statistically significant difference between RS for a longer layover, measured at the end of the outbound flight, and pre-trip.

3.3 Physiological signals

Physiological signals could be collected via measurements produced by physiological processes in human body like heart via electrocardiography (ECG), respiratory rate (RespR), skin resistance and temperature, EEG, skin temperature and many others. The signals are known to reflect changes in fatigue in individual's surrounding. Articles chosen for the meta-analysis, briefly described in table 3.3, and then analyzed in detail are just examples of some opportunities for using changes in physiological signals as indicators of fatigue presence.

Table 3.3: Meta-analysis of studies examining Physiological signals

Year	Method used	Number of participants	Conclusion	Result
2019	The possibilities of using EEG for fatigue assessment are discussed in this paper while testing a new self-learning model for fatigue status recognition in pilots.	40	The presented learning model showed a high level of accuracy, it manages to detect the increase in fatigue over time with detected changes in EEG rhythms.	**NA**
2019	In this study, multiple physiological signals were monitored in subjects to predict pilot's cognitive states. EEG, ECG, respiration and EDA signals were detected.	20	Increasing workload was reflected in decline of EEG power bands. Distraction influenced an increase in δ and γ power, mental fatigue – in δ, θ and α power and EDA parameters' increase. HR was lowered by distraction and was growing with fatigue.	**NA**
2018	EEG and fNIRS data were recorded during the tasks' in simulated and real flights. The objective of the present study is to evaluate the effectiveness of a method developed to assess levels of fatigue.	4	The accuracy for fatigue prediction for both real (1) and simulated flight (2) was the highest when both EEG and fNIRS features were combined. For the simulated flight the accuracy for EEG only (3) was higher than for fNIRS only (4). Same was true for a real flight, the fNIRs only accuracy (5) was lower than the EEG only (6).	**Percentage** (1) 87.6 % (2) 87.2% (3) 86.7% (4) 81.5% (5) 83.2% (6) 86.4%
2019	The possibility for detecting in-flight fatigue in pilots with an aviation headset was determined. PPG and ECG data were collected during flight simulations and checked for drowsiness signs.	14	The HR rating results were highly correlated across all periods (1). The HR decreased significantly (2) from the1st to the 4th period of measurement.	**Pearson coefficient** (2) $0.99 < r < 1$ **Kruskal Wallis** (1) $P < .05$
2018	The study presents results of 2 experiments on fatigue detection, where measurements of HRV and POMS during 40-hrs of SD were taken in the 1st test and BR, SBP and HR were measured under the effects of physical Fatigue in the 2nd test.	37 in the 1st test 60 in the2nd test	The RMSSD of HRV (1) decreased into the SD, HR (2) and LF/HF (3) rose in the evening on the 2nd day. Compared to control group, the test group had increased under workload HR (4), PR (5), and BR (6) parameters. PFI and SBP (7), HR (8), PR (9) and BR (10) expressed a correlation. MFI reflected changes in HR (11), SDNN (12), RMSSD (13), TP (14), VLF (15), LF (16), and LF/HF (17) parameters under the metal fatigue.	**ANOVA** (1) $F_{1,36} = 11.429$; $P = .002$ (2) $F_{1,36} = 27.624$; $P < .001$ (3) $F_{1,36} = 4.257$; $P = .047$ **t-test** (4) $t_{56} = 7.494$; $P < .001$ (5) $t_{56} = 3.995$; $P < .001$ (6) $t_{56} = 6.266$; $P < .001$ **Correlation** (7) $r = 0.300$; $F_{1,56} = 5.508$; $P = .022$ (8) $r = 0.349$; $F_{1,56} = 7.758$; $P = .007$ (9) $r = 0.383$; $F_{1,56} = 9.635$, $P = .003$ (10) $r = 0.266$; $F_{1,56} = 4.282$; $P = .043$ (11) $r = 20.170$; $F_{1,145} = 4.291$; $P = .040$ (12) $r = .286$; $F_{1,145} = 4.291$; $P < .001$ (13) $r = 0.207$; $F_{1,145} =6.483$; $P = .012$

				(14) r = 0.255; $F_{1,142}$ = 9.926; P = .002 (15) r = 0.249; $F_{1,142}$ = 9.378; P = .003 (16) r = 0.212; $F_{1,142}$ = 6.743; P = .010 (17) r = 0.200; $F_{1,141}$ = 5.831; P = .017
2016	The study aimed to outline maneuvers that require the most attention from subjects. For that purpose, an experiment with a flight simulator was conducted for subjects, during 3 hours of their session a headset was recording their EEG.	5	Changes in EEG power bands reflected actual changes in workload and performance of subjects. It was not possible to determine an effect of each singular maneuver, however on overall they all did add to workload.	**NA**
2015	The study was focused on investigating the relationship between attention and a flight fatigue. A mathematical model was created to pinpoint the focus of pilots' attention, using measurements of EDA, HR and RespR.	16	The presence of fatigue was getting more obvious with increasing time, moreover, HR (1), Resp (2) and EDA (3) also showed changes under the time into the test.	**ANOVA** (1) $F_{HR9,\ 135}$ = 3.246, p_{HR} = 0.001, η^2 = 0.178 (2) $F_{RespR9,\ 135}$ = 4.325, p_{RespR} < 0.001, η^2 = 0.224 (3) $F_{EDA9,\ 135}$ = 1.496, p_{EDA} = 0.155, η^2 = 0.091
2014	The investigation into the accuracy of predictions on fatigue of a proposed algorithm of automatic detection was launched. Recordings of EEG and EOG on LR flights were obtained.	14	Both the best false positive rate (1) and the best true positive rate (2) were recorded. The percentage of correct classifications reached 90% with β, (α+θ)/β ratio and fuzzy fusion (3), however the detection was not possible without (α+θ)/β ratio method and.	**RATE** (1) 0.02 to 0.04 (2) 0.87 to 0.92 **Kappa (κ) test** (3) κ > 0.80 and AUC > 0.9
2007	In this study the effect of caffeine on pilots' performance in a flight simulator under a 37-hrs continuous SD was investigated. As for collected measurements, body temperatures of participants were collected, only axillary temperature was used in the analyses.	13	Axillary temperature was higher for the caffeine group (1). An effect of increasing time into the SD on temperature was found (2). The main effect of caffeine was found to be (3). KSS scores and axillary temperature had a negative correlation (4).	**Kruskal-Wallis test** (1) 0.4°C; P = .08 (2) 0.3°C; P = .197 (3) 0.48°C; P = .004 **Spearman's correlation** (4) - 0.38; P = .007
2004	An investigation into the effects of a 37-hrs SD on a cognitive performance of pilots took place. The experiment was set with EEG data being collected in 2 power bands during simulated flights.	10	Time was found to drastically affect the increase in δ power in Cz (1), Pz (2) and Oz (3) electrodes placements. With closed eyes in resting period an interaction between time **x** eyes for δ power band was detected at Cz (4) and Pz (5).	**ANOVA** (1) $F_{3.13,\ 28.19}$ = 3.56; P = .0251 (2) $F_{3.70,\ 33.34}$ = 5.88; P = .0014 (3) $F_{2.42,\ 21.77}$ = 5.78; P = .0069 (4) $F_{2.43,\ 21.89}$ = 3.31; P = .0473 (5) $F_{2.20,\ 19.78}$ = 3.32; P= .0533

2001	The study investigated sleepiness and sleep in pilots on long-haul flights. EEG, skin resistance and eye movements were tracked on day and night flights for further cross-comparison of the results.	12	Generally, alertness was higher on a day flight than on a night flight (1). This effect was seen with skin resistance during the day (2) compared to night (3). SD affected levels of EEG δ (4) and θ activity (5), head (6) movement, and skin resistance (7). α (8) and θ (9) power bands were higher in wakefulness compared to sleep.	**ANOVA** (1) P < .05 (2) 79.5 - 109.5 (3) 87.4 - 134.3 (4) 0.678 - 1.257 µV; P < .05 (5) P < .05 (6) P < .001 (7) 61.85 - 63.17; P < .001 (8) 1.840 – 3.021 µV vs 4.082- 5.219 µV (9) 1.900- 3.625 µV vs 2.480 - 3.483 µV
1995	Measurements of sleepiness were conducted in a two-pilot crew, for that purpose EEG recordings were made. Sleepiness was detected by the decrease in α frequency range power and the α rhythms desynchronization.	22	The interaction *flight **x** time* showed an effect on alpha band power (1). A correlation between alpha peak frequency and the time effect (2) and the interaction *flight **x** time* (3) was significant.	**ANOVA** (1) F = 2.399 ; P = .052 (2) F = 6.845 (3) F = 4.938

EEG - electroencephalogram, ECG - electrocardiography, EDA - electrodermal activity, HR - heart rate, fNIRS - functional near infrared spectroscopy, PPG - photoplethysmogram, HRV - heart rate variation, POMS - profile of mood states, SD - sleep deprivation, BR - breathing rate, SBP - systolic blood pressure, RMSSD - square root of the mean differences of successive NN intervals, LF/HF - low frequency/high frequency, PR - pulse rate, PFI - Physical Fatigue Index, MFI - Mental Fatigue Index, SDNN - SD of normal to normal beats, TP - total power of the HRV spectrum, VLF - very low frequency, KSS - Karolinska Sleepiness Scale

3.3.1 A self-learning model for fatigue status recognition via the EEG

The presented article [32] discusses opportunities for EEG signal measurements to be used as an accurate fatigue predicting parameter. The experiment was set for 40 participants on a simulated 6-hrs flight, while their data on multiple physiological parameters was monitored.

Results of the testing were presented visually only, thus not allowing the study to be included into the statistical analysis. It was discovered that fatigue index graphs showed the same growing trend with time into the experiment across 4 measurements with different combinations of 4 rhythms of EEG signal, the change in the signal was very indicative. The fatigue status recognition using the fatigue recognition system presented was exceptional, on average reaching a success level of 91.67%.

3.3.2 Detection of fatigue in pilots using a combination of physiological signals' measurements

An investigation [33] into the potential of using measures as EEG, respiration, ECG, and electrodermal activity (EDA) for detecting states of low vigilance in pilots was launched. A total of 20 participants were asked to take part in a monotonous night flight with additional fatiguing, distracting and workload inducing factors in a flight simulator. EEG parameters were followed in 5 bands throughout the whole flight.

As the workload increased, all EEG power bands were declining under its influence. Distracted state was accurately reflected in an increase in δ and γ power, while mental fatigue state was reflected by an increase in δ, θ and α power of EEG bands. In a similar way fatigue affected EDA parameters, their values grew sharply under fatigue.

While compared to the wakefulness period, the detected heart rate (HR) was higher in a fatigue state and later decreased under distraction.

The results of this study provide some valuable information on changes in physiological parameters under fatigue, however, they cannot be used in a statistical analysis due to the way of their presentation in boxplots and topographic maps of EEG band powers solely.

3.3.3 Physiological signals-based interface for fatigue detection

As a proposed tool to measure cognitive performance of pilots in-flight a passive brain computer interface was tested on accuracy of predictions in the following study [34]. The tool is based on EEG and a functional near infrared spectroscopy (fNIRS) parameters' evaluation. 4 subjects were asked to participate in real and simulated flights with diffing traffic patterns.

For the real flight the success rate in identifying fatigue states in pilots reached its maximum by a combination of EEG and fNIRS parameters detection – 87,6%, for the EEG only the percentage was higher than for fNIRS only. The highest accuracy in detecting fatigue in a simulated flight was reached also by a combination of EEG and fNIRS parameters detection – 87,2%, the lowest it was when using fNIRS only.

3.3.4 Assessment of fatigue with an aviation headset

The study [35] focused on trustfulness of an aviation headset's predictions on detection of fatigue in-flight. For the purpose of setting an experiment 14 participants were given a simulated flight task that included take off, climb and cruise phases with extended time to allow the workload to have an effect of fatigue. In the meantime, data on ECG and photoplethysmogram (PPG) were collected and 4 periods of measurements were used in the analysis (T1-T4).

HR was found to be much lower in the 1st period, compared to the 2nd, the 3rd and the 4th period. PPG data collection was proven to be a reliable method for fatigue evaluation, as HR results across all measurements were strongly correlated.

3.3.5 A relationship between postural control and physical and mental fatigue

The article [36] discusses how are changes in postural control reflected in changes in physiological and psychological signals' measurement. A two-part experiment was set with the 1st part dedicated to a posturographic balance test with a heart rate variability (HRV), and a profile of mood states (POMS) data collected from 37 pilots under a 40-hrs SD. The 2nd part included collecting physiological signals as HR, systolic blood pressure (SBP) and breathing rate (BR) data from 60 pilots. For postuagraphic balance two indexed were calculated - the Mental Fatigue Index (MFI) and Physical Fatigue Index (PFI).

Comparing testing and control groups of subjects, test group under workload had increased HR, pulse rate (PR) and BR parameters.

Great changes that are indicative of fatigue presence were spotted in HRV already on the 2nd working day, compared to the 1st day: HR and low/high frequency (LF/HF) parameters were discovered to rise towards the night, while the square root of the mean differences of successive NN intervals (RMSSD) of HRV sharply decreased after early into SD.

The MFI and POMS showed a positive correlation. Moreover the MFI was also capable of reflecting changes in mental fatigue based on changes in physiological signals as HR, SD of normal to normal beats (SDNN), RMSSD, total power of the HRV spectrum (TP), very low, low and low/high frequency, while PFI expressed a positive correlation with SBP, HR, BR and pulse rate PR.

3.3.6 Alertness of pilots during a simulated flight

The next study [37] focused on a mechanism of a workload accumulation and investigated which maneuvers are requiring the most attention from pilots, thus being the most vigor demanding. 5 subjects were asked to participate in a 3-hrs flight in a flight simulator having a headset on, that was responsible for EEG signals collection.

Analysis of EEG recordings revealed, the 1st subject was constantly improving his performance, while the 2nd subject completely lost his awareness during the last session, which corresponds to the increase in measurements of θ power band.

In overall, every maneuver added to the collective workload, as shown by changes in β power band, however it was not possible to determine the effect and alertness consumption of each particular maneuver.

Measurements results were presented in boxplots only, besides, even the authors of the article expressed a concern, that the results might not be considered precise enough due to the small number of participants.

3.3.7 Evaluation of mental fatigue via the attention allocation

The objective of the study [38] was to examine a connection between pilots' mental fatigue and their attention distribution, as those two factors are known to commonly have a correlation. The experiment was set with 2 flight simulator tasks, while data from 16 pilots were collected for further analysis. In addition to the eye movements being tracked, heart rate HR, EDA, and RespR parameters were also monitored.

Time into the experiment was found to influence measured parameters as, HR, Resp and EDA to a great extent. Fatigue accumulation was proven to be accumulated by detection of changes in physiological parameters' measurements.

3.3.8 An automatic detection of states of low vigilance using EEG channel

14 pilots were recruited into the experiment [10] aiming at assessing fatigue and sleepiness of pilots on real LR flights. For that purpose, an algorithm that continuously detects changes in EEG was used in addition on KSS subjective ratings. Later the article discusses the successfulness of this algorithm. The accuracy of predictions was rated by a kappa (κ) test.

For 50% of subjects non-planned episodes of sleep were detected, mostly happening in periods of circadian low.

Both false and true positive rates were recorded. The accuracy of detection for low vigilance states reached 90% while using β and (α+θ)/β power bands ratio and fuzzy fusion and area under the receiver operating characteristic curve, plotted to compare results being > 0.9. The system was proven to present accurate results with k > 0.80 only with the (α+θ)/β ratio method.

3.3.9 An effect of caffeine on sleepiness in pilots

The presented study [18] aimed at investigating the impact that caffeine has on sleepiness. In order to add the workload effect, all the tests were performed in a flight simulator under a 37-hrs continuous SD. A total of 4 simulated flights took place for 13 pilots, divided into caffeine and placebo groups. In addition to subjective KSS ratings, axillary temperature, considered to represent the overall body temperature the best, was measured.

Axillary temperature has shown a great variation, it was in general found to decrease with time into SD, being the highest on the 1st testing day afternoon. Although the axillary temperature was falling in both groups, the drop was sharper for a placebo group. KSS and axillary temperature were discovered to have a negative correlation, which proves that changes in axillary temperature are indeed reflecting sleepiness states accurately.

3.3.10 An influence of a continuous 37-hrs sleep deprivation on performance

The objective of the study [39] was to investigate the influence of a continuous 35-hrs SD on pilots' performance and fatigue with a help of multiple subjective and objective fatigue assessment tools. EEG signals were measured among other parameters in intervals of 5 hours towards the end of a SD period. 10 pilots were taking part in multiple cognitive performance tasks during a simulated flight.

Absolute θ and δ powers were followed in EEG measurements at electrodes' sites Cz, Pz, and Oz according to an international system. Time into the SD significantly affected the δ power band, causing it to increase at all three placements of electrodes. The effects of time into SD and closed eyes were interacting for δ power band in both Pz and Cz.

3.3.11 A relationship between changes in physiological signals and an occurrence of sleepiness on a long range flight

The article [40] covers a topic of LR flights being especially associated with a dangerous occurrence of sleep and states of sleepiness. EEG measurements were collected from 12 pilots on LR flights that happened both during a day, on the outbound westward flight, and night hours, on the return eastward flight. Additionally, electrooculogram (EOG) parameters were recorded, as well as wrist activity, head movements and galvanic skin resistance.

Skin resistance parameter mostly showed change when pilots fell asleep, the resistance increased sharply at sleeping periods, compared to the baseline measurement both types of period on duty, after waking up the resistance was lower. The return flights' values of skin resistance were higher, than those recorded during the day, which also might be attributed to the fact that return flights were night flights.

Predictably, during sleep head movements were close to being absent, with opposite results in wakefulness period.

As for EEG measurements, the most significant effect on it was discovered of wakefulness, as in awake periods α and θ power bands showed an increase. Alertness fluctuations were detected by changes in δ and θ power bands, which corresponded to lower alertness on a night return flight. δ and θ power bands were both higher in sleep period compared to just the sleepiness one. Overall EEG increased during periods of wakefulness even with short-time arousals.

3.3.12 Sleepiness in civil pilots during LR flights

The reviewed paper [41] presents results of an experiment conducted with an aim of evaluating the occurrence of sleepiness in pilots on two night flights with a duration more than 10-hrs each. EEG recordings were collected with a headband from 22 subjects with eyes closed hourly after the take off.

An increased sleepiness was detected by following two indicators: the desynchronization of α rhythms and the α frequency range power being lower. A band power was much lower during the return flight. Both sleepiness parameters showed change during the return flight, which occurred at circadian low. Moreover, both parameters showed a dramatic change during the last hours of both flights, which is a clear indicator of fatigue accumulating with hours on duty.

3.4 Eye tracking

Various parameters acquired via an eye tracking have a potential to be used when assessing level of fatigue of states of drowsiness in pilots. The eye parameters are often subjected to involuntary change due to stimuli in subject's surrounding, much like physiological signals in human body. The table 3.4 below obtains short descriptions of analyzed studies with results of their experiments with parameters as EOG, eye opening (EO), pupil diameter (PD) and other provided.

Table 3.4: Meta-analysis of studies examining eye tracking method

Year	Method used	Number of participants	Conclusion	Result
2001	The study investigated sleepiness and sleep in aircrew during long-haul flights, aiming to evaluate the effectiveness of EOG for drowsiness detection. Eye movements were measured via EOG.	12	Late return flights were connected to greater drowsiness than early flights (1). Eye movements measurements also successfully recorded signs of sleepiness during the sleeping period, with a big influence of waking up on eye activity (2).	**ANOVA** (1) P < .05 (2) P < .001
1998	The study examined changes in sleep and in sleepiness between male and female pilots under a 40-hrs continuous SD. EOG was used to detect phases of rapid eye movements to detect sleep.	12	No correlation between gender or day and changes in sleep patterns was found, also gender **x** day interactions were not observed. The sleepiness was influenced by a 40-hrs SD interaction.	**NA**

2009	This study evaluates the impact of sustained wakefulness and night flying on a flight performance, cognition, and alertness. EEG and EOG data were collected at the same time as an objective sleepiness assessment.	10	Eye movements were detected to accurately pinpoint that subjects' eyes are closed during the resting phase of flight.	**NA**
2012	This study examined the effects of 35-hrs SD on cognitive tasks performance during a simulated flight. Eye tracking data were collected via EOG.	10	Eye tracking data were used in comparison and as a validation to PVT, OSPAN and MATB parameters measurements.	**NA**
2013	The goal of the experiment was to investigate the influence of a 32-hrs SD on cognitive performance of pilots. For that purpose, eye movements were recorded	4	Saccade amplitude of all participants subjects decreased along with SD and reached significant level (1). A consistent decreasing trend of the index of the average saccade velocity was discovered. Three subjects reached its significant level (2) at 11:00 on the 2nd day. The indicator of average fixation time increased through the SD as well.	**ANOVA** (1) $P < .05$ (2) $P < .05$
2015	The research was aiming at investigating the effects of varying levels of fatigue and workload on pilot's performance and physiological responses, additionally the accuracy of BAM predictions was tested. Polysomnography was used to collect data on eye movements and heart rate.	32	No correlation between fatigue and blink rate was discovered.	**NA**
2015	The research strived to look into the relationship of fatigue and attention of pilots during a flight. For eye tracking Smart Eye Pro was used, the following parameters were follower: the fixation distribution, PD and eyelid opening EO.	16	Comparison of PD between the 1st (1) and the 10th (2) session revealed a decrease in measured values across the sessions. The same decrease for and EO was noticed from the 1st (3) to the 2nd (4) session. A time into the experiment had a significant influence on both PD (5) and EO (6), which accurately depicts the increase of fatigued state in subjects.	**ANOVA** (1) 3.71 ± 0.85 mm (2) 3.30 ± 0.75 mm (3) 10.26 ± 1.91 mm (4) 9.66 ± 1.54 mm (5) $F_{PD\ 9,\ 135} = 2.619$; $P_{PD} = .008$, $\eta^2 = 0.149$ (6) $F_{EO\ 9,\ 135} = 2.851$; $P_{EO} = .004$, $\eta^2 = 0.160$;

2019	The study examined whether eye movements could be a reliable parameter to follow for fatigue levels assessment. Visual research and simulated flight tasks were completed by pilots, while their eye movements have with parameters as PERCLOS and BF been recorded.	14	Fatigue was correlated with both BF (1) and PERCLOS (2), while those 2 parameters were correlated between each other (3). PERCLOS and BF were at their lowest level at the wake up and rapidly increased into the testing period.	**Pearson coefficient** (1) r = 0.605 (2) r = 0.722 (3) r = 0.712

EOG - electrooculogram, SD - sleep deprivation, EEG - electroencephalography, PVT - Psychomotor Vigilance Task, OSPAN - Operation Span Task, MATB - Multi-Attribute Task Battery, BAM - Boeing Alertness Model, PD - pupil diameter, EO - eyelid opening, PERCLOS - percentage of eyelid closure over the pupil, BF - blinking frequency

3.4.1 Vigilance during long-haul flights and changes in physiological parameters

The study [40] set an aim of examining sleepiness of pilots on LR flights, testing if changes in physiological parameters indicate drowsiness accurately. 12 pilots flying a LR flight at both day and night participated in the study. Eye movements were recorded via an EOG, among other methods of objective fatigue assessments, to detect states of non-alertness.

A testing in a so-called control period was done, with eyes closed some subjects showed detected signs of sleepiness, and later wakefulness was significantly correlated with an eye response, as eye activity was much less active during the sleepiness periods compared to awake periods. Eye movements, as eyes being open to some extent, established that late return flight was connected, predictably, with lower vigilance levels. Following of the eye movements was also able to detect slight sleepiness in wakefulness periods, abrupt sleepiness and short arousals during the sleep period.

3.4.2 Effects of a sleep deprivation of 40 hours on sleep and sleepiness of male and female pilots

12 pilots participated in the study [42] aiming to measure differences in reaction and change in physiological measurements between male and female pilots under a 40-hrs SD. Rapid eye movements were successfully detected as one of the means of detecting sleep via the EOG after a flight in a flight simulator.

Further results stated that there was no correlation between gender or day and changes in sleep patterns, also gender **x** day interactions were not observed. The sleepiness was influenced by a 40-hrs interaction.

The study provides no results of measurements of rapid eye movements, except for conclusions, thus it will not be used for a statistical analysis.

3.4.3 Effects of a 37-hrs SD on cognitive performance and physiological arousal

The article [39] describes a study with and aim of measuring the effect of a SD and a night flying on wakefulness and performance of pilots. Tests were conducted on a flight every 5 hours during the last phase of the 37-hrs SD. EOG together with EEG data were collected from 10 subjects. The EOG was mostly used to detect when pilots' eyes are closed in order to continue with the EEG measurements in a resting part of a testing.

The study does not provide comprehensive data on eye movement measurements; thus, it will not be included into the statistical analysis.

3.4.4 Effects of sleep deprivation on cognitive performance

The investigation into the effect of a 25-hrs continuous SD on performance of pilots took place [30]. 10 participants took part in a simulated 3-hrs flight in a flight simulator after which EEG and eye tracking data were collected in order to assess physiological state objectively, the procedure was repeated 10 times.

Data obtained from eye tracking were used in comparison and as a validation to PVT, OSPAN and MATB parameters measurements, however, the data are not included in the paper. The study is not suitable for statistical analysis.

3.4.5 An effect of a 32-hrs continuous SD on physiological signals

The study [19] focused on measuring the effect of a 32-hrs continuous SD on physiological parameters of pilots, as well as their fatigue. Data on eye movements was collected during simulated flights on various FL in measurements that occurred at 11:00 on the 1st day, then at 15:00 and 04:00 on the 2nd day and later at 11:00 and 15:00.

Eye movement indices of pupil area decreased with time into SD, same as the index of average fixation revealed an increase along with time into SD trend, reaching the maximum value in final hours of SD. Compared to mentioned above parameters, pupil size change showed a consistent narrowing trend for all subjects, though reaching the significant level of narrowness at different times. Saccade amplitude and the index of saccade velocity were closely followed and were found to be constantly decreasing as well along with time into the SD. Saccade amplitude was significantly low in comparison to the baseline at 11:00 on the 2nd day, as well as an index of saccade for 3 subjects.

3.4.6 An impact of fatigue on commercial flight operations

The article [10] presents results of a research into the relationship between fatigue levels and performance and physiological responses of pilots, moreover, the comparison of test sessions' measurements and predictions of BAM tool on fatigue was made. Simulated flights sessions took place for 32 subjects with their eye movements being recorded with a polysomnography.

The focus was made on recordings of the blink rate, however, it, as some other objective fatigue assessment tools did not show to be affected by the fatigue, possibly due to fatigue level being different for each individual as a biological species.

3.4.7 Allocation of pilot's attention under fatigue

The current study [38] is aiming to examine how does fatigue affect a change in attention allocation in pilots, as it is well known to degrade with increasing drowsiness. 16 participating pilots took park in a simulated flight when their eye movements have been tracked to allocate the focus of attention, measuring such parameters as the fixation distribution, PD and EO.

PD and EO were greatly affected by the time into the experiment across the testing period with the pupil diameter becoming increasingly smaller at the 10th session compared to the 1st one, thus accurately reflecting the increasing fatigue.

3.4.8 Detection of fatigue with eye movement analysis

The study [43] speculated if eye movements could be a reliable parameter to track when measuring the level of fatigue in pilots. For that purpose and experiment with 14 pilots, who were asked to participate in a simulated flight with objective and subjective methods for fatigue assessment used during the test took place. The change in percentage of eyelid closure over the pupil time (PERCLOS) and blinking frequency (BF) were closely monitored.

A significant correlation was found between PERCLOS and fatigue and BF and fatigue, growing parallel to each other, as well as between BF and PERCLOS.

The change in followed parameters was such that PERCLOS was getting higher across the fatiguing period of testing, the same sharp increase was detected in BF.

Chapter 4
Statistical analysis

The current chapter is dedicated to the statistical analysis of the results achieved in described in previous chapters studies, which are collected by means of the meta-analysis.

The statistical analysis was performed on results of each separate group of studies that are dedicated to exploring and usage of one particular type for fatigue assessment, where it was possible. Later on statistical analysis was performed separately for objective and subjective fatigue evaluation tools and finally an overall cross-studies statistical analysis took place with the goal of finding similarities in accurate detection of low vigilance states in pilots by various fatigue measurement tools across all reviewed articles.

Into statistical analysis were only included articles that supported their findings by presenting clear results of their experiments in a form of actual values measured throughout the tests, those results were deemed reliable and trustworthy enough to be a subject of analysis. To that topic, quite a lot of articles that were used in the meta-analysis, though reaching significant results, that are very similar in particular points to the findings of the majority of presented articles, were, however, impossible to include in the statistical analysis, the reason for that being their choice the results' illustration of the in-test measurements with histograms, line graphs, scatterplots or radar graphs, without sharing actual measured numeric values, especially common was this type of results' presentation for an overall change of some parameters with time or when a correlation was discovered. Moreover, there was an obstacle in statistical analysis performance on some groups of articles, united by the topic of usage of a particular fatigue evaluation tool, as either the number of studies in those groups collected via the meta-analysis was too small, or there was a problem in the way the result's findings were given. However, I still deemed it appropriate to at least touch the subject of those methods for subjective fatigue measurement actually being used in aviation for sleepiness or fatigue evaluation of pilots.

In-group statistical analysis was not performed for studies in groups dedicated to Epworth Sleepiness Scale, Visual Analogue Scale, Rating of Perceived Exertion, Global Vigor Scale and NASA Task Load Index.

An absence of sufficient data in the automatic fatigue detection system study [11] narrowed down the number of analyzed articles about ESS with recorded data in them just to one, which makes it impossible to perform a statistical analysis on this current group or articles.

2 out of 3 chosen for the meta-analysis articles on VAS did not provide any numeric data regarding the results of the fatigue measurement via the VAS due to either them being identical to ones achieved via SP method or being too irrelevant. Thus, the statistical analysis could not be performed in this particular group of studies. Nevertheless, the 2nd study in this group [13] showed a potential for its results to be compared with results of other methods' studies in the same category of research. PRE and GVS methods were presented by one article only each. Out of 5 chosen articles that describe an experiment where NTLx was used to evaluate a workload of participating pilots, only 2 presented measured values of NTLx parameters and discussed the change of each of those parameters throughout the testing period, those articles have a potential to be included into the cross-studies statistical analysis, however their results were not very diverse. Unfortunately, out of 8 chosen for meta-analysis articles that describe eye tracking method for fatigue detection, 3 did not include actual measured values and 1 additional article stated that no statistically significant changes in eye movements in regard to fatigue were recorded. The rest of the articles were focusing on too different parameters connected to eye movements, thus the proper statistical analysis was not possible within the group, however, the results of the mentioned above studies could be compared with those of other fatigue evaluation tools, both objective and subjective, as even on such a small sample of experiments the successfulness and accuracy of the eye tracking method in detecting states of low vigilance was high.

The total rating for successful detection of fatigue could be seen for subjective methods in figure 2, for objective methods – in figure 4.

4.1. Statistical analysis of data provided by articles describing subjective fatigue assessment tools

In general, multiple methods used in this group of studies showed similarities in experiments' outcomes. KSS [10] and ESS sleepiness [15] both revealed that sleepiness on short and long range flights was not the same. Longer prior TST affected, in a positive way, parameters in almost all methods' studies, more precisely - in 11: it reduced KSS [1,6,7,8,9], SSS sleepiness [20], SP fatigue [2,6,7,8,9,13,21,22], VAS fatigue [13] and fatigue severity via FSS [16] and, on the contrary, increased vigor via GVS [20]. Both KSS [9] and SP ratings [2] detected a beneficial effect of on-duty sleep for sleepiness and fatigue reduction in pilots. Fatigue being higher with increasing workload was at the same time detected by SSS [20], SP [8] and NASA methods [5]. In addition to KSS sleepiness ratings greatly varying between inbound and outbound flights [1,2,13], the same effect was discovered for SP [2] and VAS fatigue [13] ratings. Moreover, both SP and KSS methods pointed out that individual characteristics of subjects might have affected sleepiness and fatigue [8]. The timing of a flight was significant for GVS vigor and SSS, evening flights were associated with lower vigor and higher sleepiness levels, in analogy with times of arrival and departure or duty start and finish times greatly influencing KSS sleepiness in 5 studies and SP fatigue in 5 studies. Detailed results of the statistical analyses in each particular group of studies dedicated to a specific subjective fatigue measurement method are presented below, including a graphical representation of effects on fatigue, detected by each individual subjective assessment tool in figure 3.

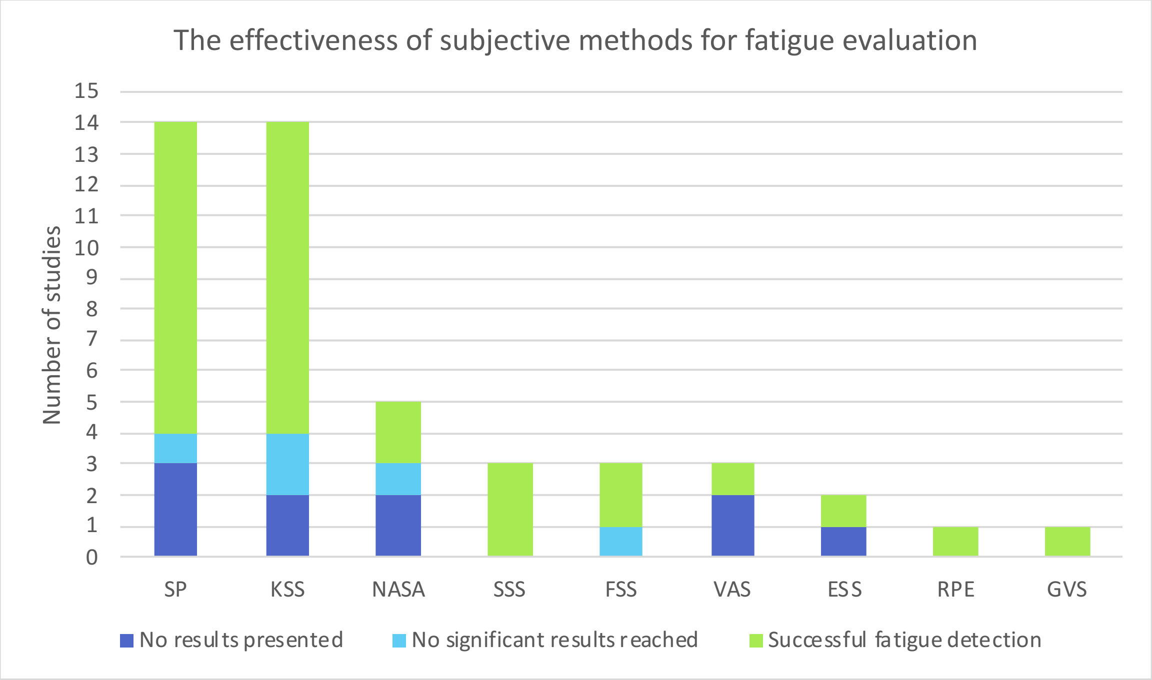

Figure 2: The effectiveness of subjective methods for fatigue evaluation

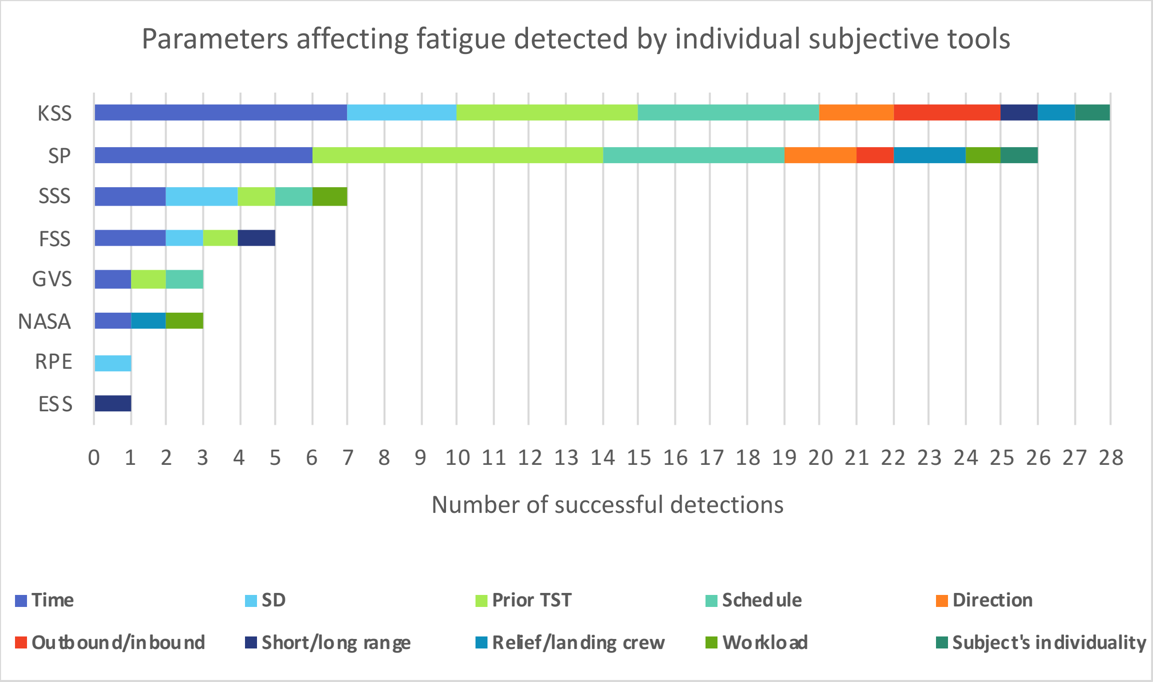

Figure 3: Parameters affecting fatigue detected by individual subjective tools

4.1.1 Statistical analysis of data provided by articles describing Karolinska Sleepiness Scale method

In total 14 studies on KSS were collected as a result of the meta-analysis with 2 studies not presenting any significant results and 2 presenting no results in terms of measured values.

The common observed change in sleepiness was its increase with pilot's time being awake, as it was noticed in 8 experiments [2,5,6,9,10,12,13,14], similar change of sleepiness with time into the flight, probably affected by the accumulation of sleepiness was discovered in further 3 studies [1,3,10]. In multiple studies was confirmed a significant beneficial effect of a longer pre-duty TST [1,6,7,8,9], a total of 5 studies revealed that result, only 1 experiment discovered no effect of a prolonged layover on fatigue [13]. Comparisons between operating and relief crews' fatigue levels were detected by KSS in 3 papers [1,9,12], as well as between inbound and outbound flights in 4 papers. Return flights were connected to higher, compared to outbound flights, level of sleepiness [1,2,13]. 2 studies assessed the same non-changing levels of sleepiness on post-flight days [1,2]. KSS ratings established, that a schedule of a flight has a significant impact on sleepiness measured either before, during or after the flight. In particular, the flight arrival time, as well as a duty finish time were influencing sleepiness in 3 studies [5,7,9], with sleepiness being at its highest for flights arriving late at night or early in the morning. Similarly, the departure or duty starting times were discovered to effect fatigue in 2 studies [6,9], here the highest KSS ratings were from pilots starting their duty between 00:00 and 06:00. The direction of a flight seemed to have an impact on KSS ratings in 2 studies [9,14].

4.1.2 Statistical analysis of data provided by articles describing Fatigue Severity Scale method

Only 3 studies that provided a description of a FSS usage as a subjective method for evaluating fatigue in pilots were included in the meta-analysis. Due to the little data obtained from all 3 studies altogether and due to the uniformity of results in terms of data recorded, it was not possible to perform an extensive statistical analysis on this particular group of studies except for underlining, that FSS in 2 studies provided results indicating a gradual increase in fatigue severity parallel to hours on duty or throughout the testing session [15,16].

4.1.3 Statistical analysis of data provided by articles describing Stanford Sleepiness Scale method

The review of the chosen 3 articles on SSS revealed an opportunity to perform a statistical analysis on how sleepiness increases across testing periods or time on duty and to study the effect of a continuous SD on sleepiness. 2 studies revealed ratings of SSS indicating higher sleepiness after a duty or testing period, compared to its beginning, with sleepiness increasing parallel to hours on duty [18,20]. Effects of the SD were also clearly reflected on sleepiness in 2 studies [18,19].

Further analysis of other results will be possible comparing data collected, for example, with SSS and KSS, as many similarities in sleepiness levels' fluctuations were found by both of those subjective sleepiness measurement methods.

4.1.4 Statistical analysis of data provided by articles describing Samn-Perelly method

As for the SP method, 14 studies representing the usage of SP method in assessing fatigue levels of pilots were chosen for the meta-analysis. Not significant enough to be included into the statistical analysis were found 4 of those papers.

The direction of flights, during which ratings of SP were collected, were discover to effect SP ratings in 2 studies [9,21]. In other 2 studies noticeable differences were found in SP ratings of relief and operating crews. 8 studies agreed, that a prior TST and layover duration were accurate predictions of collected fatigue ratings [2,6,7,8,9,13,21,22] and that fewer sleeping hours resulted into higher SP ratings. In 6 experiments SP scale revealed a significant effect of time on duty or into experiment on fatigue increase [2,8,9,12,21,22], higher fatigue being detected in the later stages of flight or a later test session. SP ratings of pilots, who started their duty or had a flight departing later in the evening or at night were higher compared to rating of pilots who had a morning duty in 3 studies [5,6,9], being the highest for departures between 00:00-06:00 and the lowest for departures between 12:00 – 18:00. SP ratings also reflected a change in fatigue in subjects with varying arrival of the flight of finish of the duty time in next 2 studies [7,9], additional study just underlines an impact of pilot flight’s schedule on fatigue [13].

4.2 Statistical analysis of data provided by articles describing objective fatigue assessment tools

As statistical analysis and results of the reviewed studies revealed, physiological signals, being very diverse, in general seem to share a high accuracy in predicting states of low vigilance, as it was discovered, they do respond to the same stimuli.

The effect of time on duty or into the experiment was probably the one inducing change in the most measured parameters in the articles chosen for the meta-analysis. 2 studies on voice detection [26,28], 8 studies on PVT [2,10,12,16,17,20,21,24], 6 studies on physiological signals measurement [35,36,38,39,40,41] and 2 studies on eye tracking [38,43], respectively, with changes in various parameters indicated a change in parameters’ values and levels under time. A SD was directly negatively affecting parameters of both eye tracking [19] and physiological signals measured [18].

Growing workload was detected by not only PVT tests, showing a decrease in performance [8], but with changes in voice [27] and in multiple physiological signals – HR, PR, BR [36]. Below are presented the results of the statistical analyses in each particular group of studies dedicated to a specific objective fatigue measurement method in details, including the graphical representation of detection frequency of effects on fatigue detected by individual objective tools in figure 5.

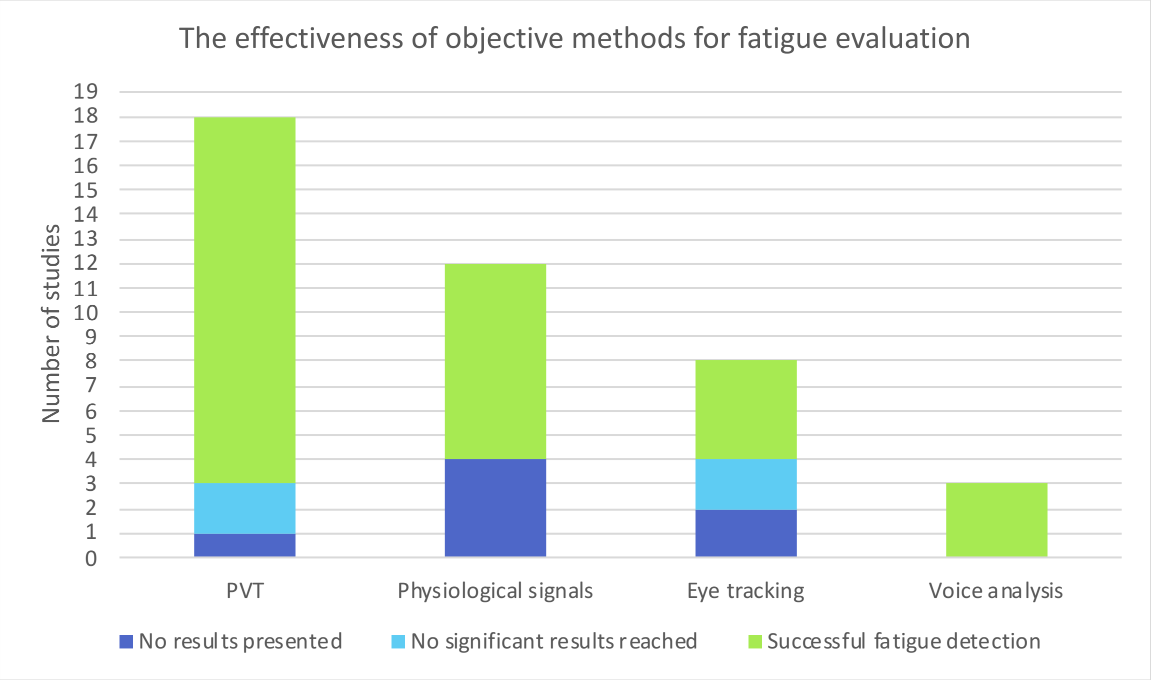

Figure 4: The effectiveness of objective methods for fatigue evaluation

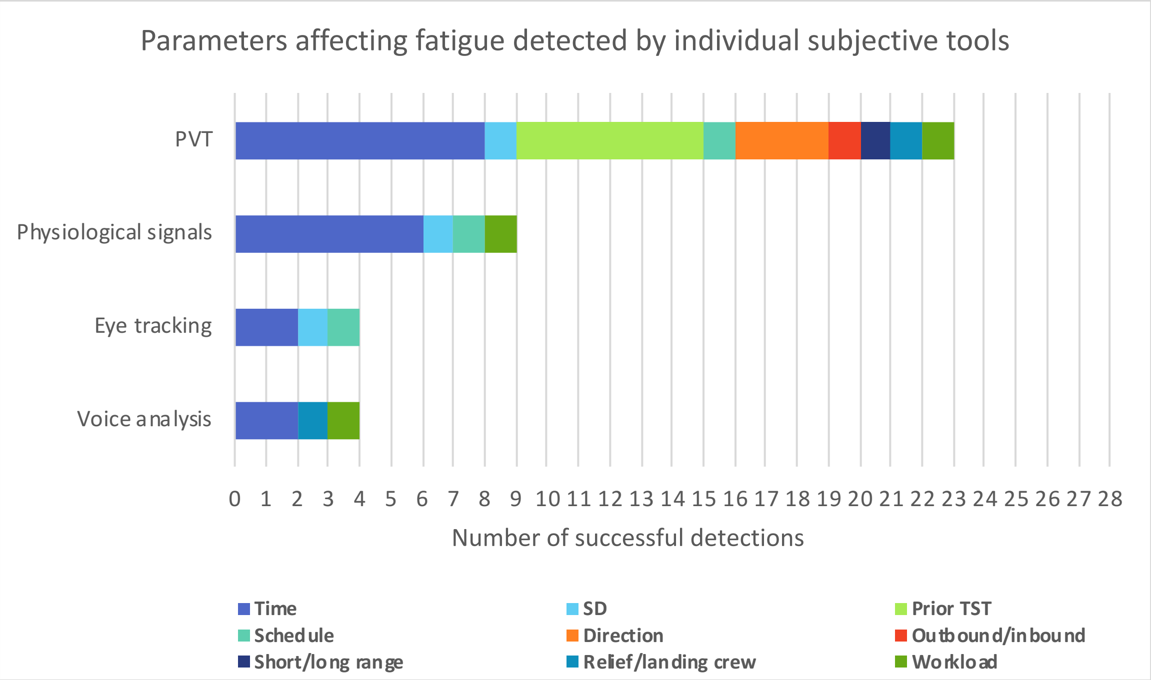

Figure 5: Parameters affecting fatigue detected by individual subjective tools

4.2.1 Statistical analysis of data provided by articles describing Voice Analysis method

Due to the fact, that multiple articles that were evaluated in order to be included into the meta-analysis were lacking results of the measurements, or having the results not being included in the study in a presentable way, or having the main focus on the study being not the fatigue detection via the speech analysis, but something else, only 3 studies were selected.

The common ground for all 3 articles was that the voice analysis systems were perfectly able to detect changes in sounds in speech with fatigue increasing. Specifically, 2 studies [26,28] revealed, that changes in voice were becoming more prominent and often with increasing time into the flight or the experiment session, being drastically different pre- and post-flight.

4.2.2 Statistical analysis of data provided by articles describing Psychomotor Vigilance Task method

A total number of the reviewed as a part of the meta-analysis articles with PVT method results being underlined was 18, however, 3 studies were not included into the statistical analyses due to either not presenting significant results or due to a PVT being used as a side measurement and not paid sufficient attention.

As for a PVT method the most common evaluated parameters were RT and lapses. RT was commonly found to get slower with more time into the flight or into the experiment, with each new test session, as was proved in 7 studies [2,10,12,16,20,21,24]. The increasing amount of errors [20] and lapses [16] in performance and lowering of the sustained attention, as well as a decrease in CFF [17] were also the results of time influence. The PVT twice showed a difference in RT between outbound and inbound flights [2,13]. The method also repeatedly indicated, that prior shorter TST negatively affects RT and sustained attention, as discovered in studies in the previous chapter [8,9,16, 20, 22, 31].

4.2.3 Statistical analysis of data provided by articles describing physiological signals measurement method

12 articles were included into the meta-analysis, that described an actual experiment for fatigue assessment with the help of physiological signals' measurement. 3 articles did not present any values measured in the study or did not include statistical analysis results. Other 2 articles only provided results on success rates and accuracy of algorithms and systems for automatic detection of low states of vigilance in pilots, that were essentially working on a mixture of multiple psychological signals measurements being evaluated. Although it does prove the efficiency of physiological signals' monitoring for detection of fatigue and shows a perspective

of them being potentially used in aviation, I did not include those 2 studies into the statistical analysis, as it is not possible to predict which particular followed parameters changed, in which way and under what influence.

The most common parameter measured – the EEG was proven to be a great marker of fatigued states, increasing, it was mostly affected by the time passed on a duty [39,40,41] or workload [39], but it also reflected higher levels of fatigue during night and return flights, decreasing [40],

HR, was discovered to slow down with an increasing time into the flight or into the experiment [35,36,38] and to be higher under workload for a test group, compared to the control one [36], thus being correlated with increasing fatigue. The effect of time was also discovered on recorded EDA [38], its parameters were growing with time, as well as the values of RespR [38].

An increase in PR and BR under a workload was detected [36]. A skin resistance showed the most prominent change between sleep and wakefulness, being higher during subject's sleep [40]. It also was lower during day the day flight, compared to the night one. Finally, axillary temperature was decreasing in a test under effect of SD and time [18].

4.2.4 Statistical analysis of data provided by articles describing eye tracking measurement method

Into the meta-analysis were included 8 studies that used eye-tracking as a method of evaluating fatigue in pilots. Out of the total amount of the reviewed articles, 3 could not be included into the statistics analysis, as they achieved particular important results, nevertheless did not provide any actual measurements taken during the test. Additional article did not discover any significant correlation between eye movements and fatigue.

In each of the reviewed studies, a totally different parameter was followed. An accumulating SD was detected accurately by the change in eye movement parameters: saccade amplitude and index of average saccade velocity, which were decreasing, while average fixation time was only increasing [19]. An EOG parameter was much higher during early morning outbound flights and reflected increased fatigue on late return ones [40]. The effect of time was consistent with changes in multiple eye tracking parameters. PD and EO were getting lower each new test session along with time [38], while PERCLOS and BF, on the contrary, were increasing [43].

4.3 Cross study statistical analysis of results

An overall analysis of articles included into the mate-analysis was quite a tricky task due to the heterogeneity of the results in those studies and parameters measured. Particular obstacle was sometimes to figure out whether 2 studies are investigating the influence of the same stimulus or condition, as there are many ways to name an object of focus and it is not always clear if two terms are similar in regards to the mechanism of fatigue accumulation. Thus, fatigue increase with time, fatigue increase in the end of duty, compared to the beginning of it (or difference between pre- and post-flight fatigue) and fatigue increase with hours being awake in different contexts could mean the same or different things, when, for instance some longer time might pass between waking up and actually starting a duty for pilots. Another example being: fatigue levels being different between morning and evening departures/arrivals, between early/late starting/finishing duty, between early/late starting/finishing flight and between morning/evening or day/night flights, again, depending on the context, those parameters could be same or different completely.

As the meta-analysis of articles proves, signals of accumulating fatigue could be hidden in changes of varied subjective and objective fatigue parameters, that do not seem to be connected at first sight. However, cross analyzing the reviewed articles and their results a revelation could be made, that actually on numerous occasions those fatigue indicators are changing in response to the same stimuli around the pilot.

In total 23 studies described usage of subjective fatigue assessment methods and 36 studies described objective fatigue evaluation tools, mostly – PVT. Multiple studies were noticed to use both types of methods, PVT was a frequent addition to subjective methods, which are seldom used alone. The graphical representation of this statistics is in figure 6.

An absolute majority for fatigue detection methods reflected an increase in fatigue parallel to increase of time into the flight. Fatigue was detected in pilots by 17 different methods, both subjectively and objectively: eye tracking, specifically measurements of PD and EO [38], PERCLOS and BF [43], EEG [39,40,41], HR, [35,36,38], EDA and RespR [38], PVT, more precisely by changes in RT [2,10,12,16,20,21,24], errors [20], lapses [16] and CFF [17], speech analysis [26,28], subjective sleepiness, KSS [1,2,3,5,6,9,10,12,13,14] and SSS [18,20], and fatigue, FSS [15,16] and SP [2,8,9,12,21,22]. A conclusion could be drawn, that changes in fatigue over time are one of the most easily and correctly recognized ones by assessment tools.

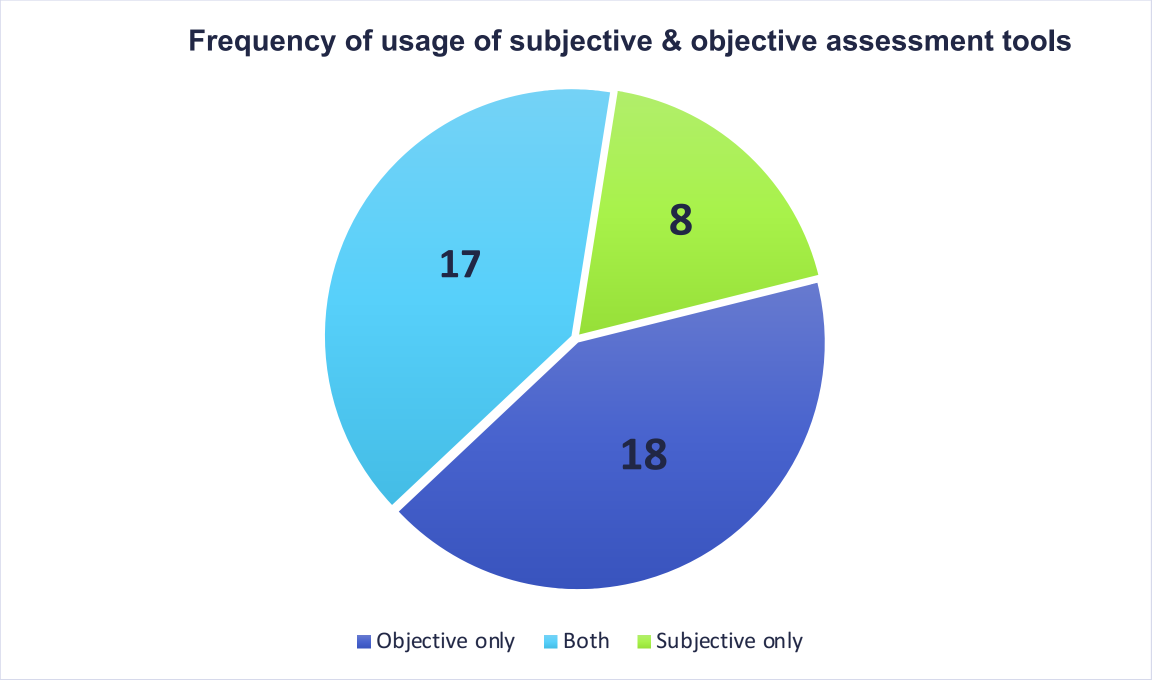

Figure 6: Frequency of usage of subjective&objective assessment tools in studies reviewed

To some extent, an effect of time on duty might be considered similar to the effect of a SD, nevertheless, I did not unite those two effects under one term, as in my understanding SD implies a greater influence of an additional fatiguing and workload infusing factor - an accumulating sleepiness. Especially when for experimental reasons in the reviewed studies SD of a duration of 24 to 40 hours was used, which could not be considered a normal and wildly spread phenomena during day-to-day flight duties in aviation, when many airlines nowadays do try to mitigate fatigue and sleepiness in pilots with FRMS and guidelines on sleep and napping. The SD effect was discovered to have influence on eye movements, as saccade amplitude, index of average saccade velocity and average fixation time [19], axillary temperature [18] and HRV [36]. The only subjective tool proved with actual results included to detect SD influence on sleepiness was SSS [18,19].

As for the workload, it was detected with a level of confidence by mostly objective methods: NASA [25], HR, BR and PR [36], PVT errors [8] and RMSE [30]. Again, SSS [20] being the only subjective tool outlining a correlation between sleepiness and workload in a sense of a more demanding work.

Relief and operating crews were estimated to have not similar levels of fatigue pre- or during the flight by SP [9], KSS [1,9,12], voice analysis [27] and RT characteristics of PVT [8,25]. This difference in fatigue levels could be explained by workload effect if an assumption is made, that relief and operating crews are being under different levels of workload.

Prior TST, and often it’s quality, measured however by methods not included in this study as they do not focus on fatigue detection, was also a parameter correlated with accurate predictions in change of many, in total of 7, subjective scales and physiological performance and signals. Thus, by the lack of sufficient prior TST were negatively affected FSS [16], VAS [13] and SP fatigue [2,6,7,8,9,13,21,22], GVS vigor [20], KSS [1,6,7,8,9] and ESS sleepiness [20] and only one objective assessment method – RT in PVT [8,9,16, 20, 22, 31].

Various features of a particular flight were discovered to have an impact on fatigue levels. Changes in fatigue influenced by the characteristics of the pilot’s schedule, both times by arrival/departure and flight timing, were only possible to detect with subjective methods of KSS [5,6,7,9] and SP [5,6,7,9,13].

To some extent this and the above paragraph might be connected, as influence of a day/night or morning/evening duty period might mean the same as an effect of the flight schedule in a particular context. Differences in parameters, attributed to changes in fatigue, between early and late day time periods are assessed by skin resistance [40], KSS [5] and GVS [20]

Pilots on inbound and outbound flights were often discovered to no have same levels of fatigue, as it was proven by experiments using EOG [40], PVT RT [2,13] and subjective SP [1,2,13]. However, there is no agreement in those results whether the most tiresome flights are outbound or inbound ones. As in the studies reviewed, the effect of the type of the flight is often considered together with its timing, thus detected changes in fatigue might be as well attributed to influence of parameters looked into in the text above.

The direction of a flight, due to its connection to various disruptions of circadian rhythms with time being on duty and multiple time zones crosses over the flight, is an important attributing to overall fatigue factor, as stated after analyzing results of SP [9,21] and KSS [9,14].

The type of a flight in a sense of being a short- or long range was revealed to have an impact on sleepiness measured by KSS [10] or ESS [15].

Only subjective scales as PS and KSS evaluated that the level of fatigue is indeed connected to pilot’s individuality and his own perception of high of low fatigue [8].

Some studies also cross-compared different methods for fatigue assessment and published results with correlations found between the tools used. SP and KSS methods, being ones of the most well-known and wildly used, were often used in addition to other subjective of objective fatigue evaluation methods, thus their ratings were numerous times compared to while proving the accuracy of achieved results. Sleepiness and fatigue were compared across a variety of their assessment tools, for the sake of statistical analysis only correlations discovered in the studies with presented results in values were reviewed. A correlation was found between higher rating of SP fatigue and amount of PVT lapses [8] and SP and VAS fatigue [23], changes in overall fatigue were proven to be associated with changes in eye movements [40] and PERCLOS and BF [43]. 1 study revealed a correlation between sleepiness and fatigue [15]. As for sleepiness by KSS, it was correlated with changes in performance via PVT [8] and in axillary temperature [18], higher KSS levels were connected to lower temperatures. Overall higher sleepiness had a relationship with greater RT and lower vigor [20]. Additionally, growing speech derivation was correlated with longer RT and pore lapses via PVT. The histogram comparing an accuracy of both subjective and objective methods in regard to changes in fatigue under a particular stimulus is depicted in figure 7.

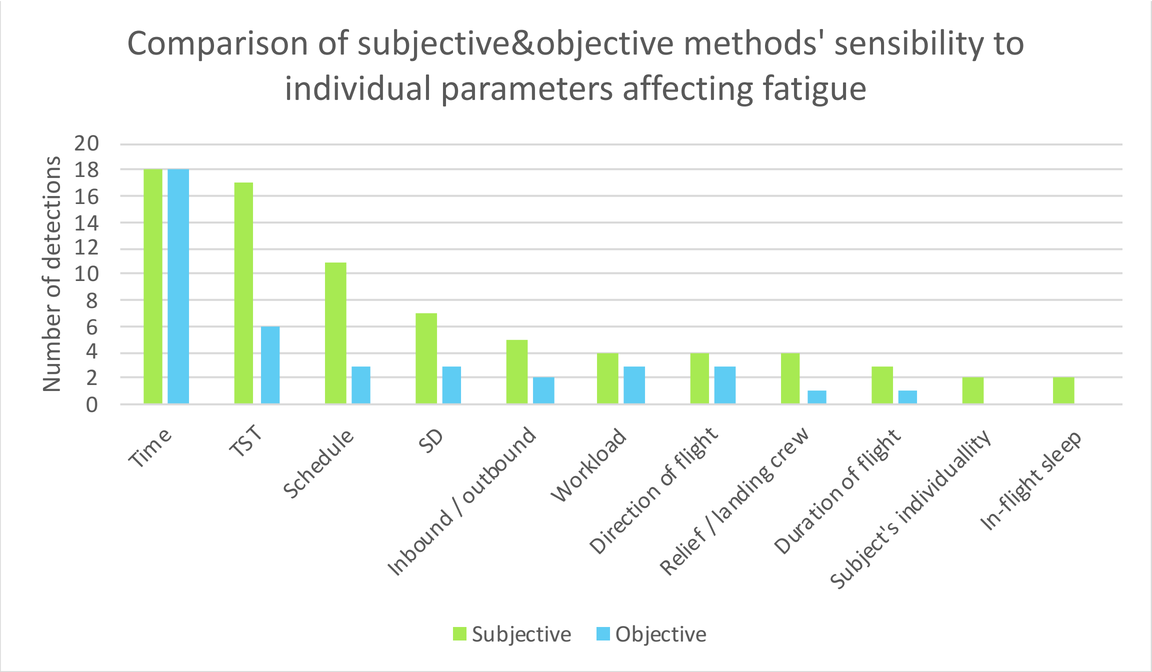

Figure 7: Comparison of subjective&objective methods' sensibility to individual parameters affecting fatigue

Conclusion

The theoretical part of the thesis contained mostly in Chapters 2 and 3 had the main goal of performing a state-of-art review of existing and currently used in aviation methods for fatigue evaluation in pilots. The following practical part was dedicated to categorization and step-by-step analysis of chosen articles by means of meta-analysis, the chosen trustworthy results achieved in those studies were later subjected to statistical analysis in groups of studies and cross-study with an aim of outlining the efficacy of discussed methods in fatigue assessment.

As it is seen in statistical analysis performed in Chapter 4, the majority of the reviewed methods for fatigue qualification, either used in a combination within one study or separately and then compared cross-study, tended to achieve similar results while evaluating states of low vigilance in response to the same stimuli, which proves their accuracy, as e

The reason why both types for fatigue assessment methods are so often still used in a combination might lie in their ability to detect fatigue under different circumstances. Fatigue is a complex phenomenon that could be described by various cognitive states and, in real life conditions, always consists of a combination and interaction of those states: subjective feelings of fatigue and sleepiness, low vigor, prolonged reaction time and more lapses, exertion, lower EEG and temperature, slower HR, elevated workload index, smaller EO and PD, more often speech deviations… Due to this reason, it seems to be quite impractical to focus on just one type for fatigue assessment tools for pilots, as it could be seem from the statistical analysis that some types of fatigue measuring tools are more or less sensitive to certain stimulus.

As it is obvious from histogram in figure 6, in terms of detecting change in pilots' fatigue with increasing time on duty, both methods were equally accurate. Nevertheless, objective methods, being usually considered more precise and trickier to be fooled by one's perceiving of drowsiness are also less sensitive to factors like schedule or type of position within the crew. Even if fatigue is only being perceived by the pilot as severe, not actually disrupting his bodily functions, it still negatively affects pilot's mentality and might lead to long-term effects as stress, depression or burnout, subjective tools are values for FRMS creation.

Taking into the consideration only validated results, an obvious successful wild-spread use of SP and KSS in numerous studies and already present FRMS can be noticed, bringing satisfactory results in subjective fatigue detection over the years. In multiple cases, however it is more beneficial to turn to objective methods that cannot be easily tricked by, sometimes inaccurate, psychological evaluation of the present mental or physiological state. Such methods could be performance tests and, most importantly, physiological signals. The only drawback of physiological signals' measurement being in present times a necessity for the pilot to be surrounded in sensors attached his body, that might affect ones' mobility and comfort. Nevertheless, there are already present some systems of automatic detection of drowsiness that are contactless or require just minimal contact.

Undoubtedly, the presented work has its flaws despite the best efforts. The work turned out to be not as extensive as it was intended, as it was mentioned in the practical part, not all of the mentioned fatigue methods reviewed had been included into the statistical analysis due to the absence of clearly or at all presented results of in-experiments measurements or due to the small sample of studies that were collected on a particular group of tools. No conclusions could be drawn on the efficacy of such fatigue evaluation tools as Epworth Sleepiness Scale, Visual Analogue Scale, Rating of Perceived Exertion, Global Vigor Scale and NASA Task Load Index, the performance of statistical analysis of those methods was impossible due to either a small sample of articles collected or due to very few to no results of in-experiments tests measurements included.

The review of a NASA Task Load Index method brought the biggest disappointment, because the method is quite a wildly used one, however, studies, that are dedicated to it, collected for the meta-analysis were very poor in terms of results presentation. A possible explanation for this could be an essence of a NASA Task Load Index itself, consisting of multiple sub-indexes that are too difficult to monitor and a change in them is presented more easily by some graph. Moreover, some amount of presented in the articles results could not be verified, as they are presented only in a form of a written statement or conclusion without actual values measured.

Despite the obstacles that this thesis was facing, the work was conducted systematically and with a great regard to the detail. All of the included articles were given a description and were analyzed to some extent, the description of current-in-use methods was created, and conclusions were drawn based on a thorough cross-studies statistical analysis. The main contribution of this study could be considered a robust recherché of means for fatigue detection and monitoring in a modern-day aviation with meta-analysis included.

Thus, the study, being one of few dedicated to qualification of methods for fatigue assessment particularly in pilots and with regards to division of methods into subjective and objective ones, could have a practical use for further studies created on the subject in terms of presenting organized information or as a guidance while making a choice of a fatigue measurement tool in an experiment. The overview of the problem of fatigue prevalence in pilots made it clear that not enough attention is focused on solving this safety-threatening issue, furthermore, the issue will become ever more pressing in years to come with aviation industry growing rapidly.

More studies should be dedicated to validation of results of existing fatigue assessment tools' results and creation of FRMS, fatigue mitigation guidelines, more advanced tools, as some of the already existing systems of automatic detection of low vigilance states. Moreover, a greater emphasis should be put on wilder implementation and improvement of objective methods, since they are less susceptible to ones' subjective judgment.

Literature

[1] HOLMES, Alexandra, Soha AL-BAYAT, Cassie HILDITCH a Samira BOURGEOIS-BOUGRINE. Sleep and sleepiness during an ultra long-range flight operation between the Middle East and United States. *Accident Analysis & Prevention* [online]. 2012, **45**, 27-31 [cit. 2020-11-12]. ISSN 00014575. Available at: doi:10.1016/j.aap.2011.09.021

[2] VAN DEN BERG, Margo J., T. Leigh SIGNAL, Hannah M. MULRINE, Alexander A. T. SMITH, Philippa H. GANDER a Wynand SERFONTEIN. Monitoring and Managing Cabin Crew Sleep and Fatigue During an Ultra-Long Range Trip. *Aerospace Medicine and Human Performance* [online]. 2015, **86**(8), 705-713 [cit. 2020-11-12]. ISSN 23756314. Available at: doi:10.3357/AMHP.4268.2015

[3] GANDER, Philippa H., Jim MANGIE, Margo J. VAN DEN BERG, A. Alexander T. SMITH, Hannah M. MULRINE a T. Leigh SIGNAL. Crew Fatigue Safety Performance Indicators for Fatigue Risk Management Systems. *Aviation, Space, and Environmental Medicine* [online]. 2014, **85**(2), 139-147 [cit. 2020-11-12]. ISSN 00956562. Available at: doi:10.3357/ASEM.3748.2014

[4] POWELL, David M. C., Mick B. SPENCER a Keith J. PETRIE. Comparison of In-Flight Measures with Predictions of a Bio-Mathematical Fatigue Model. *Aviation, Space, and Environmental Medicine* [online]. 2014, **85**(12), 1177-1184 [cit. 2020-11-12]. ISSN 00956562. Available at: doi:10.3357/ASEM.3806.2014

[5] VEJVODA, Martin, Eva-Maria ELMENHORST, Sibylle PENNIG, Gernot PLATH, Hartmut MAASS, Kristjof TRITSCHLER, Mathias BASNER a Daniel AESCHBACH. Significance of time awake for predicting pilots' fatigue on short-haul flights: implications for flight duty time regulations. *Journal of Sleep Research* [online]. 2014, **23**(5), 564-567 [cit. 2020-11-12]. ISSN 09621105. Available at: doi:10.1111/jsr.12186

[6] COSGRAVE, Jan, Lora J. WU, Margo VAN DEN BERG, T. Leigh SIGNAL a Philippa H. GANDER. Sleep on Long Haul Layovers and Pilot Fatigue at the Start of the Next Duty Period. *Aerospace Medicine and Human Performance* [online]. 2018, **89**(1), 19-25 [cit. 2020-11-12]. ISSN 2375-6314. Available at: doi:10.3357/AMHP.4965.2018

[7] VAN DEN BERG, Margo J., Lora J. WU a Philippa H. GANDER. Subjective Measurements of In-Flight Sleep, Circadian Variation, and Their Relationship with Fatigue. *Aerospace Medicine and Human Performance* [online]. 2016, **87**(10), 869-875 [cit. 2020-11-12]. ISSN 2375-6314. Available at: doi:10.3357/AMHP.4587.2016

[8] HONN, Kimberly A., Brieann C. SATTERFIELD, Peter MCCAULEY, J. Lynn CALDWELL a Hans P.A. VAN DONGEN. Fatiguing effect of multiple take offs and landings in regional airline operations. *Accident Analysis & Prevention* [online]. 2016, **86**, 199-208 [cit. 2020-11-12]. ISSN 00014575. Available at: doi:10.1016/j.aap.2015.10.005

[9] GANDER, Philippa H., Hannah M. MULRINE, Margo J. VAN DEN BERG, A. Alexander T. SMITH, T. Leigh SIGNAL, Lora J. WU a Gregory BELENKY. Effects of sleep/wake history and circadian phase on proposed pilot fatigue safety performance indicators. *Journal of Sleep Research* [online]. 2015, **24**(1), 110-119 [cit. 2020-11-12]. ISSN 09621105. Available at: doi:10.1111/jsr.12197

[10] THOMAS, Lisa C., Christopher GAST, Robert GRUBE a Kimberly CRAIG. Fatigue Detection in Commercial Flight Operations: Results Using Physiological Measures. *Procedia Manufacturing* [online]. 2015, **3**, 2357-2364 [cit. 2020-11-12]. ISSN 23519789. Available at: doi:10.1016/j.promfg.2015.07.383

[11] SAUVET, F., C. BOUGARD, M. COROENNE, et al. In-Flight Automatic Detection of Vigilance States Using a Single EEG Channel. *IEEE Transactions on Biomedical Engineering* [online]. 2014, **61**(12), 2840-2847 [cit. 2020-11-12]. ISSN 0018-9294. Available at: doi:10.1109/TBME.2014.2331189

[12] SIGNAL, T. Leigh, Hannah M. MULRINE, Margo J. VAN DEN BERG, Alexander A. T. SMITH, Philippa H. GANDER a Wynand SERFONTEIN. Mitigating and Monitoring Flight Crew Fatigue on a Westward Ultra-Long-Range Flight. *Aviation, Space, and Environmental Medicine* [online]. 2014, **85**(12), 1199-1208 [cit. 2020-11-12]. ISSN 00956562. Available at: doi:10.3357/ASEM.4034.2014

[13] POWELL, David M. C., Mick B. SPENCER a Keith J. PETRIE. Fatigue in Airline Pilots After an Additional Day's Layover Period. *Aviation, Space, and Environmental Medicine* [online]. 2010, **81**(11), 1013-1017 [cit. 2020-11-12]. ISSN 00956562. Available at: doi:10.3357/ASEM.2804.2010

[14] ANNE ERIKSEN, Claire a Torbjörn ÅKERSTEDT. Aircrew Fatigue in Trans-Atlantic Morning and Evening Flights. *Chronobiology International* [online]. 2009, **23**(4), 843-858 [cit. 2020-11-12]. ISSN 0742-0528. Available at: doi:10.1080/07420520600865574

[15] ALJURF, Tareq M., Awad H. OLAISH a Ahmed S. BAHAMMAM. Assessment of sleepiness, fatigue, and depression among Gulf Cooperation Council commercial airline pilots. *Sleep and Breathing* [online]. 2018, **22**(2), 411-419 [cit. 2020-11-12]. ISSN 1520-9512. Available at: doi:10.1007/s11325-017-1565-7

[16] O'HAGAN, Anna Donnla, Johann ISSARTEL, Eoghan MCGINLEY a Giles WARRINGTON. A Pilot Study Exploring the Effects of Sleep Deprivation on Analogue Measures of Pilot Competencies. *Aerospace Medicine and Human Performance* [online]. 2018, **89**(7), 609-615 [cit. 2020-11-12]. ISSN 2375-6314. Available at: doi:10.3357/AMHP.5056.2018

[17] MA, Jin, Ru-Meng MA, Xi-Wen LIU, et al. Workload Influence on Fatigue Related Psychological and Physiological Performance Changes of Aviators. *PLoS ONE* [online]. 2014, **9**(2) [cit. 2020-11-12]. ISSN 1932-6203. Available at: doi:10.1371/journal.pone.0087121

[18] LOHI, Jouni J., Kerttu H. HUTTUNEN, Taija M. M. LAHTINEN, Airi A. KILPELÄINEN, Arto A. MUHLI a Tuomo K. LEINO. Effect of Caffeine on Simulator Flight Performance in Sleep-Deprived Military Pilot Students. *Military Medicine* [online]. 2007, **172**(9), 982-987 [cit. 2020-11-12]. ISSN 0026-4075. Available at: doi:10.7205/MILMED.172.9.982

[19] LIU, Zhong-Qi, Qian-Xiang ZHOU a Fang XIE. Effects of Sleep Deprivation on Pilot's Cognitive Behavior in Flight Simulation. DUFFY, Vincent G., ed. *Digital Human Modeling and Applications in Health, Safety, Ergonomics, and Risk Management. Healthcare and Safety of the Environment and Transport* [online]. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, 2013, s. 45-54 [cit. 2020-11-12]. Lecture Notes in Computer Science. ISBN 978-3-642-39172-9. Available at: doi:10.1007/978-3-642-39173-6_6

[20] SIMONS, Ries, Ellen S. WILSCHUT a Pierre J. L. VALK. Sleep and Alertness in North Sea Helicopter Operations. *Aviation, Space, and Environmental Medicine* [online]. 2011, **82**(7), 704-710 [cit. 2020-11-12]. ISSN 00956562. Available at: doi:10.3357/ASEM.2960.2011

[21] PETRILLI, Renée M., Gregory D. ROACH, Drew DAWSON a Nicole LAMOND. The Sleep, Subjective Fatigue, and Sustained Attention of Commercial Airline Pilots during an International Pattern. *Chronobiology International* [online]. 2009, **23**(6), 1357-1362 [cit. 2020-11-12]. ISSN 0742-0528. Available at: doi:10.1080/07420520601085925

[22] ROACH, Gregory D., Renée M. A. PETRILLI, Drew DAWSON a Nicole LAMOND. Impact of Layover Length on Sleep, Subjective Fatigue Levels, and Sustained Attention of Long-Haul Airline Pilots. *Chronobiology International* [online]. 2012, **29**(5), 580-586 [cit. 2020-11-12]. ISSN 0742-0528. Available at: doi:10.3109/07420528.2012.675222

[23] POWELL, David, Mick B. SPENCER, David HOLLAND a Keith J. PETRIE. Fatigue in Two-Pilot Operations: Implications for Flight and Duty Time Limitations. *Aviation, Space, and Environmental Medicine* [online]. 2008, **79**(11), 1047-1050 [cit. 2020-11-12]. ISSN 00956562. Available at: doi:10.3357/ASEM.2362.2008

[24] SOCHA, Vladimi, Lenka HANAKOVA, Michal FREIGANG, Jakub KRAUS, Slobodan STOJIC, Lubos SOCHA a Peter HANAK. IMPACT OF PILOTS' TIREDNESS ON THE OUTCOME OF PSYCHOLOGICAL

TESTING. *Transport Problems* [online]. 2019, **14**(2), 55-67 [cit. 2020-11-12]. ISSN 1896-0596. Available at: doi:10.20858/tp.2019.14.2.5

[25] SUN, Jicheng, Shan CHENG, Jin MA, Kaiwen XIONG, Miao SU, Wendong HU a Richard James KEEGAN. Assessment of the static upright balance index and brain blood oxygen levels as parameters to evaluate pilot workload. *PLOS ONE* [online]. 2019, **14**(3) [cit. 2020-11-12]. ISSN 1932-6203. Available at: doi:10.1371/journal.pone.0214277

[26] GREELEY, H.P., E. FRIETS, J.P. WILSON, S. RAGHAVAN, J. PICONE a J. BERG. Detecting Fatigue From Voice Using Speech Recognition. In: *2006 IEEE International Symposium on Signal Processing and Information Technology* [online]. IEEE, 2006, 2006, s. 567-571 [cit. 2020-11-12]. ISBN 0-7803-9754-1. Available at: doi:10.1109/ISSPIT.2006.270865

[27] RUIZ, Robert, Philippe Plantin DE HUGUES a Claude LEGROS. Advanced Voice Analysis of Pilots to Detect Fatigue and Sleep Inertia. *Acta Acustica united with Acustica* [online]. 2010, **96**(3), 567-579 [cit. 2020-11-12]. ISSN 1610-1928. Available at: doi:10.3813/AAA.918307

[28] H.P. Greeley, P.G. Roma, M.M. Mallis, S.R. Hursh, A.M. Mead, T.E. Nesthus, Field Study Evaluation of Cepstrum Coefficient Speech Analysis for Fatigue in Aviation Cabin Crew. FAA report no. DOT/FAA/AM-13/19. Office of Aerospace Medicine, Washington, D.C. 2013.

[29] FERGUSON, Sally A., Nicole LAMOND, Katie KANDELAARS, Sarah M. JAY a Drew DAWSON. The Impact of Short, Irregular Sleep Opportunities at Sea on the Alertness of Marine Pilots Working Extended Hours. *Chronobiology International* [online]. 2009, **25**(2-3), 399-411 [cit. 2020-11-12]. ISSN 0742-0528. Available at: doi:10.1080/07420520802106819

[30] LOPEZ, Nadia, Fred H. PREVIC, Joseph FISCHER, Richard P. HEITZ a Randall W. ENGLE. Effects of sleep deprivation on cognitive performance by United States Air Force pilots. *Journal of Applied Research in Memory and Cognition* [online]. 2012, **1**(1), 27-33 [cit. 2020-11-12]. ISSN 22113681. Available at: doi:10.1016/j.jarmac.2011.10.002

[31] LAMOND, Nicole, Renée M. PETRILLI, Drew DAWSON a Gregory D. ROACH. Do Short International Layovers Allow Sufficient Opportunity for Pilots to Recover? *Chronobiology International* [online]. 2009, **23**(6), 1285-1294 [cit. 2020-11-12]. ISSN 0742-0528. Available at: doi:10.1080/07420520601062387

[32] WU, Edmond Q., X. Y. PENG, Caizhi Z. ZHANG, J. X. LIN a Richard S. F. SHENG. Pilots' Fatigue Status Recognition Using Deep Contractive Autoencoder Network. *IEEE Transactions on Instrumentation and Measurement* [online]. 2019, **68**(10), 3907-3919 [cit. 2020-11-12]. ISSN 0018-9456. Available at: doi:10.1109/TIM.2018.2885608

[33] HAN, Soo-Yeon, Jeong-Woo KIM a Seong-Whan LEE. Recognition of Pilot's Cognitive States based on Combination of Physiological Signals. In: *2019 7th International Winter Conference on Brain-Computer Interface (BCI)* [online]. IEEE, 2019, 2019, s. 1-5 [cit. 2020-11-12]. ISBN 978-1-5386-8116-9. Available at: doi:10.1109/IWW-BCI.2019.8737317

[34] DEHAIS, Frederic, Alban DUPRES, Gianluca DI FLUMERI, Kevin VERDIERE, Gianluca BORGHINI, Fabio BABILONI a Raphalle ROY. Monitoring Pilot's Cognitive Fatigue with Engagement Features in Simulated and Actual Flight Conditions Using an Hybrid fNIRS-EEG Passive BCI. In: *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* [online]. IEEE, 2018, 2018, s. 544-549 [cit. 2020-11-12]. ISBN 978-1-5386-6650-0. Available at: doi:10.1109/SMC.2018.00102

[35] WILSON, Nicholas, Bijay GURAGAIN, Ajay VERMA, Lewis ARCHER a Kouhyar TAVAKOLIAN. Blending Human and Machine: Feasibility of Measuring Fatigue Through the Aviation Headset. *Human Factors: The Journal of the Human Factors and Ergonomics Society* [online]. 2020, **62**(4), 553-564 [cit. 2020-11-12]. ISSN 0018-7208. Available at: doi:10.1177/0018720819849783

[36] CHENG, Shan, Jicheng SUN, Jin MA, Weitao DANG, Mengjun TANG, Duoduo HUI, Lili ZHANG a Wendong HU. Posturographic Balance’s Validity in Mental and Physical Fatigue Assessment Among Cadet Pilots. *Aerospace Medicine and Human Performance* [online]. 2018, **89**(11), 961-966 [cit. 2020-11-12]. ISSN 2375-6314. Available at: doi:10.3357/AMHP.5128.2018

[37] BINIAS, Bartosz, Dariusz MYSZOR, Michal NIEZABITOWSKI a Krzysztof A. CYRAN. Evaluation of alertness and mental fatigue among participants of simulated flight sessions. In: *2016 17th International Carpathian Control Conference (ICCC)* [online]. IEEE, 2016, 2016, s. 76-81 [cit. 2020-11-12]. ISBN 978-1-4673-8606-7. Available at: doi:10.1109/CarpathianCC.2016.7501070

[38] WU, Xu, Xiaoru WANYAN, Damin ZHUANG a Edward J. CIACCIO. Pilot's visual attention allocation modeling under fatigue. *Technology and Health Care* [online]. 2015, **23**(s2), S373-S381 [cit. 2020-11-12]. ISSN 09287329. Available at: doi:10.3233/THC-150974

[39] CALDWELL, John A., J. Lynn CALDWELL, David L. BROWN a Jennifer K. SMITH. 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* [online]. 2009, **16**(3), 163-181 [cit. 2020-11-12]. ISSN 0899-5605. Available at: doi:10.1207/s15327876mp1603_2

[40] WRIGHT, Nicola a Amanda MCGOWN. Vigilance on the civil flight deck: incidence of sleepiness and sleep during long-haul flights and associated changes in physiological parameters. *Ergonomics* [online]. 2010, **44**(1), 82-106 [cit. 2020-11-12]. ISSN 0014-0139. Available at: doi:10.1080/00140130150203893

[41] GUNDEL, Alexander, Jürgen DRESCHER, Hartmut MAAB, Alex SAMEL a Martin VEJVODA. Sleepiness of civil airline pilots during two consecutive night flights of extended duration. *Biological Psychology* [online]. 1995, **40**(1-2), 131-141 [cit. 2020-11-12]. ISSN 03010511. Available at: doi:10.1016/0301-0511(95)05107-4

[42] CALDWELL, JOHN A. a * PATRICIA A. LEDUC. Gender influences on performance, mood and recovery sleep in fatigued aviators. *Ergonomics* [online]. 1998, **41**(12), 1757-1770 [cit. 2020-11-12]. ISSN 0014-0139. Available at: doi:10.1080/001401398185956

[43] ZHANG, Liwei, Qianxiang ZHOU, Qingsong YIN a Zhongqi LIU. Assessment of Pilots Mental Fatigue Status with the Eye Movement Features. NUNES, Isabel L., ed. *Advances in Human Factors and Systems Interaction* [online]. Cham: Springer International Publishing, 2019, 2019-06-26, s. 146-155 [cit. 2020-11-12]. Advances in Intelligent Systems and Computing. ISBN 978-3-319-94333-6. Available at: doi:10.1007/978-3-319-94334-3_16