



**FAKULTA  
DOPRAVNÍ  
ČVUT V PRAZE**

---

Ústav letecké dopravy

## **Vliv únavy na výkonnost pilota**

DIZERTAČNÍ PRÁCE

Studijní program: Technika a technologie v dopravě a spojích

Studijní obor: Provoz a řízení letecké dopravy

Ing. Lenka Hanáková

---

Praha 2022

Praha, prosinec 2022

Dizertační práce byla vypracována v rámci doktorského studia na ČVUT v Praze, Ústavu letecké dopravy Fakulty dopravní.

Uchazeč: Ing. Lenka Hanáková  
Ústav letecké dopravy, České vysoké učení technické  
v Praze  
Horská 3, 128 03 Praha 2

Školitel: doc. Ing. Bc. Vladimír Socha, Ph.D.  
Ústav letecké dopravy, České vysoké učení technické  
v Praze  
Horská 3, 128 03 Praha 2

Školitel specialista: plk. doc. Ing. Jan Bořil, Ph.D.  
Katedra letectva, Univerzita obrany v Brně  
Kounicova 65, 662 10 Brno



# Abstrakt

Dizertační práce je zaměřena na studium vlivu únavy na výkonnost pilota. Aktuální stav sledování únavy v letectví je založen zejména na sběru subjektivních dat po letu, tj. primárně prostřednictvím dotazníkových šetření. Experimentálně se objevují snahy o kvantifikaci únavy s využitím fyziologických měření, tj. s využitím monitorace psychofyziologické kondice pilota. V tomto směru je však vše stále v experimentální rovině a měření se orientují pouze na kratší experimenty. Z tohoto důvodu je v práci představen koncept 24hodinových měření s cílem vytvoření proměnlivé hladiny únavy s ohledem na dobu bdělosti subjektu a s ohledem na denní dobu. V průběhu experimentu jsou prováděny simulované lety, v nichž jsou měřeny biosignály s primárním zaměřením na mozkovou a srdeční aktivitu subjektu a dále jsou sbírána letová data. V časových prostorech mezi jednotlivými lety jsou pak provedeny výkonnostní testy a dále sběr subjektivních dat. Projekt dizertační práce měl ambici vytvořit komplexní experiment s cílem studia vlivu únavy na výkonnost pilota. Výsledky naznačují, že využití fyziologických signálů může reflektovat hladinu únavy a současně, že trend poklesu výkonnosti pozorovaný na základě výkonnostních testů odpovídá trendu vycházejícím z fyziologických měření. Takováto studie by mohla přispět ke snahám o objektivizaci únavy a mohla by posloužit jako podklad při snahách o zvyšování bezpečnosti letecké dopravy v kontextu managementu únavy.

**Klíčová slova:** EEG, letectví, lidský faktor, psychofyziologické ukazatele, únava, variabilita srdečního rytmu, výkonnost, predikce

# Abstract

The thesis is focused on the study of the influence of fatigue on pilot performance. The current state of fatigue monitoring in aviation is mainly based on the collection of subjective data after flight, i.e. primarily through questionnaire surveys. Experimentally, there are attempts to quantify fatigue using physiological measurements, i.e. using monitoring of the pilot's psychophysiological condition. However, everything is still on an experimental level in this respect and measurements are oriented only to shorter experiments. For this reason, the concept of 24-hour measurements is introduced in this thesis in order to induce variable fatigue levels with respect to the subject's wakefulness time and with respect to the time of day. During the experiment, simulated flights are performed in which biosignals are measured with a primary focus on the subject's brain and heart activity, and flight data are also collected. Performance tests are then performed in the time gaps between flights, as well as the collection of subjective data. The thesis project had the ambition to create a comprehensive experiment to study the effect of fatigue on pilot performance. The results suggest that the use of physiological signals can reflect the level of fatigue and, at the same time, that the trend of performance decline observed from the performance tests is consistent with that based on physiological measurements. Such a study could contribute to efforts to objectify fatigue and could serve as a basis for efforts to improve air transport safety in the context of fatigue management.

**Keywords:** aviation, EEG, fatigue, heart rate variability, human factors, psychophysiological indicators, performance, prediction

# Poděkování

Na tomto místě bych ráda poděkovala všem lidem zapojeným do projektu souvisejícím s touto dizertační prací, ale také lidem, kteří vytvořili podmínky pro vznik této práce. V první řadě bych ráda poděkovala všem studentům, kteří se účastnili měření ve svém volném čase a dobrovolně trávili čas ve škole, za což jim byla odměnou výrazná spánková deprivace. Současně patří poděkování také všem studentům, kteří se na projektu spojeném s dizertační prací prodíleli při vypracování vlastních závěrečných prací a bez nichž by byla realizace tohoto výzkumu značně obtížnější.

Poděkování patří také kolegům z Ústavu letecké dopravy pro vytvoření podmínek pro realizaci experimentů a podporu při práci na projektu práce, specificky primárně doc. Jakubovi Krausovi. Dále pak za neustálou podporu zejména Ing. Stanislavovi Kušmírkovi, Ing. Kateřině Stuchlíkové, Dr. Romanovi Matyášovi, Ing. Michaele Kalivodové a samozřejmě Ing. Veronice Kasalové.

Za nárazové, podnětné konzultace děkuji taky Ing. Janu Štroblovi z BRAIN Teamu Fakulty biomedicínského inženýrství ČVUT v Praze, mimo jiné za to že mi vždy zvedl telefon, protože já sama bych si na jeho místě asi už přestala odpovídat. Jo a jsem první, promiň.

Největší poděkování patří doc. Vladimírovi Sochovi za všechno. Bez jeho zapojení by tento projekt vznikal a byl realizován velmi těžko. Současně bez jeho podpory by velmi těžko vznikala i tato práce. Děkuji za konzultace, podporu, mentoring, dohadování, diskuze nad kvantovou fyzikou a všechny nerdy jokes ~

Osobní poděkování patří mojí rodině za neustálou podporu a vytvoření podmínek pro studium. A také přátelům, kteří jsou vždycky poblíž se svým supportem a sklenicí vína.

## Čestné prohlášení

Prohlašuji, že jsem dizertační práci s názvem Vliv únavy na výkonnost pilota vypracovala samostatně a použila k tomu úplný výčet citací použitých pramenů, které uvádím v seznamu přiloženém k dizertační práci.

Nemám závažný důvod proti užití 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).

V Praze dne 20. prosince 2022

.....

*Podpis*

# Obsah

<b>Seznam obrázků</b>	<b>ix</b>
<b>Seznam tabulek</b>	<b>xi</b>
<b>Seznam použitých zkratk</b>	<b>xii</b>
<b>Úvod</b>	<b>14</b>
<b>1 Přehled současného stavu</b>	<b>15</b>
1.1 Únava v letectví . . . . .	16
1.2 Objektivní sledování únavy v letectví . . . . .	21
1.3 Shrnutí kapitoly . . . . .	24
<b>2 Projekt práce</b>	<b>25</b>
<b>3 Metodika</b>	<b>27</b>
3.1 Subjekty . . . . .	27
3.2 Průběh experimentu . . . . .	28
3.3 Letová data . . . . .	32
3.4 Fyzilogická měření . . . . .	34
3.4.1 Srdeční aktivita . . . . .	35
3.4.2 Mozková aktivita . . . . .	45
3.5 Pohybová aktivita . . . . .	49
3.6 Výkonnostní testy . . . . .	50
3.7 Subjektivní hodnocení . . . . .	53
3.7.1 Nasa Task Load Index . . . . .	54
3.7.2 Dotazník pro hodnocení únavy . . . . .	55
3.8 Klasifikace úrovně únavy . . . . .	58
<b>4 Výsledky</b>	<b>61</b>
4.1 Letová data . . . . .	61
4.2 Fyziologická měření . . . . .	62
4.2.1 Variabilita srdečního rytmu . . . . .	63
4.2.2 Elektrická aktivita mozku . . . . .	69
4.3 Subjektivní hodnocení . . . . .	73
4.4 Výkonnostní testy . . . . .	77
4.5 Klasifikace . . . . .	81

<b>5</b>	<b>Diskuze</b>	<b>84</b>
5.1	Variabilita srdečního rytmu . . . . .	84
5.2	Elektrická mozková aktivita . . . . .	86
5.3	Subjektivní hodnocení . . . . .	89
5.4	Výkonnostní testy . . . . .	90
5.5	Letová data . . . . .	91
5.6	Klasifikace . . . . .	91
	<b>Závěr</b>	<b>93</b>
	<b>Seznam použité literatury</b>	<b>96</b>
	<b>Přílohy</b>	<b>107</b>
A	Publikace 1 . . . . .	108
B	Publikace 2 . . . . .	115
C	Publikace 3 . . . . .	121
D	Publikace 4 . . . . .	135
E	Výsledky EKG . . . . .	142

# Seznam obrázků

3.1	Harmonogram prováděných měření . . . . .	28
3.2	Letová data – vybrané úseky pro následné hodnocení . . . . .	33
3.3	Záznam elektrické aktivity srdce . . . . .	35
3.4	Poincaré graf s vyznačenými parametry SD1 a SD2 . . . . .	42
3.5	Ukázka log-log grafu detrendované analýzy fluktuací . . . . .	45
3.6	Rozložení elektrod 1–16 na v helmě Ultracortex Mark IV . . . . .	46
3.7	Princip softwarové implementace výkonnostního testu Or-test . . . . .	52
3.8	Zobrazení zadání pro Or-test . . . . .	53
3.9	Softwarová implementace NASA Task Load Index - NTLx . . . . .	56
3.10	Subjektivní hodnocení únavy . . . . .	57
4.1	Grafické znázornění distribuce výsledků hodnocení letu . . . . .	62
4.2	Grafické znázornění distribuce výsledků časové analýzy variability srdečního rytmu u skupiny A, úsek 1 . . . . .	65
4.3	Grafické znázornění distribuce výsledků časové analýzy variability srdečního rytmu u skupiny B, úsek 1 . . . . .	66
4.4	Grafické znázornění distribuce výsledků časové analýzy variability srdečního rytmu u skupiny A, úsek 2 . . . . .	66
4.5	Grafické znázornění distribuce výsledků časové analýzy variability srdečního rytmu u skupiny B, úsek 2 . . . . .	67
4.6	Grafické znázornění distribuce výsledků frekvenční analýzy variability srdečního rytmu u skupiny A, úsek 1 . . . . .	68
4.7	Grafické znázornění distribuce výsledků frekvenční analýzy variability srdečního rytmu u skupiny B, úsek 1 . . . . .	68
4.8	Grafické znázornění distribuce výsledků nelineární analýzy variability srdečního rytmu u skupiny A, úsek 1 . . . . .	69
4.9	Proporční variabilita popsaná jednotlivými hlavními komponentami spolu se znázorněním kumulativní variability . . . . .	70
4.10	Bodový graf hlavních komponent – Score plot a grafy loadingů v závislosti na hlavních komponentách . . . . .	71
4.11	Vzdálenostní matice bodů cílové proměnné transformovaných pomocí PCA spolu s dendrogramem znázorňujícím hierarchické shlukování . . . . .	72
4.12	Grafické znázornění distribuce výsledků dotazníku NASA Task Load Index u skupiny A . . . . .	74

4.13 Grafické znázornění distribuce výsledků dotazníku NASA Task Load Index u skupiny B . . . . .	75
4.14 Grafické znázornění distribuce výsledků dotazníku zaměřeného na únavu u skupiny B . . . . .	76
4.15 Grafické znázornění distribuce výsledků jednoduchého reakčního času . . . . .	78
4.16 Grafické znázornění distribuce výsledků výběrového reakčního času	78
4.17 Bodový graf hlavních komponent - Score plot (OrTest) . . . . .	80
4.18 Grafické znázornění distribuce výsledků OrTestu . . . . .	81
4.19 Klasifikace měření s využitím neuronové sítě . . . . .	83



# Seznam tabulek

3.1	Parametry časové analýzy relevantní pro analýzu . . . . .	38
3.2	Parametry frekvenční analýzy dostupné v analytickém software. . .	41
3.3	Uspořádání dat výsledných parametrů frekvenční analýzy záznamů elektrické aktivity mozku . . . . .	48
4.1	Vnitroskupinové porovnání letových dat . . . . .	62
4.2	Vnitroskupinové porovnání jednotlivých parametrů variability srdečního rytmu . . . . .	64
4.3	Výkonové charakteristiky klasifikačních modelů . . . . .	82

# Seznam použitých zkratek

$\alpha, \alpha_1, \alpha_2$	škálovací exponenty detrendované analýzy fluktuací
4choice	čtyřvýběrový test reakčního času
ANS	autonomní nervový systém
ATCO	řídící letového provozu
DFA	analýza detrendovaných fluktuací
EEG	elektroencefalografie
EF	parametr NASA-TLX - úsilí
EKG	elektrokardiogram
FFT	rychlá Fourierova transformace - fast Fourier transform
FR	parametr NASA-TLX - frustrace
HF	výkon v pásmu vysokých frekvencí (0,15–0,4 Hz).
HR	průměrná tepová frekvence
HRV	variabilita srdečního rytmu - heart rate variability
HRVi (-)	trojúhelníkový index variability srdečního rytmu
IG	informační zisk - information gain
IMU	inerciální měřicí jednotka - inertial measurement units
LF	výkon v pásmu nízkých frekvencí (0,04–0,15 Hz).
LF/HF	poměr výkonů v pásmech nízkých a vysokých frekvencí.
LOI	linie identity, diagonála
M1–M8	měření 1 až 8
MD	parametr NASA-TLX - psychická náročnost
meanHR (bpm)	průměrná hodnota tepové frekvence
meanNN (ms)	průměrná hodnota NN intervalů
NASA-TLX	NASA Task Load Index (dotazník)
nHF	výkon v pásmu vysokých frekvencí (0,15–0,4 Hz) normalizovaný vůči celkovému výkonu spektra
nLF	výkon v pásmu nízkých frekvencí (0,04–0,15 Hz) normalizovaný vůči celkovému výkonu spektra
NN50	počet sousedících RR intervalů lišících se od sebe o více než 50 ms
NTLx	parametr NASA-TLX - celkové skóre/index
nULF	výkon v pásmu ultra nízkých frekvencí (<0,0033 Hz) normalizovaný vůči celkovému výkonu spektra
nVLF	výkon v pásmu velmi nízkých frekvencí (0,0033–0,04 Hz) normalizovaný vůči celkovému výkonu spektra
PC	hlavní komponenta z analýzy hlavních komponent

PCA	analýza hlavních komponent - principal component analysis
PD	parametr NASA-TLX - fyzická náročnost
PF	parametr NASA-TLX - výkonnost
pNN50 (%)	poměr NN50 k počtu všech RR intervalů
PSD	výkonová spektrální hustota - power spectral density
PVT	psychomotorický test bdělosti
rANOVA	analýza rozptylu pro opakovaná měření
RMSSD (ms)	druhá mocnina průměru druhých mocnin rozdílů mezi po sobě jdoucími NN intervaly (odvozeno z diferencí RR intervalů)
RT	reakční čas
sampEN	vzorková entropie
SD1	vedlejší poloosa Poincaré grafu - parametr popisující krátkodobou variabilitu srdečního rytmu
SD2	hlavní poloosa Poincaré grafu - parametr popisující dlouhodobou variabilitu srdečního rytmu
SDANN (ms)	směrodatná odchylka průměrného RR intervalu (obvykle z kratších záznamů)
SDNN (ms)	směrodatná odchylka popisující celkovou variaci v záznamu NN intervalů
SDNN index (ms)	střední hodnota směrodatné odchylky RR intervalů v pětiminutových intervalech
SRT	jednoduchý reakční čas - simple reaction time
ST	parametr upraveného dotazníku NASA-TLX pro hodnocení únavy - stres
TD	parametr NASA-TLX - časová náročnost
TINN (ms)	trojúhelníková interpolace NN intervalů
Total	celkový výkon frekvenčního spektra.
ULF	výkon v pásmu ultra nízkých frekvencí (<0,0033 Hz).
UST	ultra-krátkodobá variabilita srdečního rytmu, pro záznamy kratší než 5 minut
VLF	výkon v pásmu velmi nízkých frekvencí (0,0033—0,04 Hz).

# Úvod

Problematice únavy leteckého personálu je v leteckém prostředí věnována značná část pozornosti, a to zejména s ohledem na neustálou snahu o zvyšování bezpečnosti letecké dopravy. Vzhledem k tomu, že majoritní část leteckých nehod je zapříčiněna lidským faktorem, resp. leteckým personálem, je právě téma lidského faktoru a jeho selhání aktuálním tématem a předmětem mnoha výzkumů.

Bezpečnost letu je, mimo jiné, závislá na pilotáži, tj. nepříznivý stav pilota je pro let limitací. Výkon pilota je pak zejména ovlivněn jeho psychickými procesy, které jsou závislé nejen od profesní přípravy a zkušeností pilota, ale také od jeho aktuální psychofyziologické kondice. Je tedy předpokládáno, že únava hraje významnou roli při vykonávání pilotovy činnosti a limituje jeho výkon.

Únava jako taková bývá u profesionálních pilotů sledována, avšak nedochází k její objektivizaci, tj. únava je sledována na základě subjektivního hodnocení. S ohledem na subjektivní hodnocení únavy je pak obtížné popsat vliv únavy jako takové na pilotovu výkonnost. Pro účely objektivního posouzení vlivu únavy na výkonnost je třeba objektivního hodnocení, tj. jsou nezbytná taková data, která přímo souvisí s únavou a pomocí nichž je možné únavu co nejlépe objektivizovat a tato data dále vztáhnout k letovým datům.

Tato dizertační práce se zabývá právě vlivem únavy na výkonnost pilota, a to na základě subjektivního hodnocení únavy prostřednictvím standardizovaného dotazníkového šetření a prostřednictvím reálných měření fyziologických, výkonnostních a letových dat jako ukazatelů únavy v průběhu simulovaných letů.

Práce je psána jako monografie, věnovaná jednomu předmětu výzkumu, a to únavě. Dílčí výsledky nebo výsledky související s tématem práce však byly publikovány v průběhu zpracování práce a jsou uvedeny v přílohách a referovány v textu. Práce tedy není komentovaným souborem publikovaných článků a získaných výsledků, v každém případě tyto informace jsou uváděny pro doplnění kontextu samotné práce.

# 1. Přehled současného stavu

Do dnešního dne není objasněno, proč spíme. Je však jisté, že spánek je pro náš život nezbytný a jeho nedostatek může vést až ke smrti (Kreitzman & Foster, 2011). Potřeba spánku je pak individuální a neexistuje jasný konsensus, kolik spánku člověk skutečně potřebuje. Obecně je pak přijímán fakt, že potřeba spánku je přibližně 8 hodin (Missoni et al., 2009). Na druhou stranu pak v případě jedinců pravidelně spících 4 hodiny a méně bylo pozorováno vyšší riziko úmrtí s ohledem na věk, životní styl a zdravotní stav subjektů (Kripke et al., 2002).

V případě nedostatku spánku lze však pozorovat také krátkodobé projevy. Jedná se zejména o projevy únavy. Únava není striktně definována, resp. definic je mnoho, stejně jako forem únavy (Bendak & Rashid, 2020). Jedná se jak o stavy krátkodobé únavy, tak dlouhodobé až chronické únavy. Čeština pak neposkytuje takový rozsah pojmů, které jsou využívány v anglické literatuře a definice únavy je ještě o to náročnější. Pro účely této práce je tedy únava primárně vnímána jako krátkodobá únava (tzv. tiredness nebo sleepiness) pramenící ze spánkové deprivace, nedostatečného odpočinku a psychické zátěže.

Cyklus bdění spánku je řízen našimi vnitřními hodinami, resp. přibližně 24hodinovým cirkadiánním rytmem (circa = okolo, během; dies = den) (Kreitzman & Foster, 2011). Pacemakerem cirkadiánního rytmu je hladina hormonu melatoninu, téz označovaného jako hormon spánku. Synchronizátorem tohoto rytmu je pak sluneční svit. To tedy znamená, že v závislosti na denní fázi, dané intenzitou přirozeného světla, jsou tělesné procesy ovlivňovány v první řadě úrovní melatoninu (Cajochen et al., 2003). Jedním z takto ovlivňovaných procesů je také bdění a spánek. V praxi to tedy znamená, že v určitých denních dobách je člověk více náchylný k spánku, v jiných pak disponuje vyšší výkonností, úrovní kognitivních funkcí apod. (Goel et al., 2013).

Z hlediska cirkadiánního rytmu dochází k největšímu útlumu v ranních hodinách. Toto období je označováno jako cirkadiánní útlum nebo okno cirkadiánního útlumu (window of circadian low), jež je nejčastěji definováno jako čas mezi 2 a 6 hodinou ranní. V této denní době je tedy člověk nejvíce náchylný k spánku. Právě kombinace cirkadiánního útlumu a únavy může být poměrně nebezpečná, a to s ohledem na skutečnost, že oba tyto stavy negativně ovlivňují ostražitost a výkonnost (Dongen & Hursh, 2011).

Prostředí letectví pak přináší specifické podmínky s ohledem na charakter výkonu povolání jednotlivých leteckých odborníků, zejména pak pilotů. Dlouhé lety, noční lety, nepravidelný režim, překračování časových zón a následné

narušení cirkadiánního rytmu patří mezi jedny z významných příčin únavy (Bendak & Rashid, 2020).

## 1.1 Únava v letectví

Samotná problematika únavy pilota se dá v jistém smyslu vysledovat až k počátku dvacátého století, kdy byl proveden první transatlantický let, vykonaný Charlesem Lindberghem. Jeho let v té době trval 33,5 hodiny a Charles Lindbergh po přistání promluvil o únavě, snaze o její eliminaci a vlivu na jeho výkon, resp. o ztracení kontroly (Lindbergh, 2003).

Od tohoto historického letu došlo k poměrně významnému rozvoji v oblasti letectví a lety na velké vzdálenosti nejsou v dnešní době ničím nestandardním. Problematickým faktorem však stále zůstává člověk, který se z fyziologického hlediska prakticky nezměnil. S ohledem na tento fakt je zřejmé, že snaha o zvyšování bezpečnosti letu musí vést přes sledování letové posádky a její fyziologické (resp. psychofyziologické) kondice.

Jak již bylo zmíněno, únavu lze jen těžko jednoduše definovat. Obecně však lze říci, že únava jako taková je kombinací fyzického a psychického nedostatku. Z hlediska safety v letectví ICAO definuje únavu jako (International Civil Aviation Organization, 2014): *„Fyziologický stav snížené duševní nebo fyzické výkonnosti, který je důsledkem nedostatečného spánku nebo prodloužené bdělosti, cirkadiánního rytmu nebo pracovní zátěže (duševní a/nebo fyzická aktivita), který může ohrozit bdělost a schopnost člena posádky bezpečně provozovat letadlo nebo plnit povinnosti související s bezpečností.“*

Únava je tedy komplexní, zahrnující únavu fyzickou a psychickou. Oba tyto podtypy únavy souvisí s výkonností - únava fyzická je neschopnost zachovat výkonnost svalů/svalových skupin, resp. projevuje se snižováním svalové výkonnosti. Únava psychická je pak neschopnost zachovat stejnou úroveň kognitivní výkonnosti a je důsledkem působení cirkadiánních rytmů, nedostatečného nebo nekvalitního spánku, příp. pásmové nemoci, doby strávené ve službě, působení nadměrné psychické zátěže příp. stresu apod. S ohledem na tyto faktory je hlavní fokus výzkumů zaměřen zejména na únavu psychickou. Ta se dále může negativně projevovat zejména při vzniku nestandardních situací, kdy dochází ke zvýšenému požadavku na zapojení kognitivních funkcí (Dongen & Hursh, 2011; Reis et al., 2013).

Letectví je tedy specifickým prostředím, v němž působí mnoho z výše zmíněných faktorů ovlivňujících únavu. Posádky letadel často pracují ve směnném provozu a jsou vystaveny časovým posunům, čímž dochází k narušení

spánkového cyklu. Kromě toho je nutné vzít v úvahu nejen vliv nedostatečného spánku a celkové doby bdělosti, ale také samotného střídání dne a noci s ohledem na cirkadiánní rytmus těla, tj. vzít v úvahu také vliv nočních letů na posádku a její únavu. Je tedy zřejmé, že únava ovlivňuje výkonnost, což má (nejen) v oblasti letectví negativní důsledky na její bezpečnost (Caldwell, 2005).

Únava je pak v moderní společnosti vnímána jako naprosto běžná záležitost (Dawson et al., 2011). Tato skutečnost je problematická s ohledem na skutečnost, že negativní vlivy únavy mohou vést až k vážným chybám a následně nehodám (Caldwell et al., 2009). Z hlediska dopravy je únava chápána jako příčina nehod, které lze nejlépe předcházet (Akerstedt, 2000).

Sledování únavy v letectví, resp. u komerčních pilotů je stále založeno zejména na subjektivních kvantifikátorech, tj. únava je sledována zejména prostřednictvím dotazníkového šetření (Gregory et al., 2010; Bourgeois-Bougrine et al., 2003; Powell et al., 2007). Výsledky takovýchto šetření poukazují na fakt, že piloti reportují snížení výkonnosti právě z důvodu únavy (např. dle výsledků Gregory et al. (2010) reportuje vliv únavy na výkon více než 84 % dotazovaných pilotů). Tyto výsledky nepochybně mohou sloužit jako podklad pro další studie v této oblasti, ale stále se jedná pouze o subjektivní data. Obdobná šetření však byla prvními snahami v této oblasti a přispěla ke zvýšenému sledování letového personálu z hlediska únavy.

Vzhledem ke skutečnosti, že únava je brána jako hlavní hazard z hlediska lidského faktoru (FRMS Task Force, 2011), dostala se tato problematika také do zájmu leteckých autorit. Historicky byla únava, resp. rizika spojená s únavou regulována pouze na základě Flight and Duty Time Limitations (Omezení doby letové služby a služby a požadavky na dobu odpočinku, FTL), které upravují maximální dobu služby a minimální dobu odpočinku. Předepsané limity však v mnoha případech nebyly vyvíjeny společně s narůstajícími poznatky a tedy začaly být nedostačující. V rámci plánování je totiž např. třeba vzít v úvahu právě vliv denní doby na člověka. Je prokázáno, že stejná délka odpočinku přes noc a přes den nepřináší stejný efekt, tj. přes den je odpočinek méně účinný (Bendak, 2003).

Z těchto důvodů se začalo od standardního plánování přecházet na plánování založené na důkazech (datech) (Civil Aviation Safety Authority, 2010). Pro uvedené účely je do letectví zaváděn Fatigue Risk Management System (Systém řízení rizik spojených s únavou, FRMS). Jako první byl v nějaké míře zaváděn v Austrálii. Následně došlo k rozšíření podpoře tohoto systému. Pokyny pro implementaci FRMS plně rozvinula CASA (Civil Aviation Safety Authority, nařízení CAO 48.1 Instrument 2013) (Civil Aviation Safety Authority, 2013), spojenými silami IATA (International Air Transport Association), ICAO

(International Civil Aviation Organization) a IFALPA (International Federation of Air Line Pilots' Associations) vznikla skupina FRMS Task Force, která vydala (a nadále aktualizuje) pokyny pro zavádění FRMS u leteckých společností (FRMS Task Force, 2011).

Aktuálně tedy existují pokyny pro zavádění FRMS a systém jako takový slouží jako podpora k existujícím lokálním FTL (Civil Aviation Safety Authority, 2014). Dle FRMS Task Force (2011) je pro FRMS definováno 5 úrovní kolekce dat. První úroveň je založena pouze na subjektivním hodnocení pomocí dotazníkového šetření (standardizovaného) a jedná se o minimální požadovaná data (IATA FRMS Task Force, 2014, 2015). Úroveň 2 a 3 přidávají k základním dotazníkům také subjektivní hodnocení vztažené k spánku a spánkovému cyklu a dále data vztažená k denním a pracovním návykům subjektu. Úroveň 4 a 5 pak rozšiřují subjektivní data, která mohou být zkreslena názorem, interpretací či vnímáním subjektu, o data objektivní, získaná měřeními (IATA FRMS Task Force, 2015). Úroveň 4 tedy počítá s daty získanými aktigrafickým měřením, tj. objektivním měřením spánku (IATA FRMS Task Force, 2015; Ancoli-Israel et al., 2003). Úroveň 5 pak představuje maximální baterii testování, zahrnující navíc oproti předchozí úrovni objektivní měření výkonu, založené na tzv. Psychomotor Vigilance Task (PVT) (IATA FRMS Task Force, 2015), tedy na jednoduchém testu reakčního času (Dinges & Powell, 1985).

Součástí FRMS jsou také biomatematické modely, které nejsou brány jako základní součásti systému, ale jako doplňkové prostředky pomáhající v predikci výkonnostních úrovní, únavy apod., které jsou založeny na nasbíraných datech souvisejících se spánkovým cyklem, mentální zátěží nebo denní dobou. Ačkoli tedy dochází k objektivizaci v otázce únavy v letectví, i tyto modely a celý systém FRMS mají stále několik limitací (Civil Aviation Safety Authority, 2010).

Limitace vycházejí zejména z charakteristiky užívaného systému - systém, resp. modely, vychází z dat nasbíraných na konečném vzorku populace. Není tedy brána v úvahu individualita jedince, stejně jako nejsou kompletně popsány všechny fyziologické faktory související s únavou. Část vstupních dat je dále závislá na subjektivním hodnocení. Z uvedeného je tedy patrné, že modely jsou vytvořeny pro „průměrného“ člověka na základě dat od „průměrné“ skupiny subjektů. Není tedy zřejmé, zda jsou tyto poznatky aplikovatelné globálně a lze předpokládat, že bude existovat jisté množství odchylek (Civil Aviation Safety Authority, 2014). Individualita subjektu je závislá na demografických faktorech (Di Milia et al., 2011), ale také na individuálním zdravotním stavu (např. spánkové poruchy) (Smolensky et al., 2011). Kromě stabilních individuálních faktorů je pak třeba brát v úvahu také



krátkodobé faktory mající vliv na únavu jedince, jako je např. alergie. Dále je pak třeba vzít v úvahu faktory vyplývající z osobního života jedince, jako je narušení spánku malým dítětem, stres z rozvodu apod. (Civil Aviation Safety Authority, 2014).

Co se týká modelů samotných, aktuálně jich existuje několik (Van Dongen, 2004; Van Dongen et al., 2007). Z výzkumu však vyplývá, že je třeba dalších studií v oblasti safety a únavy na jejichž základě bude možno vyvinout nové, rozšířené postupy pro účely objektivního hodnocení únavy (Williamson et al., 2011). Nevhodné využití, příp. interpretace výstupů by v extrémním případě mohla vést dokonce ke snížení bezpečnosti (Civil Aviation Safety Authority, 2010). Kromě toho nejsou modely schopny jednoznačně odpovědět na otázky spojené s bezpečností. Je to způsobeno tím, že modely se ve skutečnosti nesnaží definovat nebo nalézt akceptovatelnou úroveň únavy. Lze je tedy využít jako oporu, ale nelze získat odpověď ve smyslu „situace je akceptovatelná/bezpečná“ (Civil Aviation Safety Authority, 2014).

Je tedy zřejmé, že ačkoli mají snahy o sledování únavy v letectví narůstající tendenci, stále neexistují konkrétní postupy, resp. návody, jak únavu sledovat. Částečně lze vycházet z předešlých výzkumů (které byly také oporou při tvorbě konceptu FRMS) a zavedených konceptů. Jako první možnost přicházejí v úvahu právě subjektivní data, tj. dotazníková šetření. Pro účely sledování zátěže byl vytvořen dotazník NASA Task Load Index (NASA-TLX), který je aktuálně standardně využíván nejen v letectví, pro které byl původně určen, ale i v řadě dalších odvětví. Dotazník je založen na váhování „proměnných“ spojených se zátěží a je navržen tak, aby byl schopen co nejlépe pokrýt zátěž bez ohledu na pracovní pozici zpovídané osoby (Hart, 2006). Dotazník je široce využíván a slouží jako základní nástroj při sledování únavy a zátěže (nejen pilotů) v mnoha studiích (Hart, 2006; Svensson et al., 1997; Samel et al., 1997).

Velmi častým přístupem kolekce subjektivních dat o únavě je také využití tzv. Karolinka Sleepiness Scale, což je 9bodová stupnice, přičemž na úrovni 1 je subjekt extrémně ostražitý a na úrovni 9 je subjekt extrémně unavený, neschopný zůstat bdělý (Shahid et al., 2011). Dalšími možnostmi subjektivního hodnocení únavy v letectví mohou být Samn–Perelli Fatigue Scale (Samn & Perelli, 1982), Epworth Sleepiness Scale (Johns, 1991), Stanford Sleepiness Scale (Hoddes et al., 1972) a další (Bendak & Rashid, 2020; Zhuzhgova, 2020).

Jak již bylo uvedeno, subjektivní hodnocení nejsou z principu schopna dostatečně objektivizovat danou problematiku a tedy mohou poskytovat nedůvěryhodné údaje, které je nezbytné brát na zřetel (Tremaine et al., 2010). Je tedy všeobecná snaha o hledání vhodných, objektivních metod hodnocení a sledování únavy. Jedná se zejména o měření fyziologických funkcí a následně

zpracování a hodnocení získaných biosignálů. Z hlediska fyziologie se pak nabízí větší množství možností měření a s ohledem na komplexitu problematiky únavy je hledání vhodných parametrů dosud výzvou. V oblasti letectví, a se zaměřením primárně na piloty, jsou pak možnosti měření dále limitovány s ohledem na specifika tohoto povolání a možnosti měření v pilotní kabině v průběhu letu. Je tedy nezbytné volit takové metody, které jsou co možná nejvíce ovlivňovány únavou a jejichž měření může probíhat v poměrně malých prostorech tak, aby jakýmkoli způsobem nebránilo nebo neovlivňovalo možnosti pilota vykonávat veškeré potřebné úkony. Nejen z těchto důvodů je jako základní měřený biosignál využíván záznam srdeční aktivity - EKG (elektrogardiogram) (Wilson et al., 2007a; Watson, 2001; Borghini et al., 2014).

Monitorování srdeční aktivity je poměrně snadné. V současné době existuje řada komerčních řešení, využitelných v této oblasti. Pro tyto účely jsou používány zejména systémy využívající jednosvodové EKG, které je nejčastěji implementováno do komerčních hrudních pásů. Dále se objevují řešení schopná snímat tepovou frekvenci pouze pomocí náramkových zařízení. Náramková zařízení jsou pak většinou využívána ne ke sledování srdeční aktivity, ale pro aktigrafická měření. I tato měření jsou součástí FRMS (od úrovně 4) a jsou využívána ve studiích zaměřených na problematiku únavy pilotů (Roach et al., 2011).

Měření EKG a aktigrafická měření jsou nejjednodušší, ne však jedinou využívanou možností. Experimentálně jsou také sledovány pohyby očí (EOG, elektrookulografie) a mozková aktivita (EEG, elektroencefalografie) (Thomas et al., 2015; Wilson et al., 2007b). Problematika sledování pohybu očí prostřednictvím EOG vyžaduje upevnění elektrod do oblasti obličeje, což společně s vedenou kabeláží nemusí být pro subjekt příjemné a může být omezující. Pro účely měření EEG je pak nezbytné umístění elektrod na hlavu subjektu. Pro tyto účely jsou využívány EEG čepice, které zajišťují přesné rozmístění elektrod. V současné době také odpadá problematika využití vodivé pasty/gelu, jejichž využití bylo pro subjekty mírně obtěžující, a je možné využít suchých elektrod. Uvedené signály jsou však stále využívány zejména experimentálně (Thomas et al., 2015; Wilson et al., 2007b; Caldwell, 2005).

Problematika fyziologických měření je dále rozšířena o výběr vhodných parametrů zvolených biosignálů, resp. o metody jejich hodnocení. Metodika hodnocení těchto signálů je nejlépe zpracována pro EKG, a to s ohledem na jeho široké využití v mnohých oblastech. Obecně však lze pro všechny signály nalézt jisté návody v jiných oblastech, než je letectví, v nichž jsou signály více využívány. Metody pak většinou zahrnují přístupy standardně využívané (pro EKG se jedná zejména o časovou a frekvenční analýzu variability srdečního

rytmu) a přístupy experimentální (např. nelineární metody) (Malik et al., 1996; Sharma, 2009).

Při snaze o sledování výkonu, resp. vlivu únavy na výkonnost, bývají využívány psychomotorické testy bdělosti (PVT, psychomotor vigilance test), tj. testy založené na reakci na předložený podnět, s nimiž počítá 5. úroveň FRMS (IATA FRMS Task Force, 2015). Tyto testy v základní konfiguraci měří pouze reakční čas jedince, např. stlačení tlačítka při rozsvícení diody apod.

## 1.2 Objektivní sledování únavy v letectví

Právě sledování výkonnosti na základě reakčního času, tedy prostřednictvím PVT jsou jednou z doporučených podpůrných metod pro sledování únavy a jsou považovány jako zlatý standard při sledování únavy (Arsintescu, 2017; Arsintescu et al., 2019). Původní design tohoto testování vychází z tzv. PVT-192 představeného v roce 1985 (Dinges & Powell, 1985). V průběhu let došlo k vytvoření a validace mnoha dalších řešení (Thorne et al., 2005; Lamond et al., 2005). Jedno z nejnovějších řešení, NASA-PVT, pak bylo představeno Národním úřadem pro letectví a vesmír (NASA, National Aeronautics and Space Administration) a jeho užití bylo následně validováno pro studium únavy při akutního omezení spánku (Arsintescu et al., 2019).

Bez ohledu na konkrétní design testu nebo přímo vytvořené zařízení je však princip vždy stejný. Subjektu je prezentován nějaký stimulus, primárně optický, na nějž je očekávána co nejrychlejší reakce. Jde tedy o měření celkového reakčního času (Socha et al., 2018) s tím, že stimuly jsou subjektu prezentovány opakovaně v pseudonáhodných rozestupech po nějaký stanovený časový interval. S ohledem na design studie se pak využívá celkové trvání testu po dobu 5–20 minut (Arsintescu et al., 2019; Veksler & Gunzelmann, 2018) nebo fixní počet opakování úlohy (Socha et al., 2020). Z hlediska samotných řešení jsou pak dostupná samostatně stojící zařízení (Thorne et al., 2005; Lamond et al., 2005; Socha et al., 2020), aplikace pro dotyková zařízení (iPad apod.) (Arsintescu, 2017) nebo jednoduchý software využitelný na vlastním počítači (Borchert, 2016, 2017).

S ohledem na jednoduchost a dostupnost těchto testů jsou PVT velmi častým prostředkem pro sledování únavy. Samotná objektivizace únavy však nelze postavit pouze na těchto jednoduchých testech. Jejich limitace souvisí zejména se skutečností, že tyto testy není možné aplikovat v průběhu vykonávání činnosti. Z hlediska letu je lze tedy využít pouze před nebo po letu. Z těchto důvodů je většinou měřena kombinace různých dat/signálů. Většinou

se jedná o signály získané monitorováním pilota v průběhu letu, a to jak data fyziologická, tak letová.

Jak již bylo zmíněno v předchozí části, jednou z nejhojněji užívaných metod, a to i s ohledem na jednoduchost její aplikace, je sledování srdečního rytmu (Wilson et al., 2007a; Watson, 2001; Borghini et al., 2014). Vhodnost využití takového signálu pak souvisí s fyziologickou odezvou těla. Procesy v těle jsou primárně řízeny autonomním nervovým systémem (ANS) (Kittnar et al., 2011). Stav jako stres, zátěž nebo právě únava ANS ovlivňují, resp. v těchto stavech se na jiné úrovni zapojují jednotlivé větve ANS – sympatikus a parasymptakus (Patel et al., 2011; Horne & Reyner, 1995). Tyto stavy tedy ovlivňují tzv. sympatovagální rovnováhu. Samotný ANS pak moduluje srdeční akci. Na základě výzkumů bylo zjištěno, že aktivita ANS, resp. aktivita jeho jednotlivých větví a popř. sympatovagální rovnováha, jsou hodnotitelné na základě analýzy srdeční činnosti (Horne & Reyner, 1995; Malik et al., 1996).

Při sledování srdeční aktivity je pořizován záznam EKG, elektrokardiogram, prezentující časovou změnu elektrického potenciálu. Výstupem takového měření je tedy EKG křivka, tj. hodnota elektrického potenciálu v čase, v níž se, u zdravého subjektu, vyskytují typické vlny a kmity. Jednotlivé vlny, příp. kmity, reprezentují šíření elektrického signálu přes srdeční sval, počínaje depolarizací síní (P vlna) a konče repolarizací komor (T vlna). Významným prvkem signálu je pak QRS komplex, reprezentující depolarizaci komor a zahrnující kmity Q, R a S (Martin et al., 2012).

Jak již bylo zmíněno, srdeční rytmus je modulován autonomním nervovým systémem (ANS), tj. sympatikem a parasympatikem. Sympatická aktivita převládá ve stresových situacích, mobilizuje systém na hrozící nebezpečí. Jeho vlivem tedy dochází k akceleraci srdeční činnosti, tj. dochází k nárůstu tepové frekvence. Parasympatikus je pak jeho antagonistou, parasympatická aktivita tedy převládá v klidu a bezpečí. V případě srdeční aktivity dochází působením parasympatiku ke snížení tepové frekvence. V případě únavy je tedy předpokládána dominantní role parasympatické části ANS (Kittnar et al., 2011).

Prostřednictvím sledování srdeční aktivity je pak možné kvantifikovat parasympatickou kontrolu nad srdcem, která může vznikat v důsledku únavy. Parasympaticky je srdeční aktivita modulována prostřednictvím *nervus vagus* (bloudivý nerv). Autonomní regulace srdeční činnosti se projevuje změnou variability srdečního rytmu (HRV, heart rate variability). Tato změna je úzce spjata s respirací. Při nádechu dochází ke zrychlení srdeční činnosti, po výdechu k jejímu zpomalení. Zapojení parasympatiku ovlivňuje také dýchání - při větším zapojení parasympatiku dochází k výraznějšímu zrychlení/zpomalení srdeční

činnosti, čímž vznikají více variabilní intervaly mezi jednotlivými údery srdce. Parasimpatikus má tedy vliv na HRV, s jeho vyšším zapojením dochází k narůstání variability srdečního rytmu (Seegerstrom & Nes, 2007).

Tohoto faktu je využíváno právě při hodnocení únavy. V letectví však v této oblasti není dostupné velké množství výzkumu - na základě meta-analýzy provedené na konci roku 2020 je v oblasti hodnocení únavy prostřednictvím fyziologických dat dostupných 12 studií (Zhuzhgova, 2020). Část těchto studií pak nedisponuje relevantními výsledky, některé studie byly provedeny na velmi malém vzorku subjektů (5 a méně subjektů), pouze 4 studie využívala sledování srdeční akce (Zhuzhgova, 2020). Na druhou stranu v pozemní dopravě je na výzkum únavy kladen velký důraz a tedy je dostupné větší množství studií, a to i s ohledem na snazší provedení výzkumu. V tomto kontextu je tedy vhodné se poučit zejména z výzkumů prováděných na řidičích a následně aplikovat danou metodologii v letectví. Právě výzkumy z pozemní dopravy potvrzují předpoklady o nárůstu HRV v ospalosti a při dlouhé monotónní jízdě (Horne & Reyner, 1995).

Hodnocení HRV pak může probíhat v několika úrovních. Standardem je analýza v časové a frekvenční oblasti (Malik et al., 1996). Stále více jsou však využívány méně standardní metody, a to zejména s ohledem na charakter hodnoceného signálu. Jedná se o metody nelineární analýzy dat, případně o analýzu v časově-frekvenční oblasti. Při využití takovéto analýzy je předpokládáno získání více informací o chování celého systému, některé z nich jsou však obtížně aplikovatelné na delší signály. Z těchto důvodů většina studií využívá právě standardní analýzu v časové oblasti a spektrální analýzu, tedy analýzu ve frekvenční oblasti.

Z hlediska časové oblasti se analýza opírá zejména o základní statistiku, tedy průměry, směrodatné odchylky apod. a metody geometrické (Malik et al., 1996). Z hlediska významnosti však většina studií ukazuje zejména na vhodnost využití spektrální (frekvenční) analýzy. V rámci techniky dochází k vytvoření frekvenčního spektra a jeho následnému rozdělení do tří specifických pásem - pásmo velmi nízkých frekvencí (VLF, very-low frequencies band), pásmo nízkých frekvencí (LF, low frequencies band) a pásmo vysokých frekvencí (HF, very-high frequencies band). V těchto pásmech je pak spočítán výkon, hodnocen je výkon zejména v LF a HF, případně jejich poměr (Malik et al., 1996).

Výsledky jednotlivých studií jsou však limitované a často si oponují. V případě některých výzkumů je tedy únava spojována právě s předpokládaným nárůstem HRV (Horne & Reyner, 1995). Při pohledu na výsledky jiných studií však autoři došli k závěru, že v důsledku únavy dochází k nárůstu výkonu v LF, tedy většímu zapojení sympatiku, což se také projevuje poklesem HRV (Apparies et al., 1998).

Úroveň bdělosti a ostražitosti ovlivňuje také naší mozkovou aktivitu. Z hlediska hodnocení EEG jsou rozlišovány 4 typy základních vln. Mezi základní vlny (rytmy) patří vlny delta, théta, alfa a beta. Delta vlny jsou nejpomalejší EEG vlny, vyskytující se při hlubokém spánku. Jako delta vlny jsou označovány vlny o frekvenci v pásmu 0,5–4 Hz. Theta vlny mají frekvenci v pásmu 4–8 Hz a je udáváno, že se vyskytují zejména u dětí, ale také u dospělých ve stresu. Alfa vlny se vyskytují u dospělých, a to v relaxované, bdělém stavu. Tyto vlny mají frekvenci v rozsahu 8–13 Hz. Beta vlny jsou ovlivňovány mentální aktivitou a jsou to vlny o frekvenci v rozsahu 13–30 Hz (Webster, 2009; Shaker, 2006).

Hodnocení EEG ve studiích spojených s únavou nebývá v leteckém prostředí hojně využíváno, a to i s ohledem na náročnost měření a zpracování tohoto signálu. Z hlediska dopravního výzkumu v oblasti pak lze nalézt několik studií věnující se tomuto signálu u řidičů pozemních dopravních prostředků. Z hlediska sledování únavy se pak studie zabývají zejména vlnami théta a alfa, případně beta. Zejména pak poměr výkonu ve frekvenčních pásmech théta a alfa (Boyle et al., 2008) se ukázal jako parametr, který dokáže indikovat změny v bdělosti (ostražitosti) řidičů (Cajochen et al., 1995). Některé studie však naznačují také užitečnost sledování pásma delta (Wright & McGown, 2001; Wilson et al., 2007b).

### 1.3 Shrnutí kapitoly

Z výše uvedeného je patrná snaha o zvyšování bezpečnosti letecké dopravy s ohledem na lidský faktor, resp. na únavu letového personálu. Ačkoli je do komerčního letectví zaváděn FRMS, který není založen čistě na subjektivním hodnocení a pracuje také s biomatematickými modely založenými na sesbíraných datech, stále existují limitace tohoto přístupu. Jako významná se jeví zejména individualita jedince, se kterou není ve FRMS příliš počítáno. Pro tyto účely se zdá jako vhodné sledování únavy, resp. její hladiny, přímo v průběhu letu a to na základě objektivně sesbíraných dat.

Předmětem práce je tedy studium únavy založené na objektivních datech, jejichž kolekce aktuálně není standardním procesem a jedná se spíše o experimentální případy. Data byla získána v průběhu simulovaných letů na Ústavu letecké dopravy, FD ČVUT.

## 2. Projekt práce

V práci bylo při stanovení cílů vycházeno ze současných limitací aktuálně využívaných postupů a dostupných studií. Parciální cíle byly zvoleny tak, aby bylo možné co nejlépe naplnit hlavní cíle práce, tj. studium únavy pilotů s ohledem na jejich výkon založené na objektivních datech. Na základě stanovených cílů pak byly formulovány výzkumné otázky a hypotézy.

### Cíle

**Návrh metodologie měření:** Analýzou současného stavu a z ní vycházejících poznatků bude navržen průběh měření, způsob kolekce dat a jejich charakter. Měření budou probíhat primárně v prostorách Ústavu letecké dopravy, FD ČVUT a měřeními subjekty budou studenti bakalářského studijního oboru Profesionální pilot.

**Návrh metod pro hodnocení dat:** Na základě naměřených signálů a jejich charakteristiky budou vybrány vhodné parametry. Tyto zvolené parametry budou dále hodnoceny pro účely studie únavy pilotů. Pro možnosti porovnání s tradičními metodami bude dále kolektován reakční čas subjektů a také subjektivní data v podobě dotazníku Nasa Task Load Index. Metodika hodnocení uvedených dat bude vycházet ze standardně užívaných metod. Objektivní data budou dále hodnocena také metodami, které nejsou standardně užívány (Rekurentní kvantifikační analýza, Pointcaré graf).

**Zpracování a analýza naměřených dat:** Data získaná v průběhu realizovaných měření budou zpracována tak, aby bylo možné na jejich základě analyzovat vliv únavy na výkonnost pilota.

**Vyhodnocení navržených metod:** Na základě analýzy dat zpracovaných pomocí navržených metod bude určena vhodnost jednotlivých metod pro využití v oblasti vlivu únavy na výkonnost pilota.

## Výzkumné otázky

Pro naplnění stanovených cílů byly formulovány následující výzkumné otázky:

- Otázka 1:** Lze prostřednictvím hodnocení fyziologických funkcí stanovit úroveň únavy?
- Otázka 2:** Lze prostřednictvím sledování výkonnosti stanovit úroveň únavy?
- Otázka 3:** Jaká metoda/metody mají největší výpovědní hodnotu s ohledem na studium únavy?

## Hypotézy

Na základě výše uvedených výzkumných otázek byly stanoveny následující hypotézy:

- H1** Změna úrovně únavy se projeví na každém ze sledovaných fyziologických parametrů.
- H2** Změna úrovně únavy se projeví na výsledcích výkonnostních testů.
- H3** Vybrané metody mohou být využity pro účely dostatečně citlivé klasifikace úrovně únavy.



## 3. Metodika

Na základě definovaných cílů a výzkumných otázek a z nich plynoucích hypotéz, byl vytvořen následující postup řešení dizertační práce. Metodika práce dále vychází z poznatků získaných v rámci přehledu současného stavu a to tak, aby bylo získáno co možná nejvíce dat vypovídajících o únavě, resp. o vlivu únavy na výkonnost. Na základě rešerše byly tedy zvolené metody založené na hodnocení biosignálů, jež jsou měřitelné prostřednictvím dostupných prostředků a zároveň jsou vhodné pro měření ve specifickém prostředí pilotní kabiny, tj. jsou mobilní, malá a pilota neobtěžují/neomezují ve vykonávání činností spojených s pilotáží. Z fyziologických signálů bylo pro tyto účely zvoleno snímání srdeční a mozkové aktivity. Dále byly zvoleny konvenčně užívané metody subjektivního hodnocení, tj. dotazník NASA Task Load Index a výkonnostní testování prostřednictvím měření reakčního času a dále prostřednictvím výkonnostního testu, tzv. OrTestu. Sledován byl také jednoduchý a výběrový reakční čas subjektů a dále byla v průběhu jednotlivých letů byla také sledována pohybová aktivita pilotů, a to prostřednictvím inerciálních měřicích jednotek (IMU, inertial measurement units). Experimentální měření probíhala ve dvou fázích, přičemž pouze u druhé fáze byl sledován tento plný výčet parametrů. Právě přidání sledovaných proměnných proběhlo na základě poznatků získaných v průběhu první fáze měření. S ohledem na skutečnost, že v druhé fázi měření bylo získáváno více datových sad a měření probíhala v jiném období, byly subjekty rozděleny do dvou skupin – A a B.

### 3.1 Subjekty

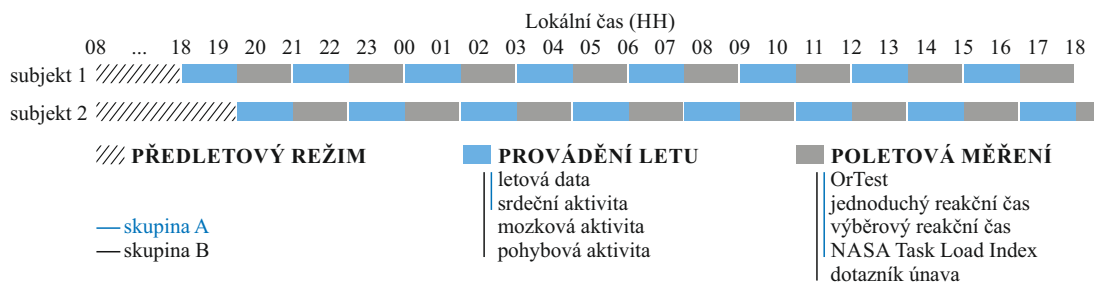
Měřenými subjekty byli studenti Ústavu letecké dopravy Fakulty dopravní Českého vysokého učení technického v Praze ze závěrečných ročníků bakalářského studijního oboru Profesionální pilot. V rámci výběru výzkumného vzorku byl kladen požadavek na co největší homogenitu skupiny a současně byly stanoveny minimální požadavky na letovou praxi. S ohledem na charakter experimentu bylo nezbytné, aby subjekty v době experimentu měly zkušenosti s létáním v podmínkách IFR, požadovaný nálet byl v rozmezí 100–150 hodin. Tím byla zajištěna obdobná úroveň teoretických i praktických znalostí jednotlivých subjektů. Experimentálních měření se zúčastnilo celkem 16 subjektů, 8 v každé fázi experimentu. V první fázi experimentu se účastnilo 7 mužů a jedna žena, subjekty byly ve věku 21–24 let (skupina A). V druhé fázi

experimentu se účastnilo 7 mužů a jedna žena, subjekty byly ve věku 21–25 let (skupina B).

Všechny subjekty splňovaly zdravotní požadavky podle lékařského osvědčení 2. třídy dle nařízení Komise (EU) č. 1178/2011, příloha IV (část MED), ve znění pozdějších předpisů. Všechny subjekty získaly základní informace o experimentu, sběru dat a způsobu anonymizace osobních a naměřených dat v souladu s etickými zásadami pro lékařský výzkum zahrnující lidské subjekty (World Medical Association and others, 2013). Subjekty podepsaly informovaný souhlas, studie byla schválena Komisí pro etiku ve výzkumu na Českém vysokém učení technickém v Praze (v současnosti Komise pro etiku vědecké práce při Vědecké radě Českého vysokého učení technického v Praze) pod jednacím číslem 0000-03/19/51902/EKČVUT.

### 3.2 Průběh experimentu

Obě skupiny subjektů absolvovaly měření dle stejného harmonogramu, přičemž experiment je rozvržen tak, aby bylo dosaženo spánkové deprivace, viz obr. 3.1, a to s ohledem na významný vliv spánku, resp. jeho nedostatku (příp. snížené kvality) na únavu. Experiment byl tedy založen právě na prodloužení doby bdělosti se současným vlivem cirkadiánního rytmu, zejména v době cirkadiánního útlumu, čímž byly vytvořeny podmínky pro navození proměnlivé úrovně únavy. Experiment je založen na 24hodinových měřeních, zahrnujících simulované lety a psychologická/výkonnostní testování. Celý koncept tedy zahrnuje sledování psychofyziologických parametrů prostřednictvím kolektování biosignálů, jako je EKG nebo EEG, výkonnostního testování a subjektivního hodnocení. Dále byla sledována pohybová aktivita subjektů, a to prostřednictvím dat získaných z IMU. Tím bylo dosaženo komplexní baterie testů/sběru dat, a to v průběhu experimentu s proměnlivou úrovní únavy, jak s ohledem na narůstající čas bdělosti subjektů, tak s ohledem na denní dobu.



Obrázek 3.1: Harmonogram prováděných měření.

Měření probíhala na letovém simulátoru letounu Beechcraft Baron využívajícím Microsoft Flight Simulator X (Microsoft Game Studios, Redmond, WA, USA). Data byla sbírána prostřednictvím programu Instructor Station by Luis Gordo (iStaion) se vzorkovací frekvencí 2,5 Hz. Experiment byl proveden jako série simulovaných letů, mezi nimiž byly prostory pro výkonnostní testy a subjektivní hodnocení pilotů, viz obr. 3.1. Každému měření předcházela předem definovaný předletový režim, s cílem zajištění co největší uniformity měření. V průběhu 24hodinových měření proběhlo celkem 8 měření na simulátoru, každé s délkou trvání přibližně 90 minut. S přihlédnutím k předletovému režimu tedy bylo dosaženo celkem 34 hodin bdělosti, díky níž docházelo k postupným změnám úrovně únavy z hlediska osmotického řízení spánku a cirkadiánního řízení spánku, nejvyšší nárůst únavy byl pak předpokládán v době cirkadiánního útlumu.

Každý experiment začínal zmíněným předletovým režimem. Subjektům bylo uloženo, aby vstávaly v 8 ráno. Během předletového režimu bylo zakázáno požívání povzbuzujících prostředků, tj. např. kávy, energetických nápojů, koly apod. Subjektům byla dále zakázána zvýšená fyzická námaha, tj. bylo jim dovoleno vykonávat pouze běžné denní činnosti. Experimentální měření začínala vždy v 18 hodin. Při každém experimentu byly souběžně měřeny dva subjekty. První subjekt zahajoval svůj let právě v 18 hodin. Po jeho ukončení byl měřen první let druhého subjektu, přičemž první subjekt plnil výkonnostní testy a vyplňoval subjektivní hodnocení. Po ukončení prvního letu druhým subjektem následoval druhý let první subjektu, přičemž druhý subjekt plnil výkonnostní testy a vyplňoval subjektivní hodnocení. Tímto způsobem došlo ke střídání jednotlivých letových fází a testovacích fází obou subjektů, viz obr. 3.1.

Sebehodnocení a výkonnostní testování tedy probíhalo v čase po každém vykonaném letu. Po absolvovaných testech pak měly subjekty v těchto „pauzách“ také možnost mentálního odpočinku, ne však spánku. Každý z letů v rámci osmi měření na simulátoru byl odlišný, lišila se však pouze trasa, tj. lety jsou srovnatelné z hlediska náročnosti, délky apod. Z hlediska letových profilů se také odlišovaly skupiny A a B. Skupina A vždy absolvovala let mezi dvěma letišti. Skupina B pak absolvovala let s mezipřistáním. Ačkoli byly letové profily odlišné mezi skupinami, byly navrženy tak, aby obsahovaly porovnatelné úseky. Letové scénáře zahrnovaly přesně definovaný profil letu včetně předepsané komunikace pilot–řídící letového provozu (ATCO), přičemž komunikace ze strany ATCO byla vedena přítomnou obsluhou simulátoru. Všechny lety probíhaly za zhoršených meteorologických podmínek, jednalo se tedy o lety dle přístrojů. Pro skupinu A byly definovány následující tratě:

**1. let: EDDC–EDDT.** Vzlet z Drážďan (EDDC, dráha 04) po odletové trati BEBEX 7E → BEBEX → VOR KLF → VOR FWE → ILS RWY 26L (Berlin Tegel, EDDT).

**2. let: EDBC–EDDT.** Vzlet z Magdeburku (EDBC, dráha 08) po odletové trati MAG 1L → VOR MAG → VOR KLF → VOR FWE → ILS RWY 26L (Berlin Tegel, EDDT).

**3. let: EDBC–EDAC.** Vzlet z Magdeburku (EDBC, dráha 08) po odletové trati MAG 1L → VOR MAG → VOR GOT → SULED → ILS RWY 22 (Lipsko–Altenburg, EDAC).

**4. let: EDAC–EDBC.** Vzlet z Lipska/Altenburgu (EDAC, dráha 22) po odletové trati GOLAT 6H → GOLAT → VOR GOT → VOR MAG → ILS RWY 26 (Magdeburk, EDBC).

**5. let: EDBC–EDDE.** Vzlet z Magdeburku (EDBC, dráha 08) po odletové trati MAG 1L → VOR MAG → VOR GOT → VOR ERF → ILS RWY 10 (Erfurt–Weimar, EDDE).

**6. let: EDAC–EDBC.** Vzlet z Lipska/Altenburgu (EDAC, dráha 22) po odletové trati GOLAT 6H → GOLAT → VOR GOT → VOR HOZ → VOR MAG → ILS RWY 26 (Magdeburk, EDBC).

**7. let: EDOP–EDBC.** Vzlet z Parchimu (EDOP, dráha 24) po odletové trati BKD 4J → VOR BKD → VOR HLZ → VOR MAG → ILS RWY 26 (Magdeburk, EDBC).

**8. let: EDBC–EDOP.** Vzlet z Magdeburku (EDBC, dráha 08) po odletové trati MAG 1L → VOR MAG → VOR HLZ → VOR BKD → ILS RWY 24 (Parchim, EDOP).

Všem subjektům byla vždy k dispozici potřebná dokumentace k letišťům (Jepessen), s níž byly zvyklí pracovat, a dále mapy odletových tratí SID, příletových tratí STAR a mapy přístrojového přiblížení IAC. Stejně tomu bylo také v případě tratí pro skupinu B:

**1. let: EDAC–EDDP–EDBC.** Vzlet z Lipska/Altenburgu (EDAC, dráha 04) po odletové trati OSKAT 1G → OSKAT → VOR GOT → SONDU → ILS

RWY 08L (Lipsko/Halle, EDDP) → odletová trať MAG 3E → PENEM → ILS RWY 26 (Magdeburk, EDBC).

**2. let: EDDC–LKKV–EDQM.** Vzlet z Drážďan (EDDC, dráha 22) po odletové trati MAREM 5W → MAREM → GOPSI → ILS RWY 29 (Karlovy Vary, LKKV) → odletová trať OKG 1F → OKG → ABERU → ILS RWY 26 (Hof-Plauen, EDQM).

**3. let: EDME–LOWS–LOWL.** Vzlet z Eggenfeldenu (EDME, dráha 08) po odletové trati BIBAG 1R → BIBAG → BADIT → ILS RWY 15 (Salcburk, LOWS) → odletová trať NEMAL 3A → NEMAL → ARASA → ILS RWY 26 (Linec, LOWL).

**4. let: LKMT–LKTb–LOWW.** Vzlet z Ostravy (LKMT, dráha 22) po odletové trati HLV 6F → HOLESOV → BUKAP → ILS RWY 27 (Brno, LKTb) → odletová trať MIKOV 2A → MIKOV → MABOD → ILS RWY 16 (Vídeň, LOWW).

**5. let: EDDC–EDAC–EDDE.** Vzlet z Drážďan (EDDC, dráha 22) po odletové trati BEBEX 7W → BEBEX → ILS RWY 22 (Lipsko/Altenburg, EDAC) → odletová trať GOLAT 7H → GOLAT → ILS RWY 28 (Erfurt, EDDE).

**6. let: LKKV–LKPR–LKPD.** Vzlet z Karlových Varů (LKKV, dráha 29) po odletové trati BALTU 9Z → BALTU → LOMKI → ILS RWY 06 (Praha-Ruzyně, LKPR) → odletová trať VOZ 3D → BAROX → USUPA → ILS RWY 27 (Pardubice, LKPD).

**7. let: LOWS–EDDM–EDDN.** Vzlet ze Salcburku (LOWS, dráha 15) po odletové trati TITG 2A → TITG → NAPSA → ILS RWY 26L (Mnichov, EDDM) → odletová trať RIDAR 6S → RIDAR → VOR DKB → ILS RWY 28 (Norimberk, EDDN).

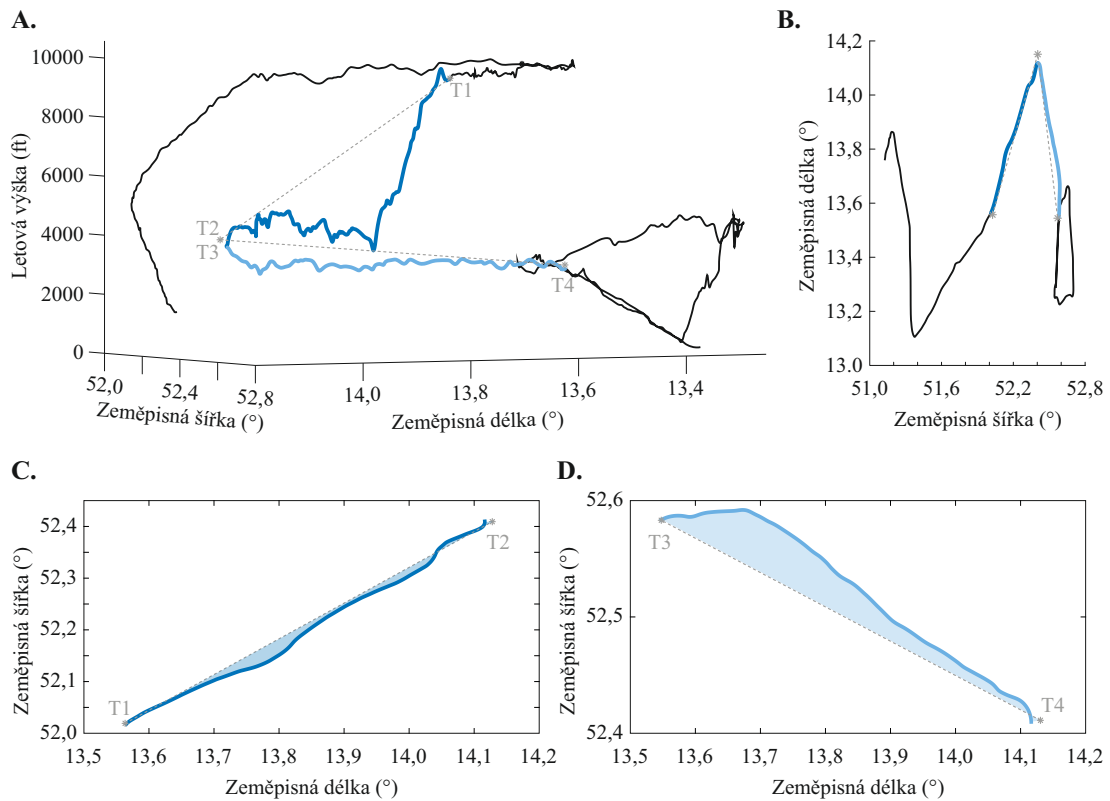
**8. let: EDDP–EDDC–EDAC.** Vzlet z Lipska/Halle (EDDP, dráha 08L) po odletové trati RUDAK 1E → RUDAK → OSKAN → ILS RWY 22 (Drážďany, EDDC) → odletová trať BEBEX 7W → GARKI → BEBEX → ILS RWY 22 (Lipsko/Altenburg, EDAC).

Na začátku měření byly subjekty seznámeny s používaným simulátorem Beechcraft Baron a jeho ovládáním. Každé měření začínalo seznámením subjektů s tratí. Simulace začínala v nastartovaném letounu na příslušném letišti na dráze v používání. Subjekty před každým letem prošly standardním postupem, jako v případě letů reálných, včetně přípravy kokpiktu s využitím checklistu a komunikace s řízením letového provozu. Měření všech dat sledovaných v průběhu simulovaných letů bylo zahájeno vždy při zahájení vzletu. Většina letu pak probíhala jako vodorovný let bez možnosti využití autopilota. Tyto úseky byly striktně definovány polohou jednotlivých bodů na trati a byly využity k následnému hodnocení, jelikož právě v těchto úsecích byly na subjekty kladeny nejnižší nároky a lze tedy předpokládat, že zátěž na subjekty byla v těchto úsecích nejnižší. Právě snížená zátěž pak mohla vést ke stavům snížené pozornosti a bdělosti, resp. snížené aktivace centrálního nervového systému (hypovigilance), což by se mohlo projevit na přesnosti pilotáže. Současně byly tyto úseky využity pro hodnocení dalších dat. Z těchto důvodů byla provedena časová synchronizace datových souborů, v rámci níž byly všechny datové vzorky označeny časem měření ve formátu Unixového času.

### 3.3 Letová data

Separování jednotlivých segmentů probíhalo na principu hledání nejbližší vzdálenosti mezi body letové trajektorie a bodem definovaným na trati. Algoritmus pro výpočet nejbližšího bodu zahrnuje výpočet Euklidovských vzdáleností všech bodů letové trajektorie vzhledem k definovanému bodu trati a následně nalezení minima vzdáleností. Nejbližší bod trajektorie pak definuje začátek, příp. konec, hodnoceného úseku letu, resp. Unixový čas  $T_i$  ( $i = 1 \dots 4$ ), viz obr. 3.2, který byl následně vymezen pro nalezení shodných úseků v dalších datových souborech (EKG, pohybová aktivita).

Pro skupiny A byly sledované úseky po sobě následující a tedy konec prvního úseku je zároveň začátkem druhého, tj.  $T_2 = T_3$ . S ohledem na skutečnost, že skupina B absolvovala v každém letovém profilu jedno mezipřistání, první hodnocený úsek se nachází mezi prvním a druhým letištem, druhý hodnocený úsek pak mezi druhým a třetím navštíveným letištem. Uvedené úseky byly pro hodnocení vybrány z již zmíněného důvodu předpokládané hypovigilance a současně pro možnost porovnatelnosti skupiny A a B. Ze stejného důvodu byla tímto způsobem hodnocena také EKG data, která byla měřena v případě obou skupin a tedy záznamy by mezi skupinami nebyly, s ohledem na charakter letových profilů, porovnatelné.



Obrázek 3.2: Letová data – vybrané úseky pro následné hodnocení. Zobrazení 3D se zvýrazněnými vybranými úseky a je ohraničujícími časy T1–T4 (A), zobrazení trajektorie v horizontální rovině se zvýrazněnými vybranými úseky (B), a zobrazení vybraného úseku 1 (C) a úseku 2 (D) v horizontální rovině. Šedá přerušovaná čára představuje ideální trajektorii, šedé hvězdičky naplánované body tratě. Modré plochy představují plochy odchylky od ideální trajektorie.

Letová data byla dále hodnocena ve smyslu přesnosti a správnosti pilotáže v kontextu předepsaných parametrů. Všechny segmenty zahrnovaly traťové lety, při nichž měl subjekt udržovat předepsaný heading a udržovat nulový náklon. S ohledem na skutečnost, že traťové úseky zahrnovaly změnu výšky bez předepsané rychlosti klesání/stoupání, nebylo možné hodnotit dodržování výšky, příp. vertikální rychlosti. Data tedy byla hodnocena v horizontální rovině. Původním záměrem byl výpočet průměrné chyby ve formě vzdálenosti od trati v každém datovém bodě jako indikátoru správnosti a směrodatné odchylky této chyby, jako indikátoru přesnosti. Z hlediska hodnocení je však vždy vhodnější využít komplexní parametr, zahrnující obě zmíněné. Z tohoto důvodu bylo přistoupeno k hodnocení formou plochy odchylky od předepsané trati, která byla definovaná počátečním a koncovým bodem. Tyto body byly fitovány polynomem prvního stupně (přímkou) pro nalezení ideální trajektorie. Body byly definovány zeměpisnou šířkou (latitude) a délkou (longitude), tedy jako  $P_1(lat_1, lon_1)$  a  $P_2(lat_2, lon_2)$ . Pro výpočet přímky  $y = ax + b$  je pak nezbytná znalost koeficientů  $a$  a  $b$ , přičemž:

$$a = \frac{lon_2 - lon_1}{lat_2 - lat_1}, \quad (3.1)$$

kde  $a$  je směrnice přímky, a koeficient  $b$  je pak získán jako:

$$b = lon_2 - a \cdot lat_2 = lon_1 - a \cdot lat_1. \quad (3.2)$$

Přímka ideální trajektorie pak byla navzorkována tak, aby její délka odpovídala délce detekovaného úseku letu. Na základě kombinace souřadnic z vektoru reálných dat a vektoru ideální trajektorie je následně vytvořen polynom, jehož plocha je zmíněnou plochou odchylky od ideální trajektorie, viz obr. 3.2C–D. Pro účely výpočtu plochy bylo využito integrovaných funkcí programového prostředí Matlab 2022a (MathWorks, Natick, MA, USA).

S ohledem na skutečnost, že úseky se mírně lišily v jejich délce, není samotná plocha vhodným indikátorem přesnosti pilotáže, jelikož je silně závislá na délce trajektorie. Z tohoto důvodu byla vypočtená plocha vždy normalizována na jednotkovou délku, tj. plocha byla podělena délkou ideální trajektorie. Pro tyto účely bylo třeba převést vzdálenost ze stupňů na metry s ohledem na sféricitu země. Uvedený výpočet byl proveden s využitím integrovaných funkcí v Mapping Toolboxu pro Matlab. U výsledné normované plochy jsou pak využívány, s ohledem na výpočet přes dvě zeměpisné koordináty o jiné škále a následnou normalizaci, arbitrární jednotky (a.u.). Takto spočtené normalizované plochy byly využity pro účely následné statistické analýzy.

V průběhu simulovaných letů byla kromě letových dat sbírána také fyziologická data ve formě záznamů elektrické srdeční aktivity, v případě skupiny B pak dále záznamy elektrické aktivity mozku a pohybové aktivity.

### 3.4 Fyziologická měření

Jak již bylo uvedeno, pro účely fyziologických měření byly zvoleny primárně EKG a EEG. Sledování těchto signálů bylo zvoleno mimo jiné i s ohledem na fakt, že mozková i srdeční činnost je únavou ovlivněna. Pro účely měření těchto signálů bylo v průběhu řešení pořízeno veškeré potřebné vybavení, a to v rámci studentských grantů orientujícího se na problematiku dizertační práce, specificky:

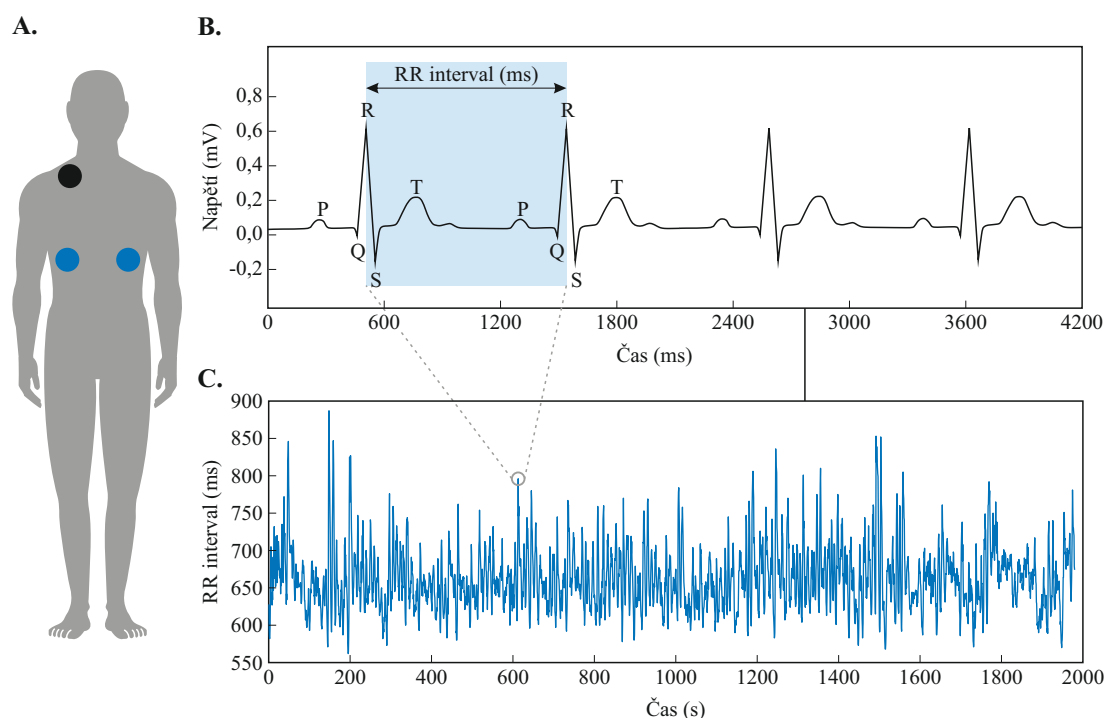
**Hanáková L.** Hodnocení psychofyziologického stavu pilota pomocí fyziologických parametrů. Interní grant Českého vysokého učení technického v Praze, SGS17/150/OHK2/2T/16. 2017–2018.



Kromě pořízení jednotlivých komponent potřebných měřicích systémů byla také v rámci řešení těchto projektů vytvořena základní část softwarových prostředků pro hodnocení sledovaných fyziologických signálů.

### 3.4.1 Srdeční aktivita

Pro účely dizertační práce byl pro získání dat s informací o elektrické srdeční aktivitě využit systém VLV Lab vyvinutý na Společném pracovišti FBMI ČVUT a 1. LF UK v Praze, který je primárně určen pro experimenty s požadavkem na kontinuální a přesné snímání biologických a technických signálů. V základní konfiguraci zařízení mezi tyto signály mimo jiné patří také EKG. Pro účely prezentované práce je využito jednosvodového EKG a jednorázových argentchloridových elektrod, umístěných v oblasti hrudníku, viz obr. 3.3A.



Obrázek 3.3: Záznam elektrické aktivity srdce – umístění měřicích elektrod (modrá) a referenční elektrody (černá) jednosvodového elektrokardiografu (A), zjednodušený záznam elektrické aktivity srdce se zvýrazněným RR intervalem a popisem jednotlivých vln (B), a tachogram RR intervalů (C).

Při měření prostřednictvím VLV Lab je získána celá EKG křivka, resp. hodnoty amplitudy elektrického napětí v čase, vzorkovací frekvence zařízení je

1000 Hz. Na základě analýzy současného stavu je zřejmé, že z hlediska EKG je nejvýznamnější analýza HRV. Primární informaci o variabilitě pak nese posloupnost RR intervalů (tzv. tachogram, viz obr. 3.3B–C), tj. v naměřeném signálu je pro účely zisku dalšího signálu třeba identifikovat kmity R a spočítat časové intervaly (ms). Vzhledem k výkonu činnosti pilota je signál dále zatížen artefakty, zejména kolísáním izoliny EKG. Pro účely srovnání izoliny a identifikace R-vln je využíváno zejména Pan-Tompkinsonovy metody (Kohler et al., 2002; Pan & Tompkins, 1985).

Prvním krokem při využití Pan-Tompkinsonovy metody je předzpracování dat, při němž dochází k převzorkování signálu na vzorkovací frekvenci 200 Hz tak, jak následně vyžadují další kroky algoritmu. Následně je využito sady filtrů tak, aby došlo k odfiltrování driftu izoliny, artefaktů spojených se svalovou činností, interference T-vln a síťového brumu (Kohler et al., 2002; Pan & Tompkins, 1985). Signál je následně zderivován s cílem zvýraznění QRS komplexu, umocněn, čímž je docíleno pozitivních hodnot v celém signálu a současně nelineárního zesílení signálu, jehož důsledkem je zvýraznění vyšších frekvencí, zejména tedy frekvencí EKG. Následně je signál integrován v pohyblivém okně, typicky o velikosti 150 ms, tedy 30 vzorků při vzorkovací frekvenci 200 Hz. Uvedené kroky slouží k potlačení šumu, zvýraznění QRS komplexů, resp. R-vln a získání charakteristiky těchto vln, jako je sklon.

Následně je implementována sada rozhodovacích kritérií sloužících pro správnou detekci jednotlivých R-vln. Prvním krokem je detekce základních vrcholů, tj. oblastí, kdy signál přechází z rostoucího v klesající. Současně je vydefinováno, že další peak nemůže být detekován v době kratší než 200 ms, což je doba odpovídající refrakterní fázi. Dalším rozhodovacím pravidlem je prahování, které je založeno na charakteristice signálu a s ohledem na daný signál definuje minimální amplitudu vrcholu. Po každém nalezeném vrcholu R-vlny je pak hodnota prahu upravena. Algoritmus také zpětně kontroluje zpracovaný signál pro potenciální detekci nedetekovaných R-vln, přičemž tato je založena na průměru 8 vybraných RR intervalů. Posledním krokem algoritmu je ochrana proti nechtěné detekci T-vlny, která je založena na sklonu detekovaného vrcholu R-vlny a vrcholu jemu předcházejícímu v případě vrcholu detekovaného do 360 ms od předchozího. Detaily popsaného algoritmu jsou k nalezení v publikaci od Pan & Tompkins (1985). Výstupem algoritmu jsou lokace detekovaných R-vln (Pan & Tompkins, 1985), z nichž jsou následně dopočítány RR intervaly. Z praktického hlediska je pak často takto získaný signál upravován do podoby tzv. NN intervalů.

Úprava tachogramu RR intervalů na NN intervaly zahrnuje eliminaci nestandardních hodnot, jako jsou ektopické (náhlé) srdeční údery. Po eliminaci

takovýchto úderů, resp. intervalů ze záznamů jsou získané intervaly často označovány jako NN (normal-normal). Standardem je, že vstupem do následné analýzy by měly být právě NN intervaly a proto jsou ektopické údery standardně detekovány a ze záznamů odstraněny. Existuje několik metod detekce ektopických úderů (Ramshur, 2010), přičemž pro účely dizertační práce byla implementována metoda založená na filtraci prostřednictvím mediánového filtru definovaného jako (McNames et al., 2004):

$$D(n) = \frac{|x(n) - med(x)|}{1,483 \cdot med\{|x(n) - med(x)|\}}, \quad (3.3)$$

kdy  $x$  je filtrovaný signál,  $n = 1 \dots N$ , kde  $N$  je délka signálu. Jestliže máme definovanou hodnotu  $\tau$ , pak

$$D(n) \geq \tau \quad (3.4)$$

je ektopický srdeční úder. Doporučená hodnota prahu je pak  $\tau = 4$  (McNames et al., 2004; Ramshur, 2010; Thuraisingham, 2006). Vlny detekované jako ektopické jsou nahrazeny mediánovou metodou v centrovaném okně definovaného počtu sousedních hodnot, tedy dle rovnice

$$x'(n) = med\{x(n+m) : |m| \leq \frac{w_m - 1}{2}\}, \quad (3.5)$$

přičemž  $x'$  je řada NN intervalů,  $w_m$  je počet sousedů ( $m = 1 \dots w_m$ ), přičemž nejsou brány v úvahu sousedi detekovaní jako ektopičtí. Počet sousedů je pak volen jako  $w_m = 5$  (McNames et al., 2004).

Touto úpravou jsou získány NN intervaly. V běžné praxi jsou však RR a NN intervaly brány jako synonymum a mezi NN a RR intervaly není prakticky rozlišováno (Malik et al., 1996). Ze získaných NN intervalů je dále možné získat tepovou frekvenci (HR, heart rate), kdy tepová frekvence 60 bpm (beats per minute, počet úderů za minutu) odpovídá NN intervalu 1000 ms. Tyto signály jsou pak dále analyzovány baterií metod.

Takovýchto metod existuje hned několik. V první řadě se jedná zejména o standardní, konvenčně užívané metody, které jsou často součástí komerčních systémů a jsou základem každé analýzy HRV. Tyto metody založené zejména na analýze v časové a frekvenční oblasti. Standardy týkající se analýzy HRV byly popsány v roce 1996 společnou evropskou a americkou pracovní skupinou (Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology) (Malik et al., 1996). Analýza v časové oblasti zahrnuje zejména parametry založené na popisné statistice. Jedná se zejména o standardní odchylky, průměry a extrémy (resp. rozpětí hodnot)

sledovaného signálu, tj. řady RR (resp. NN intervalů) a tepové frekvence. Mimo základní statistiku je variabilita srdečního rytmu popisována také prostřednictvím dalších parametrů, jako je četnost intervalů vzdálených od sebe o stanovenou hodnotu nebo parametry získané geometrickou analýzou. Z hlediska délky záznamů jsou pak definovány 3 variability - 24hodinová HRV, krátkodobá variabilita (ST – short-term, záznamy kratší než 24 hodin, ale delší než 5 minut) a ultra-krátkodobá variabilita srdečního rytmu (UST – ultra short-term, záznamy kratší než 5 minut). S ohledem na nastavení vlastního experimentu, resp. délku hodnocených úseků, je tato studie orientována na ST a v tomto kontextu byly také zvoleny vypočtené parametry. Zvolené parametry byly vypočítány prostřednictvím vlastního analytického software vytvořeném v programovém prostředí Matlab 2021a (MathWorks, Natick, MA, USA) a jsou uvedeny v tab. 3.1.

Tabulka 3.1: Parametry časové analýzy relevantní pro analýzu

Parametr	Popis
meanNN	Průměrná délka NN (RR) intervalů.
SDNN	Směrodatná odchylka NN (RR) intervalů.
RMSSD	Střední kvadratická hodnota rozdílů po sobě jdoucích RR (NN) intervalů.
meanHR	Průměrná tepová frekvence.
SDHR	Směrodatná odchylka tepové frekvence.
minHR	Minimum tepové frekvence
maxHR	Maximum tepové frekvence
NN50	Počet po sobě jdoucích NN (RR) intervalů od sebe se lišících o více než 50 ms.
pNN50	Procento NN50.

Parametr NN50 je pak silně závislý na délce hodnoceného úseku. S ohledem na skutečnost, že každý ze subjektů odletěl sledovaný úsek za odlišný čas (s ohledem na odchylku od trajektorie atp.) a současně každý z úseků nebyl přesně stejně dlouhý s ohledem na pozici definovaných bodů trati, je hodnocení tohoto parametru nerelevantní a není v následujících kapitolách dále prezentováno.

V případě analýzy ve frekvenční oblasti, je sledována výkonová spektrální hustota (PSD, power spectral density) v definovaných frekvenčních pásmech. Pro účely frekvenční analýzy HRV jsou standardně definována 3 pásma. Pásmo velmi nízkých frekvencí (VLF, very low frequencies) ve frekvenčním rozsahu 0,003–0,4 Hz. K těmto třem pásmům bývá někdy přidáno také pásmo ultranízkých frekvencí (ULF, ultra low frequencies) ve frekvenčním rozsahu 0–0,003 Hz. Pásma VLF a ULF se však v praxi moc nepoužívají, a to s ohledem na ne zcela objasněnou fyziologickou podstatu procesů reflektovaných těmito

pásmy a také s ohledem na skutečnost, že aby tyto parametry byly věrohodné, je třeba delších signálů (Ramshur, 2010). Při analýze jsou pak uplatňována zejména pásma nízkých frekvencí (LF, low frequencies) ve frekvenčním rozsahu 0,04–0,15 Hz a vysokých frekvencí (HF, high frequencies) ve frekvenčním rozsahu 0,15–0,4 Hz. Poměr parametrů LF a HF pak reflektuje sympatovagální rovnováhu. Tyto parametry mohou být vyjádřené prostřednictvím absolutních hodnot nebo prostřednictvím hodnot normalizovaných, vztažených k celkovému výkonu (Malik et al., 1996).

Odhad PSD je nejčastěji získáván prostřednictvím metod založených na rychlé Fourierově transformaci (FFT, fast Fourier transform) nebo autoregresním modelování. To však přináší jisté požadavky na charakter signálu, a to zejména ekvidistantní vzorkování signálu, které lze zajistit v případě čistého signálu EKG, nikoliv však v případě vektoru NN intervalů (Ramshur, 2010). S ohledem na tuto skutečnost se dostal do popředí tzv. Lombův (Lomb-Scargleův) periodogram, využívající metody nejmenších čtverců, který neklade požadavky na ekvidistantně vzorkovaný signál ani jeho stacionaritu (Lomb, 1976). Ostatní techniky založené na FFT nebo autoregresním modelování jsou však stále taktéž prakticky využívány, avšak při jejich využití je nezbytné data převzorkovat, což do výsledků zanáší chybu (Clifford & Tarassenko, 2005). Mezi tyto techniky odhadu PSD patří Welchův a Burgův periodogram (Ramshur, 2010).

Lombův (Lomb-Scargleův) periodogram, který patří mezi Fourierovy metody a vychází z Fourierovy transformace, ale kombinuje v sobě také další metody, jako je metoda nejmenších čtverců, a současně může být odvozen z Byasovské teorie pravděpodobnosti. Nejen z těchto důvodů tak Lombův periodogram představuje hlavní nástroj při frekvenční analýze neekvidistantně vzorkovaných signálů. Periodogram je pak statistický nástroj, využívaný pro nalezení významných periodických složek, respektive nalezení významných frekvencí v analyzované řadě, tedy může poskytnout informace o periodických složkách časových řad. Tato metoda umožňuje pozorovat spektrum a využívá se především k výpočtu odhadu složek spektrálního výkonu. Princip této metody spočívá v převedení jednotlivých členů časové řady na tvary, ve kterých vystupují goniometrické funkce sinus a kosinus.

Při výpočtu je  $x_j$  časová řada dat měřených v časech  $t_j$ , kde  $j = 1 \dots N$  a  $N$  je počet datových bodů. Průměrná hodnota je označována jako  $\bar{x}$  směrodatná odchylka jako  $\sigma^2$ . Lomb normalizovaný periodogram  $P(T)$  v periodě  $T$  označuje pravděpodobnou periodicitu a je definován jako:

$$P(T) = \frac{1}{\sigma^2} \left\{ \frac{\left[ \sum_{j=1}^N (x_j - \bar{x}) \cos \frac{2\pi(t_j - \tau)}{T} \right]^2}{\sum_{j=1}^N \cos^2 \frac{2\pi(t_j - \tau)}{T}} + \frac{\left[ \sum_{j=1}^N (x_j - \bar{x}) \sin \frac{2\pi(t_j - \tau)}{T} \right]^2}{\sum_{j=1}^N \sin^2 \frac{2\pi(t_j - \tau)}{T}} \right\}, \quad (3.6)$$

kde konstanta  $\tau$  je implicitně definována vzorcem:

$$\tan \frac{4\pi\tau}{T} = \frac{\sum_{j=1}^N \sin \left( \frac{4\pi t_j}{T} \right)}{\sum_{j=1}^N \cos \left( \frac{4\pi t_j}{T} \right)}. \quad (3.7)$$

Pro nulovou hypotézu platí, že hodnoty  $x_j$  jsou nezávislý Gaussovský náhodný šum a  $P(T)$  má exponenciální rozdělení pravděpodobnosti s UNIT MEAN. Hladina významnosti (p-hodnota) jakéhokoliv píku je dána vztahem:

$$p \equiv 1 - (1 - e^{-P(T)})^M, \quad (3.8)$$

kde  $M \approx N$ .

S ohledem na rozšíření analýzy srdeční aktivity, existují zmíněné standardy hodnocení/interpretace záznamů srdeční aktivity. Právě analýza RR (NN) intervalů ve frekvenční oblasti je přijímaná jako poměrně kvalitní nástroj pro odhad stavu ANS. V tomto kontextu se jako nejzajímavější považují pásma LF a HF, specificky výkon v těchto pásmech, resp. jejich poměr. Je obecně přijímáno, že navýšení výkonu v pásmu LF reflektuje vyšší aktivitu sympatiku a pásmo HF vyšší aktivitu parasympatiku. Jako klíčový je pak vnímán právě poměr LF/HF, přičemž nárůst tohoto poměru svědčí o vyšším zapojení sympatiku. Uvedené pak může bývá experimentálně využíváno při snahách o objektivizaci únavy.

V kontextu standardně užívané sady parametrů pro účely frekvenční analýzy HRV poskytuje analytický software mimo grafický výstup také 10 parametrů (viz tabulku 3.2), které zahrnují výkonové parametry v jednotlivých pásmech, ať už ve formě absolutních hodnot, tak i ve formě hodnot normovaných s ohledem na celkový výkon. Tyto parametry jsou dále využity pro následnou statistickou analýzu.

Kromě uvedených standardních metod jsou pro analýzu srdeční aktivity experimentálně využívány také jiné přístupy. S ohledem na charakter dat bývají voleny např. metody nelineární, které se zdají být vhodné vzhledem k nelineárním jevům podílejícím se na genezi HRV. Mezi tyto metody patří např. metody pracující s Lyapunovy exponenty, entropií, fázovými trajektoriemi apod. (Malik et al., 1996). Jednou z nejčastěji užívaných nelineárních metod analýzy HRV je tzv. Pointcaré graf (návrátový graf, návratová mapa), díky

Tabulka 3.2: Parametry frekvenční analýzy dostupné v analytickém software.

Parametr	Popis
ULF	Výkon v pásmu ultra nízkých frekvencí (<0,0033 Hz).
VLF	Výkon v pásmu velmi nízkých frekvencí (0,0033—0,04 Hz).
LF	Výkon v pásmu nízkých frekvencí (0,04—0,15 Hz).
HF	Výkon v pásmu vysokých frekvencí (0,15—0,4 Hz).
LF/HF	Poměr výkonů v pásmech nízkých a vysokých frekvencí.
Total	Celkový výkon frekvenčního spektra.
nULF	Výkon v pásmu ultra nízkých frekvencí (<0,0033 Hz) normalizovaný vůči celkovému výkonu spektra
nVLF	Výkon v pásmu velmi nízkých frekvencí (0,0033—0,04 Hz) normalizovaný vůči celkovému výkonu spektra
nLF	Výkon v pásmu nízkých frekvencí (0,04—0,15 Hz) normalizovaný vůči celkovému výkonu spektra
nHF	Výkon v pásmu vysokých frekvencí (0,15—0,4 Hz) normalizovaný vůči celkovému výkonu spektra

němuž je získávána krátkodobá i dlouhodobá variabilita (Shaffer & Ginsberg, 2017; Brennan et al., 2001).

Samotný Poincaré graf je grafickou metodou, založenou na vykreslení vektoru dat vůči jejímu časově zpožděné kopii. Jsou tedy definovány dva vektory:

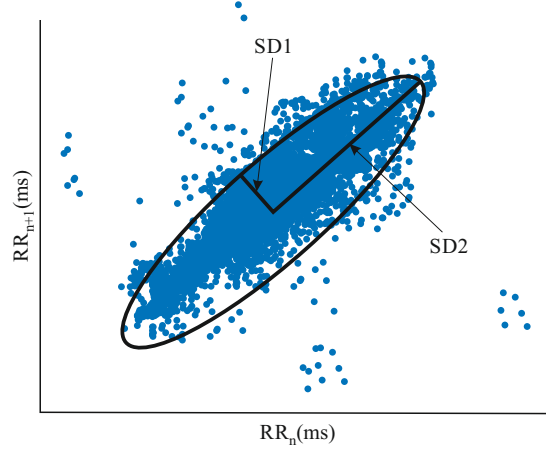
$$RRi_n = (x_1, x_2, \dots, x_{N-1}), \quad (3.9)$$

$$RRi_{n+1} = (x_2, x_3, \dots, x_N). \quad (3.10)$$

kde  $RRi_n$  je RR vektor tachogramu bez posledního datového bodu a  $RRi_{n+1}$  je časově posunutý vektor tachogramu (bez prvního datového bodu) a  $x_j$  jsou datové body tachogramu (RR intervaly), kdy  $N$  je celkový počet RR intervalů původního datasetu.

Vykreslením těchto dvou vektorů proti sobě vzniká rekurentní graf, viz obr. 3.4, což umožňuje vyhledávání vzorů skrytých v časových řadách. Poincaré graf je tedy 2D rekonstrukce fázového prostoru RR intervalů, tj. odráží dynamiku srdeční činnosti a poskytuje tak souhrnné informace o srdeční činnosti. Tato technika zpracování dat je populární zejména díky své jednoduché vizuální reprezentaci, prokázané klinické schopnosti predikce srdeční dysfunkce nebo nemocí a také proto, že na rozdíl od standardních metod používaných k analýze HRV odráží nelineární rysy variability srdeční rytmu (Malik et al., 1996; Shaffer & Ginsberg, 2017).

Poincaré graf tedy prezentuje shluk bodů roztažených podél úhlopříčky (linie identity, LOI). Pro snazší představu a kvantifikaci Poincaré grafu je do grafu



Obrázek 3.4: Poincaré graf s vyznačenými parametry SD1 a SD2.

zakreslována elipsa, jejíž hlavní poloosa je ve směru linie identity ( $x = y$ ), jejíž střed je umístěn v bodě daném průměry jednotlivých vektorů, a která je charakterizována velikostí hlavní (SD2) a vedlejší (SD1) poloosy, přičemž (Piskorski & Guzik, 2005):

$$SD1 = \sqrt{Var(x_1)}, \quad (3.11)$$

$$SD2 = \sqrt{Var(x_2)}, \quad (3.12)$$

přičemž  $Var(x)$  je variance  $x$  a  $x_1, x_2$  jsou definovány jako (Piskorski & Guzik, 2005):

$$x_1 = \frac{RRi_n - RRi_{n+1}}{\sqrt{2}}, \quad (3.13)$$

$$x_2 = \frac{RRi_n + RRi_{n+1}}{\sqrt{2}}, \quad (3.14)$$

respektive jako rotace  $RRi_n$  a  $RRi_{n+1}$  o  $\pi/4$ , tj.

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \cos(\frac{\pi}{4}) & -\sin(\frac{\pi}{4}) \\ \sin(\frac{\pi}{4}) & \cos(\frac{\pi}{4}) \end{bmatrix} \begin{bmatrix} RRi_n \\ RRi_{n+1} \end{bmatrix} \quad (3.15)$$

Parametr SD1 reflektuje krátkodobou variabilitu a je totožný s RMSSD parametrem časové analýzy, parametr SD2 reflektuje dlouhodobou variabilitu a koreluje s parametrem LF. Poměr SD1/SD2 pak koreluje s LF/HF a reflektuje autonomní rovnováhu, přičemž narůstající poměr indikuje větší zapojení sympatické větve ANS (Shaffer & Ginsberg, 2017).



Z hlediska analýzy HRV je v posledních letech také zájem o využití metod založených na výpočtu entropie. Přibližná entropie (Approximate Entropy) reflektuje míru složitosti, resp. neuspořádanosti systému. Opakování vzorů v signálu, resp. jeho předvídatelnost je odražena v nižší hodnotě entropie, vyšší hodnoty entropie pak ukazují na méně uspořádaný systém. Neuspořádanost HRV je fyziologickým projevem, naopak nízké hodnoty entropie a tedy větší uspořádanost a předvídatelnost systému mohou reflektovat patologické stavy (Francesco et al., 2012). Výpočet přibližné entropie je však limitován zejména s ohledem na požadavky na délku vstupního signálu. Z tohoto důvodu byl navržen výpočet tzv. vzorkové entropie (Sample Entropy), který se snaží eliminovat limitace přibližné entropie, avšak je přibližné entropii velmi podobný. Při vhodném nastavení vstupních parametrů a dlouhých signálech poskytují vzorková i přibližná entropie obdobné výsledky.

Vzorková entropie tedy představuje algoritmus pro odhad entropie vycházející z pravděpodobnosti, a byla vyvinuta specificky pro účely analýzy HRV jako vylepšení přibližné entropie, přičemž odstraňuje některé její limitace. Je definována, jako negativní logaritmus podmíněné pravděpodobnosti dvou náhodně vybraných sekvencí o délce  $m$  přičemž tyto zůstávají podobné také na  $m + 1$  bodech. Specificky je to definováno na základě vzdálenosti, přičemž sekvence mají mezi sebou vzdálenost menší než  $r$ , a to jak na  $m$ , tak na  $m + 1$  bodech. Vlastní shody přitom nejsou zahrnuty do výpočtu pravděpodobnosti. Je-li výsledná vzorková entropie rovná nule, jsou následné úseky shodné. Naopak narůstající hodnota entropie indikuje vyšší komplexitu signálu (Esteban, 2008). Vstupem pro analýzu je řada NN intervalů, dimenze  $m$  a vzdálenost  $r$ , přičemž doporučené hodnoty pro analýzu HRV jsou  $m = 2$  a  $r = 0,2 \cdot SD$  ( $SD$  – směrodatná odchylka). Výstupem je hodnota vzorkové entropie (sampEN).

Nelineární metody jsou často založeny také na fraktální charakteristice signálu. Jednou z takových metod je i analýza detrendovaných fluktuací (DFA, detrended fluctuation analysis), prostřednictvím níž je snaha detekovat fraktálních vlastnosti nestacionárního signálu. Tato analýza poskytuje informace o krátkodobých i dlouhodobých fluktuacích. Analýza spočívá v rozdělení signálu na fragmenty o různé délce a sledování chování signálu v těchto různě dlouhých úsecích (na různých škálách). V rámci této analýzy dochází nejprve k integraci signálu dle vztahu:

$$X(t) = \sum_{i=1}^t (x(i) - \bar{x}), \quad (3.16)$$

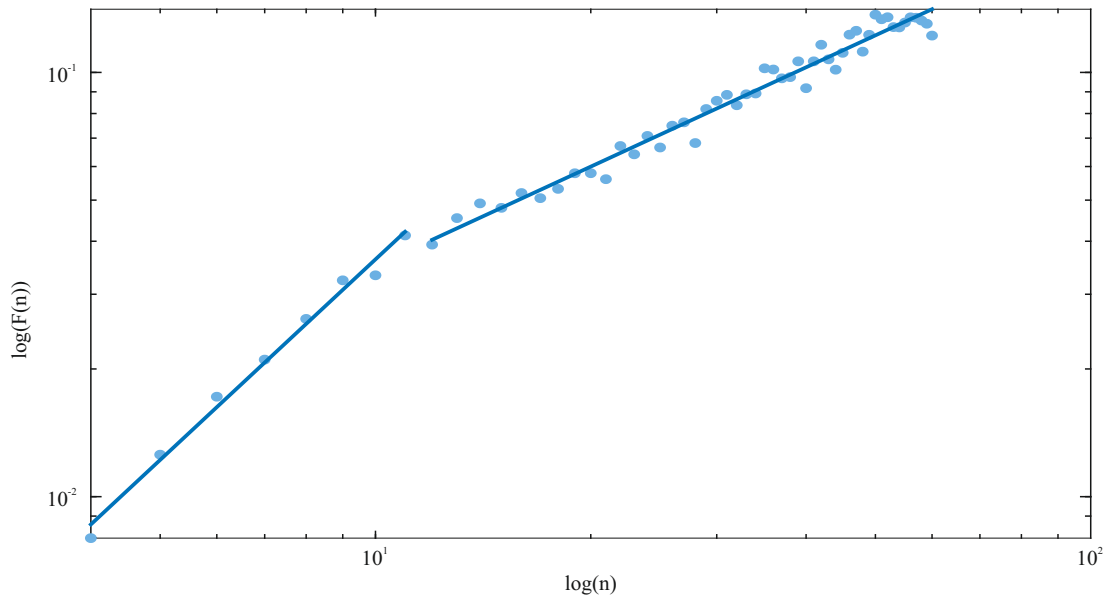
kde  $X$  je integrovaná časová řada,  $x$  je původní časová řada a  $\bar{x}$  je průměrná hodnota  $x$ . Následně je provedena lokální detrendace v úsecích délky  $n$ . Na těchto úsecích jsou data fitována polynomem prvního stupně (přímkou) metodou nejmenších čtverců, čímž je získán lokální trend, který je následně od dat odečten. Následně jsou vypočítány střední kvadratické fluktuace jako:

$$F(n) = \sqrt{\frac{1}{N} \sum_{t=1}^N (X(t) - X_n(t))^2}, \quad (3.17)$$

přičemž  $n$  je daná škála (délka úseků), jejíž interval hodnot je definován uživatelem. Fluktuace jsou tedy spočítány pro několik délek úseků a následně jsou vyneseny do log-log grafu vůči škále (délce úseku). Fluktuace bývá standardně počítána na úsecích o délce 4–400 datových bodů (Willson et al., 2002). Lineární vztah mezi  $\log(F)$  a  $\log(n)$  je reprezentován škálovacím exponentem  $\alpha$ , který je definován jako sklon vzniklé přímky (Peng et al., 1995). Současně jsou studovány také fluktuace na kratších lineárních úsecích, čímž je získán škálovací exponent  $\alpha_1$  pro krátkodobé fluktuace a  $\alpha_2$  pro dlouhodobé fluktuace. Zda se jedná o dlouhodobé nebo krátkodobé fluktuace závisí na délce jednotlivých úseků – pro analýzu HRV je bod zlomu mezi lineárními úseky typicky volen mezi délkou úseků 12–16 RR intervalů (datových bodů) (Ramshur, 2010; Peng et al., 1995). Standardně je tedy  $\alpha_1$  získán z úseků o délce 4–16 RR intervalů Willson et al. (2002). Vstupními parametry analýzy jsou horní a dolní hranice intervalu škály a bod zlomu. Výstupem jsou škálovací exponenty  $\alpha$ ,  $\alpha_1$  (škálovací exponent získaný na úsecích o délce menší nebo rovné než bod zlomu) a  $\alpha_2$  (škálovací exponent získaný na úsecích o délce větší než bod zlomu). Současně je poskytnut grafický výstup ve formě log-log grafu se zvýrazněnými lineárními úseky, viz obr. 3.5.

Pro účely dizertační práce byl tedy zvolen koncept měření EKG a následné analýzy HRV prostřednictvím prezentovaných, standardních přístupů a dále prostřednictvím experimentálně užívaných nelineárních přístupů, tj. Poincaré graf, vzorková entropie a DFA. Tento koncept byl zvolen zejména z důvodů, že analýza HRV je v současnosti jedním z nejčastěji užívaných prostředků při snahách o objektivizaci únavy a současně je tato problematika stále velmi otevřená, jelikož není uspokojivě vyřešena. I z těchto důvodů se zdálo jako vhodné využít jak standardní, tak i experimentálně využívané metody analýzy HRV, jelikož právě druhé zmíněné jsou z hlediska analýzy HRV ve stále větší pozornosti výzkumných týmů.

Parciální výsledky analýzy HRV, specificky pro skupinu A, byly prezentovány v rámci World Symposium on Applied Machine Intelligence and Informatics:



Obrázek 3.5: Ukázka log-log grafu detrendované analýzy fluktuací.

## Příloha A

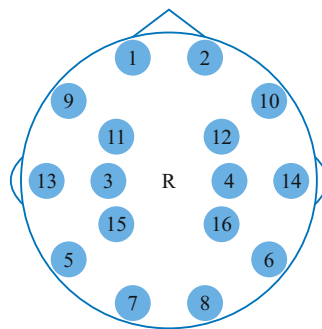
**Hanakova, L.,** Socha, V., Socha, L., Lalis, A., Kraus, J., & Malich, T. (2019). The influence of fatigue on psychophysiological indicators during 24 hours testing of pilots. In *2019 IEEE 17th World Symposium on Applied Machine Intelligence and Informatics (SAMI)* (pp. 181-186). IEEE.

### 3.4.2 Mozková aktivita

Pro sledování mozkové aktivity slouží záznam EEG. Metoda je založena na snímání elektrické aktivity mozku, při neinvazivním měření se jedná o snímání z povrchu kůže lebky. Pro správné rozmístění elektrod je využíváno EEG čepic nebo helem. Pro měření bylo dříve nezbytné využití vodivých past nebo gelů, což nebylo příjemné pro subjekty a limitovalo délku měření - gel se mohl roztékat a mohlo dojít k vodivému spojení více elektrodami. Dnes jsou dostupné systémy pracující se suchými elektrodami, které tyto limitace eliminují. Dalším problémem EEG systémů je horší dostupnost, a to zejména s ohledem na vysokou finanční náročnost. Cena standardních EEG přístrojů se pohybuje v řádech milionů. Z těchto důvodů jsou vyvíjeny experimentální systémy, přičemž snahou je vytvořit dostupné řešení pro výzkumné týmy. Dominantní postavení má v tomto odvětví společnost OpenBCI a její komunita uživatelů.

Firma nabízí dostupný systém s open-source softwarovým řešením a širokou podporou, která je dále podporována komunitou uživatelů systému, kterou

firma sdružuje a poskytuje jí prostor pro prezentaci a diskuzi. Z těchto důvodů bylo zvoleno využití právě systému OpenBCI. V rámci studentského projektu orientovaného na problematiku dizertační práce byl do Laboratoře lidského faktoru a automatizace v letectví pořízen 16kanálový EEG systém firmy OpenBCI – měřicí desky Cyton a Daisy (OpenBCI, Brooklyn, NY, USA) a helma Ultracortex Mark IV (OpenBCI, Brooklyn, NY, USA), využívající mezinárodní systém 10–20 rozmístění elektrod (viz obr. 3.6). Data byla nahrávána s nejvyšší možnou vzorkovací frekvencí – 125 Hz. V prvním kroku byl systém sestaven a byla ověřena jeho činnost. Následně byl využit při experimentálních měřeních.



Obrázek 3.6: Rozložení elektrod 1–16 na v helmě Ultracortex Mark IV dle mezinárodního systému 10–20. Bod R značí referenci.

Hodnocení mozkové aktivity je primárně založeno na frekvenční analýze. Samotný signál EEG je pak značně komplikovaný. V signálu se vyskytuje množství vln a také artefaktů. Některé typy vln jsou definovány, tj. existují základní tzv. EEG rytmy. Základní rytmy jsou 4 - každý z nich je specifický amplitudou a frekvencí, a dále je definováno, za jakých podmínek (bdělost, spánek, zavření/otevření očí apod.), a v jaké oblasti lebky se rytmy vyskytují. Dále pak existuje množství vln, které se standardně u zdravého člověka nevyskytují. V případě únavy pak převažují pomalé vlny a proto se výzkumy únavy zaměřují právě na pomalé rytmy (Thomas et al., 2015; Wilson et al., 2007b; Caldwell, 2005). Mezi základní rytmy patří vlny delta, theta, alpha a beta, které jsou definovány ve specifických frekvenčních pásmech (Webster, 2009; Shaker, 2006). Původním záměrem bylo využít právě tato pásma. Při následném zpracování se pak ukázalo vhodnější využití menších pásem, s cílem hledání frekvencí, na základě nichž by bylo možné rozlišit jednotlivá měření. Před výpočtem jednotlivých pásem však musí dojít k předzpracování dat.

Prvním krokem při zpracování dat EEG je filtrace signálu od artefaktů. Nejvýznamnější artefaktem je rušení ze sítě, z těchto důvodů je využit filtr typu pásmová propust 0,5–40 Hz. Následně je signál převeden do frekvenční oblasti prostřednictvím FFT a je spočítána plocha pod křivkou na daných frekvenčních

pásmech. S ohledem na skutečnost, že cílem bylo najít atributy, na základě nichž bude možné rozlišit různé úrovně únavy, byla vytvořena jemnější pásma, než tradičně využívaná.

Celé frekvenční spektrum bylo tedy rozděleno na malá frekvenční pásma o šířce 0,2 Hz. V těchto 0,2Hz pásmech byl následně spočítán výkon. Tímto přístupem vzniklo, s ohledem na vzorkovací frekvenci signálu, 312 atributů (výkonů v 0,2Hz pásmech, viz tab. 3.3) pro každou ze 16 použitých elektrod. Vzniklých 4992 datových bodů pro každé měření každého subjektu bylo dále využito pro následnou analýzu.

S ohledem na povahu dat a jejich velikost bylo zpracování EEG signálů podmíněno redukcí dimenzionality, a to prostřednictvím analýzy hlavních komponent (PCA, principal component analysis). Matematický základ a popis aplikace PCA lze nalézt např. v publikaci od Jolliffe & Cadima (2016). Hlavním cílem bylo zjistit, odhalí-li PCA v datech přirozené vzory odlišující jednotlivá měření mezi sebou. Jelikož byla jednotlivé měření realizována v čase, v průběhu 24 hodin, je možné, že tyto budou reflektovat únavu na základě mozkové aktivity. Za předpokladu využití všech proměnných, resp. příznaků, jako souřadnic pro vykreslení cílové proměnné (měření) do bodového grafu, existuje předpoklad, že takováto projekce bude vytvářet rozlišitelné shluky bodů. Hlavním problémem v takovémto přístupu však je, že prostor pro vykreslení takového bodového grafu by byl 4992dimenzionální a shlukování dat by fungovalo za předpokladu, že každý příznak by popisoval jistou část variability v datech, v datech by se nevyskytovala vysoká míra kolinearity atp.

I na základě výše uvedeného byl zvolen přístup s využitím PCA pro redukcí dimenzionality umožňující identifikovat závislosti v datech, transformovat data na základě znalosti těchto závislostí a kvantifikovat importanci těchto závislostí.

Jak bylo zmíněno, vstupní dataset sestával z 4992 atributů reprezentujících výkon v pásmu o šířce 0,2 Hz pro každý kanál, přičemž je využíváno celkem 16 kanálů. Cílová kategorická proměnná obsahovala 8 hodnot reprezentujících měření, a tedy 64 instancí, při zohlednění realizace měření u 8 subjektů. Generalizovaný zápis datové matice je prezentován v tab. 3.3.

Předzpracování těchto dat spočívalo zejména v jejich normalizaci před samotnou PCA prostřednictvím z-skóre (Carey & Delaney, 2009) kvůli potlačení interindividuální variability. Původní hodnoty  $x_{i,j}$  byly tedy transformovány na:

$$ZD_{i,j} = \frac{x_{i,j} - \frac{1}{\lambda} \sum_{j=1}^{\lambda} x_{i,j}}{\sqrt{\frac{\sum_{j=1}^{\lambda} (x_{i,j} - \frac{1}{\lambda} \sum_{j=1}^{\lambda} x_{i,j})^2}{\lambda}}}, \quad (3.18)$$

kde  $ZD$  je transformovaná matice z původní matice  $X$  znázorněné v tab. 3.3. Tato transformace je realizovaná po subjektech  $i$  přičemž  $i = 1 \dots N_S$ , kde  $N_S$  je celkový počet subjektů. Normalizace po subjektech je prováděna přes všechna spočtená pásma  $j$  pro daný subjekt, přičemž  $j = 1 \dots \lambda$ , kdy  $\lambda$  je počet pásem, tedy:

$$\lambda = \left\lfloor \frac{f_{vz}}{2 \cdot w_p} \right\rfloor \cdot N_M \cdot N_{Ch}, \quad (3.19)$$

kde  $f_{vz}$  je vzorkovací frekvence signálu,  $w_p$  je šířka pásma a  $\lfloor \cdot \rfloor$  je funkce dolní celá část (zaokrouhlení na celé číslo směrem dolů),  $N_M$  je počet měření a  $N_{Ch}$  je počet kanálů.

Tabulka 3.3: Uspořádání dat výsledných parametrů frekvenční analýzy záznamů elektrické aktivity mozku – spektrální výkony v 0,2Hz frekvenčních pásmech pro všechny subjekty skupiny B (S1–S8) a všechna měření (M1–M8) v 16 kanálech (Ch<sub>1</sub>–Ch<sub>16</sub>).

S	M	Ch <sub>1</sub>				...	...	Ch <sub>16</sub>	
		(0;0,2) Hz	(0,2;0,4) Hz	...	(62,2;62,4) Hz			...	...
1	1	$S_1M_1Ch_10,2$	$S_1M_1Ch_10,4$	...	$S_1M_1Ch_162,4$	...	...	$S_1M_1Ch_{16}62,4$	
1	2	$S_1M_2Ch_10,2$	$S_1M_2Ch_10,4$	...	$S_1M_2Ch_162,4$	...	...	$S_1M_2Ch_{16}62,4$	
1	3	$S_1M_3Ch_10,2$	$S_1M_3Ch_10,4$	...	$S_1M_3Ch_162,4$	...	...	$S_1M_3Ch_{16}62,4$	
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	
1	8	$S_1M_8Ch_10,2$	$S_1M_8Ch_10,4$	...	$S_1M_8Ch_162,4$	...	...	$S_1M_8Ch_{16}62,4$	
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	
8	1	$S_8M_1Ch_10,2$	$S_8M_1Ch_10,4$	...	$S_8M_1Ch_162,4$	...	...	$S_8M_1Ch_{16}62,4$	
8	2	$S_8M_2Ch_10,2$	$S_8M_2Ch_10,4$	...	$S_8M_2Ch_162,4$	...	...	$S_8M_2Ch_{16}62,4$	
8	3	$S_8M_3Ch_10,2$	$S_8M_3Ch_10,4$	...	$S_8M_3Ch_162,4$	...	...	$S_8M_3Ch_{16}62,4$	
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	
8	8	$S_8M_8Ch_10,2$	$S_8M_8Ch_10,4$	...	$S_8M_8Ch_162,4$	...	...	$S_8M_8Ch_{16}62,4$	

Ve většině případů využití PCA se předpokládá, že příznaky jsou vzájemně nezávislé a tedy je třeba je normalizovat přes sloupce tak, aby jednotlivé příznaky byly na stejné škále a byly porovnatelné. V případě spektrální analýzy jsou však příznaky závislé, a další normalizace dat je tedy nevhodná. Takováto úprava je vykonávána zejména v případě, že příznaky vstupující do PCA mají např. rozdílné jednotky, pocházejí z nezávislých měření atp.

Pro samotnou exekuci PCA byl použit algoritmus pro singulární rozklad matice (Wall et al., 2003). Výběr počtu hlavních komponent se odvíjel od celkové popsané variability hlavními komponentami (PC), avšak s ohledem na další interpretaci do maximálního počtu 3 PC.

Výše uvedené předzpracování a zpracování dat probíhalo v prostředí Orange data mining v.3 (University of Ljubljana, Lublaň, Slovinsko) a Matlab 2021a (MathWorks, Natick, MA, USA). Popsané zpracování založené na PCA

představuje hlavní zvolený přístup, který je však doplněn dalšími metodami nutnými pro interpretaci výsledků. Tyto jsou dále prezentovány v kapitole 4.2.2.

### 3.5 Pohybová aktivita

Snímání pohybové aktivity probíhalo prostřednictvím pěti IMU s centrální jednotkou připojenou přes USB do počítače. Jednalo se o senzory ProMove-mini s centrální jednotkou Basic Inertia Gateway (Inertia Technology B.V., Enschede, NL). Tento systém umožňuje měření několika veličin ve třech osách. Pro účely dizertační práce byla využita akcelerace.

Důležitou roli hraje umístění jednotlivých IMU. To bylo zvoleno s ohledem na část těla zapojené do pohybu při řízení letounu. Senzory tedy byly umístěny na obě předloktí, oba bicepsy a v oblasti pupku. Při každém měření byla zaznamenána poloha senzoru a antropometrické rozměry subjekty, definované dle (International Organization for Standardization, 2017).

Jednotlivé IMU komunikují bezdrátově s centrální jednotkou a živě jí předávají data, která mohou být zobrazena v reálném čase nebo zpětně prostřednictvím příslušného software Inertia Studio (Inertia Technology B.V., Enschede, NL). Současně senzory zaznamenávají data na svoji vnitřní paměť. To je důležité z toho důvodu, že při bezdrátovém přenosu dochází ke ztrátám. Data se tedy po ukončení měření doplňují právě z vnitřní paměti senzorů. Data jsou nahrávána se vzorkovací frekvencí 200 Hz.

Senzory měří 19 signálů, pro účely dizertační práce byly využity 3 - akcelerace v osách x, y a z. Data akcelerací byla nejprve filtrována s využitím Savitzky-Golay filtru (Karaim et al., 2019). Z takto upravených akcelerací v jednotlivých osách byla následně vypočítána 3D akcelerace  $a$  jako:

$$a = \sqrt{a_x^2 + a_y^2 + a_z^2}, \quad (3.20)$$

kde  $a_x$  je akcelerace ve směru osy x,  $a_y$  je akcelerace ve směru osy y a  $a_z$  je akcelerace ve směru osy z (Strohrmann et al., 2012).

V rámci hodnocení signálu však došlo ke zhodnocení, že data nejsou vhodná pro zpracování. Ačkoli byla data získávána profesionálním systémem, v průběhu měření docházelo k problému se snímáním dat, zejména se projevujícím výpadky signálu mezi jednotlivými IMU a centrální jednotkou, příp. kompletním zastavením přenosu signálu. Spočítané parametry v daných traťových úsecích pak disponují velkým množstvím prázdných hodnot, a proto je nebylo možné dále staticky hodnotit. Z těchto důvodů nejsou výsledky pohybové analýzy prezentovány a diskutovány v následujících kapitolách.

## 3.6 Výkonnostní testy

Výkonnostní testování bývá standardně prováděno zejména prostřednictvím PVT – psychomotorických testů bdělosti. Standardní testy jsou v podstatě základní testy reakčního času, tj. na definovaný podmět přichází požadovaná odpověď bez jakékoli nutnosti volby/rozhodování. Sofistikovanější testy pak rozšiřují úkol zejména právě o rozhodovací proces, tj. existuje několik rozdílných podmětů, na něž jsou požadovány specifické odpovědi (Dinges & Powell, 1985).

V případě reakčního času je nejčastěji sledován tzv. komplexní (celkový) reakční čas  $RT_T$ , tj. čas mezi podmětem a reakcí. Tento čas se skládá z několika dílčích reakčních časů (Board, 1998):

$$RT_T = t_P + t_D + t_M, \quad (3.21)$$

kde  $t_P$  je percepční reakční čas,  $t_D$  je dohodovací reakční čas a  $t_M$  je svalový reakční čas. Percepční reakční čas je čas, než subjekt zaznamená daný podmět. Podmět pak může být např. akustický nebo vizuální. Rozhodovací reakční čas je čas, než subjekt vyhodnotí daný podmět a rozhodne o reakci. Svalový reakční čas je pak čas nezbytný k vykonání pohybu za účelem reakce. Tento čas je dán zejména fyzickým stavem subjektu (Board, 1998).

Pro účely dizertační práce bylo využito 2 reakčních testů. První z nich byl zaměřen na jednoduchý reakční čas (SRT, simple reaction time), tedy úkol, při němž je člověku zobrazován jednoduchý vizuální podnět na nějž má reagovat, přičemž nemusí volit mezi více možnostmi. V tomto případě byl zvolen úkon, při němž měl subjekt reagovat na červený terč stiskem mezerníku. Aby bylo zamezeno zanesení chyby z hlediska vzdálenosti ruky od stlačované klávesy, musel subjekt ruku neustále držet na příslušné klávese. Test zahrnuje 20 opakování, podnět je zobrazen v čase v intervalu 2–8 sekund od předchozího podnětu. Mimo 20 měřených opakování vizuálního podnětu bylo subjektu zobrazeno dalších 5 opakování, a to na začátku testu tak, aby se subjekt seznámil s úlohou a prostředím a byl plně zaměřen na test. Pro tento test bylo využito softwaru Inquisit 5 Lab (Millisecond Software, LLC, Seattle, WA, US) a skriptu Simple Visual Reaction Time Task (Borchert, 2016). Využitým výstupem jsou reakční časy, tj. 20 reakčních časů pro každý subjekt při každém měření.

Druhý využitý reakční test (4choice) zahrnoval také rozhodovací element (Armstrong et al., 2012). Subjektu byly zobrazeny 4 prázdná pole. Ve chvíli, kdy se jedno z polí vybarvilo, bylo úkolem subjektu stisknout příslušnou klávesu (D, F, J, K). Každé ze čtyř kláves je přiřazen jeden prst, který na ní spočívá po celou dobu testování tak, aby bylo zabráněno vlivu vzdálenosti ruky



od příslušné klávesy. Na začátku testu je vždy sekvence 10 opakování, která není měřena a je určena pro nácvik a seznámení se s testem. Samotná testovací sekvence pak zahrnuje 100 opakování, přičemž neodpoví-li subjekt do 1 sekundy, je odpověď označena jako chybná. Vizualní podnět, tj. vybarvení jednoho boxu, je subjektu prezentován každých 1500 ms. V případě, že subjekt stiskne klávesu před zobrazením podnětu, je odpověď označena jako chybná. Pro tento test bylo opět využito softwaru Inquisit 5 Lab a skriptu Four Choice Reaction Time Task (Borchert, 2017). Využitým výstupem pak jsou jednotlivé reakční časy, tj. 100 reakčních časů pro každý subjekt při každém měření, a informace o chybovosti subjektu v daném testu.

Využití výběrového reakčního času, ačkoli ne s využitím tohoto přesného testu, pro účely hodnocení únavy bylo prezentováno na mezinárodní konferenci *New Trends in Civil Aviation*:

### Příloha B

Socha, V., **Hanáková, L.**, Stojić, S., Kušmírek, S., Socha, L., & Antoško, M. (2018). A fatigue influence on pilot's reaction ability during 24 hours flight simulation: A case series study. In *New Trends in Civil Aviation* (pp. 15-19). CRC Press.

Dalším výkonnostním testem využívaným pro účely dizertační práce je tzv. Or-test. Test je založen na tabulce kódů a 4 úkolech, při nichž subjekt hledá požadovaný kód. Pro účely práce byla vytvořena softwarová implementace tohoto testu, jehož princip je zobrazen na obr. 3.7

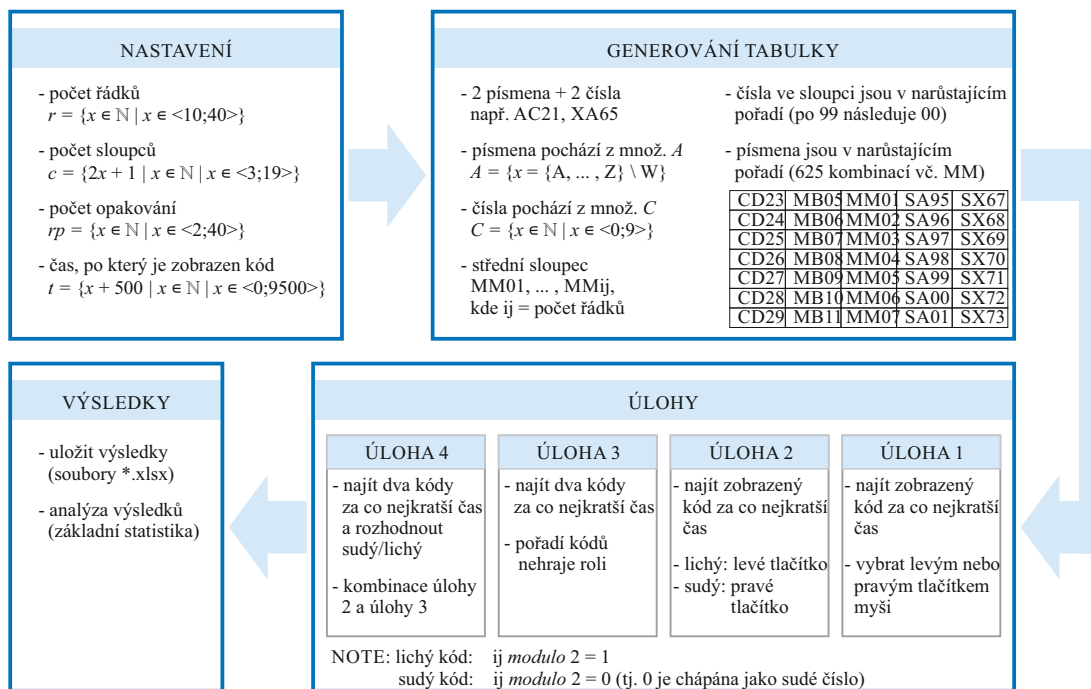
### Software

**Hanáková, L.**, Socha, V., Socha, L. & Kušmírek, S. (2018). OrTest. Software. Czech Technical University in Prague.

### Příloha C

Socha, V., **Hanakova, L.**, Freigang, M., Kraus, J., Stojic, S., Socha, L., & Hanak, P. (2019). Impact of Pilots' Tiredness on the Outcome of Psychological Testing. In *Transport Problems* (Vol. 14, Issue 2, pp. 55-67). Silesian University of Technology. DOI: 10.20858/tp.2019.14.2.5

Test se skládá se 4 podúloh s narůstající obtížností. Celkovou náročnost testu je pak možno upravit v úvodním nastavení velikostí tabulky (tj. celkovým počtem zobrazených kódů), nastavením počtu opakování a nastavením času, po němž je subjektu zobrazeno zadání, tj. kód, který musí nalézt. Tabulka

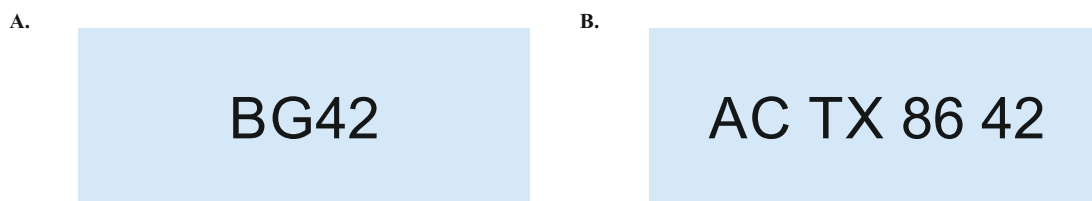


Obrázek 3.7: Princip softwarové implementace výkonnostního testu Or-test.

se skládá ze čtyřčlenných kódů (2 písmena a 2 čísla), které jsou systematicky umístěny v tabulce. Pro potlačení učebního procesu je pak při každém spuštění testu generována nová, náhodně generovaná tabulka. Tabulka má následující pravidla (1) tabulka má vždy lichý počet sloupců ( $c$ ), (2) prostřední sloupec je vždy MM, (3) v první části tabulky (sloupce 1 až  $(c - 1)/2$ ) jsou kódy začínající AA–ML, (4) v druhé části tabulky (sloupce  $(c + 1)/2$  až  $c$ ) jsou kódy začínající MN–ZZ, (5) kódy ve sloupcích jsou v abecedním pořadí, (6) číselná část kódu obsahuje na prvním řádku vždy náhodné číslo v rozmezí 00–99, (7) číselná část kódu narůstá s číslem řádku a po překročení 99 následuje 00. Z praktických důvodů program při generaci tabulky nepracuje s písmeny CH, W a s diakritikou. Softwarová implementace dovoluje následující nastavení tabulky: (1) počet řádků je 10–40, (2) počet sloupců je 7–39 (implementace umožňuje pouze lichý počet sloupců), (3) počet opakování každého úkonu je 2–40, (4) doba zobrazení každého zadání je 1000–10000 ms (s krokem 500 ms). Pro účely této práce byla zvolena tabulka o velikosti 25 řádků a 15 sloupců. Každý úkol obsahoval 15 hledaných zadání, která byla subjektu zobrazena po dobu 3000 ms. Pro účely seznámení se z testem dostaly subjekty možnost vyzkoušení si testu se 2 opakováními každé úlohy.

Před každým z úkolů subjektům jsou subjektům zobrazeny instrukce týkající se úkolu. V prvních dvou úkolech subjekt vždy hledá pouze jeden kód, viz obr. 3.8A, který označí zakliknutím myši. V případě dalších dvou úloh jsou zobrazeny dvě dvojice písmen a dvě dvojice číslic (viz obr. 3.8B), z nichž

subjekt poskládá dva kódy, které následně hledá a opět označí zakliknutím myši. V případě úloh 2 a 4 pak subjekt dále rozhoduje, zda je daný kód sudý nebo lichý. Po zakliknutí kódu/kódů je vždy ihned zobrazeno další zadání, překrývající celou tabulku, a to po nastavenou dobu. Uvedené probíhá dokud není dosaženo nastaveného počtu opakování. Následně jsou zobrazeny pokyny pro další úlohu. Výsledky testu je možné, po jeho absolvování, uložit jako soubory \*.xlsx. Výsledky obsahují reakční čas na jednotlivá zadání a správnost, resp. chybovost. V případě úlohy 3 a 4 je pak sledován celkový čas odpovědi i čas odpovědi jednotlivých úkolů. Z hlediska správnosti/chybovosti je sledován také konkrétní typ chyby. Dále je zobrazena základní statistika výsledků.



Obrázek 3.8: Zobrazení zadání pro Or-test: úlohy 1 a 2 – hledaný kód BG42 (A), úlohy 3 a 4 – hledané kódy AC86 a TX42 (B).

Z hlediska zpracování dat výběrových RT, jak pro test 4choice, tak pro časy získané v OrTestu, je pak třeba uvážit nejen průměrný RT subjektu, ale také s ním spojenou chybovost/správnost odpovědi, jelikož reakce subjektu může být rychlá, ale nikoli správná. Příp. subjekt s delším RT může vykazovat vyšší výkonnost, s ohledem na vyšší správnost odpovědi. Pro tyto účely byla navržena korekce RT o chybovost, resp. správnost (Townsend et al., 1983; Rach et al., 2010):

$$RT_1 = \frac{RT_0 \cdot p_{100}}{p_s}, \quad (3.22)$$

kde  $RT_1$  je upravený původní reakční čas  $RT_0$ ,  $p_{100}$  je maximální počet správných odpovědí a  $p_s$  je počet správných odpovědí. Upravený RT je tak nepřímo úměrný správnosti, tj. v případě 100% správnosti se  $RT_1 = RT_0$ . Vztah nepočítá s 0% úspěšností, jelikož v tomto případě je výsledek nerelevantní. V tomto případě by se  $RT_1$  limitně blížil nekonečnu. K uvedenému však v případě sbíraných dat nedošlo. Statistická analýza všech výsledků je uvedena v následující kapitole.

### 3.7 Subjektivní hodnocení

V případě subjektivního hodnocení se jedná zejména o datazníková šetření. Ačkoli je možnost vlastního sestavení dotazníku, velmi často bývá při sledování únavy využíván standardizovaný dotazník NASA Task Load Index

(NASA-TLX) (Hart, 2006; Hart & Staveland, 1988). Tento dotazník však není primárně zaměřen na sledování únavy, ale na sledování zátěže. Proto byla v rámci tohoto projektu vytvořena jeho obdoba zaměřená na únavu pilota (Herman, 2020).

### 3.7.1 Nasa Task Load Index

Dotazník je postaven na 6 kategoriích, které jsou subjektem hodnoceny v rozsahu 0 až 20. Jednotlivé kategorie jsou psychická náročnost, fyzická náročnost, časová náročnost (časový tlak), výkonnost, úsilí a frustrace. K jednotlivým kategoriím je pak uveden jejich popis (Hart, 2006; Hart & Staveland, 1988).

Psychická náročnost (MD, mental demand) odráží, kolik psychické a percepční aktivity musel subjekt vykonat, tj. nakolik musel přemýšlet, rozhodovat se a podobně. Otázka se tedy posuzuje, zda byl úkol náročný nebo obtížný, jednoduchý nebo komplexní (Hart, 2006; Hart & Staveland, 1988). V kontextu experimentu tedy subjekt reportoval, na kolik ho samotný let psychicky vyčerpával, a to například s ohledem na oblačnost a nezbytnost orientace v přístrojích, komunikaci s řídicím apod.

Fyzická náročnost (PD, physical demand) odráží, kolik fyzické aktivity musel subjekt vykonat (Hart, 2006; Hart & Staveland, 1988), tj. nakolik musel manipulovat s ovládacími prvky a podobně. Otázka tedy reflektuje náročnost úkolu z hlediska fyzické námahy (Hart, 2006; Hart & Staveland, 1988).

Časová náročnost (TD, temporal demand) odráží, jaký časový tlak subjekt pociťoval, tj. zda měl na vykonání všech úkonů dostatečný čas nebo zda pociťoval časovou tíseň (Hart, 2006; Hart & Staveland, 1988). V kontextu experimentu tedy subjekt reportoval, zda nebyl pod např. tlakem v důsledku pokynů od řídicího.

Výkonnost (PF, performance) odráží, jak si subjekt myslí, že byl úspěšný v provedení úkolu, tj. jak je spokojený s vykonáním zadaných úkolů (Hart, 2006; Hart & Staveland, 1988). V kontextu experimentu otázka reflektuje, jak spokojený byl subjekt se samotným provedením letu včetně komunikace s řídicím, pilotáží apod.

Úsilí (EF, effort) odráží, jak namáhavé pro subjekt bylo podat daný výkon (Hart, 2006; Hart & Staveland, 1988). V kontextu experimentu otázka reflektuje, jak namáhavé (fyzicky i psychicky) bylo např. dodržování pokynů řídicího, udržení přesnosti pilotáže apod.

Frustrace (FR, frustration) odráží, jak stresovaný, rozhořčený, otrávený, nejistý a odrazený nebo jistý, potěšený, spokojený a relaxovaný se subjekt cítil v průběhu plnění úkolu (Hart, 2006; Hart & Staveland, 1988). V kontextu

experimentu otázka reflektuje např. jak podrážděný a otrávený se subjekt cítil v důsledku monotónního letu nebo únavy.

Po zhodnocení indexů jednotlivých kategorií dochází k porovnání jednotlivých kategorií navzájem. Subjektu jsou představeny všechny možné dvojice kategorií (15 kombinací) přičemž subjekt vybírá tu, která jeho výkon ovlivnila nejvíce. Frekvence výběru jednotlivých kategorií je vydělena 15, čímž je spočten váhový faktor pro každou kategorii. Následně je získáno celkové skóre a hodnoty pro jednotlivé kategorie, které jsou dány právě váhovaním a upřednostněním jednotlivých kategorií (Hart & Staveland, 1988). Hodnocení kategorií na stupnici 0–20 je nejprve převedeno na procenta, přičemž každý krok stupnice představuje 5 %. Pro každou kategorii  $i$  je pak získáno celkové skóre ( $OS_i$ ) jako:

$$OS_i = \frac{x_i \cdot p_i \cdot 5}{15}, \quad (3.23)$$

kde  $x_i$  je zvolená hodnota kategorie  $i$  v rozmezí 0–20 a  $p_i$  je frekvence výběru kategorie  $i$  mezi 15 páry. Celkové skóre ( $NTLx$ ) je pak získáno jako:

$$NTLx = \sum_{i=1}^6 OS_i. \quad (3.24)$$

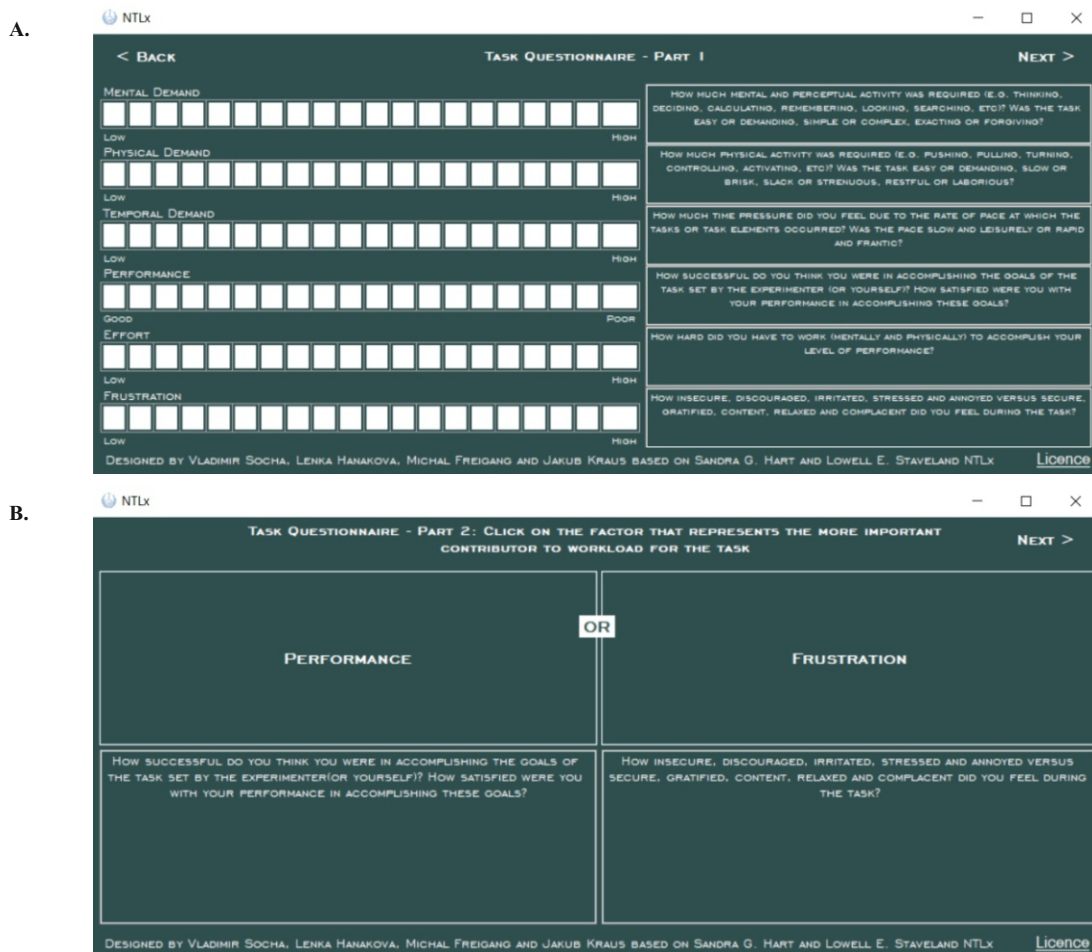
Pro každé testování je tedy získáno skóre šesti jednotlivých kategorií a celkové skóre ( $NTLx$ ). Pro účely tohoto testu byla vytvořena softwarová implementace – NASA-TLx, viz obr. 3.9.

#### Software

Socha, V., **Hanáková, L.**, Freigang, M. & Kraus, J. (2018). NTLx. Software. Czech Technical University in Prague.

### 3.7.2 Dotazník pro hodnocení únavy

Jelikož je NASA-TLX primárně určen pro sledování zátěže, zdálo se jako vhodné tento dotazník modifikovat tak, aby byl zacílen na subjektivní hodnocení únavy. Upravený dotazník (Herman, 2020) vychází z NASA-TLX, a tedy jeho princip je stejný. Taktéž zahrnuje 6 kategorií hodnocených na stupnici 0–20, které jsou následně váhovány výběrem z 15 párů. Stejným způsobem jako u NASA-TLX se také počítá celkové skóre jednotlivých kategorií ( $OS_i$ ) i celkové skóre dotazníku ( $TLx$ ). Obměna dotazníku souvisí s jednotlivými kategoriemi, resp. v jejich definici, viz obr. 3.10.



Obrázek 3.9: Softwarová implementace NASA Task Load Index - NTLx. část 1 – hodnocení kategorií na stupnici 1 - 20 (A), část 2 - porovnávání kategorií (B).

Psychická náročnost (MD, mental demand) se obdobně jako v NASA-TLX váže k psychické náročnosti úkolu a otázka byla formulována tak, že se subjektu tázala, zda se mu dařilo držet pozornost v průběhu celého měření.

Fyzická náročnost (PD, physical demand) byla opět obdobná jako v NASA-TLX, tedy zaměřená na fyzickou náročnost vykonávání úkolu a otázka byla formulována tak, že se subjektu tázala, zda se po měření cítil fyzicky vyčerpaný.

Časová náročnost (TD, temporal demand) se stejně jako NASA-TLX orientovala na časovou náročnost úkolu a otázka byla formulována tak, že se subjektu tázala, zda pocítoval časový tlak.

Výkonnost (PF, performance) se obdobně jako NASA-TLX zaměřovala na podaný výkon, zejména tedy na přesnost letu, a otázka byla formulována tak, že se subjektu tázala, zda mu nedělalo obtíže plnit pokyny řídicího.

Úsilí (EF, effort) bylo zaměřeno, obdobně jako v případě NASA-TLX, tj. jak náročné pro subjekt bylo řešit nestandardní nerutinní operace, případně při

**Subjektivní hodnocení únavy**

Jméno: \_\_\_\_\_  
Čas letu: \_\_\_\_\_

Psychické vypětí (Mental demand) – Náročnost na přemýšlení, koncentraci, při řešení situací. Vymýšlení postupů.  
Fyzická náročnost (Physical demand) – Kolik sil vás stálo vzletnutí, přistání, opravy při vylučování větru?  
Časová tíseň (Temporal demand) – Měl jste dostatek času při řešení úkolu?  
Kvalita výkonu (Performance) – Přesnost letu, přiblížení  
Náročnost úkolu (Effort) – Nestandardní, nerutinní postupy, něco, s čím nejste sžitý?  
Stres (Frustration) – Byl jste pod tlakem, nervózní, podrážděný?

Odpověď označte křížkem

Co ovlivnilo váš výkon více? (Nehodící se škrtněte)

Časová tíseň	Náročnost úkolu
Náročnost úkolu	Fyzická náročnost
Psychické vypětí	Časová tíseň
Kvalita výkonu	Náročnost úkolu
Časová tíseň	Stres
Náročnost úkolu	Psychické vypětí
Stres	Náročnost úkolu
Fyzická náročnost	Časová tíseň
Kvalita výkonu	Fyzická náročnost
Psychické vypětí	Kvalita výkonu
Stres	Psychické vypětí
Časová tíseň	Kvalita výkonu
Fyzická náročnost	Psychické vypětí
Fyzická náročnost	Stres
Stres	Kvalita výkonu

1. **Dařilo se Vám držet pozornost po celý let?**  
 Vůbec  Velmi
2. **Cítíte se fyzicky zesláblý?**  
 Vůbec  Velmi
3. **Cítíte časovou tíseň?**  
 Žádnou  Velikou
4. **Dařilo se Vám plnit pokyny řídicího?**  
 Vůbec  Bez problému
5. **Byl let náročný na techniku pilotáže?**  
 Vůbec  Velmi
6. **Byl jste během letu ve stresu?**  
 Vůbec  Velmi

Obrázek 3.10: Subjektivní hodnocení únavy prostřednictvím modifikovaného dotazníku NASA Task Load Index (Herman, 2020).

procedurách, na něž není zvyklý a otázka byla formulována tak, že se subjektu tázala, zda pro něj zadaný let byl náročný z hlediska techniky pilotáže.

Stres (ST, stress) byl obdobou kategorie FR v NASA-TLX a byl zaměřen na to, zda subjekt v průběhu měření pocítoval stres, tlak nebo frustraci. Otázka byla formulována tak, že se subjektu tázala, zda byl v průběhu letu relaxovaný.

Je tedy zřejmé, že dotazník má silnou vazbu na NASA-TLX a v podstatě pouze rozšiřuje, resp. specifikuje jednotlivé kategorie a otázky na ně vázané tak, aby byl specificky zaměřen na únavu a výkonnost pilota v průběhu letu a subjekt si nemohl otázku vyložit jiným způsobem, viz obr. 3.10. Z tohoto důvodu lze předpokládat, že výsledky obou dotazníků budou obdobné. Výsledky dotazníku byly prezentovány na mezinárodní konferenci New Trends in Civil Aviation:

## Příloha D

**Hanáková, L.,** Socha, V., Bořil, J., Herman, L., Štrobl, J., & Pilmannová, T. (2020). Subjective Evaluation of Fatigue in Pilots During 24-hour Testing. In *2020 New Trends in Civil Aviation (NTCA)* (pp. 111-116). IEEE. DOI: 10.23919/NTCA50409.2020.9291239

### 3.8 Klasifikace úrovně únavy

S cílem rozlišení jednotlivých úrovní únavy bylo přistoupeno k metodám strojového učení s cílem klasifikace jednotlivých měření (která by měla reflektovat různé úrovně únavy) na základě experimentálně sesbíraných dat. Vstupní dataset sestával z parametrů, které v základní statistické analýze prostřednictvím analýzy rozptylu pro opakovaná měření (rANOVA) nebo Friedmannova testu vykazovaly statisticky signifikantní rozdíly mezi některými měřeními. Tímto způsobem byl vybrán vstupní dataset pro následný výběr s ohledem na importanci jednotlivých parametrů, který zahrnoval parametry meanNN, minHR, maxHR, SDNN, nLF, SD2, Area (Poincaré graf), výběrový reakční čas (4choice), všechny úlohy OrTestu, všechny indexy NASA-TLX a plochu odchylky od ideální trajektorie. Cílovou kategorickou proměnnou byla jednodivá měření, tj. M1–M8.

V prvním kroku byly selektovány parametry s největší importancí pro klasifikaci jednotlivých měření. Ta byla určována na základě informačního zisku (IG, information gain), který reflektuje rozdíl entropie pro celá data a pro uvažovaný atribut, tj. snížení entropie po rozdělení dat dle daného atributu, přičemž vyšší hodnota rozdílu odpovídá vyššímu IG a tedy vyšší „importanci“ atributu (Quinlan, 1986). Jelikož je cílem umožnit implementaci přístupu do reálné praxe, výsledný výběr byl omezen tak, aby parametry pocházely z co možná nejmenšího počtu měření/testů, tj. aby případná klasifikace byla založena na co nejméně sbíraných datech a současně aby se jednalo o co možná nejjednodušší měření. Jako nejvýznamější se ukázaly jednotlivé úlohy OrTestu, MD, PD, EF a NTLx z dotazníku NASA-TLX, parametr maxHR a výběrový reakční čas 4choice. S ohledem na dříve uvedené byly vybrány zmíněné parametry NASA-TLX a 4 úlohy OrTestu, jelikož ostatní parametry nezvyšovaly přesnost klasifikace a současně rozšiřovaly sadu parametrů o další datové sady. S cílem co nejpřesnější klasifikace bylo vyzkoušeno několik relevantních technik, využívajících zvolené parametry, načež byla vybrána ta, která vykazovala nejvyšší přesnost klasifikace.

První využitou metodou byl algoritmus rozhodovacích stromů s dopředným prořezáváním. Strom je jednoduchý algoritmus, předchůdce náhodného lesa, který rozdělí data do uzlů podle informačního zisku pro kategorické a střední kvadratická chyby pro číselné cílové proměnné. Bližší popis rozhodovací stromů uvádí např. Myles et al. (2004).

Druhým přístupem byl náhodný les, využívající predikci na základě stanoveného počtu rozhodovacích stromů. Náhodný les představuje metodu učení souboru používanou pro klasifikaci, regresi a další úlohy. Náhodný les



vytváří soubor rozhodovacích stromů. Každý strom je vytvořen na základě bootstrap vzorku (výběr s opakováním) z trénovacích dat. Při vývoji jednotlivých stromů se vybírá libovolná podmnožina atributů (proto náhodný les), z níž se vybere nejlepší atribut pro rozdělení. Konečný model je založen na většinovém rozhodnutí individuálně vyvinutých stromů v lese (Breiman, 2001).

Další využitou metodou byly neuronové sítě. Využitý software používá algoritmus vícevrstvého perceptronu z Pythonovského projektu sklearn, resp. scikit-learn (Pedregosa et al., 2011). Vícevrstvý perceptron (MLP) je algoritmus učení pod dohledem, který se učí trénováním na souboru dat, přičemž může učit nelineární aproximátor funkce pro klasifikaci nebo regresi. Specifikem MLP je, že mezi vstupní a výstupní vrstvou může být jedna nebo více nelineárních vrstev, tzv. skrytých vrstev. Využitá knihovna implementuje algoritmus MLP, který se trénuje pomocí zpětného šíření. Detaily využití implementace jsou uvedeny v dokumentaci knihovny scikit (scikit-learn developers, 2019).

Algoritmus kNN, pak hledá  $k$  nejbližších trénovacích příkladů v prostoru atributů a používá jejich průměr jako predikci. Zjednodušeně, je nalezeno  $k$  nejbližších datových bodů a jejich label (kategorie) je využit k předpovědi labelů nových datových bodů. jedná se o jeden z nejjednodušších algoritmů strojového učení, jehož detaily popisuje např. Taunk et al. (2019).

Naivní Bayesův klasifikátor předpokládá, že všechny atributy jsou vzhledem ke kontextu kategorie na sobě nezávislé. Jedná se o tzv. naivní Bayesův předpoklad. Naivní Bayes využívá pravděpodobnostní model, detaily popisuje McCallum et al. (1998).

Podpůrné vektorové sítě (SVM, support vector machines) jsou technika strojového učení, která odděluje prostor atributů pomocí hyperplochy, čímž maximalizuje rozpětí mezi instancemi různých tříd nebo hodnotami tříd. Využitá implementace pochází z balíčku LIBSVM, přičemž Chang & Lin (2011) poskytují detailní popis implementace.

U všech použitých metod strojového učení byl následně proveden iterační proces učení a následné klasifikace. Počet těchto cyklů byl 100, přičemž 70 % dat bylo použito pro trénování a 30 % pro testování. Tento poměr byl zachován v každé iteraci, ale výběr dat pro trénink byl náhodný. Zbývající data (30 %) byla poté použita k testování modelu z hlediska přesnosti a správnosti. Výše uvedený náhodný výběr dat byl stratifikovaný, což znamená, že byla snaha zachovat podíl vybraných dat vzhledem k cílové kategorii. Stratifikace výběru byla provedena především proto, že při náhodném výběru se může stát, že konečný výběr dat nebude obsahovat údaje pro učení vůči určité kategorii.

Celý tento interaktivní proces trénování a testování zajistil, že bylo možné ohodnotit objektivitu konkrétního klasifikátoru. V případě vykonání např. jen

jediného natrénování modelu a jeho ověření, není možné určit, zda model pracuje správně nebo zda je jeho klasifikační přesnost zatížena náhodným výběrem dat. Model by tedy nebylo možné zobecnit.

Uvedený přístup umožňoval i výpočet průměrné správnosti klasifikace. Správnost klasifikace je metrika, která shrnuje výkon modelu klasifikace jako počet správných předpovědí dělený celkovým počtem předpovědí. Snadno se počítá a je intuitivní na pochopení, což z ní dělá nejběžnější metriku používanou pro hodnocení modelů klasifikátorů. Dalšími důležitými údaji pro popis klasifikátoru, které byly vypočítány, byla přesnost a senzitivita. Senzitivita ( $R$ ) a přesnost ( $P$ ) je vyjádřena jako:

$$P = \frac{TP}{TP + FP} \quad (3.25)$$

$$R = \frac{TP}{TP + FN}, \quad (3.26)$$

kde  $TP$  reprezentuje počet skutečně pozitivních klasifikací,  $FP$  počet falešně pozitivních klasifikací a  $FN$  počet falešně negativních klasifikací. Tyto metriky byly dále využity pro zhodnocení natrénovaných klasifikačních modelů a výběr nejlepšího modelu.

Celé výše uvedené zpracování a modelování probíhalo v prostředí Orange data mining v.3 (University of Ljubljana, Lublaň, Slovinsko) a Matlab 2021a (MathWorks, Natick, MA, USA).

## 4. Výsledky

### 4.1 Letová data

Hodnocení letových dat proběhlo pro obě skupiny ve dvou zmiňovaných úsecích prostřednictvím plochy odchylky od ideální trajektorie normalizované na jednotkovou délku. V rámci hodnocení bylo využito analýzy rozptylu pro opakovaná měření (rANOVA). Porovnání meziskupinové i vnitroskupinové vycházelo z Wilcoxonova zápisu:

$$M1 - M8 \sim GrSection, \quad (4.1)$$

kde  $M1 - M8$  označuje jednotlivá měření v čase,  $GrSection$  pak označuje kategorickou cílovou proměnnou, nabývající hodnoty  $Gr_S U_x$ , kde  $S = [A, B]$  a  $x = [1, 2]$ . S ohledem na nesplnění předpokladů sféricity dat jsou prezentovány  $p$ -hodnoty upravené Greenhouse-Geisser korekcí. Následná post-hoc analýza využívá Bonferroniho korekci.

Výsledky rANOVA indikují, že mezi kategoriemi  $GrSection$  existuje statisticky signifikantní rozdíl ( $F(21,98) < 0,001$ ;  $p < 0,001$ ). Post-hoc analýza ukazuje na 4 rozdíly mezi skupinou A a B. Pro úsek 1 byly rozdíly nalezeny u M7 ( $p = 0,013$ ; CI = (0,0001; 0,0011)) a M8 ( $p = 0,022$ ; CI = (-0,0023; -0,0001)). Pro úsek 2 pak u měření M1 ( $p = 0,033$ ; CI = (0,00004; 0,00109)) a M6 ( $p = 0,015$ ; CI = (-0,0031; -0,0003)). Kategorická proměnná  $GrSection$  umožnila také porovnání mezi jednotlivými úseky v rámci jedné skupiny. V tomto případě byl post-hoc analýzou nalezen 1 rozdíl u skupiny A, u skupiny B pak 4. V případě skupiny A se jednalo o M3 ( $p = 0,021$ ; CI = (-0,00113; -0,00008)). V případě skupiny B pak M2 ( $p = 0,006$ ; CI = (-0,0038; -0,0006)), M6 ( $p = 0,001$ ; CI = (-0,0034; -0,0009)), M7 ( $p = 0,046$ ; CI = (-0,001031; -0,000008)) a M8 ( $p = 0,031$ ; CI = (0,0001; 0,0024)).

Z hlediska porovnání jednotlivých měření v rámci příslušné skupiny/úseku, byl nalezen jeden rozdíl u skupiny A, specificky v druhém měření. Skupina B pak vykazovala nejvíce rozdílů v druhém úseku, viz tab. 4.1.

Distribuce a trendy letových dat jsou prezentovány na obr. 4.1A. Z trendů je patrný výrazný rozdíl zejména mezi úseky 1 a 2 v případě skupiny B, přičemž trend chyby je mezi těmito úseky vzájemně reverzní, viz obr. 4.1B. V případě skupiny A je pak průběh mezi oběma úseky více podobný, přičemž maximum chyby (tj. nejnižší výkonnost) nastává pro úsek 1 v M6 a M7, pro úsek M6 je

pak taktéž pozorován výrazný nárůst chyby vyjádřené normalizovanou plochou odchylky od ideální trajektorie.

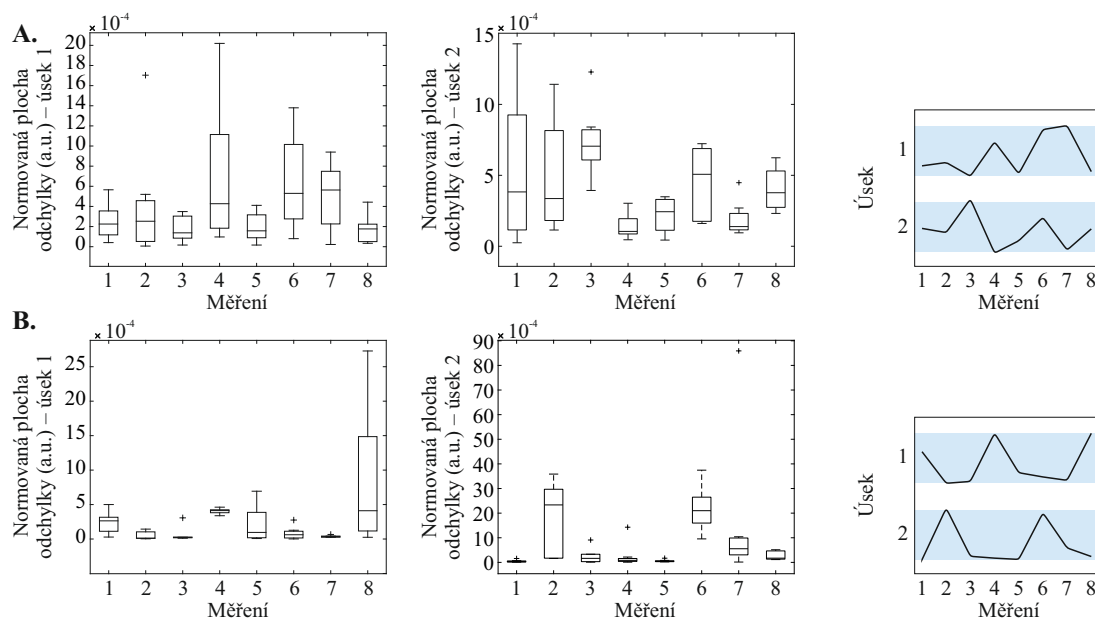
Tabulka 4.1: Vnitroskupinové porovnání letových dat pro obě skupiny prostřednictvím post-hoc analýzy. V tabulce jsou uvedeny pouze statisticky signifikantní rozdíly na hladině významnosti 5 %.

S	M	p	CI-L	CI-U	S	M	p	CI-L	CI-U
A2	3-5	0,026	< 0,001	0,001	B2	2-7	0,015	< 0,001	0,003
B1	1-8	0,019	-0,002	-0,001	B2	2-8	0,005	0,001	0,004
B1	3-8	0,009	-0,003	-0,001	B2	3-6	0,002	-0,003	-0,001
B1	7-8	0,011	-0,003	-0,001	B2	4-6	0,007	-0,003	-0,000*
B2	1-2	0,001	-0,004	-0,001	B2	5-6	< 0,001	-0,003	-0,001
B2	1-6	< 0,001	-0,003	-0,001	B2	5-7	0,035	-0,001	-0,000*
B2	2-3	0,003	0,001	0,003	B2	5-7	0,002	0,001	0,003
B2	2-4	0,004	0,001	0,003	B2	6-8	0,003	0,001	0,003
B2	2-5	0,001	0,001	0,004					

S – skupina/úsek | M – měření | p – p-hodnota

CI-U, CI-L - meze konfidenčního intervalu (horní, dolní)

\* negativní hodnota blížící se 0



Obrázek 4.1: Grafické znázornění distribuce výsledků hodnocení letu prostřednictvím normalizované plochy výchylny od ideální trajektorie a porovnání trendů pro první (A) i druhou (B) skupinu. Křížky značí odlehlé hodnoty.

## 4.2 Fyziologická měření

Fyziologická data, specificky záznamy elektrické aktivity srdeční a elektrické aktivity mozku byly kolektovány v průběhu simulovaných letů. Distribuce jednotlivých parametrů je prezentována formou boxplotů. Statistická analýza probíhala na hladině významnosti  $\alpha = 5\%$ . Pro účely redukce dimenzionality

byl v případě EEG zvolen postup s využitím analýzy hlavních komponent. Všechna hodnocená data, byla s ohledem na interindividuální variabilitu standardizována prostřednictvím z-score a to přes jednotlivé subjekty.

### 4.2.1 Variabilita srdečního rytmu

Jelikož skupina B absolvovala let s mezipřistáním, nebylo možné hodnotit EKG naměřené v průběhu celých letů, a to právě s ohledem na odlišné letové profily mezi skupinami. Z tohoto důvodu bylo EKG hodnoceno v porovnatelných úsecích – traťových letech. V každém letu byly identifikovány dva úseky traťových letů (viz kap. 3.3), které jsou vzájemně porovnatelné. Hodnocení HRV proběhlo pro obě skupiny ve dvou zmiňovaných úsecích prostřednictvím sady parametrů z časové, frekvenční a nelineární analýzy. V rámci hodnocení bylo využito analýzy rozptylu (ANOVA) pro opakovaná měření (rANOVA). Porovnání meziskupinové i uvnitř skupin vycházelo z Wilcoxonova zápisu (4.1).

Na základě statistické analýzy s využitím rANOVA (Greenhouse-Geisser korekce p-hodnoty) nebyl nalezen statisticky signifikantní rozdíl mezi středními hodnotami hodnocených úseků, a to jak pro skupinu A, tak pro skupinu B, jelikož  $p < 0,05$  pro všechny sledované parametry. Rozdíl mezi oběma skupinami byl pak nalezen pouze v M4. Specificky se jednalo o parametr LF ( $F(21,56) = 2,187$ ;  $p = 0,049$ ), přičemž rozdíl byl na základě post-hoc analýzy nalezen jak v prvním ( $p = 0,026$ ; CI = (0,169; 2,418)), tak druhém ( $p = 0,025$ ; CI = (0,263; 3,556)) sledovaném úseku.

V kontextu porovnání jednotlivých měření v rámci skupiny a sledovaného úseku byly na základě výsledků rANOVA nalezeny statisticky signifikantní rozdíly mezi měřeními u 12 parametrů. U 4 parametrů (RMSSD, pNN50, DFA  $\alpha_1$ , SD1) pak post-hoc analýza nedokázala identifikovat žádné dvojice měření, mezi nimiž by byl rozdíl, což je dáno konzervativností zvolené metody. Z tohoto důvodu nelze tvrdit, že by existoval statisticky signifikantní rozdíl mezi středními hodnotami jednotlivých měření pro daný parametr v rámci skupiny a sledovaného úseku. Výsledky post-hoc analýzy pro zbylých 8 parametrů jsou prezentovány v tab. 4.2. Nutno podotknout, že signifikantní rozdíly byly nalezeny u parametrů u všech tří metod – hodnocení v časové oblasti, frekvenční oblasti a nelineárními metodami.

Distribuce jednotlivých parametrů analýzy HRV je prezentována prostřednictvím boxplotů. S ohledem na množství sledovaných parametrů a rozdělení výsledků s ohledem na skupinu a hodnocený úsek jsou prezentovány pouze vzorové distribuce. Neprezentované boxploty zahrnující shodné trendy s daty prezentovanými jsou předmětem přílohy E.

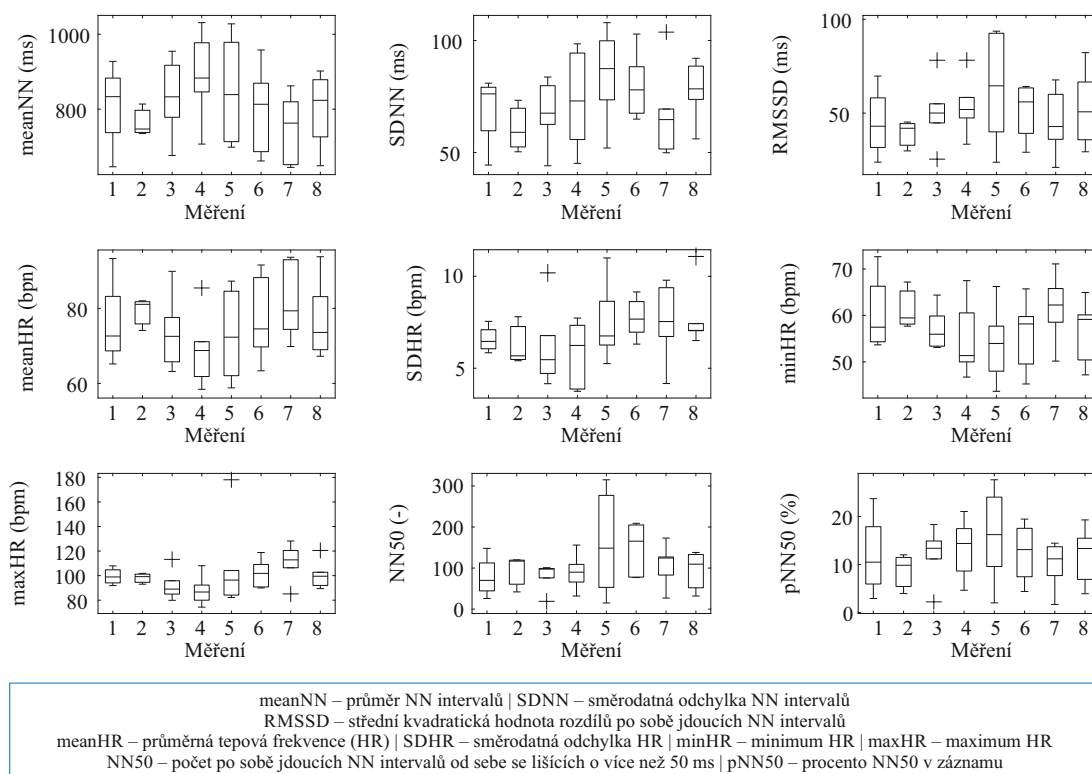
Tabulka 4.2: Vnitroskupinové porovnání jednotlivých parametrů variability srdečního rytmu pro obě skupiny prostřednictvím post-hoc analýzy. V tabulce jsou uvedeny pouze statisticky signifikantní rozdíly na hladině významnosti 5 %.

Var	S	M	p	CI-L	CI-U	S	M	p	CI-L	CI-U
meanNN	A1	2-4	0,003	-3,565	-0,852	B1	1-4	0,013	-3,859	-0,478
	A1	5-7	0,003	0,824	3,395	B1	2-4	0,007	-2,835	-0,485
SDNN	A1	2-4	0,001	-2,644	-0,875	A2	2-4	0,009	-3,612	-0,550
	A1	2-5	0,043	-4,167	-0,062					
meanHR	A1	2-4	0,004	0,756	3,619	B1	1-4	0,015	0,433	4,012
	A1	5-7	0,007	-3,269	-0,561	B1	2-4	0,010	0,407	2,886
minHR	B1	1-5	0,020	0,324	3,829					
maxHR	B1	1-4	0,030	0,154	3,157	B2	1-4	0,002	1,024	4,027
	B1	2-4	0,017	0,230	2,304	B2	2-4	0,010	0,358	2,431
Total	A1	2-4	0,000	-2,608	-0,947	A2	2-4	0,003	-3,730	-0,853
	A1	2-5	0,032	-4,314	-0,190					
nLF	B1	3-8	0,035	-2,343	-0,084					
SD2	A1	2-4	0,001	-2,533	-0,829	A2	2-4	0,007	-3,576	-0,626
Area	A1	2-4	0,017	-3,489	-0,351	A1	2-5	0,046	-5,175	-0,044

Var – parametr | S – skupina/úsek | M – měření | p – p-hodnota | CI-U, CI-L – meze konfidenčního intervalu  
meanRR (HR) – průměr NN intervalů (tepové frekvence) | SDNN – směrodatná odchylka NN intervalů  
minHR (maxHR) – minimální (maximální) tepová frekvence  
Total – celkový spektrální výkon | nLF – normalizovaný výkon v pásmu nízkých frekvencí  
SD2 – délka hlavní poloosy konfidenční elipsy (Poincaré plot) | Area – plocha konfidenční elipsy

Na základě distribucí lze konstatovat, že parametry sledují obdobný trend, jako dříveji prezentovaná data. K největším nárůstům (případně poklesům) mediánů dochází převážně okolo M4–M5, s ohledem na sledovaný parametr, u skupiny A, viz obr.4.2. Všeobecný trend tedy sleduje nejprve pokles parametru (M2) s následným postupným nárůstem ke zmíněnému vrcholu, následovaný poklesem a v případě posledního měření opětovný nárůstem (příp. inverzně s ohledem na charakter parametru). Obdobný trend je pak pozorován také v případě skupiny B, pouze bez počátečního poklesu (nárůstu) v M2 a pozvolnějším nárůstem k vrcholu, který je převážně v M5, viz obr. 4.2–4.8.

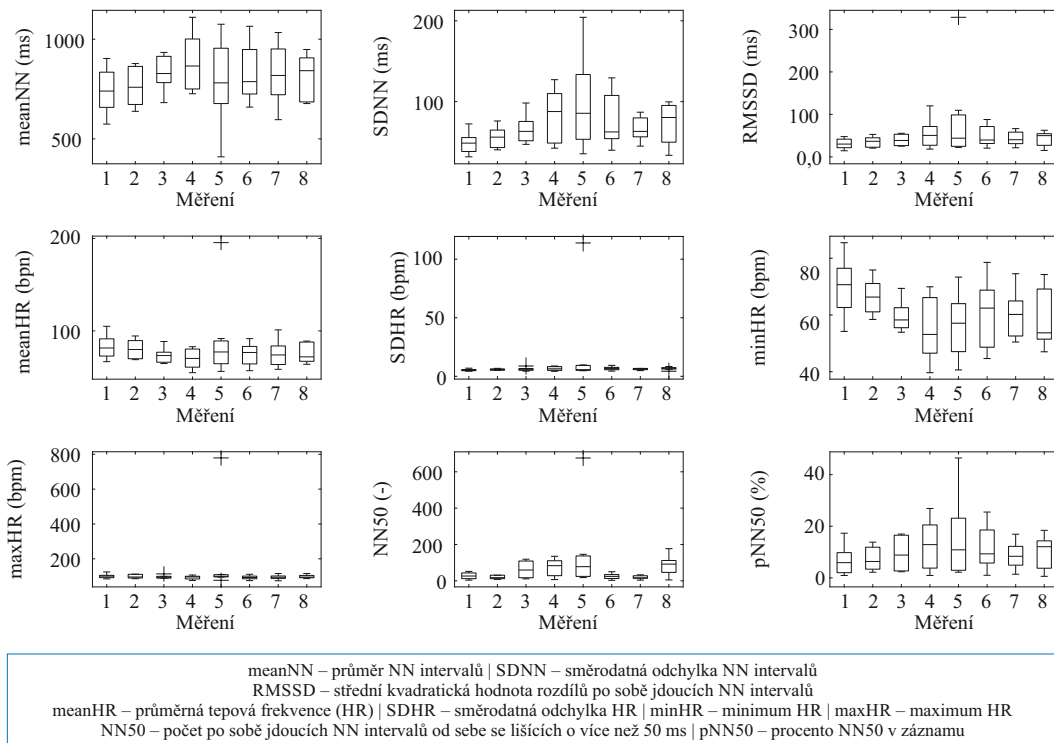
Tento trend je velmi dobře pozorovatelný např. u parametru meanNN (resp. meanHR), přičemž je možné sledovat tento trend ve všech 4 hodnocených úsecích, tj. postupný nárůst průměrného RR intervalu k M4 (přičemž u skupiny A je nejprve pozorován pokles v M2). Uvedené svědčí o snižování tepové frekvence, přičemž parametr meanHR dosahuje v M4 svého minima. Tyto projevy mohou být odrazem zvýšené aktivity parasymptiku.



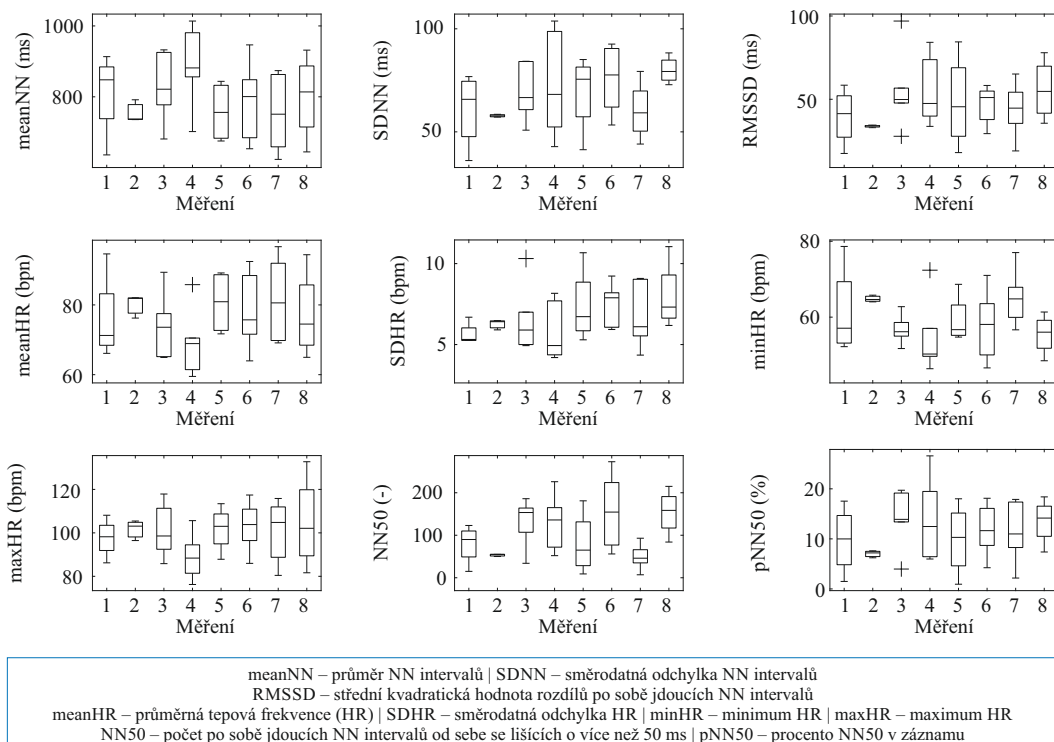
Obrázek 4.2: Grafické znázornění distribuce výsledků časové analýzy variability srdečního rytmu u skupiny A, úsek 1 v průběhu všech měření formou boxplotů. Křížky značí odlehlé hodnoty.

Následuje zkrácení průměrných NN intervalů přes M5 a M6 a následný nárůst v oblasti M7 a M8. Směrodatná odchylka hovořící o variabilitě pak sleduje ve všech 4 úsecích tento trend také, avšak maximum směrodatné odchylky, tj. největší variabilita, je pozorována v M5, přičemž právě nárůst variability svědčí o nižší aktivitě sympatiku. Minimální hodnota HR pak sleduje trend opačný k meanNN, tedy minimum v oblasti M4 (příp. M5 pro skupinu B, úsek 2), což souvisí také s poklesem meanHR, resp. nárůstem meanNN v tomto měření. Stejný trend je pozorován také pro maximální hodnotu HR, viz obr. 4.2–4.5.

Frekvenční analýza prezentovaná formou celkového výkonu ve spektru (Total) vykazuje u šech sledovaných úseků postupný nárůst až k maximálnímu výkonu, který byl pozorován u skupiny A v M5–M6, u skupiny B pak v M5–M6. Maximální hodnota parametru Total je následována postupným poklesem v M6–M7 (skupina A), resp. M5–M6 (skupina B), načež následuje opětovný nárůst směrem k M8. Parametr LF/HF, který reflektuje sympatovagální rovnováhu, pak dosahuje svého minima v M4 shodně pro obě skupiny. Parametr nLF, který primárně reflektuje aktivitu sympatiku, dosahuje svého minima v M4–M5 u skupiny A a v M4 u skupiny B, viz obr. 4.7 a 4.7.



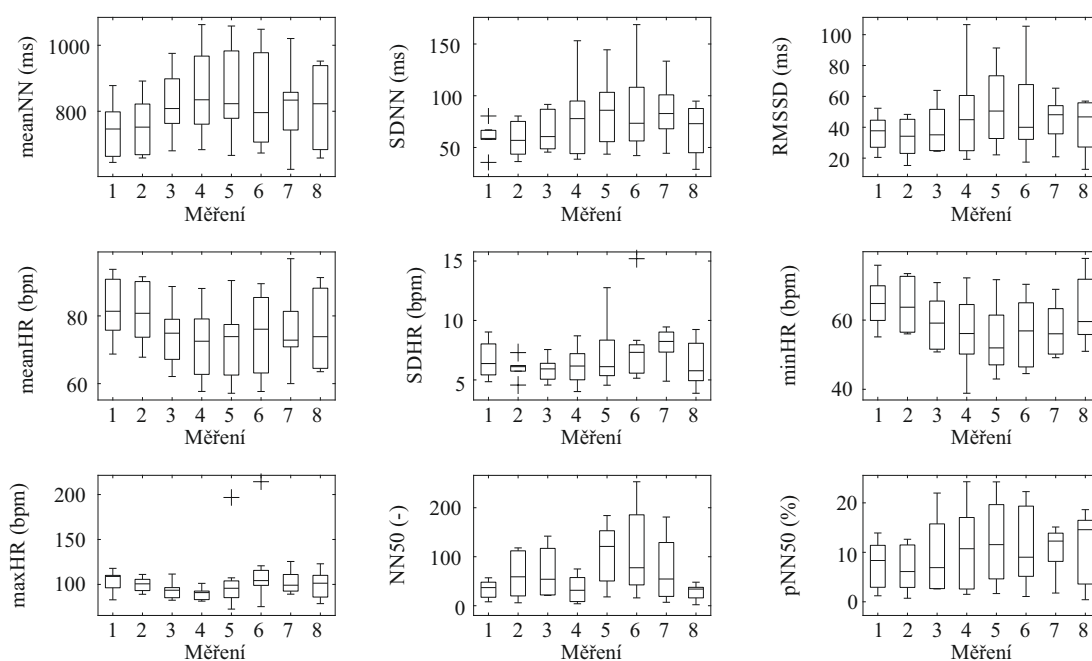
Obrázek 4.3: Grafické znázornění distribuce výsledků časové analýzy variability srdečního rytmu u skupiny B, úsek 1 v průběhu všech měření formou boxplotů. Křížky značí odlehlé hodnoty.



Obrázek 4.4: Grafické znázornění distribuce výsledků časové analýzy variability srdečního rytmu u skupiny A, úsek 2 v průběhu všech měření formou boxplotů. Křížky značí odlehlé hodnoty.

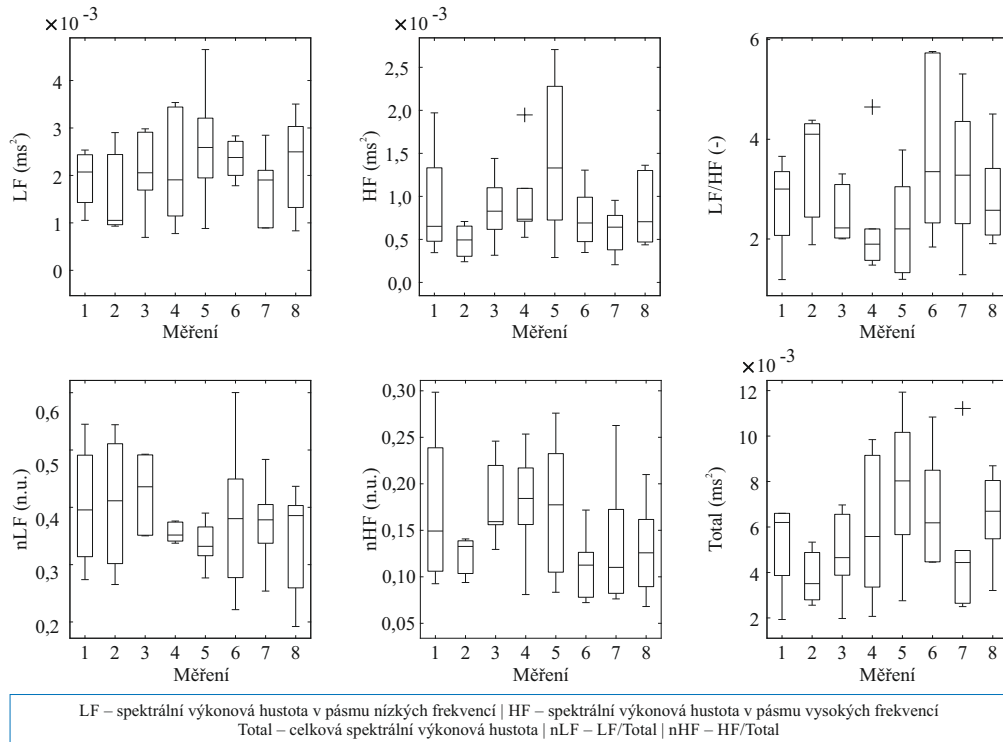


Dříve popsany trend je pozorován také v případě nelineární analýzy. Specificky pro SD2 je pozorován postupný nárůst k maximu v M5–M6 pro skupinu A a M4–M5 pro skupinu B, přičemž u skupiny A je nejprve pozorován prudší pokles v M2, což také svědčí o skutečnosti, že u skupiny A jsou M2 a M4 vzájemně odlišitelné na základě střední hodnoty dle rANOVA a následné post-hoc analýzy. Po dosažení maximální hodnoty dochází k poklesu k M7 a opětovnému nárůstu SD2 v M8, a to u obou skupin. Parametr SD2 reflektuje dlouhodobou variabilitu, koreluje s LF a je negativně ovlivňován parasympatikem. Plocha konfidenční elipsy pak reflektuje celkovou variabilitu, a v případě jednotlivých měření sleduje dříve popsany trend, přičemž maximum je pozorováno v M5, viz obr. 4.8.

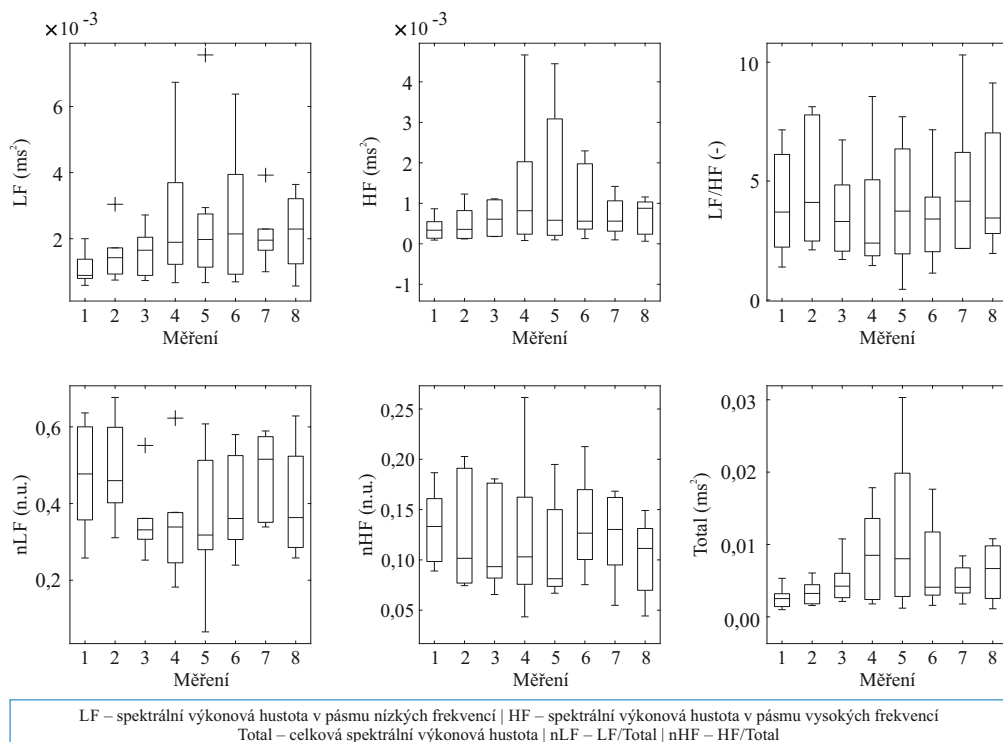


meanNN – průměr NN intervalů | SDNN – směrodatná odchylka NN intervalů  
 RMSSD – střední kvadratická hodnota rozdílů po sobě jdoucích NN intervalů  
 meanHR – průměrná tepová frekvence (HR) | SDHR – směrodatná odchylka HR | minHR – minimum HR | maxHR – maximum HR  
 NN50 – počet po sobě jdoucích NN intervalů od sebe se lišících o více než 50 ms | pNN50 – procento NN50 v záznamu

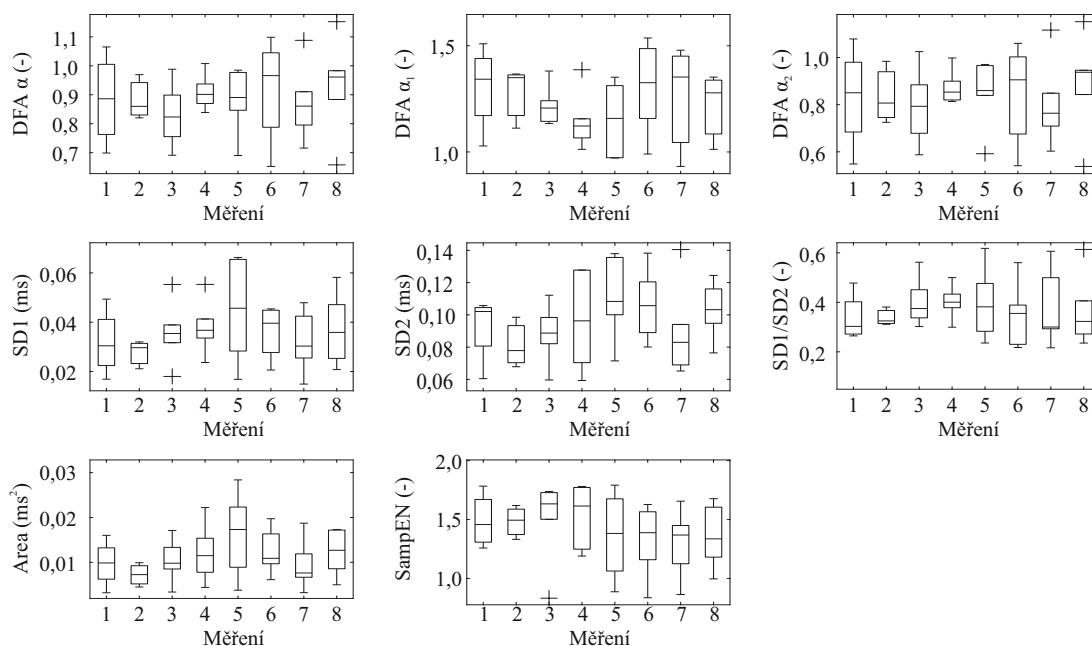
Obrázek 4.5: Grafické znázornění distribuce výsledků časové analýzy variability srdečního rytmu u skupiny B, úsek 2 v průběhu všech měření formou boxplotů. Křížky značí odlehlé hodnoty.



Obrázek 4.6: Grafické znázornění distribuce výsledků frekvenční analýzy variability srdečního rytmu u skupiny A, úsek 1 v průběhu všech měření formou boxplotů. Křížky značí odlehlé hodnoty.



Obrázek 4.7: Grafické znázornění distribuce výsledků frekvenční analýzy variability srdečního rytmu u skupiny B, úsek 1 v průběhu všech měření formou boxplotů. Křížky značí odlehlé hodnoty.



DFA  $\alpha_1$ ,  $\alpha_2$  – škálovací exponenty detrendované analýzy fluktuací | SD1 – délka vedlejší poloosy Pointcaré grafu | SD2 – délka hlavní poloosy Pointcaré grafu | Area – plocha elipsy Pointcaré grafu | SampEN – vzorková entropie

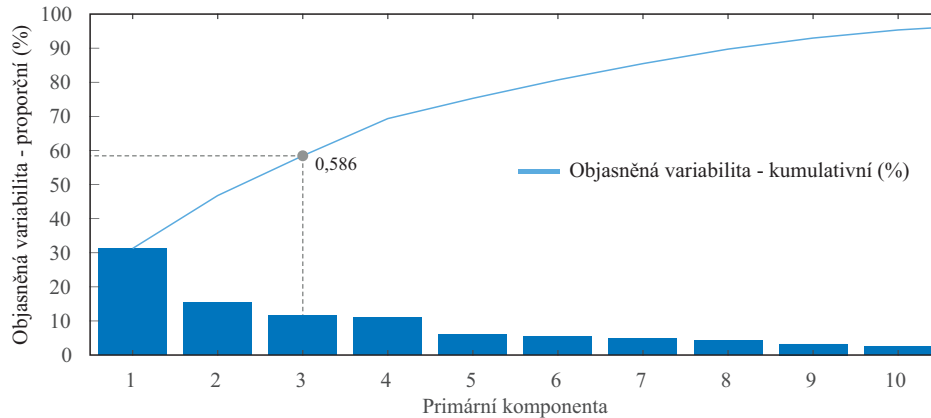
Obrázek 4.8: Grafické znázornění distribuce výsledků nelineární analýzy variability srdečního rytmu u skupiny A, úsek 1 v průběhu všech měření formou boxplotů. Křížky značí odlehlé hodnoty.

## 4.2.2 Elektrická aktivita mozku

Redukce dimenzionality byla provedena prostřednictvím PCA. Jelikož je PCA ortogonální lineární transformace, která transformuje původní data do nového souřadnicového systému takovým způsobem, že nejvíc popsané variability je vždy připisováno prvnímu koordinátu, bude PC1 jako lineární kombinace všech původních atributů vždy popisovat nejvíce variability v datech obsažené (Maćkiewicz & Ratajczak, 1993). Každá další hlavní komponenta je lineární kombinací všech atributů, avšak nekorelovaná s předcházejícími komponentami, a proto bude popisovat variability méně.

Pro popis 95 % variability obsažené v hodnocených datech by bylo třeba 11 hlavních komponent. S ohledem na skutečnost, že PCA je využívána pro redukcii dimenzionality, pro další interpretaci budou využity první tři komponenty, které kumulativně popisují 58,6 % variability obsažené v datech. Specificky pak hlavní komponenta 1 (PC1) popisuje 33,3 %, PC2 13,7 % a PC3 11,6 % variability. Grafická prezentace popsané variability obsažené v datech pomocí jednotlivých komponent, spolu se znázorněním kumulativní variability pomocí grafu úpatí (též tzv. scree plot) je zobrazena na obr. 4.9.

Vykreslením hlavních komponent do bodového grafu (tzv. scatter plot), v tomto případě trojdimenzionálního (pozn.: jedná se o projekci 3D prostoru do

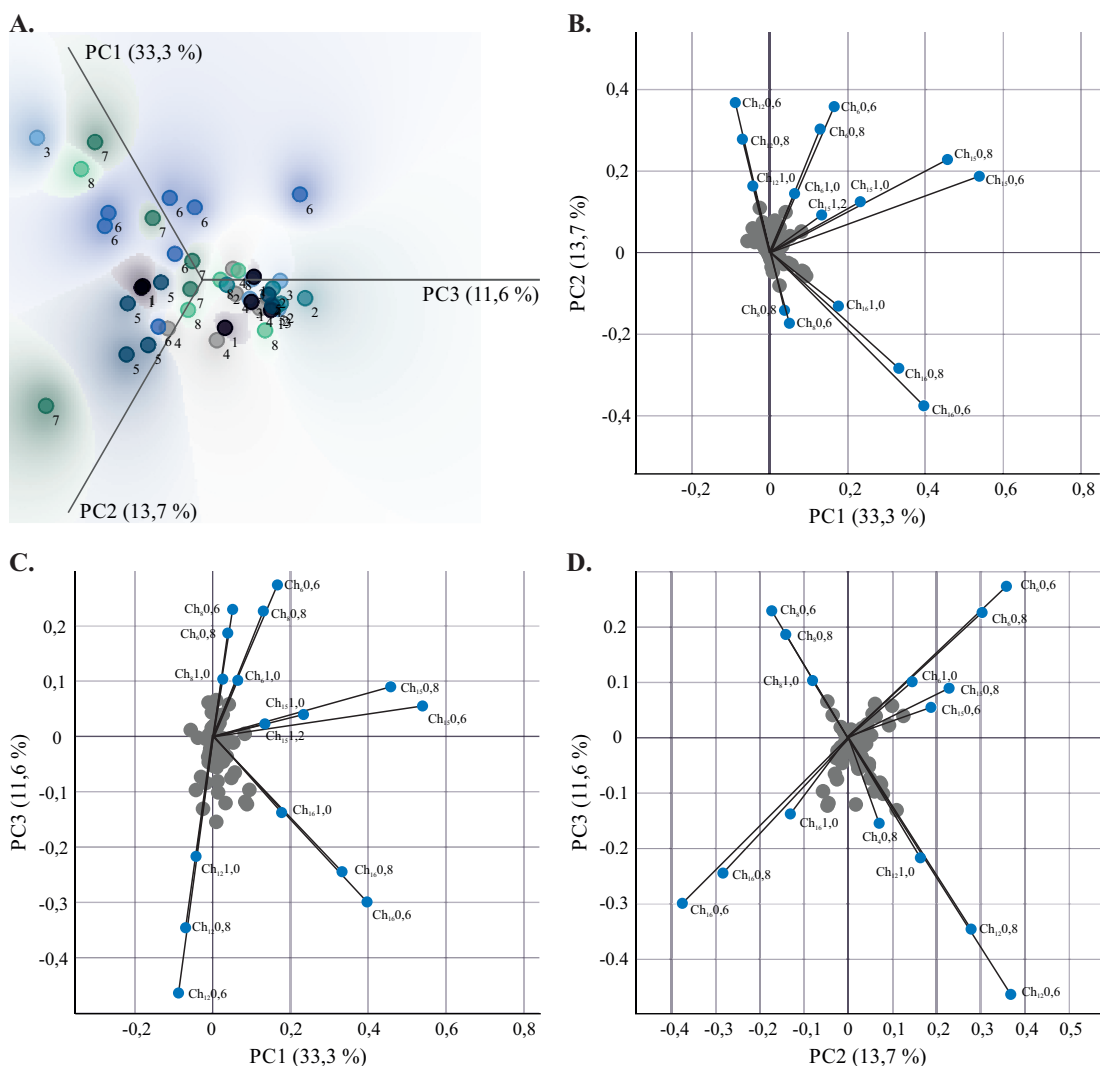


Obrázek 4.9: Proporční variabilita popsaná jednotlivými hlavními komponentami spolu se znázorněním kumulativní variability.

2D, čitelnost grafu je tedy stále závislá na jeho natočení), je možné určit, zda se v datech vyskytují shluky (clustery). Tento graf se nazývá score plot, jeho konkrétní vizualizace je umístěna na obr. 4.10A. Při vykreslení, resp. označení jednotlivých bodů hodnotami cílové proměnné, je z tohoto grafu možné pozorovat tvorbu charakteristických shluků bodů, příslušejících M6. Taktéž je možné pozorovat, že hodnoty M1 až M4 nejsou odděleny vůbec. Hodnoty náležející M5 pak mají taktéž tendenci tvořit shluk.

Jelikož jsou tyto shluky tvořeny transformovanými hodnotami původních proměnných, je možné tyto shluky popsat prostřednictvím hlavních komponent. V rámci uvedeného je nezbytné interpretovat jednotlivé PC na základě směru a magnitudy koeficientů původních proměnných. Pro tyto účely byl využit tzv. loading plot (viz obr. 4.10B–D), znázorňující vliv původní proměnné na jednotlivé PC. Loading plot znázorňuje kovarianci/korelaci původní proměnné a příslušného PC.

Výsledky ukazují, že PC1 má poměrně velkou pozitivní asociaci s původními atributy  $Ch_{15}0,6$ ;  $Ch_{15}0,8$ ;  $Ch_{15}1,0$ ;  $Ch_{15}1,2$ ;  $Ch_{16}0,6$ ;  $Ch_{16}0,8$ ;  $Ch_{16}1,0$ ;  $Ch_60,6$ ;  $Ch_60,8$ . Ve všeobecnosti je možné tvrdit, že zvýšení amplitudy v těchto frekvenčních pásmech na uvedených EEG kanálech má pozitivní vliv na PC1, a tedy i na rozložení score podél této osy (pro význam zkratk a rozložení elektrod při sběru EEG viz kapitolu 3.4.2). PC2 má potom poměrně významnou pozitivní asociaci s  $Ch_{15}0,6$ ;  $Ch_{15}0,8$ ;  $Ch_{15}1,0$ ;  $Ch_{12}0,6$ ;  $Ch_{12}0,8$ ;  $Ch_{12}1,0$ ;  $Ch_{12}1,2$ ;  $Ch_60,6$ ;  $Ch_60,8$ ;  $Ch_61,0$ . PC2 je taktéž významně negativně asociovaná s  $Ch_{16}0,6$ ;  $Ch_{16}0,8$ ;  $Ch_{16}1,0$ ;  $Ch_80,8$ ;  $Ch_80,8$ . PC3 je pozitivně asociovaná s  $Ch_80,6$ ;  $Ch_80,8$ ;  $Ch_81,0$ ;  $Ch_60,6$ ;  $Ch_60,8$ ;  $Ch_61,0$  a negativně s  $Ch_{16}0,6$ ;  $Ch_{16}0,8$ ;  $Ch_{16}1,0$ ;  $Ch_{14}0,6$ ;  $Ch_{12}0,6$ ;  $Ch_{12}0,8$ ;  $Ch_{12}1,0$ ;  $Ch_{12}1,2$ ;  $Ch_40,6$ ;  $Ch_40,8$ ;  $Ch_10,6$  a  $Ch_10,8$ .



Obrázek 4.10: Bodový graf hlavních komponent – Score plot (A) a grafy loadingů v závislosti na hlavních komponentách (C–D).

Významnost asociace atributů s jednotlivými PC byla stanovena pro hodnoty loadingů vyšších (resp. nižších) než  $\pm 0,1$ . Tento práh byl zvolen s ohledem na hlavní shluk loadingů (viz obr. 4.10), který byl považován za bezvýznamný. Pozitivní asociace znamená, že hodnoty PC rostou s rostoucími hodnotami atributu. Negativní pak, že rostoucí hodnoty atributů způsobí pokles hodnot PC.

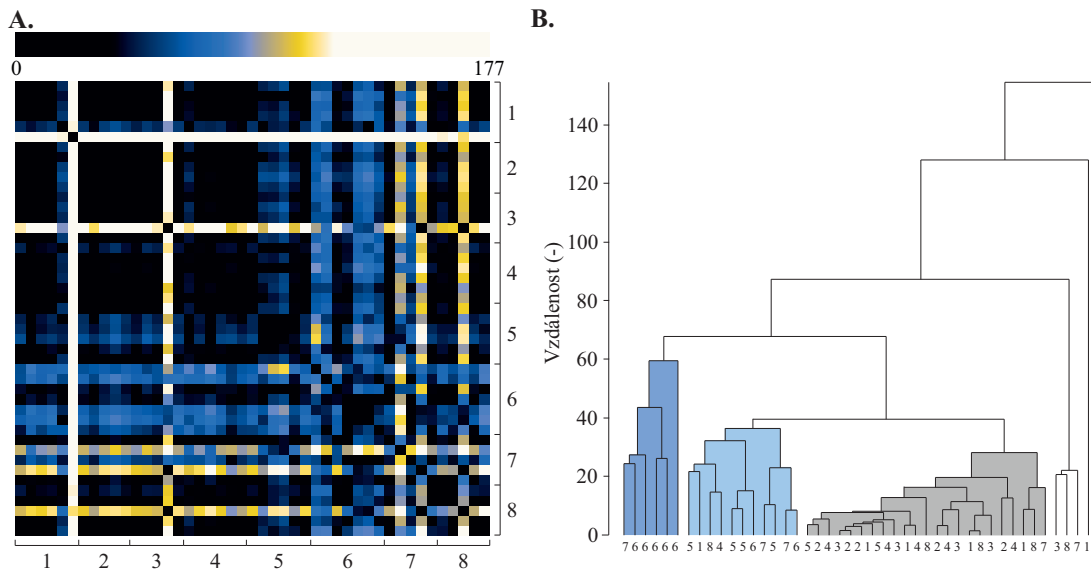
Z analýzy významných atributů, které mimo jiné přispívají k tvorbě shluků reflektujících měření, je vidět, že se jedná zejména o aktivitu v relativně nízkých frekvencích (delta pásmo) na specifických elektrodách. Uvedené bude předmětem diskuze v následující kapitole.

Jak bylo zmíněno výše, interpretace shluků je v 2D projekci problematická. Proto byla vytvořena vzdálenostní matice popisující nejkratší vzdálenosti mezi

dvěma body v 3D prostoru. Pro vyjádření rozestupů dvou bodů v prostoru byl využit výpočet Euklidovské vzdálenosti, a celá matice je tedy definována jako:

$$M_{i,j} = \sqrt{(PC1_i - PC1_j)^2 + (PC2_i - PC2_j)^2 + (PC3_i - PC3_j)^2}, \quad (4.2)$$

kde  $i = 1 \dots N_M$  a  $j = 1 \dots N_M$ , přičemž  $N_M$  je počet měření a zároveň počet bodů tvořících vektor PC1, PC2 a PC3. Takto sestavená vzdálenostní matice je zobrazena na obr. 4.11A.



Obrázek 4.11: Vzdálenostní matice bodů cílové proměnné transformovaných pomocí PCA a definovaných souřadnicemi pomocí hlavních komponent (A) spolu s dendrogramem znázorňujícím hierarchické shlukování (B).

Ze vzdálenostní matice je zřejmé, že vzájemná vzdálenost transformovaných dat na PC ukazuje, že měření 1 až 4 (M1 až M4) tvoří shluk velmi blízkých bodů. Uvedené potvrzuje score plot na obr. 4.10. Dále je možné pozorovat, že převážná většina bodů reflektujících M8 je s tímto shlukem taktéž podobná, resp. vzdálenost mezi těmito body je minimální. Pro zevšeobecnění je tedy možné tvrdit, že měření 1 až 4 a 8 není na základě EEG možné rozlišit. Body charakterizující M6 jsou ve všeobecnosti vzdálenější oproti ostatním, a tvoří samostatný shluk. Datové body pro M5 a M7 jsou poměrně složitě interpretovatelné, jelikož vzdálenostní matice naznačuje, že se nacházejí někde mezi výše popsányi shluky.

Na základě uvedeného je možné tvrdit, že v datech EEG se nacházejí přirozené vzory, které jistým způsobem reflektují měření, a tato měření pravděpodobně reflektují únavu. Každopádně, tak jako v případě subjektivního testování, EKG atp., kdy nebyly jednotlivé parametry schopny odlišit všechna

měření mezi sebou, je to v tomto případě obdobné. Jelikož úroveň únavy nemusí reflektovat měření (vice versa by to však mělo platit), byl v případě hodnocení EEG, a konkrétně pomocí vzdálenostní matice, vytvořen dendrogram hierarchického shlukování (clusteringu) (Nielsen, 2016). Principiálně se jedná o nalezení hierarchického dělení dat pro účely určení shluků v datech, a to většinou právě na základě jejich vzájemné vzdálenosti (Murtagh & Contreras, 2011; Olson, 1995). Myšlenkou tohoto přístupu bylo reklasifikovat cílovou proměnnou na takovou, která by reflektovala základní vzory obsažené v datech, a byla by využitelná pro porovnání s výsledky ostatních metod.

Výsledky hierarchického shlukování ukazují, že data je možné rozdělit do tří hlavních shluků, viz obr. 4.11B, a tedy charakteristický shluk pro M6, charakteristický shluk zahrnující body z M1–M4 a shluk zahrnující převážně body z M5. Poslední shluk se však nachází mezi dvěma významnějšími, a spadají do něj i okrajové hodnoty těchto shluků.

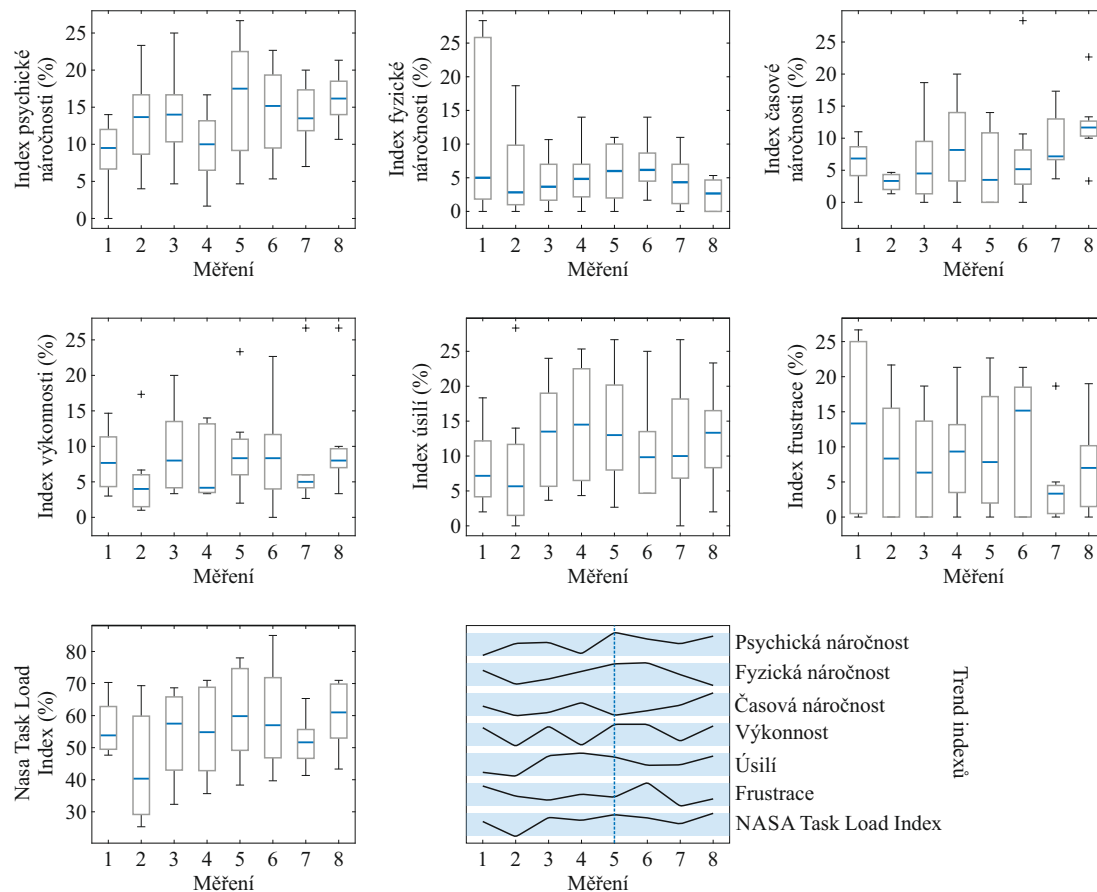
Popsaná klasifikace, resp. reklasifikace není 100%. Je možné si všimnout, že žádný z popsaných shluků neobsahuje výhradně hodnoty předmětného měření. Některá měření jsou však pravděpodobně extrémní, případně nejistota.

### 4.3 Subjektivní hodnocení

Hodnocení formou standardizovaného dotazníku NASA Task Load Index (skupina A a B) a jeho upravenou verzí zaměřenou specificky na hodnocení únavy (skupina B) je prezentováno prostřednictvím boxplotů a trendů mediánů. Ze statistické analýzy vyplývá, že v případě NASA-TLX dotazníku není možné porovnávat jednotlivé skupiny dohromady, jelikož tyto jsou od sebe, na základě analýzy prostřednictvím rANOVA, staticky signifikantně odlišné v některých sledovaných kategoriích, specificky MD ( $F(7,49) = 2,694$ ;  $p = 0,037$ ) a PD ( $F(7,49) = 3,107$ ;  $p = 0,025$ ). Hraničně pak NTLx ( $F(7,49) = 2,394$ ;  $p = 0,054$ ), kdy  $p > 0,05$ , ale hodnota je velmi blízká kritériu danému stanovenou hladinou významnosti. Jelikož data nesplňují podmínku sfericity, jsou prezentovány p-hodnoty upravené Greenhouse-Geisser korekcí. S ohledem na uvedené jsou výsledky pro jednotlivé skupiny prezentovány zvlášť.

V případě skupiny A je celková zátěž na subjekty, vyjádřená kategorií NTLx, nejvyšší při M5 a M8. Všeobecný trend pak sleduje nejdříve pokles zátěže po úvodním měření a postupný nárůst v průběhu nočních měření, až do zmíněného M5, které probíhalo v čase 06.00–09.00 místního času. Následuje výrazný pokles, a následně opětovný růst směrem k maximální pozorované zátěži, viz obr. 4.12. S ohledem na výsledky Friedmannova testu pak rozdíl mezi

jednotlivými měřeními nejsou v případě této kategorie statisticky signifikantní ( $\chi^2(7) = 12,315$ ;  $p = 0,091$ ). Celková zátěž (NTLx) pak sestává z jednotlivých kategorií, jehož distribuce a průběh jsou předmětem obr. 4.12. Kategorie MD, PD, PE dosahovaly vysokých hodnot (maxim, příp. hodnot blízkých maximu) ve zmíněném M5. Trendy jednotlivých kategorií jsou vzájemně odlišné, včetně NTLx. Statisticky signifikantní výsledky byly, na základě Friedmanova testu a následné post-hoc analýzy nalezeny pouze u kategorie TD ( $\chi^2(7) = 14,385$ ;  $p = 0,045$ ) pro dvojici měření M1-M8 ( $p = 0,035$ ; CI = (-7,477; -0,148)).

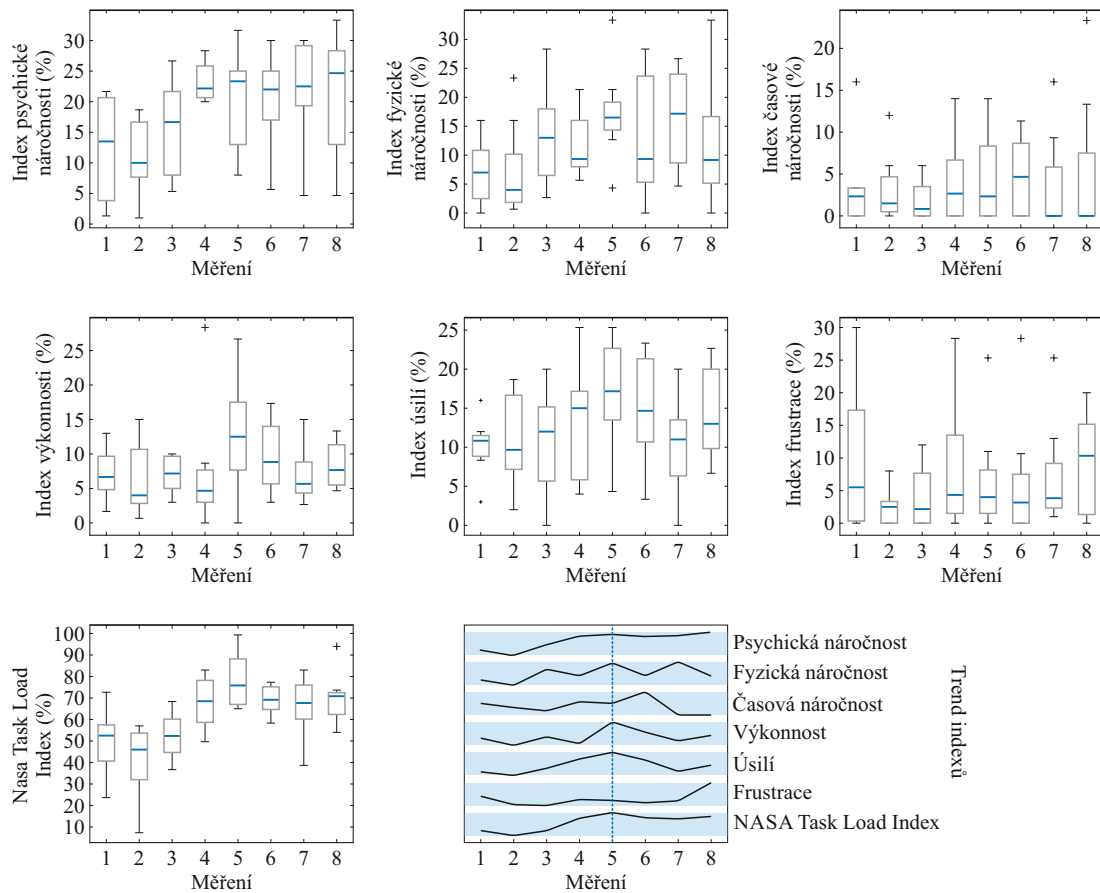


Obrázek 4.12: Grafické znázornění distribuce výsledků dotazníku NASA Task Load Index u skupiny A pro jednotlivé kategorie v průběhu všech měření formou boxplotů a porovnání trendů jednotlivých kategorií. Modrá vertikální linka zvýrazňuje označuje měření, v němž subjekty dosáhly nejvyšší reportované zátěže. Křížky značí odlehlé hodnoty.

V případě skupiny B bylo nalezeno větší množství odlišných měření a současně trendy jednotlivých kategorií sledují, ve většině případů, částečně odlišný průběh, avšak vysoké hodnoty jednotlivých indexů jsou opět pozorovány zejména v M5. Celková zátěž pak sleduje velmi podobný trend, jako u skupiny A, pouze s menší zátěží u prvních tří měření. Maximum je opět pozorováno v M5, viz obr. 4.13. Výsledky Friedmanova testu pak v případě NTLx ukazují na rozdíly ( $\chi^2(7) = 25,865$ ;  $p < 0,001$ ) mezi dvojicemi měření M1-M5 ( $p = 0,024$ ;



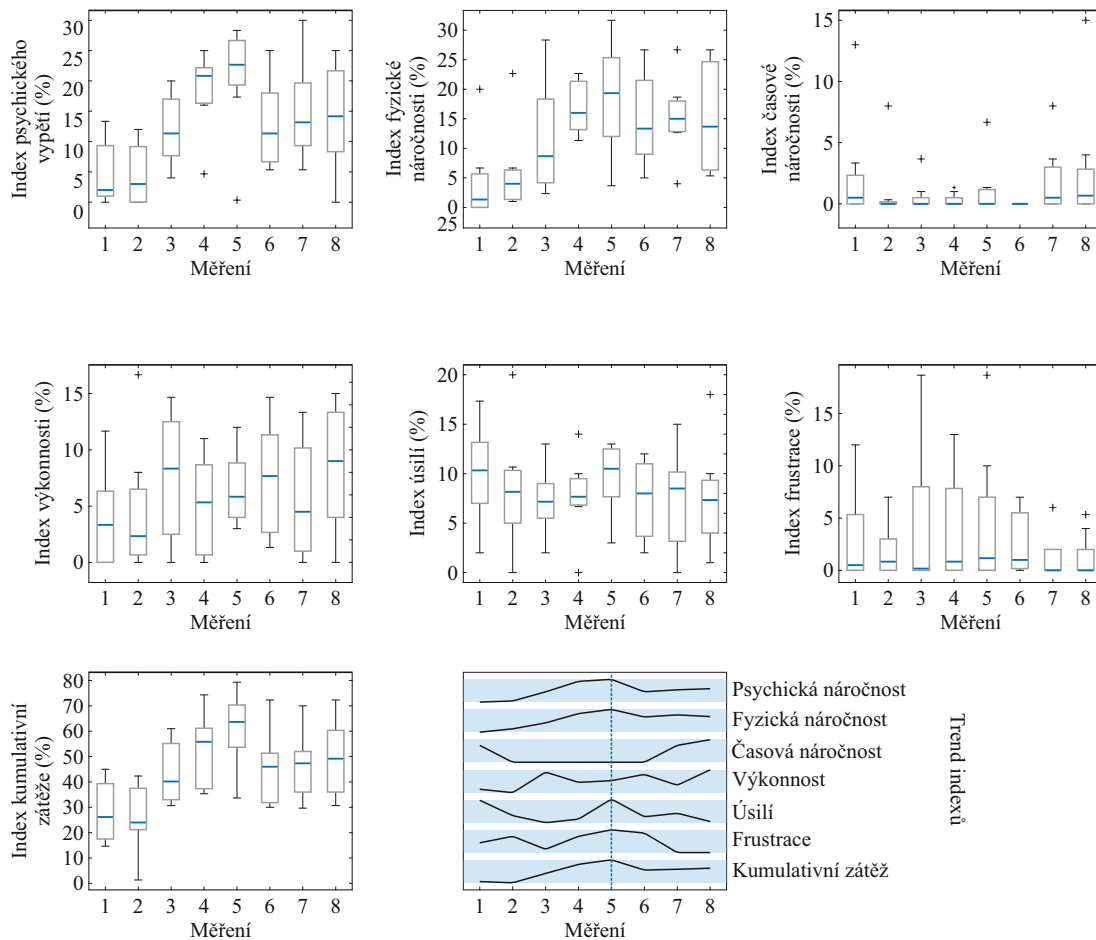
CI = (-7,704; -0,296)) a M2-M5 ( $p = 0,003$ ; CI = (-8,391; -0,984)). Statisticky signifikantní výsledky byly, na základě Friedmanova testu a následné post-hoc analýzy nalezeny u kategorie MD ( $\chi^2(7) = 18,179$ ;  $p = 0,003$ ) pro dvojici měření M2-M4 ( $p = 0,032$ ; CI = (-7,573; -0,177)) a dvojici měření M2-M8 ( $p = 0,023$ ; CI = (-7,698; -4,000)).



Obrázek 4.13: Grafické znázornění distribuce výsledků dotazníku NASA Task Load Index u skupiny B pro jednotlivé kategorie v průběhu všech měření formou boxplotů a porovnání trendů jednotlivých kategorií. Modrá vertikální linka zvýrazňuje označuje měření, v němž subjekty dosáhly nejvyšší reportované zátěže. Křížky značí odlehlé hodnoty.

Z hlediska dotazníku zaměřeného na únavu nebyly nalezeny statisticky signifikantní rozdíly mezi tímto dotazníkem a dotazníkem NASA-TLx, jelikož  $p > 0.05$  (Greenhouse-Geisser korekce) pro všechny kategorie. Při pohledu na trendy jednotlivých kategorií (indexů) je zřejmá výrazná podobnost, zejména s ohledem na výskyt maximálních hodnot při M5, avšak jisté odlišnosti v průbězích jsou zřetelné, např. v případě kategorie TD, viz obr. 4.13 a 4.14. Výsledky Friedmanova testu pak v případě indexu kumulativní zátěže ukazují na rozdíly ( $\chi^2(7) = 27,030$ ;  $p < 0,001$ ) mezi 4 dvojicemi měření, specificky M1-M4 ( $p = 0,024$ ; CI = (-7,709; -0,291)), M1-M5 ( $p = 0,003$ ; CI = (-8,459; -1,041)), M2-M4 ( $p = 0,024$ ; CI = (-7,709; -0,291)) a M2-M5

( $p = 0,003$ ;  $CI = (-8,459; -1,041)$ ). Statisticky signifikantní výsledky byly, na základě Friedmanova testu a následné post-hoc analýzy nalezeny u kategorie MD ( $\chi^2(7) = 25,412$ ;  $p < 0,001$ ) pro 4 dvojice měření, specificky M1-M4 ( $p = 0,042$ ;  $CI = (-7,429; -0,071)$ ), M1-M5 ( $p = 0,022$ ;  $CI = (-7,679; -0,321)$ ), M2-M4 ( $p = 0,008$ ;  $CI = (-8,054; -0,696)$ ) a M2-M5 ( $p = 0,003$ ;  $CI = (-8,304; -0,946)$ ), a u kategorie PD ( $\chi^2(7) = 17,786$ ;  $p = 0,004$ ) pro 2 dvojice měření, specificky M1-M5 ( $p = 0,033$ ;  $CI = (-7,579; -0,171)$ ) a M2-M5 ( $p = 0,045$ ;  $CI = (-7,454; -0,046)$ ).



Obrázek 4.14: Grafické znázornění distribuce výsledků dotazníku zaměřeného na únavu u skupiny B pro jednotlivé kategorie v průběhu všech měření formou boxplotů a porovnání trendů jednotlivých kategorií. Modrá vertikální linka zvýrazňuje označuje měření, v němž subjekty dosáhly nejvyšší reportované zátěže. Křížky značí odlehlé hodnoty.

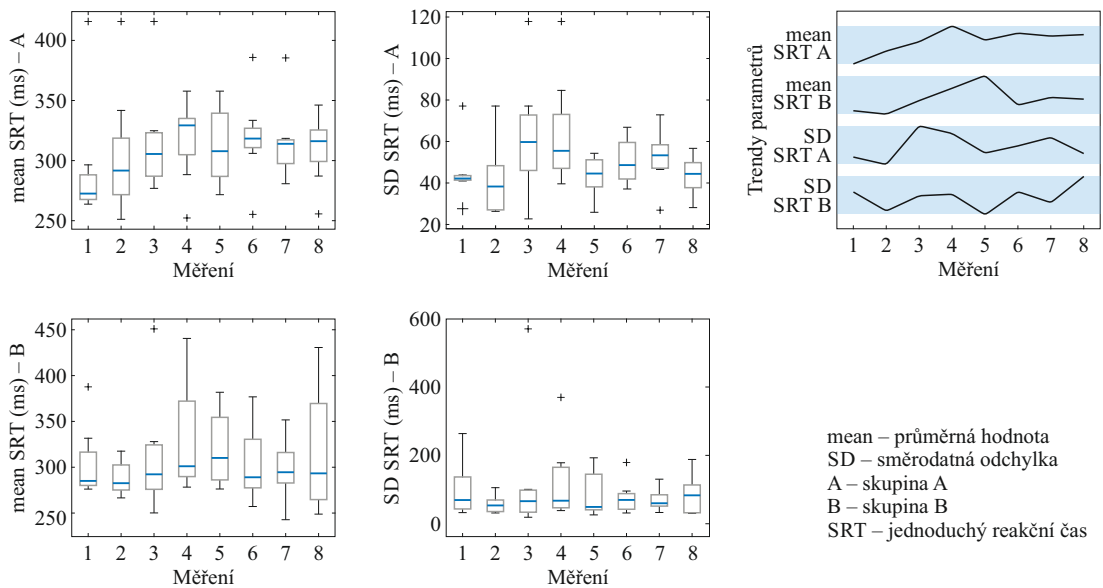
## 4.4 Výkonnostní testy

Výkonnostní testování sestávalo z několika úloh/testů zaměřených na reakční čas (jednoduchý/výběrový) subjektů a z OrTestu. Výsledné parametry získané z jednotlivých testů jsou prezentovány formou boxplotů a trendů mediánů.

V případě testu zaměřeného na jednoduchý reakční čas nejsou na základě Friedmanova testu pozorovány žádné statisticky signifikantní rozdíly mezi jednotlivými měřeními v průměrných hodnotách reakčních časů u skupiny A ( $\chi^2(7) = 11,068$ ;  $p = 0,137$ ), tak u skupiny B ( $\chi^2(7) = 7,042$ ;  $p = 0,425$ ). Stejně výsledky byly pozorovány také pro směrodatné odchylky reakčních časů, jak u skupiny A ( $\chi^2(7) = 15,755$ ;  $p = 0,027$ ), tak u skupiny B ( $\chi^2(7) = 4,083$ ;  $p = 0,770$ ). U skupiny A v tomto případě je sice  $p < 0,05$ , ale následná post-hoc analýza využívající konzervativní Bonferroniho korekce neodhalila žádné rozdílné měření. Porovnání skupin A a B prostřednictvím rANOVA neprokázalo statisticky signifikantní rozdíl mezi skupinami jak pro průměrné hodnoty SRT ( $F(7,49) = 0,659$ ;  $p = 0,595$ ), tak pro jeho směrodatné odchylky ( $F(7,49) = 0,340$ ;  $p = 0,752$ ). Jelikož data nesplňují podmínku sphericity, jsou prezentovány p-hodnoty upravené Greenhouse-Geisser korekcí. S ohledem na uniformitu prezentovaných výsledků jsou výsledky pro jednotlivé skupiny prezentovány zvlášť i v případě SRT.

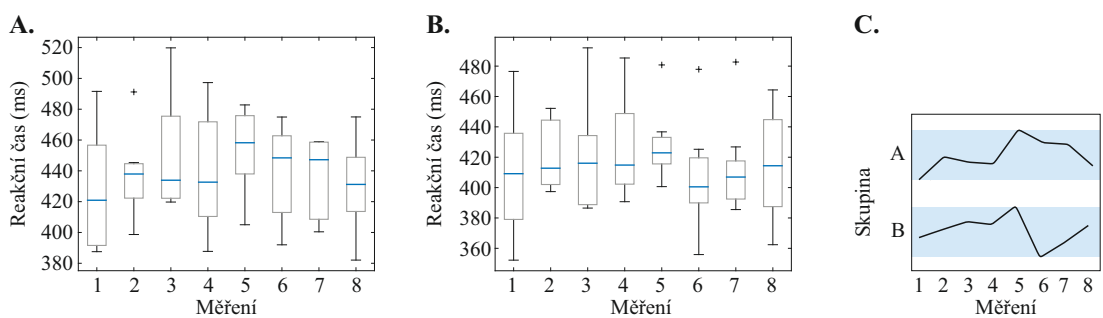
Z prezentovaných trendů je zřejmé, že skupina A vykazovala nejnižší výkonnost (nejvyšší RT) při M4. Tento vrchol byl dosažen postupným nárůstem RT od M1 až ke zmíněnému M4. Následoval pokles, přičemž v M5–M8 je poté sledována relativně stabilní hodnota srovnatelná s M3. V případě skupiny B následoval po úvodním poklesu M2 nárůst RT směrem k maximu RT (nejnižší výkonnost) v M5. V M6–M8 je pozorována stabilní hodnota RT, která je na úrovni M1–M4, viz obr. 4.15.

Výběrový reakční čas ukázal na rozdíly mezi jednotlivými měřeními pouze v případě průměrných reakčních časů, a to jak u skupiny A, tak i B. tyto hodnoty však mají minimální informační hodnotu s ohledem na skutečnost, že chybovost (resp. správnost) může být signifikantním faktorem. Z těchto důvodů byly hodnoty průměrného času opraveny o chybovost. Na základě Friedmanova testu byl nalezen statisticky signifikantní rozdíl mezi měřeními u skupiny A ( $\chi^2(7) = 14,250$ ;  $p = 0,047$ ). Následná post-hoc analýza však neodhalila žádnou dovojití měřeními, mezi nimiž by byl rozdíl, jelikož pro všechny dvojice měření bylo  $p > 0,05$ . Zvolená post-hoc analýza využívá konzervativní Bonferroniho korekci, která byla zvolena z důvodu, aby byla co nejvíce potlačena možnost nesprávné identifikace rozdílů. S ohledem na tuto skutečnost lze konstatovat, že ve skupině A nebyl nalezen žádný signifikantní rozdíl mezi měřeními, stejně jako



Obrázek 4.15: Grafické znázornění distribuce výsledků jednoduchého reakčního času pro obě skupiny v průběhu všech měření formou boxplotů a porovnání trendů jednotlivých parametrů. Křížky značí odlehlé hodnoty.

u skupiny B ( $\chi^2(7) = 14,250$ ;  $p = 0,047$ ). Stejně tak nebyl na základě výsledků rANOVA odhalen statisticky signifikantní rozdíl mezi oběma skupinami ( $F(7,49) = 6,167$ ;  $p = 0,520$ ). Při pohledu na trend je zřejmé, že průběh je až po maximum (nejnižší výkonnost) v M5 obdobný. Stejně jako u SRT pak v případě skupiny B dochází k výraznému poklesu při M6 a následnému nárůstu přes M7 a M8. Oproti tomu u skupiny A dochází v M6 pouze k mírnému poklesu, přičemž tato hodnota je srovnatelná v M7, načež v M8 dochází k dalšímu poklesu, viz obr. 4.16.



Obrázek 4.16: Grafické znázornění distribuce výsledků výběrového reakčního času (průměrný čas opravený o chybovost) pro první (A.) i druhou (B.) skupinu v průběhu všech měření formou boxplotů a porovnání trendů jednotlivých parametrů (C.). Křížky značí odlehlé hodnoty.

Při hodnocení OrTestu bylo třeba vzít v úvahu kromě samotných skupin také jednotlivé úlohy. Výsledky byly vypracovány pro meziskupinové a vnitroskupinové porovnání, tedy pro porovnání skupiny A a skupiny B mezi sebou a průběhu

jednotlivých měření v rámci konkrétní skupiny subjektů. Tato analýza vycházela z porovnání distribuce hodnot měření realizovaných v konkrétních časech, pro konkrétní úlohu a u konkrétní skupiny subjektů. Meziskupinové a vnitroskupinové porovnání tedy vycházelo z Wilcoxonovy notace:

$$M1 - M8 \sim GrTask, \quad (4.3)$$

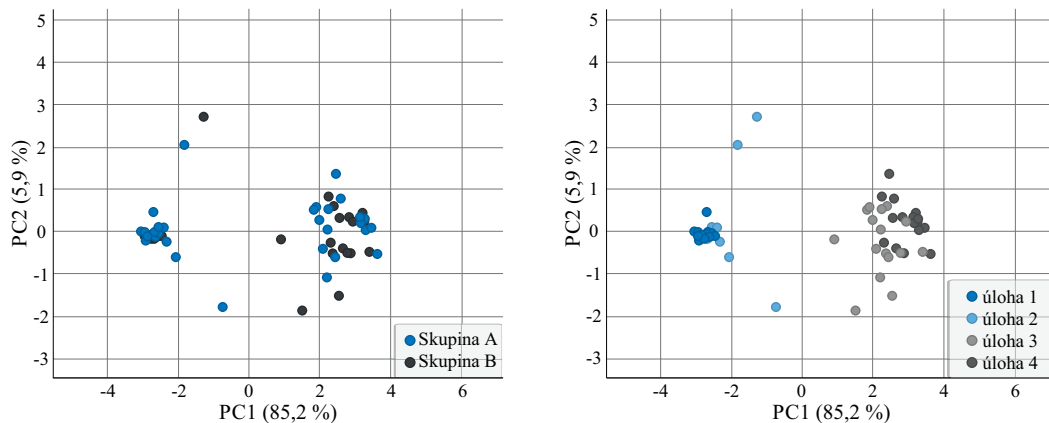
kde  $M1 - M8$  reprezentuje jednotlivá měření v čase a  $GrTask$  je kategorická cílová proměnná reprezentující úlohu náležející konkrétní skupině. Proměnná  $GrTask$  tedy nabývala kategorické hodnoty  $Gr_S U_x$ , pro  $S = [A, B]$  a  $x = [1, 2, 3, 4]$ , přiřazené každé hodnotě zaznamenaného měření u konkrétního subjektu. Pro účely analýzy byla využita rANOVA s post-hoc analýzou využívající Bonferroniho korekci.

Výsledky rANOVA ukázaly, že mezi kategoriemi  $GrTask$  existuje statisticky významný rozdíl ( $F(49, 392) = 1,3949$  a  $p = 0,047$ ). Post-hoc analýza pro porovnání rozdílů mezi skupinami v rámci stejných úloh v podstatě neukázala žádný signifikantní rozdíl. Jediným identifikovaným signifikantním rozdílem mezi skupinou A a B je M7, specificky v úloze 3 ( $p = 0,017$ ;  $CI = (-1,351; -0,079)$ ), přičemž průměrná hodnota je u skupiny B signifikantně vyšší oproti skupině A.

Výsledky taktéž ukazují, že u skupiny A nejsou střední hodnoty měření v úloze 1 statisticky signifikantně odlišné od těch získaných v úloze 2. Uvedené taktéž platí pro porovnání měření v úloze 3 vůči úloze 4. Střední hodnoty měření v úloze 1 naopak vykazují signifikantní rozdíly oproti těm z úlohy 3 a 4. Uvedené obdobně platí pro úlohu 2 v porovnání s úlohami 3 a 4. Stejně výsledky je pak možné pozorovat i v rámci skupiny B. Tyto výsledky tedy ukazují, že obtížnost, resp. chybovost v úloze 1 a 2, a též v úloze 3 a 4, jsou porovnatelné. Uvedené výsledky taktéž poukazují na to, že hodnoty měření v úloze 1 a 2 budou tvořit odlišný datový shluk oproti úloze 3 a 4.

Uvedené je možné vizualizovat stejnou technikou, jaká byla využita při hodnocení EEG, tj. analýzou hlavních komponent. I v tomto případě je třeba zredukovat dimenzionalitu v datech. Pro vykreslení závislosti mezi měřeními, pro účely vyšetření shlukování datových bodů, by totiž byl potřebný 8dimenzionální prostor (hodnoty každého měření by musely být vyneseny na jednu osu). V tomto konkrétním případě postačují 2 PC pro popis 90 % variability obsažené v datech. Vykreslením dvou prvních komponent je tedy možno zobrazit body v dvousouřadnicovém systému tak, jak je zobrazeno na obr. 4.17. Na obrázku je možno vidět přirozené shlukování datových bodů

náležících úlohám, a taktéž nerozlišitelnost skupin. Uvedené je v souladu s předcházejícím popisem výsledků.

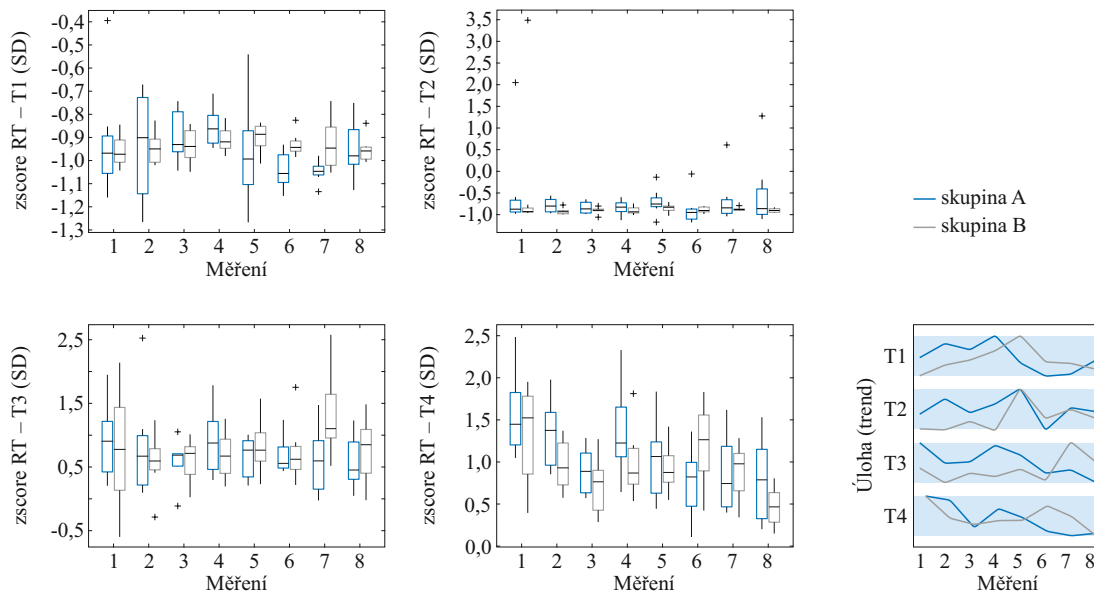


Obrázek 4.17: Bodový graf hlavních komponent - Score plot, s označením bodů dle příslušnosti ke skupinám subjektů a skupinám úloh OrTestu.

V rámci vnitroskupinového hodnocení průběhů střeňích hodnot (sledování rozlišitelnosti mezi měřeními u konkrétní skupiny a konkrétní úlohy) nebyl, v rámci úlohy 1 u skupiny A nalezen žádný statisticky signifikantní rozdíl. Uvedené znamená, že neexistuje dostatek důkazů na to, aby bylo v tomto případě možno oddělit jednotlivá měření realizovaná v čase mezi sebou, a tedy není možné tvrdit, že únava má na výsledky vliv. Uvedené platí i v případě úlohy 2 a 3. V případě úlohy 4 lze pozorovat nesignifikantní kontinuální nárůst výkonnosti při měřeních M1–M3 (nižší hodnoty odrážejí vyšší výkonnost), následovaný signifikantním poklesem výkonnosti v měření M4, kde je pozorován rozdíl M3 oproti M4 ( $p = 0,019$ ;  $CI = (0,058; 1,099)$ ). Poté následuje nevýznamný pokles hodnot na hodnoty reprezentované měřeními M5 a další pokles na M6. M6 pokleslo na úroveň M3, M7 a M8, mezi nimiž není statisticky signifikantní rozdíl. Průměrná hodnota M6 je však statisticky významně nižší (výkonnost je vyšší) než M2 a M4 ( $p = 0,013$ ;  $CI = (-0,872; -0,062)$ ) a  $p = 0,002$ ;  $CI = (0,146; 1,052)$ ). Grafická prezentace distribuce naměřených hodnot pro jednotlivá měření spolu s vykreslením průběhů pro úlohy skupiny A (a také skupiny B) je uvedeno na obr. 4.18.

V rámci vnitroskupinového hodnocení průběhů a středních hodnot u úlohy 1 a 2 pro skupinu B nebyl nalezen žádný statisticky signifikantní rozdíl. Interpretace by tedy byla obdobná, jako v případě skupiny A. Je tedy možné tvrdit, že subjekty v těchto úlohách nevykazovaly významné výkonnostní fluktuace v průběhu 24hodinového experimentu.

V případě úlohy 3 u skupiny B je potom možné pozorovat konstantní, nesignifikantně měnící se výkon subjektů mezi měřeními M5–M6. Následně je možné pozorovat prudký negativní propad výkonnosti v měření M7, které je



Obrázek 4.18: Grafické znázornění distribuce výsledků OrTestu (z-score průměrného času opraveného o chybovost) pro skupinu A (modrá) i B (šedá) pro všechny typy úloh T1–T4 v průběhu všech měření formou boxplotů a porovnání trendů. Křížky značí odlehlé hodnoty.

statisticky signifikantně rozdílné od M2 ( $p = 0,010$ ;  $CI = (-1,355; -0,112)$ ), M3 ( $p = 0,001$ ;  $CI = (-1,211; -0,194)$ ) a M6 ( $p = 0,004$ ;  $CI = (-1,055; -0,123)$ ). V měření M8 je potom možné pozorovat opět významný nárůst výkonnosti oproti M7 ( $p = 0,012$ ;  $CI = (0,074; 1,012)$ ). Měření M8 však není statisticky signifikantně odlišné od ostatních měření.

V případě úlohy 4 u skupiny B je průběh měření podobný jako v případě skupiny A. Je možné pozorovat kontinuální pokles hodnot (narůstající výkonnost) mezi měřeními M1-M3. Tato měření však nejsou statisticky signifikantně odlišná. Následuje statisticky signifikantní pokles výkonnosti v M4 oproti M3 ( $p = 0,010$ ;  $CI = (-0,912; -0,074)$ ). Měření M5 není signifikantně odlišné od ostatních měření, hodnoty tohoto měření tedy jakýsi průnik s hodnotami ostatních měření. Měření M6 indikuje nárůst výkonnosti, a to s ohledem na negativní posun střední hodnoty. Tato změna však není signifikantní a ke změně nedochází ani v případě M7. Statisticky významný nárůst výkonnosti je možné zaznamenat až v případě M8, které je statisticky významně odlišné od M6 ( $p < 0,001$ ;  $CI = (0,305; 1,184)$ ).

## 4.5 Klasifikace

Natrénované klasifikační modely vykazovaly obecně uspokojivou klasifikační schopnost, která se pohybovala nad úrovní 60 %, s výjimkou klasifikátoru Naive

Bayes. Z tabulky 4.3 je patrné, že nejvyšší kvantifikovanou úspěšnost mají neuronové sítě. Přestože interpretace neuronové sítě je ze své podstaty nemožná vzhledem k tomu, že klasifikace probíhá v tzv. černé skříňce, bude další demonstrativní popis věnován právě tomuto výsledku.

Natrénovaný klasifikátor na bázi neuronové sítě je schopen vykonat správnost klasifikace k cílové proměnné na úrovni 80 %. To znamená, že na základě výsledků OrTestu, celkového indexu NASA Task Load Index a subjektivních indikátorů psychické náročnosti (MD), fyzické náročnosti (PD) a úsilí (EF) na základě dotazníku NASA-TLX je možné jednotlivá měření poměrně přesně klasifikovat. Tato měření v kontextu uspořádání experimentu a také předchozích prezentovaných výsledků odrážejí úroveň únavy.

Tabulka 4.3: Výkonové charakteristiky natrénovaných klasifikačních modelů, vycházející z jejich iterativního trénování a testování.

Model	Správnost	Přesnost	Senzitivita
Neuronová síť	0,815324675	0,815886898	0,815324675
kNN	0,776623377	0,778521832	0,776623377
Náhodný les	0,717012987	0,717778254	0,717012987
Rozhodovací strom	0,701168831	0,701552196	0,701168831
Podpůrné vektorové sítě	0,609220779	0,61563791	0,609220779
Naivní Bayes	0,416493506	0,417464011	0,416493506

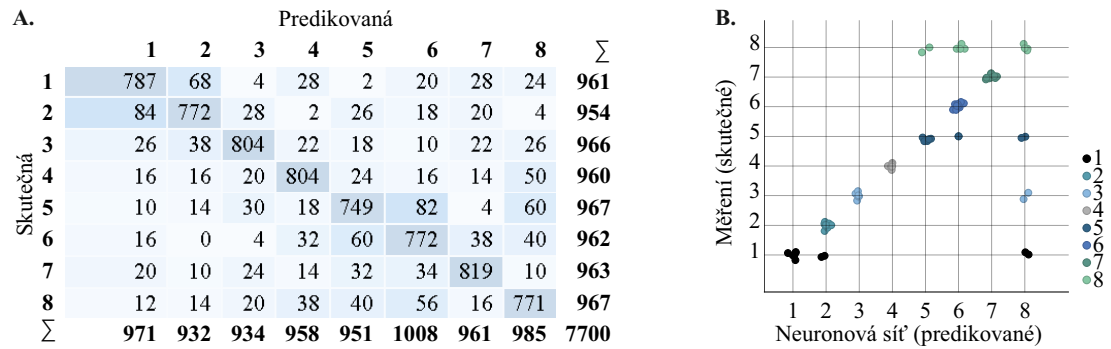
Diskutovaný klasifikátor má také přesnost a senzitivitu (citlivost) 80 %. Přesnost udává, jaký podíl pozitivních predikcí je správný. Naproti tomu senzitivita udává, kolik skutečných (cílových) bodů bylo klasifikováno správně. V obou případech se jedná o velmi uspokojivý výsledek, a to i s ohledem na počet vstupních dat použitých pro strojové učení.

Matice popisující počty správně a nesprávně klasifikovaných prvků v případě neuronové sítě je znázorněna na obrázku 4.19A. Z této matice, označované také jako matice záměny nebo konfuzní matice, lze pozorovat kumulativní četnosti predikovaných hodnot vůči hodnotám skutečným. Zobrazená matice platí pro výsledky iterativního trénování a testování, a tedy odráží všeobecný výkon klasifikátoru.

Dosud popisovaný výkon klasifikátoru je zobecněný, a je tedy výsledkem iterativního trénování/testování. Proto bylo dále provedeno trénování právě jedné výsledné neuronové sítě, která by svou povahou měla reflektovat dříve popsané výsledky. Výsledný klasifikátor byl natrénován na náhodně vybraných datech (70 %). Výsledek schopnosti klasifikovat cílovou proměnnou na základě výsledného klasifikátoru založeného na neuronové síti je uveden na obrázku 4.19B. Bodový graf znázorňuje vztah mezi skutečnou hodnotou měření a hodnotou predikovanou. K vytvoření tohoto grafu bylo jako vstup pro klasifikátor použito zbývajících 30 % dat, která nebyla využita pro trénování. Z



bodového grafu je patrné, že klasifikace byla úspěšná i přes chybnou klasifikaci některých hodnot.



Obrázek 4.19: Klasifikace měření s využitím neuronové sítě – bodový graf zobrazující predikované a skutečné hodnoty (A) a matice záměn (B).

## 5. Diskuze

Představená práce je zaměřena na hodnocení únavy u pilotů, a to se snahou o její objektivizaci, pro tyto účely bylo sbíráno několik datasetů v průběhu 24hodinových experimentů, které byly dále podstoupeny analýze.

### 5.1 Variabilita srdečního rytmu

Objektivizace únavy je v letectví stále otevřeným tématem, kterému je věnována značně malá pozornost. Z hlediska využití EKG pak existuje velmi limitované množství studií, které jsou dále limitovány např. počtem subjektů, experimentálním designem nebo přístupem ke zpracování dat. Poznatky aplikovatelné do letectví se pak částečně dají přebírat také z oblasti pozemní dopravy, v níž je na toto téma kladen větší důraz. Ačkoli tomu tak je, těchto studií je stále velmi malé množství, jsou velmi limitované zejména z výše uvedených důvodů, plně nereflektují specifika leteckého prostředí a jejich výsledky si často oponují. Lu et al. (2022) provedli systematickou analýzu v databázích PubMed, Scopus a Web Of Science, přičemž relevantních článků týkajících se využití HRV v kontextu únavy u řidičů bylo pouze 18, v letectví pak takové studie téměř neexistují. Na základě meta-analýzy těchto článků pak autoři konstatují, že výsledky jednotlivých studií, stejně jako interpretace výsledků a pochopení fyziologické odezvy organismu z hlediska únavy a v kontextu sledovaných parametrů HRV jsou vzájemně v rozporu. Z uvedeného tedy plyne, že tato problematika je značně nevyřešená a je jí třeba věnovat další pozornost, přičemž zájem je směřován směrem k detekčním systémům únavy.

Zásadním problémem v jakýchkoli detekčních systémech založených na fyziologických datech je individualita subjektu, a to nejen s ohledem na klidové hodnoty jednotlivých signálů, ale také na specifika i intenzitu fyziologické reakce, citlivost k okolním podmínkám a projevy. Ve většině dostupných studiích je uvedené diskutováno jako limitace a faktor, který hraje významnou negativní roli při snaze o generalizaci populace. Problematiku individuality projevů únavy a úrovně únavy v daných podmínkách pak adresují Bendak & Rashid (2020), kteří na základě systematické rešerše a následné meta-analýzy zaměřené na únavu v letectví konstatují, že všechny stávající modely pro predikci únavy předpovídají únavu jako bodovou, dikrétní hodnotu a žádný z nich nezjišťuje pravděpodobnou variabilitu únavy. V rámci této práce byla snaha co nejvíce eliminovat individualitu subjektů tak, aby sesbíraná data mohla sloužit pro co nejvíce precizní rozlišení úrovně únavy. Uvedené bylo

provedeno s využitím z-score, které je využíváno pro normalizaci dat s cílem porovnání různých typů proměnných nebo právě potlačení individuálních vlivů.

S tímto postupem je však spojen problém související se zpracováním dat v reálném čase. Z-score je vytvářeno z celého datasetu (resp. ze všech měření pro daný parametr) pro jeden subjekt. Takovýto přístup tedy neumožňuje hodnocení v reálném čase, ale pouze zpětně. V systémech indikujících změnu psychofyziologického stavu jedince, jako jsou zařízení typu smartwatch, je pak využíváno techniky učení. Detekce je tedy ze začátku nespolehlivá, a až s narůstajícím datasetem dochází ke zpřesnění predikce systému. Ve většině studií zaměřených na vytvoření detekčního systému únavy je z těchto důvodů dosaženo poměrně nízké predikce, jelikož experimentální měření jsou omezena jejich délkou, a tedy i rozsahem datasetu.

Představená práce je limitována právě rozsahem datasetu, a tedy přístup s využitím z-score se zdál jako nejvhodnější, ačkoli není ve studiích běžně využíván. I s ohledem na nízký počet subjektů, který je způsoben zejména náročností měření z hlediska času, dostupnosti subjektů a obsluhujícího personálu, pak není možné výsledky generalizovat, ale na výzkum lze pohlížet jako na pilotní studii, jelikož obdobná problematika nebyla v takovémto rozsahu zatím řešena a problematika objektivizace únavy v letectví není, jak již bylo uvedeno, uspokojivě vyřešena. Nedostatek důkazů o odlišnosti jednotlivých měření pak může být způsoben s největší pravděpodobností limitovaným výzkumným vzorkem. Ačkoli nebylo možné na základě statistické analýzy odlišit většinu měření od sebe, výsledky přinášejí nový náhled do této problematiky a rozšiřují znalostní bázi.

Obecně byl pozorován trend, který odpovídá počátečním předpokladům o navození proměnlivé úrovně únavy v kontextu interference osmotického a cirkadiálního řízení spánku, tedy prodloužení doby bdělosti a vlivu cirkadiálního rytmu. Analýza elektrické aktivity srdce prostřednictvím HRV tedy poukázala na společný trend, který byl přítomný u většiny měření, bez ohledu na úsek a skupinu. Obecný trend sledoval postupný růst k maximu, které nastávalo při měřeních prováděných v ranních hodinách, po čemž následoval pokles a následný nárůst parametrů (případně se jednalo o inverzní průběh, s ohledem na charakter parametru). V ranních hodinách, nejčastěji v období M4, příp. M5–M6 docházelo k poklesu parametrů, které jsou pozitivně ovlivňovány aktivitou sympatiku. Je pak obecně přijímáno, že stav snížené pozornosti, ospalosti, resp. únavy je spojen právě se sníženou aktivitou sympatiku a zvýšenou aktivitou parasympatiku (Byeon et al., 2006).

V kontextu sledování sympatovagální rovnováhy je nejvíce využíván parametr frekvenční analýzy LF/HF. Právě tento parametr dosahuje minima v

měření M4, které probíhalo v brzkých ranních hodinách, specificky mezi 03:00 a 06:00. Toto měření tedy celé probíhalo v době okna cirkadiánního útlumu. Dle Byeon et al. (2006) je právě pokles poměru LF/HF spojován s poklesem pozornosti a ospalostí. V souladu s těmito výsledky pak Patel et al. (2011) reportují pokles poměru LF/HF u unavených řidičů oproti běžnému stavu, a konstatují, že právě LF/HF může být využíván jako indikátor únavy. Ačkoli mezi měřeními nebyly nalezeny statisticky signifikantní rozdíly, při pohledu na průběh středních hodnot parametru je trend, kdy nejprve dochází k postupnému poklesu směrem k minimu následovaný opětovným nárůstem a zakončený poklesem, zřejmý. Uvedené potvrzují také další parametry, které tento trend kopírují a u nichž byly rozdíly nalezeny. Je třeba konstatovat, že většina rozdílů byla spojena právě s M4, přičemž tyto bylo pozorovány oproti měřením M1 a M2.

Z výsledků se tedy zdá, že by mohly reflektovat únavu, jelikož sledují předpokládaný trend, tj. maximální útlum v období WOCL, následné snížení únavy po rozednění s opětovným nástupem únavy v důsledku prodloužené doby bdělosti v kontextu osmotické regulace spánku.

## 5.2 Elektrická mozková aktivita

Záznamy elektrické aktivity mozku byly pořizovány pouze v druhé fázi experimentu, tj. pouze u skupiny B. Měření jako taková pak byla využita pouze pro validaci únavy, tj. se snahou klasifikovat únavu, resp. určit relativní úroveň únavy u jednotlivých měření tak, aby tato mohla být následně využita v případech dalších sledovaných dat. Cíl a předpoklad klasifikace únavy prostřednictvím EEG byl založen na skutečnosti, že většina studií věnujících se únavě reportuje nárůst výkonu v alfa pásmu. Uvedené mimo jiné souvisí s výskytem tzv. spánkových vřeten (alpha spindles) při zvyšující se únavě. Někteří autoři však reportují v kontextu únavy změny v pásmech delta a/nebo theta (Wright & McGown, 2001; Wilson et al., 2007b). Samotné alfa pásmo je pak významně ovlivněno hlukem/šumem a výsledky založené na hodnocení alfa pásma nemusí být relevantní. Uvedené mohlo být příčinou toho, že právě oblast alfa nebyla pro separaci měření významná, jelikož experiment probíhal v leteckém simulátoru, přičemž pro co největší reálnost podmínek byly po celou dobu měření subjekty vystaveny hluku simulujícímu reálné zvukové podmínky za letu.

Stejně jako v případě EKG, i studie zabývající se EEG a únavou v letectví jsou velmi limitované, jak do počtu, tak s ohledem na velikost výzkumného vzorku, limitace experimentálního protokolu a zpracování dat, mimo jiné v

kontextu individuality subjektů, která byla diskutována u EKG. Většina studií zabývajících se problematikou klasifikace únavy prostřednictvím EEG je tedy opět z domény silniční dopravy. Jednotlivé studie jsou pak opět primárně zaměřeny na detekční systémy únavy a často kombinují více sledovaných proměnných, zejména pak mrkání. Uvedený přístup však vykazuje obdobné limitace, jako tomu bylo v případě EKG. Zejména se jedná o zmíněnou individualitu, která opět musí být potlačena a komplikuje tak snahy o real time detekci únavy.

EEG data pro účely této práce byla tedy nejprve standardizována prostřednictvím z-score přes jednotlivé subjekty právě s cílem potlačení individuality. Výsledky PCA úplně exaktně nereflktují výsledky získané prostřednictvím odlišných dat, avšak korespondují s nimi v kontextu nalezení shluků reprezentujících specifická měření, která spolu souvisí, a která by mohla reflektovat stejnou úroveň únavy. Významnými frekvencemi pro toto rozdělení měření pak byly velmi nízké frekvence, reprezentující pásmo delta. To může být způsobeno zmíněným hlukem při měření, který mohl ovlivnit pásmo alfa, které dále nevykazovalo žádné sinifikantní změny v kontextu únavy. Na druhou stranu, ačkoli se velká část výzkumu shoduje na významu pásma alfa z hlediska monitoringu únavy, Wilson et al. (2007b) reportují nárůst aktivity v pásmu delta, jako důsledku únavy. V kontrastu s tímto poznatkem pak Jap et al. (2009) v případě únavy nepozorují změny v pásmu delta, naopak signifikantní je pásmo beta. Naopak Lal et al. (2003) vytvořili algoritmus pro detekci tří úrovní únavy na všech 4 standardně sledovaných pásmech, přičemž zejména v prvotních fázích únavy sledují signifikantní změny ve velmi nízkých rytmech, primárně v pásmech delta a theta, přičemž aktivita v těchto pásmech s narůstající únavou stoupá. Projevy v pásmu alfa jsou pak pozorovány až ve velmi vysoké úrovni únavy. Je tedy zřejmé, že výsledky studií jsou nekonzistentní. Současně omezení na 4 pásma se zdá být příliš hrubé. Původním záměrem v případě hodnocení prezentované v této práci bylo omezení se na tato pásma. Zpracování těchto dat však nepřineslo žádné zajímavé informace. Při zpracování dat se tak zdálo vhodné zvýšit „vzorkování“ pásem. Původní myšlenou bylo využít citlivost kroku frekvenčního spektra vytvořeného FFT. Uvedené nebylo možné vykonat, jelikož by byl dataset příliš obsáhlý pro následnou PCA. Klasické podvzorkování se pak nezdálo jako vhodný přístup, jelikož při takovém dochází ke ztrátě informace. Z těchto důvodů bylo přistoupeno k vytvoření 0,2Hz pásem, v nichž byla spočítána spektrální výkonová hustota, která byla následně využita pro PCA. Tímto přístupem bylo měření možné rozdělit do 3 skupin, které by mohly reprezentovat obdobnou úroveň únavy, přičemž k tomu rozdělení přispívají pouze velmi pomalé delta

vlny (0,6–1,2 Hz). Současně, exaktně oddělenou skupinou bylo M6, které tvořilo samostatný shluk. Uvedené koresponduje i s výsledky EKG, pro skupiny B byly extrémny pozorovány zejména v M5–M6. Zdá se tedy, že únava se projevila právě v pásmu delta tak, jak reportují někteří autoři (Wilson et al., 2007b; Lal et al., 2003). Je třeba zmínit, že přístup hodnocení celého pásma není standardně využíván ačkoli se zdá, že právě užší spektra mohou lépe reflektovat případnou změnu stavu, v tomto případě specificky únavy.

Význam elektrod, na kterých byla identifikována signifikantní změna výkonu (v pásmu delta) a které také nejlépe odrážely únavu ve vztahu k provedeným měřením, je vzhledem k malému množství nálezů v publikované literatuře sporný. Každopádně lze pozorovat, že identifikované elektrody patří pravé hemisféře, a to především v temporální a parietální oblasti. Výše uvedené by mohlo reflektovat výsledky některých studií, které poukazují na zvýšení průtoku krve mozkem a okysličení v těchto částech během mentální únavy, a které popisuje a shrnuje např. Qi et al. (2019). Také při svalové únavě bylo pozorováno zvýšené střídání mozkové aktivity právě na pravé hemisféře (Taghizadeh et al., 2017). Ve výše uvedeném případě však nebylo sledováno pásmo delta a zvýšená aktivita byla pozorována v pásmu alfa a gama.

Každopádně, zmíněné rozdělení měření do shluků bylo právě využito pro ověření, že byl navozen proměnlivý stav únavy. Z výsledků se zdá, že uvedené bylo potvrzeno a i v kombinaci s analýzou HRV, kde sice nebylo možné exaktně statisticky oddělit většinu měření, avšak byl pozorován předpokládaný trend s útlumem v ranních hodinách, lze pak konstatovat, že fyziologická měření jsou určitým způsobem schopna reflektovat únavu, což potvrzují i výše uvedené studie prováděné mimo oblast letectví, a že proměnlivá úroveň únavy byla skutečně navozena. To potvrzuje také chování subjektů v průběhu měření, které nebylo možné kvantifikovat, avšak lze konstatovat, že zejména v ranních hodinách byla na subjektech pozorována zvýšená únava projevující se zejména změnami chování, hraničně až agresivitou, demotivací k provádění experimentu, resp. u subjektů byl pozorovatelný diskomfort, resp. frustrace. Na druhou stranu, je třeba pořád uvažovat skutečnost, že ačkoli byla snaha co nejvíce eliminovat individualitu subjektů, tak tato nebude potlačena nikdy a s ohledem na menší velikost výzkumného vzorku je třeba brát tyto výsledky spíše jako nový náhled do problematiky a znalostní bázi pro navazující měření, která jsou nezbytně nutná pro další aplikaci výsledků do praxe.

## 5.3 Subjektivní hodnocení

Z hlediska subjektivního hodnocení, specificky dotazníku NASA Task Load Index, ukazuje celkový trend zátěže na kontinuální vysokou zátěž při měření od půlnoci (M3) a dále, přičemž maximum je pozorováno v M5 u skupiny A. U skupiny B pak dochází k nárůstu pocíťované zátěže až při měřeních od 03:00 (M4), přičemž maximum nastává shodně v M5. Nejvyšší zátěž byla tedy u obou skupin pocíťována mimo jiné v období WOCL, a také s prodlužující se dobou bdělosti. Rozsah absolutních hodnot zátěže je pak vyšší u skupiny B, kdy obě skupiny na začátku experimentu reportovaly přibližně stejnou úroveň únavy, přičemž maximum u skupiny A dosahovalo hodnot NTLx okolo 60 % (medián), u skupiny B pak přibližně 75 %. Upravená verze NASA-TLX dotazníku pro specifické účely hodnocení únavy pak reportuje stejné výsledky, jako samotný NASA-TLX, a tedy se zdá, že ačkoli je NASA-TLX obecnější a zaměřený na zátěž, lze jej využít také pro subjektivní hodnocení únavy při podobném nastavení, kdy jednotlivé úlohy jsou v průběhu času srovnatelné, a jediným zátěž ovlivňujícím faktorem je právě úroveň únavy. Uvedené by mohlo znamenat, že zvýšená subjektivní zátěž souvisí s menšími dostupnými zdroji organismu v důsledku únavy. Hart (2006) pak uvádí, že únava jako kovariát k NASA-TLX byla sledována přibližně ve 2 % případů využití dotazníku. Výsledky korespondují s fyziologickými měřeními, kde extrémní hodnoty byly pozorovány právě v oblasti M4–M5 pro skupinu A a oblasti M5–M6 pro skupinu B. Zdá se tedy, že zmiňovaný trend ve fyziologických datech, který je velmi obdobně pozorován v případě NTLx, by skutečně mohl reflektovat únavu.

Výsledky subjektivního hodnocení mimo jiné poukazují na skutečnost, že skupina A a B od sebe byly mírně odlišné, přičemž se zdá, že vzájemný posun dat je přibližně jedno měření. Nejvíce je uvedené zřetelné v kategoriích MD a PD, jež byly odlišeny na základě výsledků rANOVA, přičemž tento fenomén potvrzují také výsledky HRV. S ohledem na stanovenou hladinu významnosti pak lze považovat za rozdílnou také celkovou zátěž reprezentovanou NTLx, jelikož p-hodnota je na hranici kritické hodnoty. Zejména právě rozdíl v celkové vnímané zátěži je zajímavý v kontextu nastavení experimentu. Měření skupiny A probíhala v letních měsících, specificky od května do července. Měření skupiny B pak probíhala v měsících zimních, od prosince do února. Právě rozdíl v ročních obdobích, projevující se výrazným rozdílem mezi dobami východu a západu slunce (fotoperiodami) se zdá být přispěvatelem k rozdílu mezi celkovou pocíťovanou zátěží. Uvedené je také v kontextu se studii zaměřenými na pracovníky ve směnných industriálních provozech, přičemž např. Ogiński et al. (1976) na základě studie provedené 70. letech minulého století na

vzorku 1664 pracovníků železáren konstatují, že změny denní doby v kontextu ročního období představují signifikantní faktor z hlediska úrovně únavy s ohledem na interferenci těchto změn s cirkadiánními rytmy. Uvedené také podporují závěry Friberg et al. (2011), kteří porovnávali mimo jiné subjektivně pocíťovanou únavu mezi letním a zimním a lrtním období v Tromsø (severní Norsko, velké sezonní rozdíly ve fotoperiodě) a v Accra (Ghana, téměř žádné sezonní rozdíly ve fotoperiodě). Na základě dotazníkového šetření mezi 180 subjekty v Ghaně a 150 subjekty v Norsku jsou zřejmé výrazné rozdíly v pocíťované únavě v závislosti na ročním období (leden vs. srpen) u subjektů v Norsku, a téměř žádné nebo minimální u subjektů v Ghaně. Problematika sezónnosti, resp. délky fotoperiody, pak není, nejen v oblasti letectví dostatečně adresována a studie věnující se únavě v letectví se problematice délky fotoperiody nevěnují.

## 5.4 Výkonnostní testy

Ačkoli v případě jednoduchého reakčního času nebyly nalezeny žádné statisticky signifikantní výsledky při pohledu na trendy je zřejmá zejména minimální výkonnost v M4, u skupiny A, resp. M5 u skupiny B. Tyto výsledky opět podkládají předpokad o vysoké úrovni únavy v ranních hodinách projevující se sníženou výkonností. Současně je zřetelný trend posunu měření v jednotlivých skupinách, který podporuje diskutovaný vliv roční doby, resp. délky fotoperiody. Uvedené je pak méně zřetelné v případě výběrového reakčního času (4choice), v němž je sice viditelný rozdíl mezi trendy, avšak nejhorší výkonnost nastává shodně v M5. Výrazný posun minimální výkonnosti je ale zřetelný u komplexnějších úloh, specificky u OrTestu. Z výsledků OrTestu je zřejmý velmi podobný průběh výkonnosti mezi jednotlivými úlohami, s výjimkou úlohy 2. Ve skupině B je pak sledován největší pokles výkonnosti při pátém měření v případě jednodušších úloh (T1–T2), v pozdějších měření pak u komplexních úloh (T3–T4). Právě u komplexních úloh je pak předpoklad vyššího negativního vlivu únavy, což konstatují také Wilson et al. (2007b) s ohledem na skutečnost, že pozorovali rozdílnou výkonnost v kontextu náročnosti úlohy v případě únavy. Rozdílná výkonnost mezi jednoduchými a komplexními úlohami je patrná také z PCA analýzy, která umožňuje rozdělení měření do dvou shluků. Výsledky naznačují, že indikované trendy poklesu výkonnosti ve většině případů odpovídají výsledkům fyziologických měření a subjektivního hodnocení, přičemž nejnižší výkonnost u skupiny A je pozorována v M4, kde je, na základě HRV a NASA-TLX také indikované pravděpodobné maximum



únavy. Shodně lze diskutovat také pro skupinu B, s výjimkou T3, kde je minimum výkonnosti pozorováno až v M7. V ostatních úlohách se však nejnižší výkonnost projevuje v M5, resp. M6, což opět velmi dobře koresponduje s předcházejícími výsledky. Na základě výsledků lze tedy pozorovat negativní vliv únavy na výkonnost, zejména v období nejvyšší únavy. Současně se zdá, že v případě jednoduchých úloh jsou projevy únavy mírnější.

## 5.5 Letová data

Letová data pak tuto skutečnost nepotvrzují, ale ani nevyvracejí. Zejména u skupiny B jsou mezi jednotlivými úseky pozorovány vzájemně opačné výsledky, tj, snižuje-li se výkonnost v jednom úseku, v druhém narůstá a vice versa. U skupiny A pak výsledky mezi úseky poskytují více synchronizované trendy, přičemž zejména v úseku 1 je patrný výrazný pokles výkonnosti v M4, M6 a M7. Výsledky však nejsou přesvědčivé a s ohledem na jejich nekonzistenci lze jen těžko konstatovat vliv únavy. Takto neprůkazné výsledky jsou pak pravděpodobně způsobeny malou zkušeností subjektů s instrumentálními lety, jelikož se všichni nacházeli v počátečních fázích instrumentálního výcviku. Při analýze bylo zjevné, že se některé subjekty často ztrácely, musely se vracet, neletěly předepsanou trať apod. Uvedené je pak těžko přisuzovat únavě, jelikož tyto fenomény se vyskytovaly průřezově přes všechna měření. Uvedené pak zanáší nejistotu do výsledků a ty nejsou pro hodnocení relevantní.

## 5.6 Klasifikace

S cílem potvrzení předpokladu o navození různých úrovní únavy a jejich vzájemné odlišitelnosti bylo využito metod strojového učení, přičemž výsledná výkonnost klasifikátoru dosahovala úrovně okolo 80 %. Z hlediska klasifikace se jedná o velmi dobrý výsledek a bez ohledu na dříve diskutované limitace je zřejmé, že jednotlivá měření, reflektující odlišnou úroveň únavy jsou vzájemně odlišitelná na základě subjektivního hodnocení a výkonnostního testu – OrTest. Tento výsledek podkládá hypotézu o významném vlivu únavy na výkonnost, jelikož právě s ohledem na subjektivně pocítovanou zátěž, která byla ovlivněna pouze prodlouženou dobou bdělosti a denní dobou, a výkonnost v jednoduchých i komplexních testech bylo možné velmi přesně klasifikovat jednotlivá měření. Současně, při pohledu na konfuzní matici je zřejmé, že dalšího zvýšení by bylo dosaženo spojením dvojic měření M1–M2, M3–M4, M5–M6 a M7–M8 do čtyř

kategorií. To by znamenalo, že obdobná úroveň únavy byla pozorována vždy přes dvě měření, tj. v přibližně v tříhodinových blocích.

Důležité je zmínit, že i přes limitace studie se projevil negativní vliv únavy na výkonnost, a že klasifikace, resp. rozlišení úrovně únavy je možné na základě dvou jednoduchých testů, resp. jednoho testu a jednoho dotazníku. To poskytuje případnou možnost poměrně snadné implementace do praxe, bez nutných obsáhlých znalostí a bez potřeby nákladného vybavení. Na druhou stranu, studie byla provedena na poměrně malém vzorku subjektů a pro potvrzení, příp. upřesnění výsledků by bylo třeba dalších měření. To však nesnižuje přínos studie z hlediska nového náhledu do problematiky únavy a jejího vlivu na výkonnost v leteckém prostředí se specifickým zaměřením na piloty. Studie byla prováděna jako pilotní, jelikož uvedený koncept nebyl dříve využit a koncept 24hodinových měření je značně unikátní.

# Závěr

Dizertační práce je zaměřena na problematiku únavy v letectví, zejména s ohledem na snahu o její objektivizaci a vliv na výkonnost se zaměřením na piloty. Na základě identifikace současného stavu byly identifikovány limitace současného přístupu, které pak sloužily jako podklad pro vytvoření experimentálního konceptu. Současně byly vzaty v úvahu také současné poznatky související s objektivizací a detekcí únavy v dopravním prostředí. Experimentální design zahrnoval měření s maximální dobou bdělosti 34 hodin, přičemž samotné testování probíhalo v rozsahu 24 hodin, zbylých 8 hodin bylo vyhrazeno pro předletový režim s cílem snahy o co nejuniformější měření. V průběhu měřených 24 hodin subjekty absolvovaly celkem 8 měřených simulovaných letů, které byly proloženy bezletovými periodami vyhrazenými pro provádění poletových výkonnostních testů a subjektivního hodnocení. Experiment byl rozdělen do dvou fází, přičemž první probíhala v letních měsících (květen–červenec) a druhá v zimním období (prosinec–únor). V druhé fázi také byla rozšířena baterie využitých testů a sbíraných dat. U subjektů byly sledovány fyziologické signály (elektická aktivita srdce – obě skupiny, elektrická mozková aktivita – skupina B), které byly současně s letovými daty a pohybovou aktivitou (skupina B) měřeny v průběhu letu. V poletových pauzách byla měřena výkonnost prostřednictvím testů zaměřených na jednoduchý a výběrový reakční čas a výkonnostního testu, tzv. OrTestu. Dále bylo sbíráno subjektivní hodnocení prostřednictvím standardizovaného dotazníku NASA Task Load Index a v případě skupiny B také prostřednictvím jeho upravené verze zaměřené na únavu. Z výsledků vyplývá, že všechna hodnocená data vykazují obdobný trend, který reflektuje předpokládaný průběh proměnlivé hladiny únavy, primárně reprezentovaný maximem únavy v době cirkadiálního útlumu a ranních hodinách. V tomto období je primárně pozorován útlum organismu, a snížení výkonnosti. Uvedené potvrzuje také sebehodnocení, které sleduje obdobný trend v případě celkové zátěže, která je ovlivněna pouze denní dobou a dobou bdělosti. Na základě výsledků je možné klasifikovat jednotlivá měření, reflektující různé úrovně únavy s vysokou přesností, přičemž pro klasifikaci jsou využívány pouze data z NASA-TLX a OrTestu. Z výsledků tedy vyplývá, že je možné rozlišit určitou úroveň únavy na základě jednoduchého dotazníku a výkonnostního testu.

Vyhodnocení testu hypotéz je následující:

### **H1 Změna úrovně únavy se projeví na každém ze sledovaných fyziologických parametrů.**

Hypotéza byla odmítnuta. Změna úrovně únavy se projevila pouze na vybraných parametrech, přičemž většinu měření nebylo možné, na základě statistické analýzy vzájemně rozlišit ani u těchto. Současně však byl, u většiny parametrů pozorován jednotný, s únavou souhlasný trend.

### **H2 Změna úrovně únavy se projeví na výsledcích výkonnostních testů.**

Hypotéza byla odmítnuta. Změna úrovně únavy nebyla ve většině případů statisticky signifikantní, avšak na základě dat je zřetelný trend korespondující s úrovní únavy a současně výkonnostní test – OrTest – je využit pro klasifikaci jednotlivých měření reflektujících odlišné úrovně únavy.

### **H3 Vybrané metody mohou být využity pro účely dostatečně citlivé klasifikace úrovně únavy.**

Hypotéza byla přijata. Vybrané metody, dokáží dostatečně kvalitně klasifikovat různé úrovně únavy. Současně se ukázalo, že pro tuto klasifikaci stačí pouze dvě jednotlivé metody, kombinující subjektivní a objektivní hodnocení. Klasifikace založená na takto malém počtu jednoduchých, na vybavení a expertní znalost nenáročných metod pak nabízí velký potenciál pro praktické využití, jelikož může sloužit k exaktnímu hodnocení úrovně únavy a může tak rozšířit současné metody založené pouze na subjektivním hodnocení o objektivní data.

Nejvýznamnější limitací studie je zřejmě počet subjektů, jelikož v jednotlivých fázích bylo měřeno vždy 8 subjektů. Uvedené bylo způsobeno zejména vysokou časovou a personální náročností experimentů, přičemž byl limitován jak počet dostupných subjektů, tak i kapacity obsluhujícího personálu. Některé sledované signály se pak neukázaly jako vhodné pro hodnocení, resp. jejich zařazení do diskuze, a to s ohledem na technické problémy vzniklé na straně měřicího hardware v případě pohybové aktivity, a letových dat pravděpodobně zatížených chybou v podobě nízkých zkušeností subjektů s přístrojovými lety. Současně probíhala měření v jednotlivých fázích v odlišná roční období, tj. ve dnech s odlišnou délkou fotoperiody. Výsledky naznačují, že mezi skupinami je malý, avšak pozorovatelný posun v datech.

Uvedené však nelze brát pouze jako limitaci, ale také jako příležitost. Délka fotoperiody se zdá jako signifikantní faktor ovlivňující úroveň únavy, jak také naznačují studie z jiných odvětví. Současně tato problematika není v letectví

dostatečně adresována. Výsledky studie pak poukazují na jednoduchou možnost klasifikace úrovní únavy spojením výkonnostního testování a subjektivního hodnocení a s ohledem na unikátnost studie také nabízejí nový náhled do problematiky únavy.

S ohledem na limitace je však nezbytné v problematice pokračovat a validovat koncept navržené klasifikace na dalších měřeních. V tomto případě by měl být zejména rozšířen výzkumný vzorek. Současně by bylo vhodné uvažovat sezónnost jako faktor. Dále by bylo vhodné rozšířit sbíraná data o subjektivní hodnocení o jednu z dostupných stupnic ospalosti.

# Seznam použité literatury

- Akerstedt, T. (2000). Consensus statement: fatigue and accidents in transport operations. *Journal of sleep research*, 9(4), 395–395.
- Ancoli-Israel, S., Cole, R., Alessi, C., Chambers, M., Moorcroft, W., & Pollak, C. P. (2003). The role of actigraphy in the study of sleep and circadian rhythms. *Sleep*, 26(3), 342–392.
- Apparies, R. J., Riniolo, T. C., & Porges, S. W. (1998). A psychophysiological investigation of the effects of driving longer-combination vehicles. *Ergonomics*, 41(5), 581–592.
- Armstrong, L. E., Ganio, M. S., Casa, D. J., Lee, E. C., McDermott, B. P., Klau, J. F., Jimenez, L., Le Bellego, L., Chevillotte, E., & Lieberman, H. R. (2012). Mild dehydration affects mood in healthy young women. *The Journal of nutrition*, 142(2), 382–388.
- Arsintescu, L. (2017). Psychomotor vigilance task evaluation for touchscreen devices. In *International Conference on Managing Fatigue*, ARC-E-DAA-TN40373.
- Arsintescu, L., Kato, K. H., Cravalho, P. F., Feick, N. H., Stone, L. S., & Flynn-Evans, E. E. (2019). Validation of a touchscreen psychomotor vigilance task. *Accident Analysis & Prevention*, 126, 173–176.
- Bendak, S. (2003). 12-h workdays: current knowledge and future directions. *Work and stress*, 17(4), 321–336.
- Bendak, S., & Rashid, H. S. (2020). Fatigue in aviation: A systematic review of the literature. *International Journal of Industrial Ergonomics*, 76, 102928.  
URL <https://doi.org/10.1016/j.ergon.2020.102928>
- Board, N. R. C. U. T. R. (1998). *Transportation Research Record*. Transportation Research Record. Transportation Research Board, Commission on Sociotechnical Systems, National Research Council, National Academy of Sciences.  
URL <https://books.google.cz/books?id=GtZSAAAAMAAJ>
- Borchert, K. (2016). Simple reaction time task - visual.  
URL [https://www.millisecond.com/download/library/v6/simplereactiontime/simplereactiontime\\_visual/srtvisual.manual](https://www.millisecond.com/download/library/v6/simplereactiontime/simplereactiontime_visual/srtvisual.manual)

- Borchert, K. (2017). Four choice reaction time task - keyboard input.  
URL <https://www.millisecond.com/download/library/v6/fourchoicereactiontimetask/fourchoicereactiontimetask/fourchoicereactiontimetask.manual>
- Borghini, G., Astolfi, L., Vecchiato, G., Mattia, D., & Babiloni, F. (2014). Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. *Neuroscience & Biobehavioral Reviews*, *44*, 58–75.
- Bourgeois-Bougrine, S., Carbon, P., Gounelle, C., Mollard, R., & Coblenz, A. (2003). Perceived fatigue for short-and long-haul flights: a survey of 739 airline pilots. *Aviation, space, and environmental medicine*, *74*(10), 1072–1077.
- Boyle, L. N., Tippin, J., Paul, A., & Rizzo, M. (2008). Driver performance in the moments surrounding a microsleep. *Transportation research part F: traffic psychology and behaviour*, *11*(2), 126–136.
- Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32.  
URL <https://doi.org/10.1023/a:1010933404324>
- Brennan, M., Palaniswami, M., & Kamen, P. (2001). Do existing measures of poincare plot geometry reflect nonlinear features of heart rate variability? *IEEE Transactions on Biomedical Engineering*, *48*(11), 1342–1347.  
URL <https://doi.org/10.1109/10.959330>
- Byeon, M.-K., Han, S.-W., Min, H.-K., Wo, Y.-S., Park, Y.-B., & Huh, W. (2006). A study of hrv analysis to detect drowsiness states of drivers. In *Proceedings of the 24th IASTED international conference on Biomedical engineering*, (pp. 153–155).
- Cajochen, C., Brunner, D. P., Krauchi, K., Graw, P., & Wirz-Justice, A. (1995). Power density in theta/alpha frequencies of the waking eeg progressively increases during sustained wakefulness. *Sleep*, *18*(10), 890–894.
- Cajochen, C., Kräuchi, K., & Wirz-Justice, A. (2003). Role of melatonin in the regulation of human circadian rhythms and sleep. *Journal of neuroendocrinology*, *15*(4), 432–437.
- Caldwell, J. A. (2005). Fatigue in aviation. *Travel medicine and infectious disease*, *3*(2), 85–96.

- Caldwell, J. A., Mallis, M. M., Caldwell, J. L., Paul, M. A., Miller, J. C., & Neri, D. F. (2009). Fatigue countermeasures in aviation. *Aviation, space, and environmental medicine*, *80*(1), 29–59.
- Carey, J. J., & Delaney, M. F. (2009). T-scores and z-scores. *Clinical Reviews in Bone and Mineral Metabolism*, *8*(3), 113–121.  
URL <https://doi.org/10.1007/s12018-009-9064-4>
- Chang, C.-C., & Lin, C.-J. (2011). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, *2*, 27:1–27:27. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- Civil Aviation Safety Authority (2010). Biomathematical fatigue modeling in civil aviation fatigue risk management: Application guidance. Version 1.0.
- Civil Aviation Safety Authority (2013). Civil aviation order 48.1 Instrument 2013.
- Civil Aviation Safety Authority (2014). Biomathematical fatigue models: Guidance document.
- Clifford, G. D., & Tarassenko, L. (2005). Quantifying errors in spectral estimates of hrv due to beat replacement and resampling. *IEEE transactions on biomedical engineering*, *52*(4), 630–638.
- Dawson, D., Noy, Y. I., Härmä, M., Åkerstedt, T., & Belenky, G. (2011). Modelling fatigue and the use of fatigue models in work settings. *Accident Analysis & Prevention*, *43*(2), 549–564.
- Di Milia, L., Smolensky, M. H., Costa, G., Howarth, H. D., Ohayon, M. M., & Philip, P. (2011). Demographic factors, fatigue, and driving accidents: An examination of the published literature. *Accident Analysis & Prevention*, *43*(2), 516–532.
- Dinges, D. F., & Powell, J. W. (1985). Microcomputer analyses of performance on a portable, simple visual rt task during sustained operations. *Behavior research methods, instruments, & computers*, *17*(6), 652–655.
- Dongen, H. P. V., & Hursh, S. R. (2011). Fatigue, performance, errors, and accidents. In *Principles and Practice of Sleep Medicine*, (pp. 753–759). Elsevier.  
URL <https://doi.org/10.1016/b978-1-4160-6645-3.00067-0>
- Esteban, R. G. (2008). Heart rate variability characterization using entropy measures. *Faculdade de Engenharia da Universidade do Porto*.



- Francesco, B., Grazia, B. M., Emanuele, G., Valentina, F., Sara, C., Chiara, F., Riccardo, M., & Francesco, F. (2012). Linear and nonlinear heart rate variability indexes in clinical practice. *Computational and Mathematical Methods in Medicine*, 2012, 1–5.  
URL <https://doi.org/10.1155/2012/219080>
- Friborg, O., Bjorvatn, B., Amponsah, B., & Pallesen, S. (2011). Associations between seasonal variations in day length (photoperiod), sleep timing, sleep quality and mood: a comparison between Ghana (5°) and Norway (69°). *Journal of Sleep Research*, 21(2), 176–184.  
URL <https://doi.org/10.1111/j.1365-2869.2011.00982.x>
- FRMS Task Force (2011). Fatigue risk management systems: Implementation guide for operators.
- Goel, N., Basner, M., Rao, H., & Dinges, D. F. (2013). Circadian rhythms, sleep deprivation, and human performance. *Progress in molecular biology and translational science*, 119, 155–190.
- Gregory, K. B., Winn, W., Johnson, K., & Rosekind, M. R. (2010). Pilot fatigue survey: Exploring fatigue factors in air medical operations. *Air Medical Journal*, 29(6), 309–319.  
URL <https://doi.org/10.1016/j.amj.2010.07.002>
- Hart, S. G. (2006). Nasa-task load index (nasa-tlx); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting*, vol. 50, (pp. 904–908). Sage Publications Sage CA: Los Angeles, CA.
- Hart, S. G., & Staveland, L. E. (1988). Development of nasa-tlx (task load index): Results of empirical and theoretical research. In *Advances in psychology*, vol. 52, (pp. 139–183). Elsevier.
- Herman, L. (2020). Porovnaní subjektivního a objektivního hodnocení únavy pilotůu.
- Hoddes, E., Zarcone, V., & Dement, W. (1972). Stanford sleepiness scale. *Enzyklopädie der Schlafmedizin*, 1184.
- Horne, J. A., & Reyner, L. A. (1995). Sleep related vehicle accidents. *Bmj*, 310(6979), 565–567.
- IATA FRMS Task Force (2014). Fatigue safety performance indicators (SPIs): A key component of proactive fatigue hazard identification.

- IATA FRMS Task Force (2015). Common protocol for minimum data collection variables in aviation operations.
- International Civil Aviation Organization (2014). Fatigue risk management systems: Manual for regulators.
- International Organization for Standardization (2017). ISO 7250-1:2017 Basic human body measurements for technological design — Part 1: Body measurement definitions and landmarks.
- Jap, B. T., Lal, S., Fischer, P., & Bekiaris, E. (2009). Using eeg spectral components to assess algorithms for detecting fatigue. *Expert Systems with Applications*, *36*(2), 2352–2359.
- Johns, M. W. (1991). A new method for measuring daytime sleepiness: the epworth sleepiness scale. *sleep*, *14*(6), 540–545.
- Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *374*(2065), 20150202.  
URL <https://doi.org/10.1098/rsta.2015.0202>
- Karaim, M., Noureldin, A., & Karamat, T. B. (2019). Low-cost imu data denoising using savitzky-golay filters. In *2019 International Conference on Communications, Signal Processing, and their Applications (ICCSPA)*, (pp. 1–5). IEEE.
- Kittnar, O., et al. (2011). *Lékařská fyziologie*. Grada.
- Kohler, B.-U., Hennig, C., & Orglmeister, R. (2002). The principles of software QRS detection. *IEEE Engineering in Medicine and Biology Magazine*, *21*(1), 42–57.  
URL <https://doi.org/10.1109/51.993193>
- Kreitzman, L., & Foster, R. (2011). *The Rhythms Of Life: The Biological Clocks That Control the Daily Lives of Every Living Thing*. Profile.  
URL <https://books.google.cz/books?id=X1UyNPDMY8sC>
- Kripke, D. F., Garfinkel, L., Wingard, D. L., Klauber, M. R., & Marler, M. R. (2002). Mortality associated with sleep duration and insomnia. *Archives of general psychiatry*, *59*(2), 131–136.
- Lal, S. K., Craig, A., Boord, P., Kirkup, L., & Nguyen, H. (2003). Development of an algorithm for an EEG-based driver fatigue countermeasure. *Journal of*

*Safety Research*, 34(3), 321–328.

URL [https://doi.org/10.1016/s0022-4375\(03\)00027-6](https://doi.org/10.1016/s0022-4375(03)00027-6)

Lamond, N., DAWSON, D., & Roach, G. D. (2005). Fatigue assessment in the field: validation of a hand-held electronic psychomotor vigilance task. *Aviation, space, and environmental medicine*, 76(5), 486–489.

Lindbergh, C. A. (2003). *The Spirit of St. Louis*. Scribner.

Lomb, N. R. (1976). Least-squares frequency analysis of unequally spaced data. *Astrophysics and space science*, 39(2), 447–462.

Lu, K., Dahlman, A. S., Karlsson, J., & Candefjord, S. (2022). Detecting driver fatigue using heart rate variability: A systematic review. *Accident Analysis & Prevention*, 178, 106830.

URL <https://doi.org/10.1016/j.aap.2022.106830>

Maćkiewicz, A., & Ratajczak, W. (1993). Principal components analysis (PCA). *Computers & Geosciences*, 19(3), 303–342.

URL [https://doi.org/10.1016/0098-3004\(93\)90090-r](https://doi.org/10.1016/0098-3004(93)90090-r)

Malik, M., Bigger, J. T., Camm, A. J., Kleiger, R. E., Malliani, A., Moss, A. J., & Schwartz, P. J. (1996). Heart rate variability: Standards of measurement, physiological interpretation, and clinical use. *European Heart Journal*, 17(3), 354–381.

URL <https://doi.org/10.1093/oxfordjournals.eurheartj.a014868>

Martin, E., Alan, B., & Martin, F. (2012). *Základy srdeční elektrofyzologie a katérových ablací*. Grada Publishing as.

McCallum, A., Nigam, K., et al. (1998). A comparison of event models for naive bayes text classification. In *AAAI-98 workshop on learning for text categorization*, vol. 752, (pp. 41–48). Madison, WI.

McNames, J., Thong, T., & Aboy, M. (2004). Impulse rejection filter for artifact removal in spectral analysis of biomedical signals. In *The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 1, (pp. 145–148). IEEE.

Missoni, E., Nikolić, N., & Missoni, I. (2009). Civil aviation rules on crew flight time, flight duty, and rest: comparison of 10 icao member states. *Aviation, space, and environmental medicine*, 80(2), 135–138.

- Murtagh, F., & Contreras, P. (2011). Algorithms for hierarchical clustering: an overview. *WIREs Data Mining and Knowledge Discovery*, 2(1), 86–97.  
URL <https://doi.org/10.1002/widm.53>
- Myles, A. J., Feudale, R. N., Liu, Y., Woody, N. A., & Brown, S. D. (2004). An introduction to decision tree modeling. *Journal of Chemometrics: A Journal of the Chemometrics Society*, 18(6), 275–285.
- Nielsen, F. (2016). Hierarchical clustering. In *Introduction to HPC with MPI for Data Science*, (pp. 195–211). Springer International Publishing.  
URL [https://doi.org/10.1007/978-3-319-21903-5\\_8](https://doi.org/10.1007/978-3-319-21903-5_8)
- Ogiński, A., Koźlakowska-Swigoń, L., & Pokorski, J. (1976). Diurnal and seasonal variations in industrial fatigue of shift workers. *Proceedings of the Human Factors Society Annual Meeting*, 20(22), 515–518.  
URL <https://doi.org/10.1177/154193127602002203>
- Olson, C. F. (1995). Parallel algorithms for hierarchical clustering. *Parallel Computing*, 21(8), 1313–1325.  
URL [https://doi.org/10.1016/0167-8191\(95\)00017-i](https://doi.org/10.1016/0167-8191(95)00017-i)
- Pan, J., & Tompkins, W. J. (1985). A real-time QRS detection algorithm. *IEEE Transactions on Biomedical Engineering*, BME-32(3), 230–236.  
URL <https://doi.org/10.1109/tbme.1985.325532>
- Patel, M., Lal, S., Kavanagh, D., & Rossiter, P. (2011). Applying neural network analysis on heart rate variability data to assess driver fatigue. *Expert Systems with Applications*, 38(6), 7235–7242.  
URL <https://doi.org/10.1016/j.eswa.2010.12.028>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Peng, C.-K., Havlin, S., Stanley, H. E., & Goldberger, A. L. (1995). Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series. *Chaos: an interdisciplinary journal of nonlinear science*, 5(1), 82–87.
- Piskorski, J., & Guzik, P. (2005). Filtering poincare plots. *Computational methods in science and technology*, 11(1), 39–48.

- Powell, D., Spencer, M. B., Holland, D., Broadbent, E., & Petrie, K. J. (2007). Pilot fatigue in short-haul operations: effects of number of sectors, duty length, and time of day. *Aviation, Space, and Environmental Medicine*, *78*(7), 698–701.
- Qi, P., Ru, H., Gao, L., Zhang, X., Zhou, T., Tian, Y., Thakor, N., Bezerianos, A., Li, J., & Sun, Y. (2019). Neural mechanisms of mental fatigue revisited: New insights from the brain connectome. *Engineering*, *5*(2), 276–286.  
URL <https://doi.org/10.1016/j.eng.2018.11.025>
- Quinlan, J. R. (1986). Induction of decision trees. *Machine learning*, *1*(1), 81–106.
- Rach, S., Diederich, A., & Colonius, H. (2010). On quantifying multisensory interaction effects in reaction time and detection rate. *Psychological Research*, *75*(2), 77–94.  
URL <https://doi.org/10.1007/s00426-010-0289-0>
- Ramshur, J. T. (2010). *Design, evaluation, and application of heart rate variability analysis software (HRVAS)*. Ph.D. thesis, University of Memphis Memphis, TN.
- Reis, C., Mestre, C., & Canhão, H. (2013). Prevalence of fatigue in a group of airline pilots. *Aviation, Space, and Environmental Medicine*, *84*(8), 828–833.  
URL <https://doi.org/10.3357/asem.3548.2013>
- Roach, G. D., Darwent, D., Sletten, T. L., & Dawson, D. (2011). Long-haul pilots use in-flight napping as a countermeasure to fatigue. *Applied ergonomics*, *42*(2), 214–218.
- Samel, A., Wegmann, H.-M., & Vejvoda, M. (1997). Aircrew fatigue in long-haul operations. *Accident Analysis & Prevention*, *29*(4), 439–452.
- Samn, S. W., & Perelli, L. P. (1982). Estimating aircrew fatigue: a technique with application to airlift operations. Tech. rep., School of Aerospace Medicine Brooks Afb tx.
- scikit-learn developers (2019). 1.17. neural network models (supervised).  
URL [https://scikit-learn.org/stable/modules/neural\\_networks\\_supervised.html](https://scikit-learn.org/stable/modules/neural_networks_supervised.html)
- Segerstrom, S. C., & Nes, L. S. (2007). Heart rate variability reflects self-regulatory strength, effort, and fatigue. *Psychological science*, *18*(3), 275–281.
- Shaffer, F., & Ginsberg, J. (2017). An overview of heart rate variability metrics and norms. *Frontiers in public health*, *5*, 258.

- Shahid, A., Wilkinson, K., Marcu, S., & Shapiro, C. M. (2011). Karolinska sleepiness scale (kss). In *STOP, THAT and One Hundred Other Sleep Scales*, (pp. 209–210). Springer.
- Shaker, M. M. (2006). Eeg waves classifier using wavelet transform and fourier transform. *brain*, 2(3).
- Sharma, V. (2009). Deterministic chaos and fractal complexity in the dynamics of cardiovascular behavior: Perspectives on a new frontier. *The Open Cardiovascular Medicine Journal*, 3(1), 110–123.  
URL <https://doi.org/10.2174/1874192400903010110>
- Smolensky, M. H., Di Milia, L., Ohayon, M. M., & Philip, P. (2011). Sleep disorders, medical conditions, and road accident risk. *Accident Analysis & Prevention*, 43(2), 533–548.
- Socha, V., Kušmírek, S., Hanáková, L., Lališ, A., Pereverzeva, A., Průcha, J., Socha, L., & Hanák, P. (2018). Design of reaction time measuring device. In *Proc. of 22nd International Scientific on Conference Transport Means 2018*, (pp. 580–584). Kaunas University of Technology.
- Socha, V., Kušmírek, S., Hanáková, L., Olexa, P., & Šimerka, J. (2020). Portable reaction time measurement device: Prototype proposal and validation. In *2020 New Trends in Civil Aviation (NTCA)*, (pp. 135–138). IEEE.
- Strohrmann, C., Harms, H., Kappeler-Setz, C., & Troster, G. (2012). Monitoring kinematic changes with fatigue in running using body-worn sensors. *IEEE transactions on information technology in biomedicine*, 16(5), 983–990.
- Svensson, E., Angelborg-Thanderez, M., Sjöberg, L., & Olsson, S. (1997). Information complexity-mental workload and performance in combat aircraft. *Ergonomics*, 40(3), 362–380.
- Taghizadeh, S., Pirouzi, S., Zamani, A., Motealleh, A., & Bagheri, Z. (2017). Does muscle fatigue alter EEG bands of brain hemispheres? *Journal of Biomedical Physics and Engineering*.  
URL <https://doi.org/10.31661/jbpe.v0i0.621>
- Taunk, K., De, S., Verma, S., & Swetapadma, A. (2019). A brief review of nearest neighbor algorithm for learning and classification. In *2019 International Conference on Intelligent Computing and Control Systems (ICCS)*, (pp. 1255–1260). IEEE.

- Thomas, L. C., Gast, C., Grube, R., & Craig, K. (2015). Fatigue detection in commercial flight operations: Results using physiological measures. *Procedia Manufacturing, 3*, 2357–2364.
- Thorne, D. R., Johnson, D. E., Redmond, D. P., Sing, H. C., Belenky, G., & Shapiro, J. M. (2005). The walter reed palm-held psychomotor vigilance test. *Behavior Research Methods, 37*(1), 111–118.
- Thuraisingham, R. A. (2006). Preprocessing RR interval time series for heart rate variability analysis and estimates of standard deviation of RR intervals. *Computer methods and programs in biomedicine, 83*(1), 78–82.
- Townsend, J. T., Ashby, F. G., et al. (1983). *Stochastic modeling of elementary psychological processes*. CUP Archive.
- Tremaine, R., Dorrian, J., Lack, L., Lovato, N., Ferguson, S., Zhou, X., & Roach, G. (2010). The relationship between subjective and objective sleepiness and performance during a simulated night-shift with a nap countermeasure. *Applied ergonomics, 42*(1), 52–61.
- Van Dongen, H. (2004). Comparison of mathematical model predictions to experimental data of fatigue and performance. *Aviation, space, and environmental medicine, 75*(3), A15–A36.
- Van Dongen, H. P., Mott, C. G., Huang, J.-K., Mollicone, D. J., McKenzie, F. D., & Dinges, D. F. (2007). Optimization of biomathematical model predictions for cognitive performance impairment in individuals: accounting for unknown traits and uncertain states in homeostatic and circadian processes. *Sleep, 30*(9), 1129–1143.
- Veksler, B. Z., & Gunzelmann, G. (2018). Functional equivalence of sleep loss and time on task effects in sustained attention. *Cognitive science, 42*(2), 600–632.
- Wall, M. E., Rechtsteiner, A., & Rocha, L. M. (2003). Singular value decomposition and principal component analysis. In *A practical approach to microarray data analysis*, (pp. 91–109). Springer.  
URL <https://doi.org/10.48550/arXiv.physics/0208101>
- Watson, D. W. (2001). Physiological correlates of heart rate variability (hrv) and the subjective assessment of workload and fatigue in-flight crew: a practical study. In *Human Interfaces in Control Rooms, Cockpits and Command Centres, 2001. People in Control. The Second International Conference on (IEE Conf. Publ. No. 481)*, (pp. 159–163). IET.

- Webster, J. G. (2009). *Medical instrumentation: application and design*. John Wiley & Sons.
- Williamson, A., Lombardi, D. A., Folkard, S., Stutts, J., Courtney, T. K., & Connor, J. L. (2011). The link between fatigue and safety. *Accident Analysis & Prevention, 43*(2), 498–515.
- Willson, K., Francis, D. P., Wensel, R., Coats, A. J., & Parker, K. H. (2002). Relationship between detrended fluctuation analysis and spectral analysis of heart-rate variability. *Physiological measurement, 23*(2), 385.
- Wilson, G. F., Caldwell, J. A., & Russell, C. A. (2007a). Performance and psychophysiological measures of fatigue effects on aviation related tasks of varying difficulty. *The international journal of aviation psychology, 17*(2), 219–247.
- Wilson, G. F., Caldwell, J. A., & Russell, C. A. (2007b). Performance and psychophysiological measures of fatigue effects on aviation related tasks of varying difficulty. *The international journal of aviation psychology, 17*(2), 219–247.
- World Medical Association and others (2013). World medical association declaration of Helsinki: ethical principles for medical research involving human subjects. *Jama, 310*(20), 2191–2194.
- Wright, N., & McGown, A. (2001). Vigilance on the civil flight deck: incidence of sleepiness and sleep during long-haul flights and associated changes in physiological parameters. *Ergonomics, 44*(1), 82–106.  
URL <https://doi.org/10.1080/00140130150203893>
- Zhuzhgova, M. (2020). Methods for evaluation of fatigue in pilots. Bakalářská práce. České vysoké učení technické v Praze.



# Přílohy

<b>A</b>	<b>Publikace 1</b>	.....	108
<b>B</b>	<b>Publikace 2</b>	.....	115
<b>C</b>	<b>Publikace 3</b>	.....	121
<b>D</b>	<b>Publikace 4</b>	.....	135
<b>E</b>	<b>Výsledky EKG</b>	.....	142

## A Publikace 1

Hanakova, L., Socha, V., Socha, L., Lalis, A., Kraus, J., & Malich, T. (2019).  
**The influence of fatigue on psychophysiological indicators during 24  
hours testing of pilots.**

2019 IEEE 17th World Symposium on Applied Machine Intelligence and  
Informatics (SAMI) (pp. 181-186). IEEE.

# The influence of fatigue on psychophysiological indicators during 24 hours testing of pilots

Lenka Hanakova

*Department of Air Transport  
Czech Technical University in Prague  
Prague, Czech Republic  
hanakle1@fd.cvut.cz*

Vladimir Socha

*Department of Air Transport  
Czech Technical University in Prague  
Prague, Czech Republic  
sochavla@fd.cvut.cz*

Lubos Socha

*Department of Air Transport Management  
Technical University of Kosice  
Kosice, Slovak Republic  
lubos.socha@tuke.sk*

Andrej Lalis

*Department of Air Transport  
Czech Technical University in Prague  
Prague, Czech Republic  
lalisand@fd.cvut.cz*

Jakub Kraus

*Department of Air Transport  
Czech Technical University in Prague  
Prague, Czech Republic  
krausjak@fd.cvut.cz*

Tomas Malich

*Department of Air Transport  
Czech Technical University in Prague  
Prague, Czech Republic  
malictom@fd.cvut.cz*

**Abstract**—This study presents the issue of fatigue and its quantification by the means of psychophysiological indicators. For the purpose of the study, 24 hours experiment was proposed consisting of eight about an hour long simulated IFR flights. Eight students of the Department of Air Transport, Faculty of Transportation Sciences, Czech Technical University in Prague participated in the study, all with comparable theoretical and practical skills. During individual flights, ECG curve was recorded as a heart activity signal. From the ECG signal, individual RR intervals were extracted and subsequently an analysis in three areas was performed: time and frequency analysis and analysis by the means of Pointcaré plot. The acquired data were statistically evaluated and compared. Based on Friedman ANOVA and subsequent post-hoc analysis, statistically significant differences between first and sixth measurement were identified, which corresponded to the circadian low. It is also possible to claim that Pointcaré plot appears as a suitable means for fatigue quantification.

**Index Terms**—aviation safety, fatigue, heart rate variability, human factor, psychophysiological indicators, Pointcaré plot

## I. INTRODUCTION

In modern aviation, safety is one of the most important aspects. Consequently, it is important to deal with human factors as a part of risk understanding and its management so as to integrate the resulting knowledge into aviation and its legislation.

It is not unusual that air carriers require their pilots to fly maximum flight hours allowed by the regulations. In summer season the volume of traffic reaches maximum with most of airlines. Many flights are scheduled during night or early morning hours and considering high frequency and workload of the crews, symptoms of fatigue frequently appear.

Fatigue can be divided into physical and psychological. Physical fatigue implies incapability of a muscle or a group of muscles to maintain required performance. Psychological fatigue is then inability to maintain cognitive performance. This work deals with psychological fatigue and its indications. The fatigue results from the effect of circadian rhythms, poor or

insufficient rest, duty duration, shifts operation, psychological stress and strain. It is also important to distinguish between subjective feeling of stress or exhaustion and objective reduction in subject's performance as these are often not the same. Limitation in the assessment of the fatigue influence on non-standard situations is then its nature as it results in random intervals of inattention, reduced performance and similar. To meet the precondition to err, the intervals have to overlap with events of increased cognitive demand. The effect of fatigue is variable and depends on physical and psychological condition of an individual, his or her experience etc. [1]–[4].

Fatigue is a serious problem in all high-risk industries where human factors matter, especially in transportation. Pilots, air traffic controllers, drivers, ground handling and manufacturing facilities technology operators and others are all subject to fatigue. One of the most popular cases where fatigue played significant role were Three Mile Island and Chernobyl nuclear installation accidents, Challenger space shuttle accident or the collision of two Boeing 747 aircraft on Tenerife island. According to the U.S. National Transportation Safety Board (NTSB), fatigue has been identified as likely cause of accident or contributing factors in 20 % of all major investigations between the years 2001 and 2012. Fatigue reduces productivity and contributes to the influence of other factors affecting safety. For instance, it limits capability to assess situations and make adequate decisions, reduces situational awareness, decreases reaction time and degrades capability to maintain vigilance. At the same time, it impacts team cooperation as it increases apathy, irritation and willingness to take risks and reduces self-control. Fatigue itself then hinders respective individual to realistically evaluate its own state. [1], [2], [5], [6].

Aviation is specific environment where many of the aforementioned factors come into play. Flight crews often work in shifts and cross time zones, both disrupting their sleep cycles. Their workplace is typical for increased noise, vibrations, low

humidity and reduced level of oxygen, supporting even further the development of fatigue [7].

To date, most of aviation incidents and accidents are attributed to human factors. Fatigue is a significant factor affecting human performance thus subject of the discussions among aviation authorities (ICAO, FAA, EASA, etc.) so as among the aviation professionals (trade and professional unions) who all attempt to decrease its influence on incidents and accidents development, typically by regulating duty duration. This approach, however, is not fully effective as it does not account for the time spent off duty, such as transfer from airport to hotel when crews do not rest. Questionnaire surveys among pilots reveal that more than 80 % of the asked are faced with the effects of fatigue in cockpit and more that 40 % experienced at least one time on their duty a microsleep or fell asleep. From this an effort follows to record significant effects of fatigue and alert to them in real time [1], [2], [5], [7], [8].

Monitoring of aviation personnel fatigue is currently done subjectively, i.e. by the means of questionnaire surveys or self-evaluation [9]–[11]. Overall efforts to objectify fatigue seek methods of its assessment which would allow its quantification. In addition to that, it is important to consider the specifics of pilot profession. It is important to find a solution, which would not influence or bother pilot during his duty and which would be compact and mobile. It follows that the most suitable appears to be heart activity monitoring by the means of ECG curve, which is easily measurable and capable of reflecting fatigue [12], [13].

Heart activity, or more precisely heart rate variability (HRV), can be evaluated by many methods. Standard methods based on time and frequency analysis of RR intervals, however, do not appear to be sufficient with respect to the nature of ECG signal. Methods allowing deeper insight into the dynamics of a system are considered more suitable. One of such methods is a method based on non-linear dynamics, the so-called Pointcaré plot, which is a form of recurrent graph. It looks for self-similarity of given signal and provides new information compared to standard methods, i.e. information not only about long-term, but also about short-term variability [14], [15].

The goal of this paper is an objectification of pilots' fatigue by the means of simple HRV measurement, evaluated by the Pointcaré plot and its parameters, i.e. by method taking into account fluctuations present in the signal, providing not only information about long-term, but also short-term variability. A tool utilizing the method could contribute to increased safety in the context of human factors.

## II. MATERIALS AND METHODS

### A. Participants

For the purpose of the research, students of the study field Professional Pilot at the Department of Air Transport, Czech Technical University in Prague were measured. The subjects were selected to achieve maximum possible uniformity of the measured sample, especially with regard to the experience in flight hours. The sample consisted of 7 men and 1 woman, in the age of 21 - 25 years. All subjects successfully completed

VFR phase of integrated ATPL training. Some subjects were undergoing IFR training using FNPT II flight simulator, others had minimum flight hours as SEP IFR SPIC. Theoretical and practical background of the pilots was thus comparable.

### B. Experimental Setup and Data Acquisition

For the purpose of the measurement, 24 hours experiment consisting of eight flights was conducted. Each measurement started at 18:00 of local time, where both subjects participating were familiarized with the experiment, i.e. the simulator functioning, options for briefing using iPad and cockpit instrumentation setting were explained. The first subject commenced simulator measurement. The measurement lasted for about an hour with 20 minutes pre-flight briefing. The task was to depart, fly on pre-defined route and land with the aircraft on destination airport. The flight was conducted with IFR flight rules and ECG was measured during the flight. After landing, the subject was disconnected from ECG and familiarized with the course of the flight, i.e. he or her could examine the trajectory of the flight. Subsequently, the second subject was measured whilst the first subject completed series of psychological and intelligence tests and in the remaining time could rest, but not sleep.

All subjects had to follow predefined regime before the experiment started. They had to wake up at 08:00, length of sleep and eating habits were prescribed, such as avoidance of energy drinks and similar. These measures were taken to support fatigue and assure highest possible uniformity of individual measurements. It follows that during the 24 hours experiment, all subjects were awake for 36 hours.

Data about heart activity were collected by VLV-LAB measurement device (Joint Research Centre of the Czech Technical University and the Charles University in Prague, Prague, CZ), see Fig. 2, a device primarily intended for research experiments. The device serves as wireless polygraph, i.e. it is capable of recording multiple physiological signals. For the purpose of this research, only single channel EKG was measured with the utilization of one-time contact silver chloride (Ag/AgCl) electrodes. The signal was captured with sampling frequency of 1000 Hz and AD converter resolution of 24 bits.

### C. Data Processing and Data Analysis

VLV-LAB device was used to acquire complete ECG curve. For further analysis, it was necessary to extract individual RR intervals from the curve. To detect individual R waves, Pan-Tomkins method was used which includes a set of filters for comparison of ECG curve isoline and peak detector identifying individual R waves in compared curve, detecting QRS complex [16]. From the identified R waves, Pointcaré plots for measured signals were acquired and their SD1 and SD2 characteristics computed (their ratio). For the purpose of comparison with standard methods, selected parameters of time and frequency analysis were obtained.



Fig. 1. VLV-LAB measurement device with Ag/AgCl electrodes.

1) *Standard methods:* For the purpose of HRV analysis, time and frequency analysis parameters are used as standard. Methods of time analysis rely on base descriptive statistics. Most frequently used parameters are mean RR interval (meanRR), standard deviation of RR intervals or NN (normal-to-normal) intervals (SDNN), or alternatively mean heart rate (meanHR) with its standard deviation (stdHR) [17]. For the purpose of this study, these two pairs of parameters were selected.

Another approach is evaluation in frequency domain. There are three base parts of HRV frequency spectrum. The first is very low frequency (VLF) band ranging from 0.0033 Hz to 0.04 Hz, which is not sufficiently described from the perspective of physiology. The second is low frequency (LF) band ranging from 0.04 Hz to 0.15 Hz, which reflects the activity of textcolorredsympathetic and parasympathetic nervous system. The third is high frequency (HF) band ranging from 0.15 Hz - 0.40 Hz. The HF band is influenced only by parasympathetic nervous system [17], [18]. With respect to the nature of the signal, i.e. data not sampled equidistantly, Lomb-Scargle method based approach was selected. The method is based on fitting the signal by sinusoids using the smallest squares method. The signal is then expressed as the sum of sines and cosines with different amplitudes and frequencies, which is essentially an equivalent of the Fourier transform. [19], [20]. The application of frequency analysis on HRV data led to acquisition of individual bands performance. Consequently LF and HF parameters were selected for this study, including their ratio (LF/HF) which serves as an indicator of sympathovagal balance, i.e. reflecting the balance between long- and short-term HRV [21].

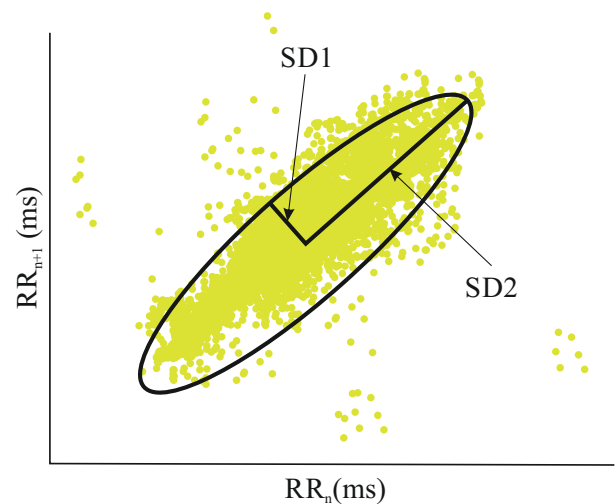


Fig. 2. Pointcaré plot with SD1 and SD2 parameters.

2) *Pointcaré plot:* The main purpose of Pointcaré plot (also return map) is acquisition of recurrent graph, i.e. plot of each RR interval against subsequent RR interval (see Fig. ??), allowing visual search for patterns hidden in time-series. Pointcaré plot is thus 2D reconstruction of RR intervals phase space, i.e. it reflects the dynamics of heart activity. It provides summary information about heart activity. This technique of data processing is becoming popular due to its simple visual representation, proved clinical capability of predicting heart dysfunction or diseases and also because, unlike the standard methods used to HRV analysis, it reflects nonlinear features of heart rate variability [17], [22].

As already mentioned, Pointcaré plot depicts RR interval (x-axis) against the subsequent RR interval (y-axis). The result is a cluster of points stretched along the diagonal (line of identity, LOI). For the purpose of quantification, Pointcaré plot is fitted by ellipse [23]. Next, the length of semi-major and semi-minor axes is determined (SD1 and SD2). SD1 parameter indicates the short-term variability, SD2 parameter indicates the long-term variability. Their ratio (SD2/SD1) then indicates the balance between long-term and short-term variability. This ratio, similar to LF/HF parameter, shows the balance between long- and short-term HRV [21].

#### D. Statistical analysis

For the purposes of statistical analysis of the acquired parameters, Friedman's test (Friedman non-parametric one-way ANOVA with repeated measures) [24] was used and implemented in Matlab program environment. Based the test results, post-hoc analysis was carried out with the usage of Dunn-Sidak test [25]. Distribution of individual parameters was depicted by boxplots. The analysis was carried out by own-designed software in Matlab environment (MATLAB R2017a, MathWorks, Inc., Natick, MA, USA). The testing was performed with 5% level of significance.

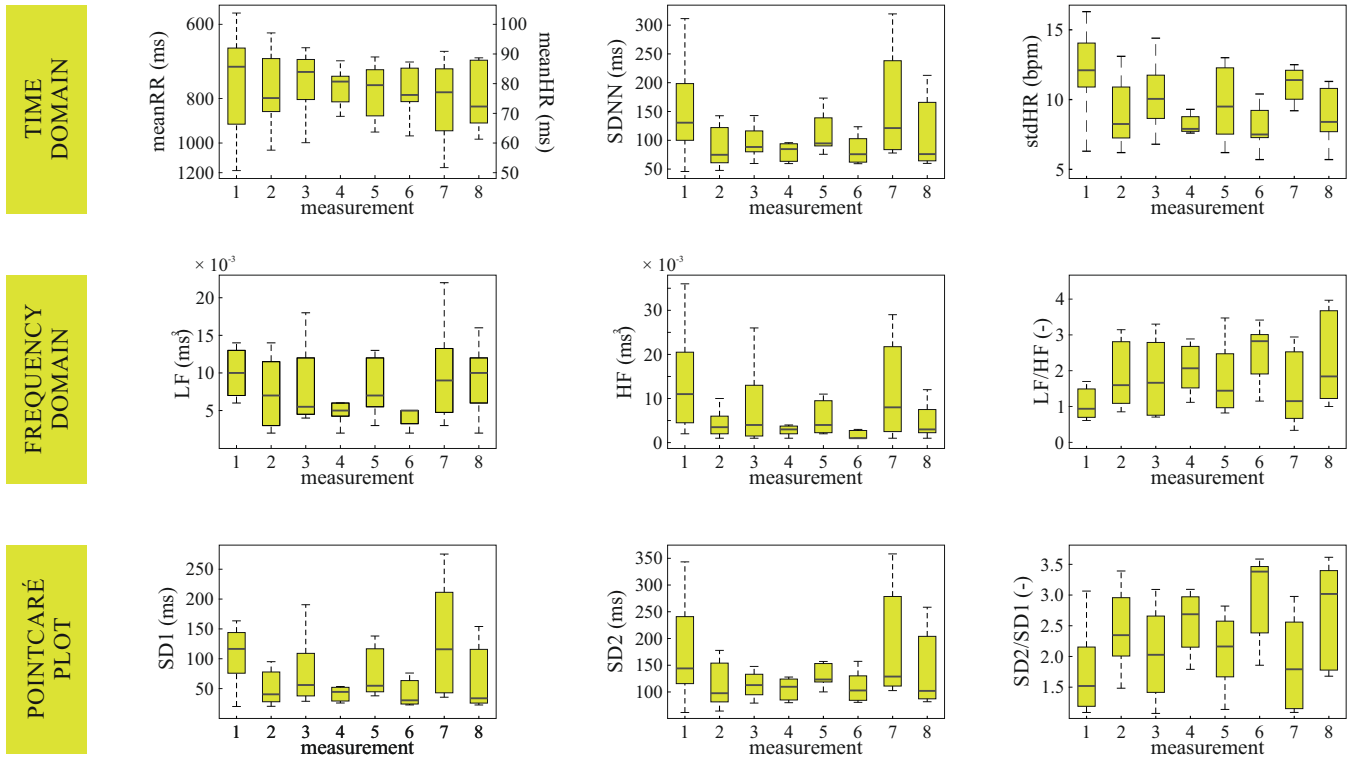


Fig. 3. Graphical representation of the distribution of the parameters.

### III. RESULTS

Friedman’s test served the comparison of individual 1 - 8 measurements. Chi-square ( $\chi^2$ ) expresses the calculated statistics. Next,  $p$ -value is presented or alternatively the confidence interval. Parameter distribution is depicted graphically, using boxplots.

Parameters of time analysis related to RR intervals show no statistically significant difference between individual measurements. In case of meanRR parameter for overall comparison,  $\chi^2(7) = 2.272$ ,  $p = 0.943$ . In case of SDNN parameter,  $\chi^2(7) = 10.152$  and  $p = 0.180$ . Distribution of individual measurements parameters are depicted in Fig. 3.

The parameter meanHR of time analysis shows no statistically significant difference between individual measurements. For overall comparison,  $\chi^2(7) = 2.185$ ,  $p = 0.949$ . In case of stdHR parameter,  $\chi^2(7) = 15.994$ ,  $p = 0.025$ . During the test of pairs from 1. and 6. measurement,  $p$ -value was close to 0.05, namely  $p = 0.059$ . Confidence interval was (-56.167, -0.761). Therefore, it is possible to claim that there was statistically significant difference between 1. and 6. measurement. Distribution of individual measurements parameters are depicted in Fig. 3.

The parameters of frequency analysis shows no statistically significant difference between individual measurements. In case of LF parameter for overall comparison,  $\chi^2(7) = 12.094$ ,  $p = 0.098$ , in case of HF parameter,  $\chi^2(7) = 10.270$  with  $p = 0.174$  and in case of LF/HF parameter,  $\chi^2(7) = 7.014$

with  $p = 0.427$ . Distribution of individual measurements parameters are depicted in Fig. 3.

Base parameters of Pointcaré plot shows no statistically significant difference between individual measurements. In case of SD1 parameter for overall comparison,  $\chi^2(7) = 11.826$  with  $p = 0.106$  and in case of SD2 parameter,  $\chi^2(7) = 8.478$  with  $p = 0.292$ . Distribution of individual measurements parameters are depicted in Fig. 3.

In case of SD2/SD1 parameter, Friedman’s test shows statistical significance difference between measurements,  $\chi^2(7) = 16.060$  and  $p = 0.025$ . Post-hoc analysis identified statistically significant difference between 1. and 6. measurements. Confidence interval was (-56.167, -0.761) and  $p$ -value for this pair of measurements was 0.038. Distribution of individual measurements parameters are depicted in Fig. 3.

### IV. DISCUSSION

From the results, it follows that in case of majority of parameters, no statistically significant differences between individual measurements were identified for the monitored subjects. Difference was identified only in case of the pair of 1. and 6. measurements concerning SD2/SD1 and stdHR parameters. Both these parameters indicate signal variability. More significant difference was identified for SD2/SD1 parameter, which accounts for both short- so as long-term variability. The parameter stdHR reflecting only long-term variability achieved  $p$ -value close to 0.05. Significant difference between these

measurements was assumed, especially with respect to considerable fatigue of the subjects. During 6. measurement, the subjects exhibited the highest level of fatigue and sleepiness what was clear from their subjective self-evaluation. With respect to the results it is possible to pronounce that SD2/SD1 Pointcaré plot could be suitable for this type of studies.

Despite the fact that from the data it is impossible to determine any concrete parameter which would sufficiently change with fatigue, we can observe certain trend. The biggest changes in measured parameters appeared between 1. and 6. measurement. As already mentioned, the subjects from 6. measurement reported greater fatigue likely because this measurement fitted the period corresponding to attenuation of circadian rhythm, i.e. about 30 hours without a sleep.

Overall, not many significant differences were identified, i.e. it was not possible to detect fatigue. This may be due to the fact that the experiment duration could be insufficient, i.e. there was no significant fatigue state induced. This may also relate to the fact that one-time fatigue was attempted for young subjects. We can assume that sufficient fatigue could not be induced with respect to the group of measured subjects.

## V. CONCLUSION

This paper introduces experiment which aimed at identification of pilot fatigue with the utilization of heart activity records, more precisely one of its parameters, with the focus on Pointcaré plot, which accounts for both short- and long-term signal fluctuations. The paper introduces the experiment, which consisted of 8 simulated IFR flights during 24 hours. 8 students of the Department of Air Transport, Faculty of Transportation Sciences, Czech Technical University in Prague participated in the experiment. During individual flights, ECG curve of subjects was measured and used to identify R waves so as to extract RR intervals. Subsequently, standard methods were used (time and frequency analysis) and also Pointcaré plot parameters.

The results indicate that statistically significant differences were identified only for some parameters (stdHR, SD2/SD1) and only between 1. and 6. measurements. Measurement 6 corresponds to the time of circadian low and the subjects during the measurement exhibited considerable fatigue, so the results can be considered as anticipated. It is possible to pronounce that Pointcaré plot applied on data about heart activity could be used for identification of pilot fatigue.

The limitation of the study was low number of participants especially due to time- and personal-demanding nature of the measurement. The experiment duration can also be considered a limitation, owing to which there could be insufficient degree of fatigue induced. This could be also due to the age of participants and the fact that the fatigue was only one-time experience.

Such complex study was never done before. Therefore it could bring a new insight into fatigue of pilots. The study will be used as a base for a next research since the presented approach seems to be correct, i.e. psychophysiological indicators could be used to determine a fatigue of pilots. With

respect to the limitations, the next research should be extended in terms of the subjects sample size, but so that the group of subjects remains uniform as much as possible. Further, it is advisable to extend the experiment duration and repeat the measurements during several consecutive days to assure inducing persistent fatigue, not just one-time fatigue as during the presented experiment.

Despite the mentioned limitations, the paper shows parameters which appear suitable for the purpose of quantification of pilot fatigue. The results could be used as pilot study serving as a base for next research, which could further contribute to the improvement of flight safety.

## ACKNOWLEDGMENT

This work was supported by the Grant No. SGS17/150/OHK2/2T/16 "Evaluation of psychophysiological state of pilots based on physiological parameters" of the Czech Technical University in Prague. This work was also supported by the Ministry of Industry and Trade of the Czech Republic within execution of the project No. FV10329 "Transportable personalized medical device for vacuum-compression therapy" and also by the Slovak Research and Development Agency under the grant No. APVV-17-0167 "Application of the Self-regulatory techniques for the Flight Crew Preparation".

## REFERENCES

- [1] H. P. V. Dongen and S. R. Hursh, "Fatigue, performance, errors, and accidents," in *Principles and Practice of Sleep Medicine*. Elsevier, 2011, pp. 753–759. [Online]. Available: <https://doi.org/10.1016/b978-1-4160-6645-3.00067-0>
- [2] C. Reis, C. Mestre, and H. Canhão, "Prevalence of fatigue in a group of airline pilots," *Aviation, Space, and Environmental Medicine*, vol. 84, no. 8, pp. 828–833, aug 2013. [Online]. Available: <https://doi.org/10.3357/ASEM.3548.2013>
- [3] L. C. Thomas, C. Gast, R. Grube, and K. Craig, "Fatigue detection in commercial flight operations: Results using physiological measures," *Procedia Manufacturing*, vol. 3, pp. 2357–2364, 2015. [Online]. Available: <https://doi.org/10.1016/j.promfg.2015.07.383>
- [4] I. C. A. Organization, "Doc 8984, manual of civil aviation medicine, third edition," 2012.
- [5] N. Wright and A. McGown, "Vigilance on the civil flight deck: incidence of sleepiness and sleep during long-haul flights and associated changes in physiological parameters," *Ergonomics*, vol. 44, no. 1, pp. 82–106, jan 2001. [Online]. Available: <https://doi.org/10.1080/00140130150203893>
- [6] NTSB, "Reduce fatigue-related accidents," 2017.
- [7] D. HOLLEY, B. SUNDARAM, and D. WOOD, "Shift work and aviation safety," *Clinics in Occupational and Environmental Medicine*, vol. 3, no. 2, pp. 231–262, may 2003. [Online]. Available: [https://doi.org/10.1016/s1526-0046\(03\)00073-6](https://doi.org/10.1016/s1526-0046(03)00073-6)
- [8] E. C. Association, "Pilot fatigue rrometer," 2012.
- [9] K. B. Gregory, W. Winn, K. Johnson, and M. R. Rosekind, "Pilot fatigue survey: Exploring fatigue factors in air medical operations," *Air Medical Journal*, vol. 29, no. 6, pp. 309–319, nov 2010. [Online]. Available: <https://doi.org/10.1016/j.amj.2010.07.002>
- [10] S. Bourgeois-Bougrine, P. Carbon, C. Gounelle, R. Mollard, and A. Coblenz, "Perceived fatigue for short- and long-haul flights: a survey of 739 airline pilots," *Aviation, space, and environmental medicine*, vol. 74, no. 10, pp. 1072–1077, 2003.
- [11] D. Powell, M. B. Spencer, D. Holland, E. Broadbent, and K. J. Petrie, "Pilot fatigue in short-haul operations: effects of number of sectors, duty length, and time of day," *Aviation, Space, and Environmental Medicine*, vol. 78, no. 7, pp. 698–701, 2007.
- [12] C. Zhao, M. Zhao, J. Liu, and C. Zheng, "Electroencephalogram and electrocardiograph assessment of mental fatigue in a driving simulator," *Accident Analysis & Prevention*, vol. 45, pp. 83–90, 2012.

- [13] S.-J. Jung, H.-S. Shin, and W.-Y. Chung, "Driver fatigue and drowsiness monitoring system with embedded electrocardiogram sensor on steering wheel," *IET Intelligent Transport Systems*, vol. 8, no. 1, pp. 43–50, 2014.
- [14] M. Brennan, M. Palaniswami, and P. Kamen, "Do existing measures of poincare plot geometry reflect nonlinear features of heart rate variability?" *IEEE transactions on biomedical engineering*, vol. 48, no. 11, pp. 1342–1347, 2001.
- [15] L. Mourot, M. Bouhaddi, S. Perrey, S. Cappelle, M.-T. Henriot, J.-P. Wolf, J.-D. Rouillon, and J. Regnard, "Decrease in heart rate variability with overtraining: assessment by the poincare plot analysis," *Clinical physiology and functional imaging*, vol. 24, no. 1, pp. 10–18, 2004.
- [16] J. Pan and W. J. Tompkins, "A real-time QRS detection algorithm," *IEEE Transactions on Biomedical Engineering*, vol. BME-32, no. 3, pp. 230–236, mar 1985. [Online]. Available: <https://doi.org/10.1109/tbme.1985.325532>
- [17] M. Malik, J. T. Bigger, A. J. Camm, R. E. Kleiger, A. Malliani, A. J. Moss, and P. J. Schwartz, "Heart rate variability: Standards of measurement, physiological interpretation, and clinical use," *European Heart Journal*, vol. 17, no. 3, pp. 354–381, mar 1996. [Online]. Available: <https://doi.org/10.1093/oxfordjournals.eurheartj.a014868>
- [18] V. Socha, J. Schlenker, P. Kal'avksy, P. Kutilek, L. Socha, S. Szabo, and P. Smrcka, "Effect of the change of flight, navigation and motor data visualization on psychophysiological state of pilots," in *2015 IEEE 13th International Symposium on Applied Machine Intelligence and Informatics (SAMII)*. IEEE, jan 2015. [Online]. Available: <https://doi.org/10.1109/sami.2015.7061900>
- [19] N. R. Lomb, "Least-squares frequency analysis of unequally spaced data," *Astrophysics and Space Science*, vol. 39, no. 2, pp. 447–462, feb 1976. [Online]. Available: <https://doi.org/10.1007/bf00648343>
- [20] J. D. Scargle, "Studies in astronomical time series analysis. II - statistical aspects of spectral analysis of unevenly spaced data," *The Astrophysical Journal*, vol. 263, p. 835, dec 1982. [Online]. Available: <https://doi.org/10.1086/160554>
- [21] P. Guzik, J. Piskorski, T. Krauze, R. Schneider, K. H. Wesseling, A. Wykretowicz, and H. Wysocki, "Correlations between the poincare plot and conventional heart rate variability parameters assessed during paced breathing," *The Journal of Physiological Sciences*, vol. 57, no. 1, pp. 63–71, 2007.
- [22] F. Shaffer and J. Ginsberg, "An overview of heart rate variability metrics and norms," *Frontiers in public health*, vol. 5, p. 258, 2017.
- [23] M. P. Tulppo, T. Makikallio, T. Takala, T. Seppanen, and H. V. Huikuri, "Quantitative beat-to-beat analysis of heart rate dynamics during exercise," *American journal of physiology-heart and circulatory physiology*, vol. 271, no. 1, pp. H244–H252, 1996.
- [24] J. J. Higgins, "Introduction to modern nonparametric statistics," 2003.
- [25] O. J. Dunn, "Multiple comparisons using rank sums," *Technometrics*, vol. 6, no. 3, pp. 241–252, aug 1964. [Online]. Available: <https://doi.org/10.1080/00401706.1964.10490181>



## **B Publikace 2**

Socha, V., Hanáková, L., Stojíc, S., Kušmírek, S., Socha, L., & Antoško, M.  
(2018).

**A fatigue influence on pilot's reaction ability during 24 hours flight  
simulation: A case series study.**

In *New Trends in Civil Aviation* (pp. 15-19). CRC Press.

# A fatigue influence on pilot's reaction ability during 24 hours flight simulation: A case series study

V. Socha, L. Hanáková, S. Stojić & S. Kušmírek

*Faculty of Transportation Sciences, Czech Technical University in Prague, Czech Republic*

L. Socha & M. Antoško

*Faculty of Aeronautics, Technical University of Košice, Slovak Republic*

**ABSTRACT:** Currently, the most erroneous element during the flight operation is a human. Regarding to continuous efforts spent on increase of the air transport safety level, a greater attention is now focused on human factors and comprehension of the factors having a negative impact on pilots and other crew members. The main goal of this study is to evaluate the fatigue effect on pilot performance during at least 24-hour measurements. The whole measurement consisted of two-hour time blocks, during which the pilot was flying, then the reaction time was measured and at the end psychological tests were conducted. The reaction time was chosen as an indicator of short-term performance. The study involved 4 pilots (men aged 21–27 years) with a given pre-flight regime. The results show that even though the pilots experienced a gradual fatigue during the measurements, it did not manifest in reaction times measurements. It can be stated, that despite the fatigue, the pilots are able to maintain short-term attention and maintain constant performance.

## 1 INTRODUCTION

Together with the increasing technological development and continuous improvement in aviation and aerospace technology, it gets clear that the weakest and most erroneous element of the whole flying process is a human. A human is a system element considered to be the most erroneous one (Martinussen & Hunter 2017, Shappell & Wiegmann 2012). Therefore, in order to improve aviation safety, these errors and mistakes should be eliminated. With the year-to-year air traffic volume increase, it is important to maintain and increase the level of safety within the whole air transport process. Due to the mentioned reason it is required to examine the impact of the human factor on the system as a whole.

Generally speaking, a large amount of factors, both negative and positive, can influence a human activity. In order to ensure the highest level of aviation safety, it is necessary to recognize and understand factors, which have a negative impact on flightcrew activity. According to ICAO, aviation safety is defined as a state, where the risk of injury or damage to property is reduced and maintained at an acceptable or lower level, through a continuous hazard identification process (Luxhoj 2003, Lee 2006, Goldstein, Demak, Northridge, & Wartenberg 1992). Similarly, a hazard is a state, which creates the preconditions for an air accident or safety occurrence (Halpert 2009).

As mentioned above, it is clear that a focus of examination must be placed on understanding and elimination of a hazard. One of the factors that could have a negative impact on safety from the human factor failure perspective is a fatigue. ICAO, IATA and IFALPA have developed the Fatigue Risk Management System (FRMS), which is a part of the ICAO Standard and Recommended Practice (SARP), and it serves for monitoring and prevention of a fatigue. A fatigue is considered to be one of the most significant risks causing human factor failure, due to a fact that it affects many aspects of the flight crew members' ability to work and has a significant impact on safety.

According to the facts mentioned above, the article is focused on the assessment of the reaction ability during a continuous, 24-hour flight operation. For this purpose, an experiment was performed, applying methods for the reaction times measurement of the pilots in a continuous 24-hour workload period.

## 2 MATERIALS AND METHODS

### 2.1 *Participants and experiment description*

The experiment was conducted at the Faculty of Aeronautics of Technical University of Košice at the Department of Flight Preparation.

Four experimental pilots, aged twenty one to twenty seven years, participated in the experiment. The pilots had completed instrument flight rules (IFR) training. Task schedule was filled within a two hours cycle. Each person completed 10–12 cycles. Within a one cycle, the Beechcraft Baron Be-58 simulator flight took around 60 minutes. Flight was conducted according to the prepared flight plan for IFR conditions. All flights performed within each cycle over a 24-hour period were unified. After completing the flight, the pilot performed tasks during which an ability to concentrate was monitored through standardized psychological tests developed by the Institute of Aviation Medicine. Together with these, the speed and ability to perform several tasks simultaneously were tested through the reaction time test. During the experiment, test activities were scheduled in a way to prevent a pilot relaxation. During one test cycle, pilots had several minutes for basic physiological needs (food, hygiene). The pilots were not allowed to consume stimulating substances in the form of energy drinks, coffee or food during the experiment. The schedule of tasks is illustrated in Figure 1. For the purpose of this study, the main focus will be on the reaction time.

The pilots (monitored persons) were not prepared for the experiment in any way, in terms of awareness of the tasks being performed. The pilots had to strictly adhere to pre-flight regime. This regime included a pre-flight sleep time of minimum 8 hours, no drugs and psychotropic substances usage, including alcoholic beverages 24 hours before flight. The same regime had to be followed by the supervisors and measurement supervisors, responsible for the measurement equipment.

The entire study was performed under professional supervision, in accordance with the Helsinki Declaration (World Medical Association 2013) and with the consent of the participants of the experiment.

## 2.2 Reaction time measuring device and procedure

Although the entire setup of the experiment consisted of a number of particular parts, the object of

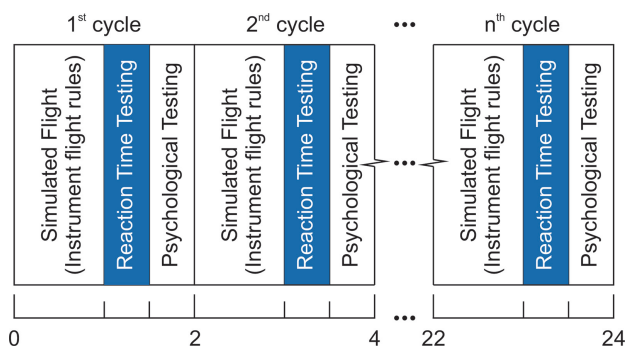


Figure 1. Schematic experimental setup.

the survey was focused on the evaluation of the reaction times. The reaction times were measured using the device proposed at the Faculty of Aeronautics of the Technical University of Kosice shown at Figure 2 (see Antosko et al. (2014)). The device consists of the hardware and software parts. The hardware section consists of the three primary buttons and two pedals used to confirm the response to the visual stimuli. It includes earphone and keyboard for evaluation of acoustic stimuli. Visual and acoustic stimuli were generated randomly by the software. The graphical user interface consists of the three active fields on which the randomly distributed geometric shapes of different colours are displayed. The pilots had to evaluate in which field the two matching points of the same colour are the closest, and to press the button corresponding to the given visualization field (left, middle, right). In the case when a red dot was displayed in the left field, the pilot had to press a right pedal with a right foot, and vice versa. When the red dot was in the right field, he had to use left foot and press left pedal. In addition, during a measurement a beep sound was generated with randomly generated time difference. The pilot had to count beeps and when the red dot is displayed on the middle field of a screen, he had to enter the exact number of beeps using the numeric keypad. After this task (whether the beep was correctly or incorrectly entered), the system reset the counter and the parallel sound generation started from the beginning.

The total number of tasks was 6 (left button, middle button, right button, left foot, right foot, and sound stimuli). Each of these were generated 50 times by software in random order. The total number of tasks that the subject had to correctly evaluate and respond to was 300. For all the mentioned stimuli, the reaction time and the number of possible errors in the reaction were measured.

## 2.3 Data processing and statistical analysis

For a simpler interpretation of the following descriptions, all data characterizing the reaction time



Figure 2. Measuring device used for collection of reaction times of tested subjects.

when the upper limbs were used (i.e., the first three types of tasks) will be named as Task 1. This characterized data set contains the vector of 150 reaction times on visual stimuli. Data characterizing lower limbs engagement will be characterized as Task 2. This dataset contains a vector of 100 reaction times in case when the pilot had to engage the right foot for reaction on visual stimuli and vice versa. The response time for the sound stimulus will be labeled as Task 3 and it contains a vector of 50 values.

For each task performed in the cycle, analysis was performed to confirm normal distribution hypothesis. Testing was performed using a single sample Kolmogorov-Smirnov goodness-of-fit hypothesis test. The assumption of normal distribution was not confirmed for all datasets. Therefore, non-parametric data evaluation was used for further evaluation.

Distributions of a single data characterizing individual tasks were plotted in the form of box-plots and distributed on the timeline, according to the time of measurement. Each box-plot represents the minimum, the first quartile, the median, the third quartile and the maximum of the given distribution. The outlier values (values that are significantly distant from the distribution boundaries) are then characterized as:

$$Q_1 - 1.5 \cdot IQR > X_{out} > Q_3 + 1.5 \cdot IQR \quad (1)$$

where  $Q_1$  is a first quartile,  $Q_3$  is third quartile,  $IQR$  is interquartile range defined as  $Q_3 - Q_1$  and  $X_{out}$  is outlier (extreme) value.

In addition to the actual distribution of the individual reaction times, it is possible to display the course of the changes in the reaction time for the pilots separately.

For the evaluation of the statistical significance in the change of the reaction times between the two time periods, a non-parametric sign-rank test was used. This test performs a paired, two-sided test of the hypothesis that the difference between the matched samples in the vectors  $X$  and  $Y$  comes from a distribution whose median is zero. The interpretation for  $p < 0.05$  means that the test rejects a zero hypothesis that there is no difference between the median values between the two data sets. If  $p < 0.05$  it can be argued that there is enough statistical evidence to conclude that the value of the median of one data set is statistically significantly greater/smaller than the median of the second dataset.

According to the fact that there is no scope for evaluating and interpreting of all intergroup tests as part of the presentation of the results, testing will be performed only between time adjacent Tasks. This means that for purposes of this article only statistical significances between  $Task_i$  vs.  $Task_{i+1}$  will be checked, where  $i = 1 \dots n$  and  $n$  is the number of measurements or cycles (see Fig. 1).

### 3 RESULTS

The individual reaction times gathered during 12 hours of measurement are present in a box-plot on the Figures 3–5. The second pilot did not complete a whole set of 12 measurements – after the 10 measurements the test was terminated due to fatigue.

The graphical representation of the reaction times distribution, measured during the first task is presented on the Figure 4. It is clear that in case of the first pilot, during measurement process only small changes of the reaction times occurred which

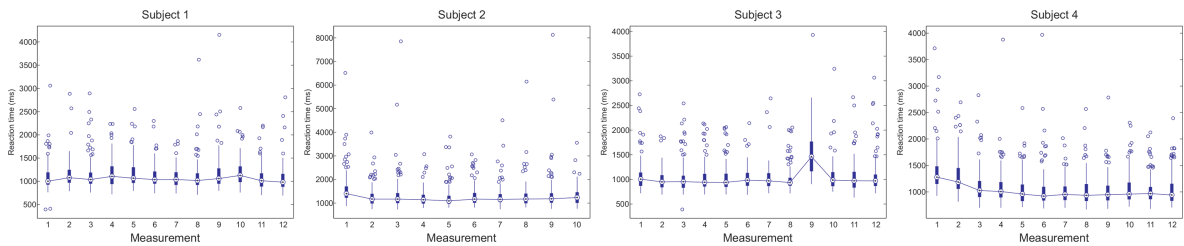


Figure 3. The Task 1 reaction times distribution for all subjects separately.

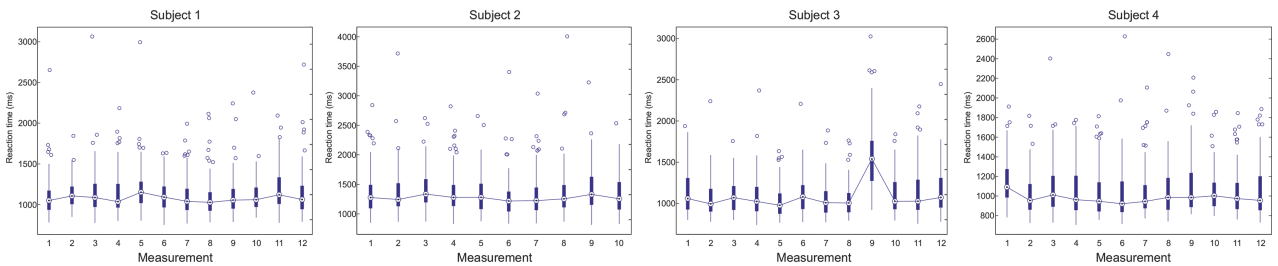


Figure 4. The Task 2 reaction times distribution for all subjects separately.

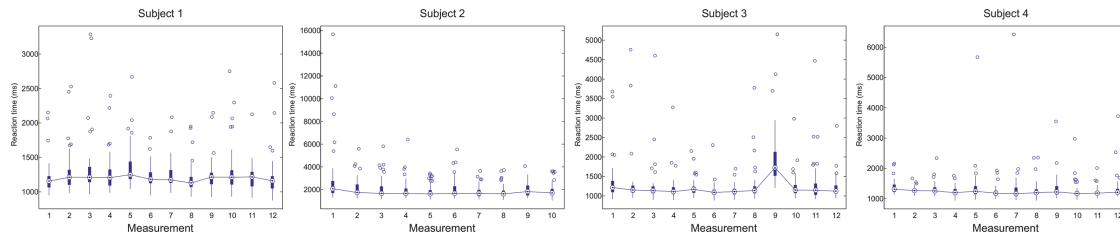


Figure 5. The Task 3 reaction times distribution for all subjects separately.

is considered to be statistically non-significant. The data showed one statistically significant change ( $p = 7.15 \cdot 10^{-5}$ ), where drop of the median of the reaction times between the 10th and 11th measurement was from 1,113.91 ms to 1,008.70 ms.

Also, in the case of the second pilot, the distribution of the measured reaction times was balanced throughout all measurements. In the measured data, only one statistically significant change was detected, between the 1st and 2nd measurement when a statistically significant decrease happened ( $p = 7.65 \cdot 10^{-9}$ ) – from a 1,382.79 ms to 1,149.50 ms of median value.

In the case of the third pilot, the distributions of the reaction times were in balance. From the Figure 3 a significant increase was detected within 9th measurement. Between the 8th and 9th measurement a statistically significant increase ( $p = 1.57 \cdot 10^{-20}$ ) of the reaction time median happened— from 925.96 ms to 1457.96 ms and then between 9th and 10th measurement, a statistically significant decrease occurred ( $p = 1.10 \cdot 10^{-18}$ ) – a drop to the value of 978.68 ms. Statistically significant decrease ( $p = 0.04$ ) was detected between the 1st and 2nd measurement— from 976.21 ms to 946.19 ms.

Also, in the case of the fourth pilot, the distribution was balanced during the measurement. Data showed only one statistically significant decrease ( $p = 3.40 \cdot 10^{-9}$ ) between the 2nd and 3rd measurement— from 1,170.93 ms to 1,012.08 ms of the reaction times median value.

The graphical representation of the reaction times measured during the second task was presented on the Figure 4. In the case of the first pilot, there was several statistically significant changes between particular measurements. Firstly, an increase ( $p = 0.03$ ) between the 1st and 2nd measurement— from 1,049.00 ms to 1,103.80 ms, after that an increase ( $p = 1.88 \cdot 10^{-3}$ ) between the 4th and 5th measurement— from 1,039.51 ms to 1,149.27 ms and followed by decrease ( $p = 0.01$ ) to 1,090.43 ms. Subsequently, a statistically significant increase ( $p = 0.04$ ) was detected between 10th and 11th measurement— from 1,057.27 ms to 1,119.81 ms and subsequent decrease occurred ( $p = 0.03$ ) to 1,049.77 ms.

In the case of the second Task for the second pilot, a distribution of the measured reaction times during whole measurement process was balanced. Within a measured data, there was no statistically significant changes of the reaction times.

In case of the third pilot several statistically significant changes were found during Task 2. First, there was a decrease ( $p = 0.04$ ) of the median of the reaction times between the 1st and 2nd measurement— from 1,056.22 ms to 992.33 ms, then the increase ( $p = 0.02$ ) between the 5th and 6th measurement— from the value 967.75 ms to 1,077.93 ms. Similarly as in the Task 1, an increase ( $p = 3.57 \cdot 10^{-16}$ ) between the 8th and 9th measurement was observed— from 992.28 ms to 1,519.73 ms and subsequent a decrease occurred ( $p = 1.68 \cdot 10^{-13}$ ) to 1,020.25 ms.

In the case of fourth pilot, Task 2 brought only one statistically significant change, a decrease ( $p = 1.81 \cdot 10^{-5}$ ) of the response times median between 1st and 2nd measurements— from 1,084.54 ms to 946.39 ms.

A graphical representation of the reaction times distribution measured in Task 3 is presented on Figure 5. It is obvious that even in Task 3, the distribution of the reaction times was balanced during the measurement. For the first pilot, several statistically significant changes were found. First, the increase of the reaction time median between the 1st and 2nd measurements— from 1,147.56 ms to 1,196.61 ms, then the decrease ( $p = 4.75 \cdot 10^{-3}$ ) between the 5th and 6th measurement— from 1,219.99 ms to 1,177.38 ms and then increasing ( $p = 0.05$ ) between the 8th and 9th measurements— from 1,118.31 ms to 1,211.88 ms.

In the case of the Task 3 for second pilot, the distribution of measured reaction times throughout the measurement was balanced. No statistically significant change in the reaction times was found in the measured data.

In the case of the third pilot, several significant changes were found in the Task 3. The first is the decrease ( $p = 0.02$ ) of the reaction times median between the 5th and 6th measurement— from 1,153.28 ms to 1,086.74 ms, then the increase ( $p = 0.03$ ) between the 7th and 8th measurements— from 1,105.87 ms to 1,109.82 ms, then increase ( $p = 9.14 \cdot 10^{-8}$ ) to 1,694.97 ms and finally decrease ( $p = 1.59 \cdot 10^{-8}$ ) to 1,131.03 ms.

In case of the fourth pilot, during Task 3 the following statistically significant changes were found. Firstly, a decrease ( $p = 0.02$ ) between the 1st and 2nd measurement— from 1,292.13 ms to 1,195.96 ms, then the decrease ( $p = 0.05$ ) between 3rd and 4th measurement— from 1,246.41 ms to 1,158.19



ms and finally an increase between 11th and 12th measurement—from 1,179.80 ms to 1,203.77 ms.

#### 4 DISCUSSION

Due to a fact that increasing fatigue brings a deterioration of concentration, it was assumed that with the longer and increasing time of the measurement, the reaction times of the pilots will be increased. The results presented in Figures 3–5 show that the measured reaction times, their medians respectively, did not create the expected trend, but their course was balanced in the most cases.

In the case of the first pilot, several significant changes were found. However, these changes had not been reflected across the different tasks and therefore it could not be considered as a single trend. Similar conclusions can also be presented for the fourth pilot. It can therefore be assumed that the reaction time was not negatively influenced by the pilot's fatigue.

In the case of the second pilot, a significant change was observed only during the Task 1. In other cases, the median progressions were balanced and no significant differences were found. This pilot terminated the measurement before expected time due to fatigue. It can be stated therefore, that the attention of the pilot in the short-term tasks was not affected by its fatigue and the anticipated increase in a reaction time did not occur.

In the case of the third pilot, there was a significant increase in the 9th measurement. This increase was reflected in all 3 tasks as a strong peak in data. It manifested through a one-off decline of the pilot's attention and, with regard to the other data, it can also be stated that even in the case of the third pilot the short-term attention was not affected by fatigue.

According to the results presented in Figures 3–5, it could be stated that the short-term attention that was tested during the reaction time measurement is not significantly influenced by the fatigue of tested pilots.

#### 5 CONCLUSION

The goal of the experiment was to observe and assess the impact of the fatigue on pilot's performance during twenty-four hours. Performance changes were evaluated on the graphs that were created based on the results of individual stress tests. After twenty-four hours a certain fatigue influence was expected. It was assumed that increasing fatigue would extend test completion time. For some pilots, a slight deterioration in test results was detected. For some pilots, the results were improved at the end of the experiment. As mentioned above, the experiment consisted of several parts, with sixty minutes flight in average in

every cycle, and the stress tests together around twenty minutes. Since the stress tests were of short duration, the experiment demonstrated that even in the absence of rest, a person is able to keep the attention at the desired level in such short-term activities.

A one-off experiment lasting twenty four hours proved that within this time-frame, one is able to maintain approximately the same level of performance, especially for short-term tasks. It can be assumed that this finding is also related to the fact that the tested persons were young men in good physical condition. In personal interaction with subjects during the experiment, it was found that even if subjective feelings of fatigue appeared, it had little or no effect on the test results. A fatigue was observed during a personal interaction in the form of yawning, more frequent gurgling, scattered attention, and at the end of the experiment a mild mood worsening. We can say that the fatigue was observed at each of the participating pilots, but it only slightly affected their short-term performance.

#### ACKNOWLEDGEMENT

This work has been supported by the project SGS17/150/OHK2/2T/16 "Evaluation of psychophysiological state of pilots based on physiological parameters".

#### REFERENCES

- Antosko, M., J. Pila, P. Korba, & P. Lipovsky (2014). Psychological readiness of air traffic controllers for their job. *Nase More* 61(1–2), 5–8.
- Goldstein, B.D., M. Demak, M. Northridge, & D. Wartenberg (1992). Risk to groundlings of death due to airplane accidents: a risk communication tool. *Risk analysis* 12(3), 339–341.
- Halpert, M. (2009). From a plane crash to the conviction of an innocent person: Why forensic science evidence should be inadmissible unless it has been developed as a safety-critical system. *Hamline L. Rev.* 32, 65.
- Lee, W.-K. (2006). Risk assessment modeling in aviation safety management. *Journal of Air Transport Management* 12(5), 267–273.
- Luxhoj, J.T. (2003). Probabilistic causal analysis for system safety risk assessments in commercial air transport.
- Martinussen, M. & D.R. Hunter (2017). *Aviation psychology and human factors*. CRC Press.
- Shappell, S.A. & D.A. Wiegmann (2012). *A human error approach to aviation accident analysis: The human factors analysis and classification system*. Ashgate Publishing, Ltd.
- World Medical Association (2013, nov). World medical association declaration of helsinki. *Journal of the American Medical Association* 310(20), 2191.

## C Publikace 3

Socha, V., Hanakova, L., Freigang, M., Kraus, J., Stojic, S., Socha, L., &  
Hanak, P. (2019).

**Impact of pilots' tiredness on the outcome of psychological testing.**

Transport Problems, 14. (pp. 55-67). Wydawnictwo Politechniki Slaskiej.

**Keywords:** human factors; fatigue; psychological testing; performance; pilot; flight simulator

**Vladimir SOCHA, Lenka HANAKOVA, Michal FREIGANG,  
Jakub KRAUS, Slobodan STOJIC**

Czech Technical University in Prague, Faculty of Transportation Sciences  
Horska 3, 128 03 Prague, Czech Republic

**Lubos SOCHA\*, Peter HANAK**

Technical University of Košice, Faculty of Aeronautics  
Rampová 7, 041 21 Košice, Slovakia

\*Corresponding author. E-mail: [lubos.socha@tuke.sk](mailto:lubos.socha@tuke.sk)

## **IMPACT OF PILOTS' TIREDNESS ON THE OUTCOME OF PSYCHOLOGICAL TESTING**

**Summary.** Presented work is primarily oriented on the experimental verification of the influence of fatigue on the psychological condition of the flying personnel, using psychological and performance tests. For the evaluation of a pilot performance, the 24 hours experiment was conducted. In total, eight subjects participated in the experimental measurements. Eight participants went through several tests, including simulator flights, to investigate the effects of the fatigue on the results of psychological measurements. Measurements included workload evaluation, using NASA task load evaluation concept and performance testing, using the so-called OR-test. Significant statistical differences between measurements performed during 24 hours were not found in the case of NASA task load Scores. In the case of OR-test, Friedman ANOVA and subsequent post-hoc analysis show that the greatest decrease in performance was observed in approximately 22 hours of wakefulness, i.e. approximately in half of the measuring process. The concept of 24-hour measurements for the quantification of fatigue is not commonly used yet as well as objectivization using performance testing. As the apparent effect of fatigue is mainly on performance testing results, it can be argued that this work could serve as a basis for further studies on fatigue. Also, it could serve as a support for introducing new pilots' psychological testing procedures in the future, which could contribute to current efforts to improve aviation safety.

### **1. INTRODUCTION**

The role of fatigue in the human organism has extraordinary significance. It can be considered as an important natural indicator of the overloading of the body that is responsible for its protection [1]. As the beginning of fatigue is slow and unnoticeable, the risk is that a pilot will not realize a decrease in his performance and abilities [2]. Fatigue is a highly subjective matter, and it manifests in different ways. An actual level of fatigue could be best evaluated by the person who feels it. It is, therefore, difficult to measure or detect it. However, its characteristics are the same for one particular person, who knows what the reasonable forms of rest and regeneration time are [1,3]. In the case of objectivization, the fatigue could be detected through behavior and facial expression or, for example, using voice analysis [4]. Fuzzy face expression along with irritability, malaise, passivity and slow reactions can be considered as typical signs of fatigue, indicating a decrease in overall human performance [3].



Fatigue, as one of the possible changes in the current physical state of the body, has a direct impact on the mental and physical performance of a pilot. One of the basic requirements for pilots is physical and mental ability to perform their duties in the cockpit [1]. Typical fatigue symptoms in the cockpit include inattention, laxity, impaired situational awareness, and mistakeness [2,4-8], and a pilot is easily disturbed, he responds irritably, his communication with other crew members is rather restrained, and his mood varies. The problems with piloting precision, fine locomotor skills, short-term memory, and seeing could be observed [2,8]. Furthermore, fatigue reduces pilot resistance to hypoxia, overload, or kinetosis [1]. In extreme cases, a workload could be so high that a pilot does not have enough capacity to deal with a complex situation. Exhaustion of the functional reserve of an organism then brings a risk of failure [9].

Pilots and crew members are constantly confronted with long working days, early departures, late arrivals, and non-standard working hours, which include night shifts and specific timetables [10,11]. Commercial pilots on long journeys, as well as military pilots, frequently pass through many time zones, which contribute to circadian disorders and sleeping problems [9].

Safety is one of the most important factors in aviation. A negative impact on the safe conduction of the flight may be related to the family relationships problems, working shifts, financial problems or problems derived from dissatisfaction from career growth. Stress and anxiety have a demonstrable impact on air incidents [12-15]. Sleeping disorders could also have an impact on pilot performance, and these are proved, in context of increased fatigue, to be the cause of some air accidents [4].

A monitoring of fatigue in aviation, especially, among commercial pilots, is based mainly on subjective quantifiers, i.e. a fatigue is monitored mainly through questionnaire survey [16-18]. The results of such surveys point to the fact that pilots report a decrease in performance mainly due to a fatigue (for example, according to the results [5] an impact of fatigue on the performance is reported by more than 84% of the pilots questioned). Standard methods for tracking the performance of the pilots include the NASA Task Load Index (NTLx). NTLx is a standardized questionnaire, i.e. a subjective pilot evaluation after performed tasks [19].

At the same time, not only in aviation, there is a general effort leading to an objectification of the data. In recent years, it has been realized that questionnaire surveys are not the most appropriate way for the quantification of fatigue and other techniques appeared. One of these techniques is performance testing. These tests are part of the standard testing performed at the Institute of Aviation Medicine.

Based on what is mentioned above, the aim of this study is an evaluation of the influence of pilot's fatigue on the outcome of psychological and performance tests. Such testing could further contribute in enhancing air transport safety, in the context of the introduction of the Fatigue Risk Management System in aviation.

## **2. MATERIALS AND METHODS**

### **2.1. Participants and measuring procedure**

The experiment was prepared, and the person who was the subject of the measurement was awake at least 36 hours after the measurement. This was achieved in a way that subjects got up on the measurement day at 6:00 or earlier. During the following 12 hours (until the beginning of measurements), the subjects were not allowed to carry out strenuous physical activities, use substances that cause fatigue, take alcohol, etc. On this basis and according to the predefined rule, a goal was to standardize the daily regime of the attended participants. Altogether, eight subjects took part in experiment, from the age range of 21–24 years and all being a professional pilots, with experience in Instrument Flight Rules (IFR) flying in Instrumental Meteorological Conditions (IMR) of around 100–150 flight hours. The theoretical and practical base of the pilots was comparable.

The measurements, in the sense of a 24-hour experiment, began from 18:00 Central European Time in the summer months between 1. 6. 2018 and 31. 8. 2018. The subjects were instructed in detail how to use the navigation devices effectively, and they were also reminded about the controls and settings

of the onboard devices that were available. It was necessary to use an iPad device in which the JeppFD (Jeppesen, Englewood, CO, USA) application was installed. It is important to note that these air maps are currently used by many airlines, and it is not easy for a layman to understand such maps. All subjects knew these maps, as they were commonly used for flights during practical training, whether on a simulator or on an airplane. After a detailed explanation, the subjects got answers for all their questions related to the test.

Flights took place on the Beechcraft simulator, a twin-engine propeller aircraft. The cockpit of the simulator was equipped with standard flight, navigation and communication devices along with an engine and structure control equipment. Instruments using mechanical indicators, integrated indicators and electronic display EFIS were available. During the flights, the emphasis was placed on the use of instruments using mechanical indicators – mechanical speedometer, artificial horizons, barometric mechanical altimeter, CDI pointer, turn indicator in combination with the relative tilting indicator (turn coordinator), directional flywheel, mechanical variometer, VOR / LOC / GPS indicator, ADF indicator (MDI) and magnetic compass.

There were 8 measurements on the simulator for all participants (see Fig. 1), each of which lasted approximately one and a half hours. The briefing before the flight took about 20 minutes. All the measurements took place in Germany, from several airports through different radionavigation facilities to several other airports. Subjects never took off and landed to one airport, and it was always a navigational flight. All flights were under IFR for IMC. Even though the flights were not exactly the same, all were of the same nature, with the main concern to ensure the uniformity of the course of measurement. Therefore, each flight was conducted over three radionavigation points under defined conditions, which ensured the same length and character of the individual flights.

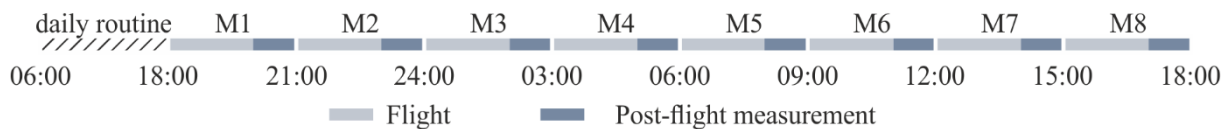


Fig. 1. Overall schedule and concept of experimental measurements

The subjects had a simulation of a real IFR flight, including engine start, requests and clearances, listening of information from the automatic terminal information service, etc. All communication with the air traffic controller was conducted in English. The surrounding traffic, including communication with them was not simulated.

After the flight, a post-flight measurement took place where the subjects were questioned through questionnaire survey, where developed software was used for the calculation of the NTLx. The subjects also underwent the performance test, the OR-test. This testing is described in the chapters below. In the presented study, the emphasis was placed on these two tests, which were carried out on the basis of the individual flights that were realized. The overall schedule and concept of experimental measurements is shown in Fig. 1.

## 2.2. Psychological testing

The two types of testing were used in the experiment. The first of the chosen methods was the performance test, called OR-test, based on a standard test used at the Institute of Aviation Medicine. The test was performed in approximately three-hour intervals and was repeated eight times for each subject.

The test was performed on a PC, and it was actually a program designed in C# language. The user started the program, generating a table of 25 rows and 15 columns (a total of 375 cells).

Each cell of the table contained one code consisting of 4 characters, see Fig. 2. The first two characters were letters and the other two were numbers. The combination of letters was always retained throughout the column, the column started with a random number, and increased with each row with a value of 1. If the value of 99 was exceeded within the column, the value 00 was then

followed by the value of 01. The table in terms of columns consisted of three parts. In the central part, i.e. the middle column, the letter part of the code always composed of the letters MM. The numeric code in this column then started with 01 and went further to the value of 25. The first part of the table contained codes starting with letters A–M. In individual columns, the codes were in an alphabetical order, i.e., the first column will always be the first combination in alphabetical order. There cannot be a situation where the first column begins with the letter C and the second with the letter A. Similarly, there cannot be a situation in which one column is the combination of AC and the following the combination of AA. It is clear, therefore, that in the 7th column there may be the combination of ML. The third part of the table (columns 9–15) worked on a similar principle, but with the second part of the alphabet, i.e. the letters N–Z. For this part of the table, the 9th column could be the combination of MN, and the last one may be the highest alphabetical combination, i.e., ZZ. For the purpose of the program, letters of the alphabet containing the diacritics, the letter CH and the letter W, were omitted. The basic principle of the test is shown in Fig. 2.

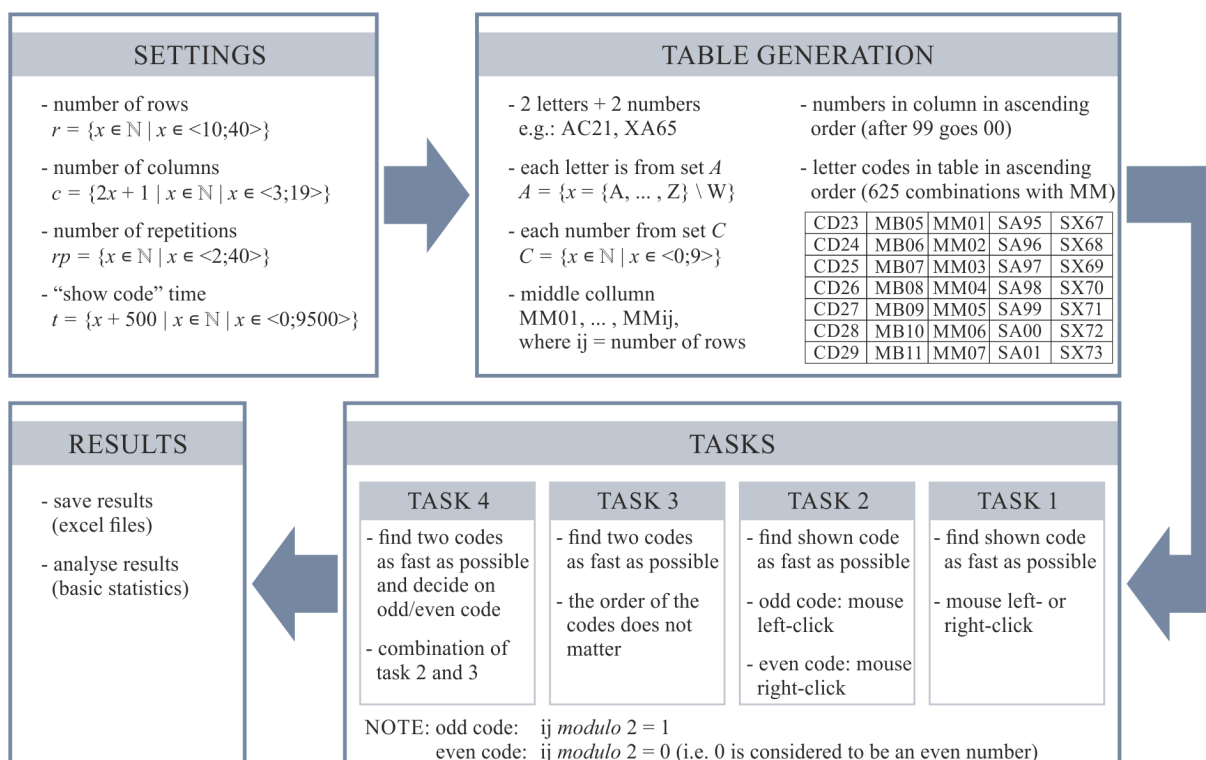


Fig. 2. Block diagram of OR-test

The experiment was divided into four parts, of which each task was repeated fifteen times. After starting the program, task instructions in English were displayed in front of each section, detailing exactly what the task is about.

The first part of the experiment included a field with two-letter and two-digit code displayed on the monitor for three seconds, see Fig. 2. Subsequently, it was the task of the subjects to locate and check this field as quickly as possible. The correctness and time from the disappearance of the text input field after clicking on any field by the user were evaluated.

The second part was the same as part one; however, the user also had to evaluate whether the number behind the letters was even or odd. If it was even, it was clicked with the right mouse button, and if it was odd, it was clicked with the left mouse button. If the 00 combination was displayed, it was considered to be an even number.

The third part had an increased demand on users compared to the previous parts. On the screen, two pairs of letters and two pairs of digits appeared for 3 seconds. The objective of the subject was to

combine the first pair of letters with the first pair of digits and the second pair of letters with the second pair of digits (e.g. SX MB 67 09). According to Fig. 2, the task was to find the SX67 and MB09 fields and check them in the table as soon as possible.

The last part of the test was based on the previous part but was further extended with classifying the numbers on even and odd. In the case shown in Fig. 2, correct number was clicked with the left mouse button on both fields, as both numbers are odd. Again, the number 00 was required to be numbered as even.

At the end of the fourth part, the results were shown to the subjects (the average response time and correctness). The data were saved in \*.xlsx format for subsequent analysis. The exported data carried information about the required codes and the real codes, response time information and a list of correct/incorrect answers.

### 2.3. NASA Task Load Index

For purposes of subjective performance evaluation, the so-called NASA Task Load Index (NTLx) was used. Therefore, a software tool for collecting and evaluating data was created using the standard test, see also [19,20]. Subjects passed through this questionnaire after each simulated flight, where their mental demand, physical demand, temporal demand, performance, effort and frustration were evaluated on a scale from 1 to 20.

In the case of mental demand, the subjects were assessed on the basis of the individual load from the point of view of psychological actions, e.g. how much he had to think, count, navigate in space, how much he was influenced in flight by clouds, the IMC conditions, and also searching for instruments, or unclear arrangement of devices (orientation in the instrument order), communication with the controller (another accent), whether it was simple or exhausting for him.

In terms of physical demand, the subjects were determined on the basis of how much the task was physically demanding, e.g., airplane control, whether the flight was relaxing or there was a constant need for flight controls movement.

In the case of temporal demand, subjects were evaluated on the basis of the exposure to time pressure, e.g. whether they were instructed progressively and logically by the ATCO, whether there was need to hurry during the take-off, did they have time to check all on-board instruments, to perform a motor test, check the amount of fuel or pressure, or whether there was not enough time to finish all tasks in the given time frame.

Performance scores were self-ratings from an individual's perspective, i.e., how successful an individual thinks he was and how much he was generally satisfied with his performance during the flight, course of communication, navigation, take-off, and landing.

In the case of the evaluation of the effort, the subjects were evaluated on the scale from 1 to 20, i.e., how demanding it was to satisfy the required level of performance e.g. whether holding the given altitude for them was simple or complicated, or whether they were able to keep the altitude in the tolerance after being warned by the controller.

Regarding frustration, subjects were assessed on how insecure, demotivated, stressed, angry or outraged by the flight they were, no matter whether due to a complexity of the flight tasks performed or because of their own mistakes, fatigue or long stereotypical flight.

After completing this questionnaire (after evaluation), the second part followed. It represented a comparison of these six categories to each other. Each category was compared with all other categories; therefore, it was possible to prove what affected the subjects the most. The subjects gradually marked the item that affected them more from their impression. Based on this approach, it is possible to specify weight for individual items.

The result of the test, then, indicates the magnitude of the significance of each monitored category to the total Task Load Index (TLx). The evaluation procedure consists of the conversion of the first part of the questionnaire into a percentage form. The scale from one to twenty represents 5%–100% with a 5% step (see Fig. 3-A). Through the second questionnaire, it is possible to assign the weight to individual factors based on the particular summation of their designations or preferring it to another category. In total, from six to two combinations, the number of occurrences is divided by 15, which

gives weight for each category (see Fig. 3-B). If this weight is multiplied by percentages from the first questionnaire, the score is obtained. The level of significance of each of the monitored categories implies from the obtained score (see Fig. 3-C), therefore, we can talk about overall scores. If all these categories are added, a total TLx [20] can be obtained. However, this value is inherently meaningful and does not carry any information on the overall score. Therefore, overall scores will be used for further evaluation within the submitted paper.

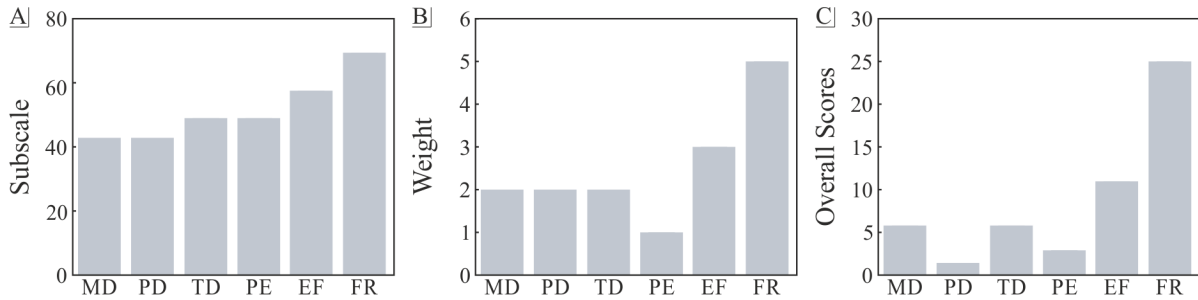


Fig. 3. Example of partial results for the Nasa Task Load Index

## 2.4. Data processing and statistical evaluation

The results of psychological and performance testing were further processed statistically. In view of the results of the Jarque–Bera test, in which the zero hypothesis was rejected, i.e., the data did not come from the normal distribution, nonparametric statistical testing was chosen. Based on measured data, i.e., data from eight subjects collected in eight measurements, the Friedman test (also Friedman's non-parametric one-way ANOVA with repeated measures) as a test for more than two dependent selections was chosen to test the differences between individual measurements [21]. For the post-hoc analysis, the Dunn–Sidak test [22] was used. Statistical testing was carried out using own-design software in Matlab environment (MATLAB R2017a, MathWorks, Inc., Natick, MA, USA).

## 3. RESULTS

### 3.1. NTLx results

In the case of the mental demand parameter, statistically significant differences between the distributions in the measured measurements were not found using the Friedmann test,  $X^2=11.72$  a  $p=0.11$ . Therefore, post-hoc analysis did not show any statistically significant differences between the individual pairs of measurements at the significance level of  $\alpha=5\%$ .

Distribution of individual measurements for mental demand is shown in Fig. 4. There was an increase in the subjective sensation of psychological stress from 1 to 3, followed by a slight decrease and a rapid increase, gradual increase to the maximum at the measurement 6. During the 7th measurement, a slight decrease was evident, followed by a renewed increase. However, the differences were not so great and the Friedman test did not show any statistically significant differences

In the case of the physical demand parameter, statistically significant differences were not found,  $X^2=3.61$ , and the resulting p-value was 0.82. Therefore, post-hoc analysis did not show any statistically significant differences between the individual pairs of measurements at the significance level of  $\alpha=5\%$ .

No statistically significant differences were found for the temporal demand parameter,  $X^2=8.61$  and  $p=0.28$ . Post-hoc analysis in this case also did not show any statistically significant differences between the individual pairs of measurements at the significance level of  $\alpha=5\%$ . Despite the absence of significant differences between observed measurements, it is possible from the data presented in Fig. 4 to see a characteristic trend. The subjective score of the temporal demand given by

the candidates from the second measurement (at 24:00) continually increased to the measurement number 4 (at 06:00), and it decreased during fifth measurement and subsequently increased again continually.

No statistically significant differences were found for the performance parameter,  $X^2=4.86$ ,  $p=0.67$ . The progress of this parameter is similar to the temporal demand score.

In case of effort, statistically significant differences were not found as well, as  $X^2=4.86$  and  $p=0.67$ . Post-hoc analysis in this case also did not show any statistically significant differences between the individual pairs of measurements at the significance level of  $\alpha=5\%$ . The distribution of the results of the test related to the effort, shown in Fig. 4, shows a slight increase in the subjective feeling of effort from the 3rd (at 03:00) to the 5th measurement (at 09:00), followed by a slight decrease of the median in 6th measurement (at 12:00).

In the case of the frustration parameter, statistically significant differences were not found,  $X^2=2.87$  and  $p=0.89$ . From the distribution of this parameter in Fig. 4, it is possible to see the biggest decrease in the level of frustration during the 7th measurement.

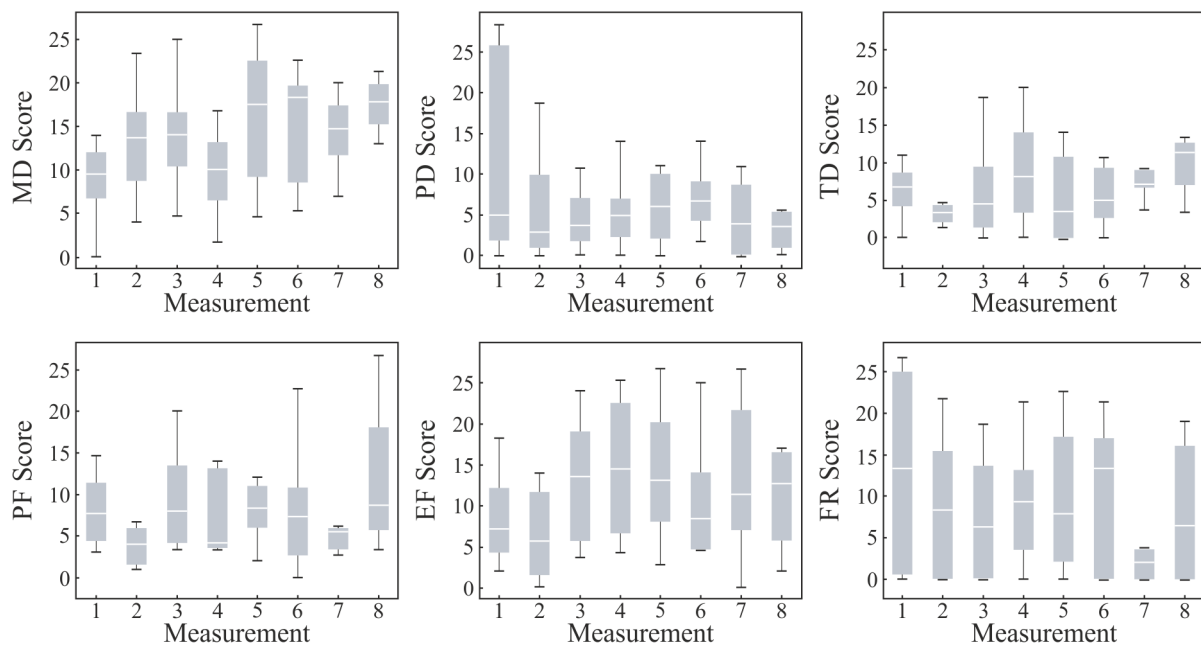


Fig. 4. Presentation of the results of overall scores for individual task load categories (MD - mental demand, PD - physical demand, TD - temporal demand, PF - performance, EF - effort, FR - frustration)

### 3.2. OR-Test results

For the time required to check the code in task 1, using the Friedmann test, statistically significant differences between distributions in the observed measurements were found,  $X^2=33.66$  and  $p=1.99 \times 10^{-5}$ . Post-hoc analysis, therefore, showed statistically significant differences between measurements 3 and 6 ( $p=0.03$ , confidence interval (6.24, 219.15)), measurements 3 and 7 ( $p=0.02$ , confidence interval (9.49, 231.67)), measurement 4 and 5 ( $p<0.01$ , confidence interval (18.36, 224.05)), measurements 4 and 6 ( $p<0.01$ , confidence interval (33.96, 246.88)), and measurements 4 and 7 ( $p<0.01$ , confidence interval (37.22, 259.40)) at the significance level of  $\alpha=5\%$ . For graphical presentation of mentioned results, see Fig. 5-A.

For the time required to check the code in task 2, using the Friedmann test, statistically significant differences between distributions in the observed measurements were found,  $X^2=30.07$  and  $p=9.20 \times 10^{-5}$ . The post-hoc analysis, therefore, showed statistically significant differences between measurements 1 and 6 ( $p=0.03$ , confidence interval (3.68, 216.95)), measurements 3 and 6 ( $p=0.02$ , confidence interval (4.12, 216.55)), measurement 5 and 6 ( $p<0.01$ , confidence interval (34.77,



247.20)), measurements 5 and 7 ( $p=0.02$ , confidence interval (7.61, 229.28)), and measurements 5 and 8 ( $p<0.01$ , confidence interval (2.02, 223.69)) at the significance level of  $\alpha=5\%$ . For graphical presentation of mentioned results, see Fig. 5-B.

In the case of the total time required for checking the code in task 3, Friedmann's test found statistically significant differences between the distributions of the measured values,  $X^2=12.95$  and  $p=7.33\times 10^{-2}$ . The post-hoc analysis showed a statistically significant difference between measurements 4 and 7 ( $p=0.04$ , confidence interval (2.37, 224.55)) at the significance level of  $\alpha=5\%$ . For graphical presentation of mentioned results, see Fig. 5-C.

In the case of the total time required to check the code in task 4, using the Friedmann test, statistically significant differences were found between the distributions of the measured values,  $X^2=67.28$  and  $p=5.22\times 10^{-12}$ . Post-hoc analysis, therefore, showed statistically significant differences between measurements 1 and 5 ( $p=0.02$ , confidence interval (9.05, 214.74)), measurements 1 and 6 ( $p<0.01$ , confidence interval 83.73, 296.64)), measurements 1 and 7 ( $p<0.01$ , confidence interval (44.04, 266.21)), measurements 1 and 8 ( $p<0.01$ , confidence interval (95.13, 317.31)), measurements 2 and 6 ( $p<0.01$ , confidence interval (43.96, 256.87)), measurements 2 and 7 ( $p=0.03$ , confidence interval (4.27, 226.44)), measurements 2 and 8 ( $p<0.01$ , confidence interval (55.35, 277.53)), measurement 3 and 8 ( $p=0.02$ , confidence interval (11.15, 233.32)), measurements 4 and 6 ( $p<0.01$ , confidence interval (53.83, 266.74)), measurements 4 and 7 ( $p=0.01$ , confidence interval (14.14, 236.31)), and measurements 4 and 8 ( $p<0.01$ , confidence interval (65.23, 287.40)) at the significance level of  $\alpha=5\%$ . Graphical presentation of results is shown in Fig. 5-D.

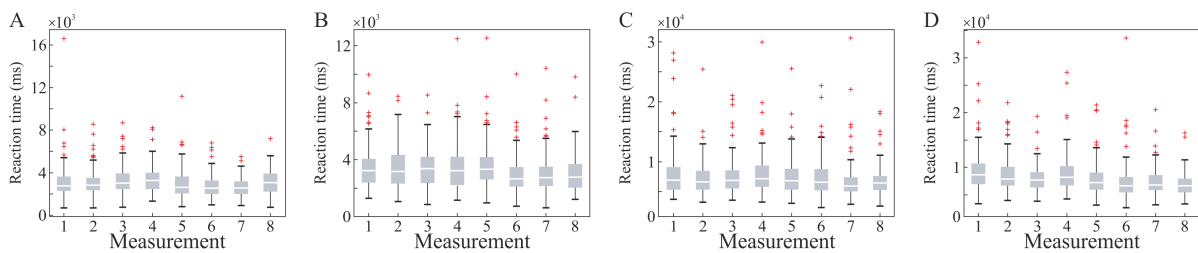


Fig. 5. Presentation of the OR-test results of total time required to check the codes in respective task (A – task 1; B – task 2; C – task 3; D – task 4)

For the time required to check the first code in task 3, using the Friedmann test, statistically significant differences were found between distributions in the measured values,  $X^2=148.48$  and  $p=8.43\times 10^{-29}$ . Post-hoc analysis, therefore, showed statistically significant differences between measurements 1 and 3 ( $p<0.01$ , confidence interval (-274.77, -50.73)), measurements 1 and 4 ( $p<0.01$ , confidence interval (-352.43, -128.39)), measurements 1 and 5 ( $p<0.01$ , confidence interval (-406.80, -182.76)), measurements 1 and 6 ( $p<0.01$ , confidence interval (-251.28, -27.24)), measurements 1 and 7 ( $p<0.01$ , confidence interval (-250.16, -26.12)), measurements 2 and 3 ( $p<0.01$ , confidence interval (-325.90, -101.86)), measurement 2 and 4 ( $p<0.01$ , confidence interval (-403.57, -179.53)), measurements 2 and 5 ( $p<0.01$ , confidence interval (-457.93, -233.89)), measurements 2 and 6 ( $p<0.01$ , confidence interval (-302.41, -78.37)), measurements 2 and 7 ( $p<0.01$ , confidence interval (-301.30, -77.26)), measurements 3 and 5 ( $p<0.01$ , confidence interval (-244.04, -20.01)), measurements 4 and 8 ( $p<0.01$ , confidence interval (76.88, 300.92)), measurements 5 and 6 ( $p<0.01$ , confidence interval (43.51, 267.54)), measurements 5 and 7 ( $p<0.01$ , confidence interval (44.61, 268.65)), and measurements 5 and 8 ( $p<0.01$ , confidence interval (131.24, 355.28)) at the significance level of  $\alpha=5\%$ . Results are graphically presented in Fig. 6-A.

For the time required to check the first code in task 4, using the Friedmann test, statistically significant differences were found between distributions in the monitored measurements,  $X^2=86.28$  and  $p=7.14\times 10^{-16}$ . Post-hoc analysis, therefore, showed statistically significant differences between measurements 1 and 4 ( $p<0.01$ , confidence interval (-262.37, -39.26)), measurements 1 and 5 ( $p<0.01$ , confidence interval (-264.03, -40.92)), measurements 2 and 3 ( $p<0.01$ , confidence interval (-317.28, -94.16)), measurements 2 and 4 ( $p<0.01$ , confidence interval (-364.58, -141.46)), measurement 2 and 5

( $p < 0.01$ , confidence interval (-366.24, -143.12)), measurements 2 and 6 ( $p < 0.01$ , confidence interval (-309.02, -85.91)), measurements 2 and 8 ( $p = 0.04$ , confidence interval (-225.36, -2.25)), measurements 4 and 7 ( $p < 0.01$ , confidence interval (40.01, 263.12)), measurements 4 and 8 ( $p < 0.01$ , confidence interval (27.66, 250.77)), measurements 5 and 7 ( $p < 0.01$ , confidence interval (41.67, 264.78)), and measurements 5 and 8 ( $p < 0.01$ , confidence interval (29.31, 252.43)) at the significance level of  $\alpha = 5\%$ . Results are graphically presented in Fig. 6-B.

In case of mistakenness in task 1, Friedmann's test did not show statistically significant differences between distributions in the monitored measurements,  $X^2 = 13.62$  and  $p = 0.06$ . Therefore, post-hoc analysis did not show any statistically significant differences between the individual pairs of measurements at the significance level of  $\alpha = 5\%$ .

In the case of mistakenness in task 2, Friedmann's test did not show statistically significant differences between distributions in the monitored measurements,  $X^2 = 10.77$  and  $p = 0.15$ . Therefore, post-hoc analysis did not show any statistically significant differences between the individual pairs of measurements at the significance level of  $\alpha = 5\%$ .

In both cases, errors were minimal, i.e., as the extremes of respective distributions. The graphical presentation of these results would, therefore, show a common distribution median, minimum, maximum, and quartiles, along with extraordinary values. Therefore, these results are not graphically presented.

In the case of mistakenness in task 3, Friedmann's test did not show statistically significant differences between the distributions in the measured measurements,  $X^2 = 5.02$  and  $p = 0.65$ . Therefore, post-hoc analysis did not show any statistically significant differences between the individual pairs of measurements at the significance level of  $\alpha = 5\%$ . Results are graphically presented in Fig. 6-C.

In the case of mistakenness in task 4, Friedmann's test did not show statistically significant differences between distributions in the monitored measurements,  $X^2 = 7.81$  and  $p = 0.35$ . Therefore, post-hoc analysis did not show any statistically significant differences between the individual pairs of measurements at the significance level of  $\alpha = 5\%$ . Results are graphically presented in Fig. 6-D.

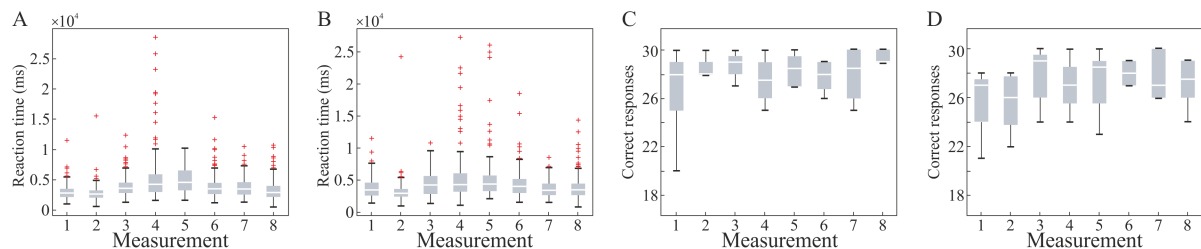


Fig. 6. Presentation of the OR-test results of time required to check first code in task 3 (A) and task 4 (B) along with distribution of correct answers in task 3 (C) and task 4 (D)

#### 4. DISCUSSION

Results of psychological testing through NTLx showed no statistically significant differences. The subjects were most affected by the frustration, the least by time pressure. In the case of performance tests, statistically significant differences were observed from the time perspective, but not from the correctness of the answers. It is, therefore, likely that the subjects have the priority to tick the correct field over the priority of filling the box as soon as possible. With increasing experiment time, pilot performance in the test did not change much, and the scores of individual subjects were almost unchanged throughout the measurement. Therefore, the effect of fatigue on the correctness of responses in these tests was not observed.

Success in OR-test was generally high, the best subject achieved an average score of 88.17 out of a total score of 90 possible points, and the least successful subject got 76.80 out of 90 possible. The average score was 83.98 points out of a total of 90.



The average response time for task 1 across all measurements was 3040 ms, with the highest average time of 3382 ms being found in the 4th measurement, and the shortest average time was 2673 ms, detected at the 7th measurement. These results can be attributed to the fact that subjects were awake after the 4th measurement for about 20 hours in the early morning hours, and thus, considerable fatigue with impaired ability to perceive and process information was observed. In the case of the 7th measurement, daylight was already present, and the subjects were well informed about the test. This can also be supported by a statistical test, the results of which show a statistically significant difference between the 4th and the 5th measurements, with the measurement 5 taking place at the time of transition from night to day.

The average response time for task 2 across all measurements was 3330 ms, with the highest average time of 3583 ms being found in the 5th measurement, and the shortest average time was 2877 ms found in the 6th measurement. This is supported by statistical testing, the results of which show a statistically significant difference between these two measurements.

The average cumulative time for checking the two fields in task 3 was 7442 ms, with the highest average time of 7922 ms being found in the 4th measurement and the shortest average cumulated time was 6915 ms found in the 8th measurement. Statistically significant differences were found between measurements 4 and 7. It is, therefore, obvious that in the case of the 4th measurement, subjects were experiencing the greatest fatigue manifestations.

The average cumulative time for checking the two fields in task 4 was 8204 ms, with the highest average time of 9324 ms being found for the 1st measurement, and the shortest cumulative average time was 7088 ms for the 8th measurement. In our opinion, in this task, the ability to learn this test had manifested, and therefore, the best performance in the case of the last measurement was the worst in the case of measurement 1. Based on statistical testing a statistically significant difference was found between these measurements.

The average time for checking one field in task 3 was 3738 ms, and in task 4, it was 4106 ms. From an overall view of the results, it is clear that the subjects experienced the greatest fatigue in the morning, with no daylight yet.

Despite the effort to suppress the learning process that was developed during the OR-test preparation, a certain degree of learning was observed and a system for a successful solution was obviously found. The correctness of the marked fields was probably greatly influenced by the reduced ability to perceive and understand the information. NTLx results do not show differences between measurements. However, the results show that the least influential factor was the time load and the biggest one was frustration, which is one of the contributing factors of air accidents [23, 24].

Looking at the results of performance tests, it is clear that with increasing fatigue, performance is decreasing. At the same time, it is obvious that fatigue is not only related to the total wakefulness time, but pilots, like other people, are also affected by the daily cycle, especially with regard to the presence of daylight and also other factors influencing the circadian rhythm, body temperature, heart rhythm, blood pressure, sensory or adrenal activity [2,25,26]. This may be the basis for degraded performance scores, in the early morning hours.

Overall, the results indicate that measurement using objective methods, i.e. performance tests, seems to be appropriate for pilots' fatigue monitoring and can provide a new insight into this issue, thus enhancing the knowledge of new information not only through subjective surveys.

## 5. CONCLUSION

The aim of this article was to evaluate the impact of fatigue on the outcome of psychological and performance tests. For the purpose of the study, a 24-hour measurement experiment with a predefined pre-flight mode was performed in order to cause fatigue. One of the simplest possible reasons for inducing fatigue was the disruption of the normal daily rhythm with the absence of night sleep. Due to the cockpit environment, the pilot occupation and the measurement needs, this choice seemed to be convenient. The study was attended by 8 students of the Faculty of Transportation Sciences, department of Air Transport, CTU in Prague with comparable theoretical and practical knowledge.

During the course of the measurements, the subjects underwent 8 simulated IFR flights. After each flight, they have undergone psychological testing through a standardized NASA Task Load Index (NTLx) and performance testing through the OR-test, which is similar to a test used at the Institute of Aviation Medicine. During the preparation of this test, the goal was to reduce the learning process as much as possible. For the purpose of this study, both of these tests were implemented in their own-developed software. The results of both tests were then statistically evaluated.

From the results of testing, it is obvious that the greatest fatigue subjects achieved in the early morning hours. In the case of these measurements, which were at the transition from night to day light, there was an increase in the time required to perform the performance test tasks, and, at the same time, an increase of error rate of the subjects was observed, i.e., a decrease in performance. In the case of the standardized NTLx questionnaire, the results show that the largest number of subjects was burdened with frustrations, whereas the smallest number of subjects was burdened with time constraints, which can be attributed to the fact that no time limits were set for the subjects to meet individual flight tasks. No statistically significant differences between the measurements were found in NTLx results. Thus, it is clear that objective testing that was performed in this case through performance testing is able to provide more information than standard questionnaires that are represented by NTLx. At the same time, it is obvious that not only the time of vigilance but also the time of day and the presence of daylight play a role in the case of fatigue, and it is also necessary to take this factor into consideration when planning flights.

One of the limitations of this study is the lack of subjects involved in the measurement. This was, however, caused by time-consuming experimental setup that was a limiting factor for subjects' participation. Moreover, it was also necessary to provide additional testing personnel for the duration of the measurement, to act as air traffic controllers, to communicate with the crew of the airplane, and to monitor the uniformity of the subjects' regime and the measurement in the context of activities performed outside the simulated flights. Therefore, the high personnel and time requirements for the measurement are obvious, which are the basis for above-mentioned limitation. Another limitation could be the measurement length with respect to the age of the subjects. As the experiment was attended by relatively young persons (aged 21–24 years), it can be assumed that such one-time fatigue did not have to be sufficiently limiting for them.

On the basis of what is mentioned above, it can be stated that it would be further advisable to extend the sample of the measured subjects. In addition, it would be advisable to perform more measurements to induce longer fatigue that could be much more observable during the test. It should be also taken into consideration possibilities for the extension of the tests performed during the experiment, which could contribute to a further understanding of the pilot fatigue.

Although this is a pilot study on a relatively small sample of subjects, the apparent effect of fatigue is mainly on performance testing results. It can, therefore, be argued that this work could serve as a basis for further studies to provide a more detailed description of fatigue and could serve as a support for introducing new pilots' psychological testing procedures in the future, which could contribute to current efforts to improve aviation safety.

## Acknowledgements

This work was also supported by the Slovak Research and Development Agency under the grant No. APVV-17-0167 “Application of the Self-regulatory techniques for the Flight Crew Preparation”.

## References

1. Abd-Elfattah, H.M. & Abdelazeim, F.H. & Elshennawy, S. Physical and cognitive consequences of fatigue: A review. *Journal of Advanced Research*. 2015. Vol. 6. No. 3. P. 351-358.
2. *Human performance and limitation: JAA ATPL training. Second Edition*. Neu-Isenburg: Jeppesen. 2007.

3. Sundelin, T. & Lekander, M. & Kecklund, G. & Van Someren, J. W. & Olsson, A. & Axelsson, J. Cues of Fatigue: Effects of Sleep Deprivation on Facial Appearance. *Sleep*. 2013. Vol. 36. No. 9. P. 1355-1360.
4. Creeley, H. & Nesthus, T. Predicting fatigue using voice analysis. *Aviation, Space, and Environmental Medicine*. 2007. Vol. 78. No. 7. P. 730-743.
5. Kozuba, J. & Pil'a, J. Aircraft automation systems versus pilot situational awareness (SA) - Selected aspects. In: *Proceedings of 19th International Conference "Transport Means"*. Kaunas: Kaunas University of Technology. 2015. P. 688-693.
6. Kozuba, J. & Pil'a, J. Chosen aspects of pilots situational awareness. *Nase More*. 2015. Vol. 62. No. SI. P. 175-180.
7. Madej, K. & Kozuba, J. Technology as a capability enhancement in the air training. In: *Proceedings of 6th International Conference on "Military Technologies"*. Brno: University of Defence. 2017. P. 467-476.
8. Morris, M.B. & Wiedbusch, M.D. & Gunzelmann, G. Fatigue Incident Antecedents, Consequences, and Aviation Operational Risk Management Resources. *Aerospace Medicine and Human Performance*. 2018. Vol. 89. No. 8. P. 708-716.
9. Bennett, S.A. Pilot workload and fatigue on short-haul routes: an evaluation supported by instantaneous self-assessment and ethnography. *Journal of Risk Research*. 2016. Vol. 21. No. 5. P. 645-677.
10. Weiland, M. & Nesthus, T. & Compatore, C. & Popkin, S. & Mangie, J. & Thomas, L. C. & Flynn-Evans, E. Aviation fatigue: Issues in developing fatigue risk management systems. In: *Proceedings of the Human Factors and Ergonomics Society*. San Diego. 2013. P. 1-5.
11. Avers, K. & Hauck, E.L. & Blackwell, L.V. & Nesthus, T.E. A qualitative and quantitative analysis of fatigue countermeasures training in the aviation industry. *International Journal of Applied Aviation Studies*. 2010. Vol. 10. No. 2. P. 51-66.
12. Caldwell, J.A. Fatigue in aviation. *Travel Medicine and Infectious Disease*. 2005. Vol. 3. No. 2. P. 85-96.
13. Mehta, R.K. & Peres, S.C. & Steege, L.M. & Potvin, J.R. & Wahl, M. & Stanley, L.M. & Nesthus, T.E. Fatigue monitoring and management across different industries. In: *Proceedings of the Human Factors and Ergonomics Society*. Washington. 2016. P. 993-996.
14. Antoško, M. & Sabo, J. & Hovanec, M. & Aviation, Sspace, P. & Sekelová, M. How to evaluate the actual psychological readiness of ATCO. In: *Proceedings of 21st International Conference "Transport Means"*. Juodkrante: Kaunas University of Technology. 2017. P. 1062-1065.
15. Antoško, M. & Pil'a, J. & Korba, P. & Lipovský, P. Psychological readiness of air traffic controllers for their job. *Nase More*. 2014. Vol. 61. No. 1-2. P. 5-8.
16. Gregory, K.B. & Winn, W. & Johnson, K. & Rosekind M.R. Pilot fatigue survey: exploring fatigue factors in air medical operations. *Air medical journal*. 2010. Vol. 29. No. 6. P. 309-319.
17. Bourgeois-Bougrine, S. & Carbon, P. & Gounelle, C. & Mollard, R. & Coblenz, A. Perceived fatigue for short-and long-haul flights: a survey of 739 airline pilots. *Aviation, space, and environmental medicine*. 2003. Vol. 74. No. 10. P. 1072-1077.
18. Powell, D.M. & Spencer, M.B. & Holland, D. & Broadbent, E. & Petrie, K.J. Pilot fatigue in short-haul operations: effects of number of sectors, duty length, and time of day. *Aviation, Space, and Environmental Medicine*. 2007. Vol. 78. No. 7. P. 698-701.
19. Hart, S.G. Nasa-Task Load Index (NASA-TLX): 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. 2006. Vol. 50. No. 9. P. 904-908.
20. Hart, S.G. *NASA Task load Index (TLX) Volume 1.0. Paper and Pencile Package*. 1986. Available at: <http://humansystems.arc.nasa.gov/groups/TLX/downloads/TLX.pdf>
21. Gibbons, J.D. *Nonparametric statistics: An introduction*. Newbury Park: Sage Publications. 1993. 96 p.
22. Dunn, O.J. Multiple comparisons using rank sums. *Technometrics*. 1964. Vol. 6. No. 3. P. 241-252.

23. Loewenthal, K.M. & Eysenck, M. & Harris, D. & Lubitsh, G. & Gorton, T. & Bicknell, H. Stress, distress and air traffic incidents: job dysfunction and distress in airline pilots in relation to contextually-assessed stress. *Stress Medicine*. 2000. Vol. 16. No. 3. P. 179-183.
24. Grandjean, E.P. & Wotzka, G. & Schaad, R. & Gilgen, A. Fatigue and stress in air traffic controllers. *Ergonomics*. 1971. Vol. 14. No. 1. P. 159-165.
25. Graeber, R.C. Aircrew fatigue and circadian rhythmicity. 1988. *Human factors in aviation*. P. 305-344.
26. Hursh, S.R. & Redmond, D.P. & Johnson, M.L. & Thorne, D.R. & Belenky, G. & Balkin, T.J. & Storm, W.F. & Miller, J.C., Eddy, D.R. Fatigue models for applied research in warfighting. *Aviation, space, and environmental medicine*. 2004. Vol. 75. No. 3. P. 44-53.

Received 03.12.2017; accepted in revised form 03.06.2019

## D Publikace 4

**Hanáková, L., Socha, V., Bořil, J., Herman, L., Štrobl, J., & Pilmannová, T.**  
(2020, November).

**Subjective Evaluation of Fatigue in Pilots During 24-hour Testing.**

2020 New Trends in Civil Aviation (NTCA) (pp. 111-116). IEEE.

# Subjective Evaluation of Fatigue in Pilots During 24-hour Testing

Lenka Hanáková  
*Department of Air Transport*  
*Czech Technical University in Prague*  
Prague, Czech Republic  
hanakle1@fd.cvut.cz

Vladimír Socha  
*Department of Air Transport*  
*Czech Technical University in Prague*  
Prague, Czech Republic  
sochavla@fd.cvut.cz

Jan Bořil  
*Department of Air Force*  
*University of Defence in Brno*  
Brno, Czech Republic  
jan.boril@unob.cz

Lukáš Herman  
*Department of Air Transport*  
*Czech Technical University in Prague*  
Prague, Czech Republic  
hermaluk@fd.cvut.cz

Jan Štrobl  
*Department of Biomedical Technology*  
*Czech Technical University in Prague*  
Kladno, Czech Republic  
strobja1@fbmi.cvut.cz

Terézia Pilmannová  
*Department of Air Transport*  
*Czech Technical University in Prague*  
Prague, Czech Republic  
pilmater@fd.cvut.cz

**Abstract**—With developing efforts to increase aviation safety, the focus on human factors is being emphasized as well. One of the phenomena affecting the performance of the pilot or other aviation specialists is undoubtedly the fatigue. Dangerous situations can occur, especially when fatigue accumulates. Monitoring fatigue is therefore in the interest of all stakeholders as well as the aviation authorities, which have developed Fatigue Risk Management System. In spite of the fact that the system itself at its highest level also uses the objective data, to this day, the monitoring of fatigue in aviation has mainly been based on a questionnaire survey. Although, the standardized NASA Task Load Index (NASA TLX) is often applied, it remains questionable whether such a questionnaire realistically reflects the fatigue. Based on previously mentioned, the NASA TLX was modified to focus on pilot fatigue and then applied during 24-hour measurements for the purposes of this paper. Eight pilots took part in the measurement, each of them performed 8 simulated flights. From the results of the questionnaire survey, a window of circadian low can be observed, in accordance with the evaluated theta/alpha parameter obtained by spectral analysis of EEG records. The results therefore indicate that such a questionnaire is able to indicate fatigue.

**Index Terms**—aviation, electroencephalography, fatigue, human factors, pilot, questionnaire, safety

## I. INTRODUCTION

Fatigue is a psychophysiological phenomenon having an effect on the ability to perform certain activities. In most cases, it is preceded by excessive, mental or physical strain on the individual. Symptoms of fatigue vary slightly from other causes. In general, there may be a decrease in concentration, deterioration of perception, or an increase in reaction time. Mood swings, excitation may also occur. An individual can solve problems carelessly without analyzing them in the necessary detail as he does not have the capacity to do so.

In the aviation industry, fatigue is an even more serious aspect, as the safety of the entire flight depends on the performance of the pilots. Pilots must not be irritated as

they have to cooperate in a team. Their attention must not be reduced, due to flight monitoring and the communication with the air traffic controller. They must not have a reduced cognitive performance due to the possible occurrence of an abnormal situation [1], [2].

Pilots fatigue has been a discussed topic since the first half of the 20th century. Although there has been some progress since then, aviation fatigue is of a growing concern. Airlines exploit crews to the fullest, therefore they often suffer from fatigue. With regards to accident statistics, fatigue is the among the most common causes of the incidents caused by the pilot [3]. Pilots themselves take fatigue seriously. Frequent passing through time zones, circadian rhythm disturbances, night flights, etc. cause great fatigue and contribute to the reduced mental capacity of the pilot [4]. Disruption of the circadian rhythm appears to be one of the biggest issues when it comes to fatigue. The human body is set for an approximately 24-hour cycle based on external stimuli, especially daylight levels [4]. Circadian low occurs in the morning, approximately between 2 and 6 o'clock in the morning [4]. If this time coincides with low workload, the crew is very prone to microsleep. Given the current nature of flights and crew planning, fatigue is inevitable and must be considered a risk factor [3].

Nowadays, the airline's approach to the fatigue of its pilots is very important. Properly conducted research on fatigue and consequent fatigue management systems are crucial. Training of both crews and flight dispatcher conducting flight planning is necessary [3]. For example, dispatchers need to know that flights to the east are more demanding and therefore require an adjustment of the schedule in such a way that pilots are able to handle it safely.

The simplest way to collect data concerning fatigue is to conduct a questionnaire survey, which is regularly used in aviation for these purposes [5]–[7]. At the basic level,

the Fatigue Risk Management System [8], [9] utilizes a questionnaire survey. Concerning data collection, there is no uniform standard for such a questionnaire survey, only recommendations regarding the focus of individual questions are available [9]. In addition to the questionnaire relating to sleep and work deployment also only very recent questionnaires related directly to in-flight fatigue or to the task performance have also been applied for the research [8], [9]. By default, the workload is monitored mainly via the standardized NASA Task Load Index (N-TLX) questionnaire [8]. Due to the ease of use of this questionnaire and its benefits in terms of workload monitoring, the questionnaire is used worldwide, in a number of sectors, including aviation [10]. However, the questionnaire does not target a specific application and therefore the available generalized instructions may be misunderstood by the subject. On the other hand, there is no other standardized questionnaire for the purpose of studying aviation fatigue, and the relevant recommendations related to the investigations conducted via questionnaire only point out a possible direction of the questions. Therefore, comparison of results and their possible objectification is limited.

Due to the ease of use of this questionnaire and its benefits in terms of workload monitoring, the questionnaire is used worldwide, in a number of sectors, including aviation [10]. However, the questionnaire does not target a specific application and therefore the available generalized instructions may be misunderstood by the subject. On the other hand, there is no other standardized questionnaire for the purpose of studying aviation fatigue, and the relevant recommendations related to the investigations conducted via questionnaire only point out a possible direction of the questions. Therefore, comparison of results and their possible objectification is limited.

Based on the above, the subject of this research paper is the application of a questionnaire survey based on a standardized N-TLX questionnaire, which will be specifically focused on the study of pilots' fatigue during the flight. The questionnaire is then used in an initial study applying experiments with variable levels of pilots' fatigue during simulated flights. In order to validate the questionnaire survey, the EEG was also monitored during the flight, especially focusing on fatigue monitoring.

## II. MATERIALS AND METHODS

### A. Measuring Procedure

The fatigue was monitored during an experiment with variable levels of fatigue. The experiment was attended by 8 students of Professional Pilot field of study. All subjects held a PPL (A) - a private pilot of aircraft license, and during the experiment were in IR (Instrument Rating) training to obtain a qualification for flight conduction according to the instrumental flight rules. The average age of the subjects was 22 *pm* 3 years and their flight hours ranged from 100–150 hours in airplanes and 30–50 hours on a certified FNPT II simulator in the Approved Training Organization. The subjects were classmates one the professional pilot field of study and

thus had comparable theoretical and practical knowledge in flight operations in instrumental flight rules (IFR) conditions.

During the measurement, 8 simulated flights were conducted, the first of which started at 18.00. After the first flight was finished by the first pilot, the first flight of the second pilot followed. In the meantime, the first pilot passed written tests, i.e. filled in a questionnaire and then waited for the second pilot to complete his flight. At the same time, however, he was not allowed to sleep. As described above, these 2 pilots were measured during each part of the experiment, while flying, being tested or waiting for the next flight. The experiment ended after the eighth flight of the second pilot, at approximately 6 p.m. the following day. The total measurement took approximately 24 hours. Also, the pilots had to follow a uniform 10 hours long regimen before the experimental measurement itself, i.e. they woke up at the same time of day, were forbidden to have higher physical activity, etc. in order to obtain greatest possible uniformity of the measurement. Pilots also had to adhere to this regime during the measurement, i.e. in total the subjects were awake for 34 hours at the end of the experiment.

All flights consisted of 2 parts and although they took place between different airports, they had a uniform course, where after takeoff from the first airport it was necessary to follow the standard instrument departure (SID), then follow the standard instrument arrival (STAR) at the second airport, perform arrival, instrument approach, land and then immediately begin to configure the aircraft for the second part of the flight, which also led through the SID and STAR routes to the ILS approach to the third airport. During the whole measurement, the subject configured the aircraft, worked with navigation documentation, set up navigation instruments, performed the flight itself and switched between individual radio frequencies on which he communicated with the air traffic controller, who successively issued flight permits adequate for the part of the flight. This way, conditions very similar to those that prevail on board of a real aircraft when flying according to instrumental flight rules were created.

### B. Questionnaire Design

For the purposes of the design of the questionnaire survey, a standardized N-TLX questionnaire has been applied, which uses six categories and their evaluation by users. Respective categories include mental demand, physical demand, temporal demand, performance, effort and frustration. These categories have been adjusted in a way that it is possible to quantify fatigue while focusing on specific problematic activities, resp. to evaluate the performance of the flight with regards to its accuracy and precision, or the psychophysiological condition of the pilot. The questions were always accompanied by an explanation of the state or situation the questions target to quantify and formulated in such a way that the meaning of the question was clear to the subject, i.e. in accordance with the principles for creating a questionnaire survey.





demanding. The prerequisite for this question is the gradually increasing fatigue, which can cause simple tasks to be perceived as more demanding. Specifically, the question asks whether the flight was demanding on piloting techniques. In order to apply an objective approach to evaluation of this question the analysis of flight data can be used.

The last category - Frustration - focused on feelings of stress, whether the subject felt under pressure, nervous or irritated. Specifically, the question asked whether the subject felt stress during the flight. This question is formulated in a manner that an objective approach to its evaluation can be applied while assessing the psychophysiological state of the pilot, using, for example, cardiac activity records.

Each of the questions has a scale of 0 - 20, see Fig. 1. The scale can be further converted to percentages, as each step represents 5 %. The questionnaire is then evaluated with respect to the weighting performed each time it is being completed. First, the number of times each of the options was selected during the initial selection of 15 pairs is summed and frequency of individual categories is obtained. This value is then divided by 15. This gives a weighting factor, which is then multiplied by the score for the category and the Overall score for the category. Therefore, for each  $i$  category, the corresponding Overall Score ( $OS_i$ ) is obtained as a percentage as [11]:

$$OS_i = \frac{x_i \cdot p_i \cdot 5}{15}, \quad (1)$$

where  $p_i$  is the frequency of the  $i$  category at initial weighting by selecting from 15 pairs. By summing the individual  $OS_i$ , the Total Task Load Index (Total TLX) can then be obtained. Further, an objective approach to its evaluation can be applied and the methods of fatigue can be quantified, such as parameters based on the analysis of records of brain or cardiac activity.

### C. Objective Measures

In order to obtain objective data, electroencephalography (EEG) fatigue monitoring was chosen, assuming that fatigue affects both EEG and the results of questionnaire survey, especially in the Total TLX parameter. Monitoring the electrical activity of the brain is one of the means of studying fatigue. Different numbers and locations of electrodes are used for the measurement, as well as different monitored parameters based on the spectral analysis of the EEG recording. By default, the EEG recording is processed using frequency analysis and power monitoring in specific frequency bands. By default, 4 frequency bands are used - alpha (8–13 Hz), beta (13–30 Hz), delta (0.5–4 Hz) and theta (4–8 Hz). When studying fatigue, the alpha and theta or beta bands are mainly used. In terms of monitoring individual parameters, the theta/alpha ratio [12] was used in transport applications, which decreases in case of vigilance [13].

A 16-channel Cyton + Daisy EEG with an Ultracortex Mark IV helmet (OpenBCI, New York, NY, US) with a sampling frequency of 125 Hz was used for EEG monitoring. During

processing, a highpass filter was used in order to remove the DC component and a notch filter (50 Hz). Data processing was done in the Matlab environment (MATLAB R2017a, MathWorks, Natick, MA, US).

### D. Statistical Analysis

The parameters, whether subjective or objective, were compared with each other during the individual measurements via the Friedmann test followed by a multi-comparative post-hoc analysis (Dunn-Sidak method), taking into account the nature of the data in which normal distribution was not observed based on the Jarque-Berra test and considering the fact that the data is paired. The distribution of individual parameters was shown through box fences. The statistical analysis took place in matlab (MATLAB R2017a, MathWorks, Natick, MA, US).

## III. RESULTS

In the case of the Mental demand Score, a significant difference was observed in measurement pairs 1-5 ( $p = 0.026$ ), 2-4 ( $p = 0.012$ ) and 2-5 ( $p = 0.004$ ) based on post-hoc analysis. In the case of the distribution of this parameter (see Fig. 2) a gradual increase of the median is apparent until the measurement No. 5, during which it reached a maximum. Subsequently, a significant decrease is observed followed by a slight increase and stabilization.

A similar median trend is observed for the Physical demand Score parameter (see Fig. 2). For this parameter, post-hoc analysis showed a significant difference in the 5% level of significance only between pairs 1-5 ( $p = 0.041$ ).

The Temporal demand Score parameter did not show any significant differences between the tested measurement pairs. The median course shows no significant trend (see Fig. 2).

In the case of the Performance Score, the maximum was again observed in the measurement No. 5. Even in the case of this measurement, no significant difference was observed between the individual measurements at the 5% level of significance as well as in case of Stress Score.

In terms of fatigue, Total TLX is the most interesting parameter. In the case of this parameter, the same trend is observed as in the previous ones, i.e. an increase from the 2nd measurement with a maximum in the fifth measurement and a subsequent decrease. In terms of post-hoc analysis, a significant difference was identified for pairs 1-4 ( $p = 0.030$ ), 1-5 ( $p = 0.003$ ), 2-4 ( $p = 0.030$ ) and 2-5 ( $p = 0.003$ ).

For the purposes of validation of the questionnaire survey, the theta / alpha parameter was chosen as an indicator of fatigue, which should show a similar trend as the Total TLX parameter. For this purpose, the C3 and C4 electrodes were selected from the standard electrode placement of the 10-20 system, which had previously shown significant results in fatigue monitoring.

In the case of the C3 electrode, a statistically significant difference was observed for the theta/alpha parameter based on post-hoc analysis in the pair of measurements 1-3 ( $p = 0.003$ ), 1-4 ( $p = 0.049$ ) and 2-3 ( $p = 0.005$ ).

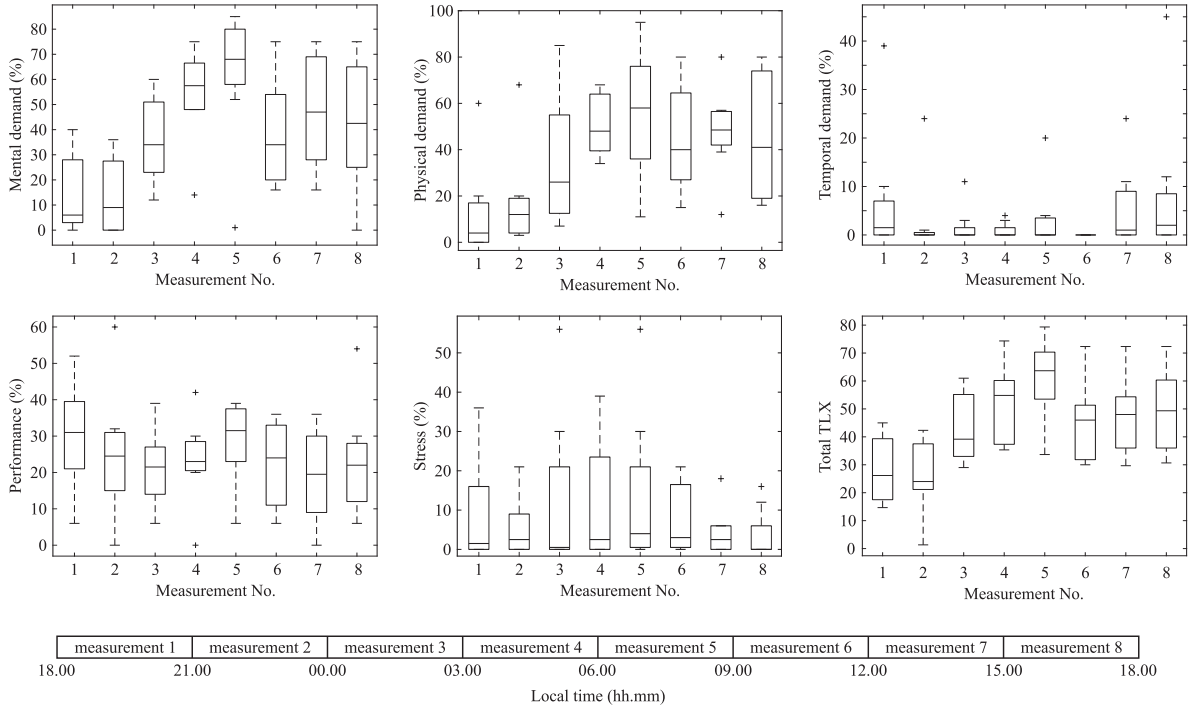


Fig. 2. Graphical representation of the Overall scores distribution for each questionnaire category, including the Total Task Load Index (Total TLX) during all eight measurements.

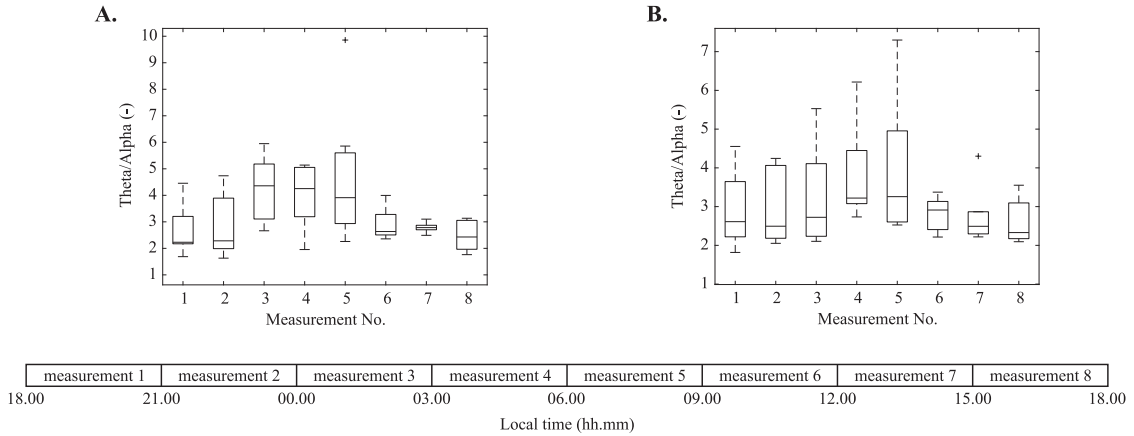


Fig. 3. Graphical representation of the distribution of the theta / alpha ratio in the electrodes C3 (A) and C4 (B) during all eight measurements.

The distribution shows a gradual increase from the 2nd measurement to the 4th measurement, at which the parameter reached its maximum, followed by a gradual decrease, see Fig. 3A.

Only one significant transition at the significance level of 5 % was observed for the C4 electrode, for the pair of measurements 1-4 ( $p = 0.034$ ). In the case of the parameter distribution (with Fig. 3B) a similar trend is observed as for the electrode C3, i.e. a gradual increase from the second measurement with a maximum in the 5th measurement.

#### IV. DISCUSSION

Significant differences were observed in the case of the questionnaire survey only for the parameters Mental demand Score and Total TLX. This is consistent with the initial assumption that fatigue is most evident on the psychological side of subjects. The observed significant differences were mainly related to the first measurements and measurements taking place in the morning. Measurement 4 took place around 6 am, measurement 5 then around 9 am. It is therefore clear that, based on the results of the questionnaire survey, variable

fatigue is evident, as well as circadian low, which is observed in the morning.

The EEG results are then consistent with these conclusions. The highest values of the theta/alpha parameter are observed for both evaluated electrodes in the morning hours of the measurement. Such a result was expected because in the case of fatigue, there is an increase in performance in the theta band and suppression of performance in the alpha band [12], [14].

As such, the results suggest the possible usefulness of subjective evaluation, as its results correspond to the results of the EEG. On the other hand, it was only a one-day fatigue, which can be considered as a limitation of the study. It is clear that although fatigue was observed, both on the basis of the individual parameters evaluated and on the basis of the observed behavior of subjects who were especially in the morning hours aggressive and visibly tired, for a further research on fatigue and experiment on the longer induced fatigue should be included. However, performing such an experiment is very demanding in terms of time and human resources. Therefore, the experiments that lasted 24 hours were, as such, already difficult to perform. This was also reflected in a smaller number of subjects that participated on the experiment.

#### V. CONCLUSION

In this study, pilot fatigue was monitored during 24-hour measurements, in the form of a subjective evaluation through a questionnaire survey based on the standardized Nasa Task Load Index, and through the evaluation of EEG records. The study, in which a variable level of fatigue was induced, involved pilots - students in IR training.

The results show that the questionnaire survey showed similar results as the EEG, and thus indicate that the questionnaire survey really reflects fatigue. On the other hand, the study has several limitations. The main one is the lower number of subjects, which is a result of the time and personnel demands of such an experiment. Furthermore, a higher level of fatigue could be induced, but this would presuppose, particularly, a longer experiment, the execution of which would be demanding with regards to human resources. On the other hand, the results may indicate a predicted trend in fatigue during this measurement and may be considered as an initial study and provide a knowledge base for follow-up research, with respect

to the fact that research involving this type of experiment has not been performed previously.

#### ACKNOWLEDGMENT

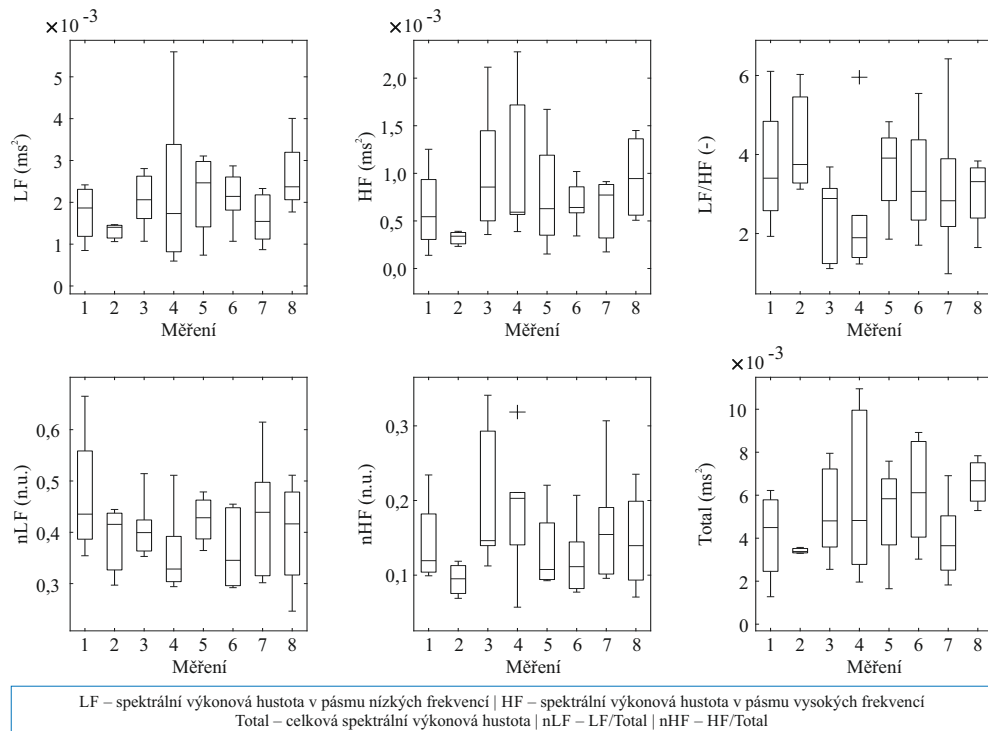
This work has been supported by the project SGS19/124/OHK2/2T/16 "Study of Fatigue Influence on Pilots' Performance".

#### REFERENCES

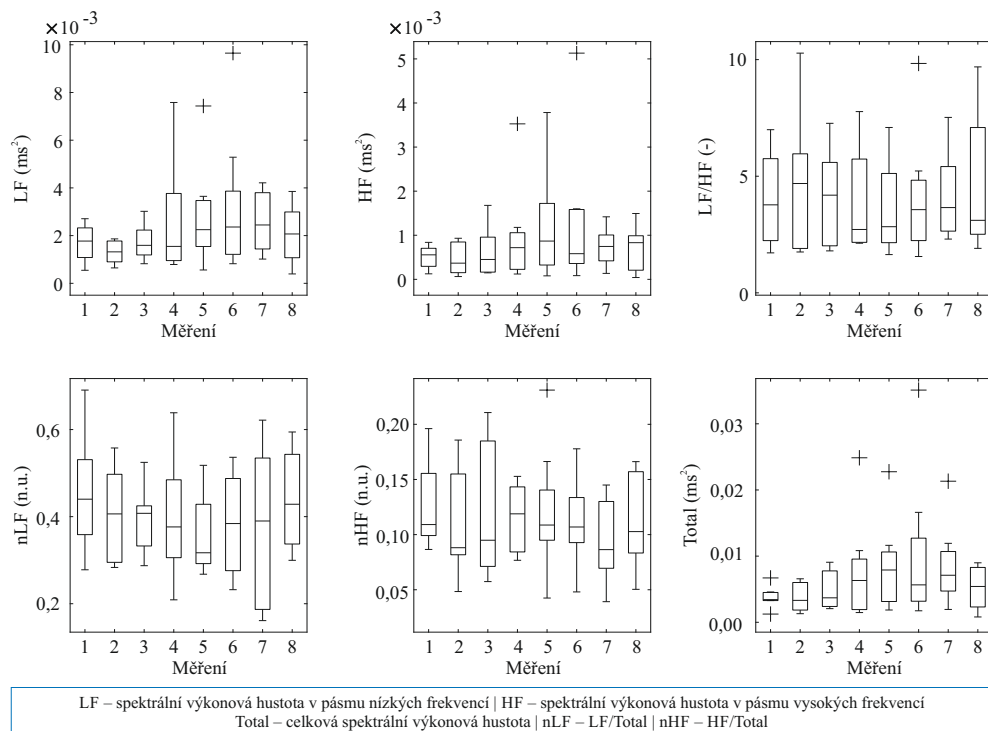
- [1] H. P. V. Dongen and S. R. Hursh, "Fatigue, performance, errors, and accidents," in *Principles and Practice of Sleep Medicine*. Elsevier, 2011, pp. 753–759. [Online]. Available: <https://doi.org/10.1016/b978-1-4160-6645-3.00067-0>
- [2] C. Reis, C. Mestre, and H. Canhão, "Prevalence of fatigue in a group of airline pilots," *Aviation, Space, and Environmental Medicine*, vol. 84, no. 8, pp. 828–833, aug 2013. [Online]. Available: <https://doi.org/10.3357/ase.m.3548.2013>
- [3] J. A. Caldwell, "Fatigue in aviation," *Travel Medicine and Infectious Disease*, vol. 3, no. 2, pp. 85–96, 2005. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S1477893904001012>
- [4] O. Academy, *ATPL Ground Training Series: Human Performance & Limitations*. Oxford Aviation Academy, 2009. [Online]. Available: <https://books.google.cz/books?id=f5fFwQEACAAJ>
- [5] K. B. Gregory, W. Winn, K. Johnson, and M. R. Rosekind, "Pilot fatigue survey: Exploring fatigue factors in air medical operations," *Air Medical Journal*, vol. 29, no. 6, pp. 309–319, nov 2010. [Online]. Available: <https://doi.org/10.1016/j.amj.2010.07.002>
- [6] S. Bourgeois-Bougrine, P. Carbon, C. Gounelle, R. Mollard, and A. Coblentz, "Perceived fatigue for short-and long-haul flights: a survey of 739 airline pilots," *Aviation, space, and environmental medicine*, vol. 74, no. 10, pp. 1072–1077, 2003.
- [7] D. Powell, M. B. Spencer, D. Holland, E. Broadbent, and K. J. Petrie, "Pilot fatigue in short-haul operations: effects of number of sectors, duty length, and time of day," *Aviation, Space, and Environmental Medicine*, vol. 78, no. 7, pp. 698–701, 2007.
- [8] FRMS Task Force, "Fatigue risk management systems: Implementation guide for operators," 2011.
- [9] IATA FRMS Task Force, "Common protocol for minimum data collection variables in aviation operations."
- [10] S. G. Hart, "NASA-task load index (NASA-TLX); 20 years later," in *Proceedings of the human factors and ergonomics society annual meeting*, vol. 50, no. 9. Sage publications Sage CA: Los Angeles, CA, 2006, pp. 904–908.
- [11] S. G. Hart and L. E. Staveland, "Development of nasa-tlx (task load index): Results of empirical and theoretical research," in *Advances in psychology*. Elsevier, 1988, vol. 52, pp. 139–183.
- [12] L. N. Boyle, J. Tiffin, A. Paul, and M. Rizzo, "Driver performance in the moments surrounding a microsleep," *Transportation research part F: traffic psychology and behaviour*, vol. 11, no. 2, pp. 126–136, 2008.
- [13] C. Cajochen, D. P. Brunner, K. Krauchi, P. Graw, and A. Wirz-Justice, "Power density in theta/alpha frequencies of the waking eeg progressively increases during sustained wakefulness," *Sleep*, vol. 18, no. 10, pp. 890–894, 1995.
- [14] K.-C. Huang, C.-H. Chuang, Y.-k. Wang, C.-Y. Hsieh, J.-T. King, and C.-T. Lin, "The effects of different fatigue levels on brain-behavior relationships in driving," *Brain and Behavior*, vol. 9, no. 12, p. e01379, 2019.

## **E Výsledky EKG**

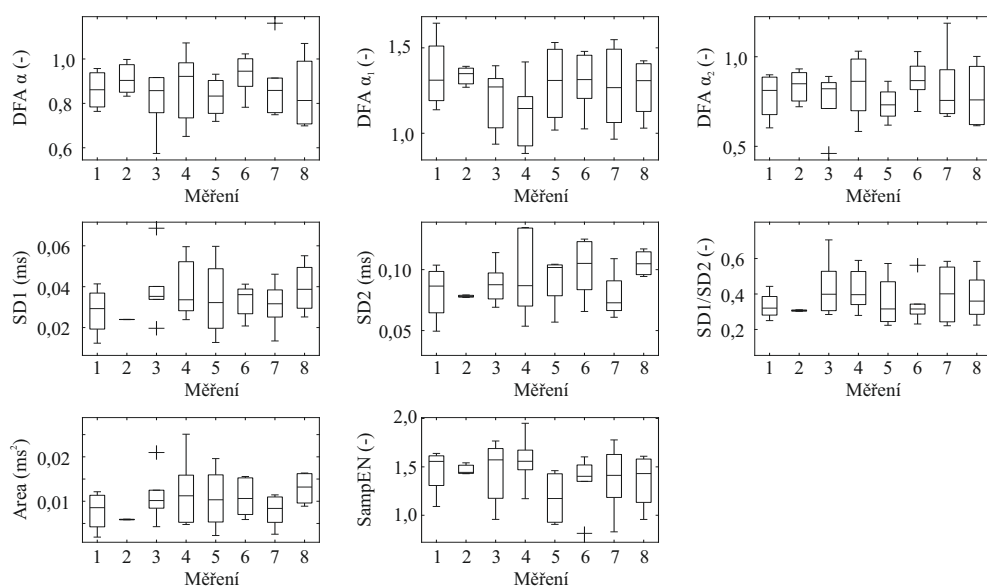
Grafické znázornění distribuce výsledků analýzy variability srdečního rytmu.



Grafické znázornění distribuce výsledků frekvenční analýzy variability srdečního rytmu u skupiny A, úsek 2 v průběhu všech měření formou boxplotů. Křížky značí odlehle hodnoty.

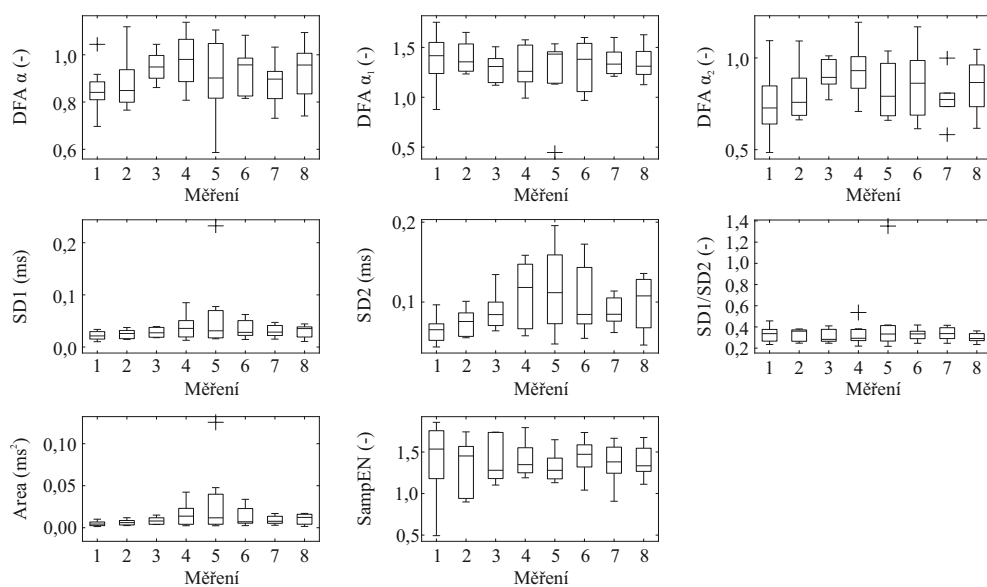


Grafické znázornění distribuce výsledků frekvenční analýzy variability srdečního rytmu u skupiny B, úsek 2 v průběhu všech měření formou boxplotů. Křížky značí odlehle hodnoty.



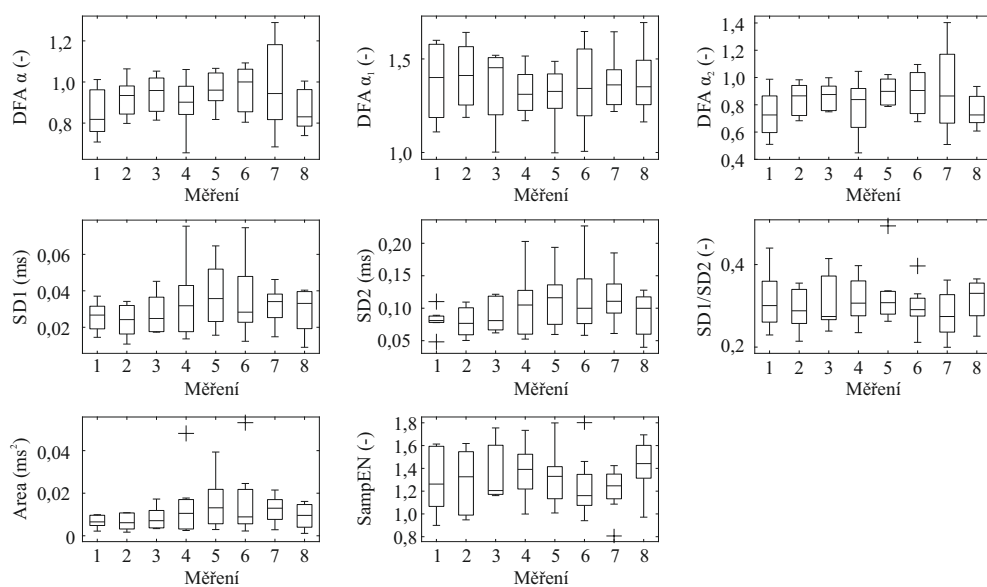
DFA  $\alpha$ ,  $\alpha_1$ ,  $\alpha_2$  – škálovací exponenty detrendované analýzy fluktuací | SDI – délka vedlejší poloosy Pointcaré grafu  
 SD2 – délka hlavní poloosy Pointcaré grafu | Area – plocha elipsy Pointcaré grafu | SampEN – vzorková entropie

Grafické znázornění distribuce výsledků nelineární analýzy variability srdečního rytmu u skupiny A, úsek 2 v průběhu všech měření formou boxplotů. Křížky značí odlehle hodnoty.



DFA  $\alpha$ ,  $\alpha_1$ ,  $\alpha_2$  – škálovací exponenty detrendované analýzy fluktuací | SDI – délka vedlejší poloosy Pointcaré grafu  
 SD2 – délka hlavní poloosy Pointcaré grafu | Area – plocha elipsy Pointcaré grafu | SampEN – vzorková entropie

Grafické znázornění distribuce výsledků nelineární analýzy variability srdečního rytmu u skupiny B, úsek 1 v průběhu všech měření formou boxplotů. Křížky značí odlehle hodnoty.



DFA  $\alpha$ ,  $\alpha_1$ ,  $\alpha_2$  – škálovací exponenty detrendované analýzy fluktuací | SD1 – délka vedlejší poloosy Pointcaré grafu  
SD2 – délka hlavní poloosy elipsy Pointcaré grafu | Area – plocha elipsy Pointcaré grafu | SampEN – vzorková entropie

Grafické znázornění distribuce výsledků nelineární analýzy variability srdečního rytmu u skupiny B, úsek 2 v průběhu všech měření formou boxplotů. Křížky značí odlehlé hodnoty.