Masterarbeit

Measurements of Physiological Parameters in Flight Simulator Studies for Stress Detection and Analysis Through Deep Learning Applications

Patrick Lorrig

Zeitraum: 01.11.2022 – 01.05.2023

Betreuer: Dipl.-Ing. Matthias Lehmann
Dipl.-Ing. Marco Dupper

Institut für Luftfahrtsysteme
Universität Stuttgart
Professor Dr.-Ing. Reinhard Reichel

Nr. M-100

Universität Stuttgart

Institut für Luftfahrtsysteme

Institutsleitung
Prof. Reinhard Reichel

Kontakt
Dipl. Ing. Matthias Lehmann
Pfaffenwaldring 27
70569 Stuttgart
T +49 711 685-62703
F +49 711 685-62964
e-mail:
matthias.lehmann@ils.uni-stuttgart.de

https://www.ils.uni-stuttgart.de/

01.11.2022

Aufgabenstellung

Masterarbeit

Measurements of Physiological Parameters in Flight Simulator Studies for Stress Detection and Analysis Through Deep Learning Applications

Die Mensch-Maschine-Interaktion stellt auch mit modernsten Luftfahrtsystemen immer wieder neue Herausforderungen dar. Hoher Stress kann dazu führen, dass Personal an seine Leistungsgrenzen gebracht wird, wodurch potenziell gefährliche Situationen entstehen können. Daher soll mittels Simulatorstudien untersucht werden wie sich bei Luftfahrzeugführer*innen in kritischen Flugsituationen die physiologische Reaktion darstellt.

Aufgabe

Es soll eine Flugsimulatorstudie vorbereitet und durchgeführt werden mit dem Ziel physiologische Reaktionen von Pilot*innen in stressigen und kritischen Flugsituationen aufzuzeichnen. Ebenso soll zum Aufbau einer geeigneten Datenbasis der momentane Stresslevel mit geeigneten Mitteln erhoben werden.

Mittels dieser Flugsimulatorstudien soll eine geeignete Datenbasis zum Trainieren eines Machine-Learning-Systems aufgebaut werden.

Basierend auf dieser Datenbasis soll die Möglichkeit der Entwicklung eines Echtzeitsystems untersucht und evaluiert werden.

Dieses soll EKG und ggf. andere sinnvolle physiologische Daten bei Pilot*innen analysieren und einen Rückgabeparameter bezüglich mentalen Stresses ausgeben.

Arbeitsschritte:

- Literaturrecherche nach Stand der Forschung/Technik
- Evaluation von Daten aus vorherigen Flugsimulatorstudien
    - Future Sky Safety Studie
    - Single Pilot Operation April 2022
- Konzepterstellung für weitere Flugsimulatorstudie
- Durchführung der Flugsimulatorstudie
- Aufbereitung und Analyse der Daten
- Evaluation der Flugsimulatorstudie sowie der entstandenen Daten
- Konzepterstellung Machine-Learning-System
- Anwendung und Erprobung des Machine-Learning-Systems

Beginn: 01.11.2022

Abgabe: _____

Betreuer: Dipl. Ing. Marco Dupper (ILS), Marcus Biella (DLR)

Prüfer: Dipl. Ing. Matthias Lehmann

Datum, Unterschrift Student: _____

Rechtliche Bestimmungen: Der/die Bearbeiter/in ist grundsätzlich nicht berechtigt, irgendwelche Arbeits- und Forschungsergebnisse, von denen er/sie bei der Bearbeitung Kenntnis erhält, ohne Genehmigung des/der Betreuers/in dritten Personen zugänglich zu machen. Bezüglich erreichter Forschungsleistungen gilt das Gesetz über Urheberrecht und verwendete Schutzrecht (Bundesgesetzblatt I/S. 1273, Urheberschutzgesetz vom 09.09.1965). Der/die Bearbeiter/in hat das Recht, seine/ihre Erkenntnisse zu veröffentlichen, soweit keine Erkenntnisse und Leistungen der betreuenden Institute und Unternehmen eingeflossen sind. Die von der Studienrichtung erlassenen Richtlinien zur Anfertigung der Bachelor-/Masterarbeit sowie die Prüfungsordnung sind zu beachten.

Selbständigkeitserklärung

Hiermit versichere ich, dass ich diese Masterarbeit selbständig mit Unterstützung des Betreuers/ der Betreuer angefertigt und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe. Die Arbeit oder wesentliche Bestandteile davon sind weder an dieser noch an einer anderen Bildungseinrichtung bereits zur Erlangung eines Abschlusses eingereicht worden.
Ich erkläre weiterhin, bei der Erstellung der Arbeit die einschlägigen Bestimmungen zum Urheberschutz fremder Beiträge entsprechend den Regeln guter wissenschaftlicher Praxis[1] eingehalten zu haben. Soweit meine Arbeit fremde Beiträge (z.B. Bilder, Zeichnungen , Testpassagen etc.) enthält, habe ich diese Beiträge als solche gekennzeichnet (Zitat, Quellenangaben) und eventuell erforderlich gewordenen Zustimmungen der Urheber zu Nutzung dieser Beiträge in meiner Arbeit eingeholt. Mir ist bekannt, dass ich im Falle einer schuldhaften Verletzung dieser Pflichten die daraus entstehenden Konsequenzen zu tragen habe.

Stuttgart, den Thursday 6th April, 2023

Patrick Lorrig

[1] Nachzulesen in den DFG-Empfehlungen zur "'Sicherung guter wissenschaftlicher Praxis"' bzw. in der Satzung der Universität Stuttgart zur "'Sicherung der Integrität wissenschaftlicher Praxis und zum Umgang mit Fehlverhalten in der Wissenschaft"'

Nutzungsrechterklärung

Hiermit erkläre ich mich damit einverstanden, dass meine Masterarbeit zum Thema:

Measurements of Physiological Parameters in Flight Simulator Studies for Stress Detection and Analysis Through Deep Learning Applications

in der Institutsbibliothek des Institutes für Luftfahrtsysteme mit sofortiger Wirkung öffentlich zugänglich aufbewahrt und die Arbeit auf der Institutswebseite sowie im Online-Katalog der Universitätsbibliothek erfasst wird. Letzteres bedeutet eine dauerhafte, weltweite Sichtbarkeit der bibliographischen Daten der Arbeit (Titel, Autor, Erscheinungsjahr, etc.).
Nach Abschluss der Arbeit werde ich zu diesem Zweck meinem Betreuer neben dem Prüfexemplar eine weitere gedruckte sowie eine digitale Fassung übergeben.
Der Universität Stuttgart übertrage ich das Eigentum an diesen zusätzlichen Fassungen und räume dem Institut für Luftfahrtsysteme an dieser Arbeit und an den im Rahmen dieser Arbeit von mir erzeugten Arbeitsergebnissen ein kostenloses, zeitlich und örtlich unbeschränktes, einfaches Nutzungsrecht für Zwecke der Forschung und der Lehre ein. Falls in Zusammenhang mit der Arbeit Nutzungsrechtsvereinbarungen des Instituts mit Dritten bestehen, gelten diese Vereinbarungen auch für die im Rahmen dieser Arbeit entstandenen Arbeitsergebnisse.

Stuttgart, den Thursday 6th April, 2023

Patrick Lorrig

Kurzzusammenfassung

Messung physiologischer Parameter in Flugsimulatorstudien zur Erkennung von kognitivem Stress mittels Deep-Learning-Anwendung

Die Automatisierung in der Luftfahrt hat sich im Laufe seiner Existenz nicht nur etabliert, sondern nimmt immer noch weiter stetig zu. Der Arbeitsplatz der Pilot*innen hat sich damit bereits ebenso massiv verändert und wird sich wohl genauso weiterhin wandeln. Dabei haben sich die Aufgaben des Flugpersonals weg von handwerklichen und mechanischen hin zu beobachtenden Tätigkeiten und Systemanwendung bewegt. Mit weiterer Automatisierung des Flight Decks werden die mentale Leistungsfähigkeit sowie die verbleibenden mentalen Ressourcen zu einer wertvollen Metrik. In dem Fall würde die Systemtiefe sowie die Arbeitsbeanspruchung sich an eben diese menschliche Leistungsfähigkeit der Pilot*innen anpassen. Im Rahmen des 'Human Performance Envelope' hat sich Stress als eines der Kriterien herausgestellt, welches die Leistungsfähigkeit maßgeblich beeinflusst.
Ziel der Arbeit war es physiologische Daten von Pilot*innen unter stressigen Flugszenarien zu messen. Auf Basis dieser sollte dann ein Deep Learning Model erstellt werden, welches Stresslevel anhand von EKG-Signalen erkennt. Dieses soll, zumindest theoretisch, dazu in der Lage sein in Echtzeit das Eingangssignal zu analysieren und eine entsprechende Einschätzung ausgeben.
Um dieses Deep Learning Model trainieren zu können, wurden in zwei Flugsimulatorstudien physiologische Daten gemessen. Während einer ersten Studie im Rahmen einer anderen Masterarbeit ging es vornehmlich um die Fragestellung, wie sich die Anfluggenauigkeit bei Single Pilot Operation verändert. Dabei wurde ein 1-Kanal-EKG aufgezeichnet sowie von den Versuchspersonen nach dem Flug händisch ein Stressverlauf erstellt. Diese Daten wurden dann zunächst analog-digital-gewandelt und anschließend mit dem EKG-Signal zusammengeführt. Aufbauend auf den Erkenntnissen dieser Studie wurde eine weitere Flugsimulatorstudie erstellt, die sich auf die Messung von EKG und Atemkurve fokussierte. Dazu wurden Szenarien konstruiert, die bei den Pilot*innen ein möglichst hohes Maß an Stress hervorrufen sollten. Hierfür wurden bewusst Systemausfälle gewählt, die zum einen realistisch und zum anderen im Simulator AVES umsetzbar waren. Um das hervorgerufene Maß an Stress einschätzen zu können, wurde ein Online Stress Assessment Tool entwickelt, welches dies möglichst genau erfassen soll. Der Vorteil dieses Tools ist es, dass der angegebene Stresswert des Piloten zeitlich genau mit dem EKG-Signal synchronisiert werden kann.
Auf Basis der Daten dieser beiden Studien wurde dann versucht ein Long-Short Term Memory (LSTM) zu trainieren, um Stresslevel vorherzusagen. Dazu wurde zunächt eine relativ einfache LSTM-Struktur erstellt, die auf Basis der mittleren quadratischen Abweichung und dem Adam-Algorithmus optimiert wurde. In vier Iterationsschritten wurden die Hyperparameter verändert, um das Konvergenzverhalten zu verbessern. Bei der letzlich gefundenen Modellarchitektur stellte sich bei Betrachtung der Vorhersagen heraus, dass dieses eher versucht das EKG nachzuahmen als Stresslevel zu erkennen. Aufgrund der langen Rechenzeit des Modells konnten leider keine weiteren Iterationen mit Anpassungen und Verbesserungen an der Modellarchitektur untersucht werden. Die gewonnenen Daten aus den Flugsimulatorstudien eröffnen jedoch weitere Möglichkeiten zum Aufbau eines Stress-Erkennungssystem für Flight Crews.

Abstract

Measurements of Physiological Parameters in Flight Simulator Studies for Stress Detection and Analysis Through Deep Learning Applications

Automation in aviation has not only established itself over the course of its existence, but is still advancing. The workplace of pilots has changed massively since their inception and will probably continue to change in the same way. Tasks done by flight personnel have moved away from manual and mechanical tasks to observational activities and system applications. With further automation of the flight deck, the mental capacity as well as the remaining mental resources become a valuable metric. In that case system depth as well as imposed workload onto flight crews would be adapted to this very human performance of the pilots. In the context of the 'Human Performance Envelope', stress has emerged as one of the criteria that significantly influences this.

The aim of this thesis was to measure physiological data of pilots under stressful flight scenarios. On the basis of this data, a deep learning model was to be created which recognizes stress levels on the basis of ECG signals. This model should, at least theoretically, be able to analyze the input signal in real time and output a corresponding stress level assessment.

In order to train this deep learning model, physiological data were measured in two flight simulator studies. During the first study, which was part of another Master's thesis, the primary question was how approach accuracy changes in single-pilot operations. A 1-lead ECG was recorded as well as an analog stress level tracker filled out after each flight scenario. These data were then first analog-to-digital converted and then merged with the ECG signal. Building on the findings of this study, another flight simulator study was created and conducted which focused on ECG and respiratory curve measurements. For this purpose, scenarios were constructed that were intended to induce the highest possible level of stress in the pilots. System failures were deliberately chosen which were both feasible to implement in the simulator AVES and realistic. In order to be able to assess the level of stress caused, an Online Stress Assessment Tool was developed to record stress levels as accurately as possible. The advantage of this tool is that the indicated stress value of the pilot can be precisely time synchronized with the ECG signal.

Based on the data from these two studies, an attempt was then made to train a Long-Short Term Memory (LSTM) to predict stress levels. To do this, a relatively simple LSTM structure was first created and optimized based on Mean Square Error and Adam algorithm. In four iteration steps, the hyperparameters were changed to improve the convergence behaviour. With the final model architecture found, looking at the predictions revealed that this system attempts to mimic the ECG rather than detect stress levels. Unfortunately, due to the long computation time of the model, no further iterations with adjustments and improvements to the model architecture could be investigated. However, the data obtained from the flight simulator studies opens up further possibilities for building a stress detection system for flight crews.

Acknowledgements

It is with deep gratitude that I acknowledge the contributions of those who supported me undertaking this research journey, which has been both rewarding and challenging. I am grateful to have had the privilege.
First and foremost, I extend my heartfelt thanks to my supervisors, *Dipl. Ing. Marco Dupper* and *Dipl. Ing. Matthias Lehmann*, for their invaluable guidance, expertise, and constructive feedback. Their mentorship not only during this thesis but also over the course of my studies has helped me to grow as a researcher and an individual.
I am also grateful to my family and friends, whose unwavering encouragement and support kept me motivated throughout the research process. Especially and foremost, to *Dr. Steven Koenig* who agreed to proof-read this thesis and gave me the unwavering support to get through challenges and stay focused on my goals.
My research would not have been possible without *Michael Ritzau-Jost* who simulated the air traffic controller during both simulator studies. Special thanks to the incredible work of *Marc Ilic* whose dedicated support with any AVES-related topics was invaluable.
I would like to express my appreciation to *Marcus Biella* whose guidance and support were invaluable throughout the entire process. Also, this extends to the whole Institute of Flight Guidance at the DLR Brunswick, Germany for providing me with the resources and facilities that I needed to undertake this research. Anyone I met was always willing to help and provided their expertise.
I am thankful to the participants of this study, whose time, experience and contribution made this research possible. Their willingness to share their knowledge and perspectives has helped me to gain a deeper understanding of the research topic.
Additionally, I would like to acknowledge the incredible work of the open source software contributors who have created and maintained the tools and technologies that made this research possible. Their commitment to creating accessible and innovative software has been crucial in advancing research and technology across many fields. Special thanks to all contributors of those projects and technologies.
Lastly, I extend my gratitude to everyone who played a role, big or small, in my research journey. Their contributions, advice, and encouragement have been instrumental in shaping this thesis.
Thank you all so much!

Contents

List of Figures

List of Tables

LIST OF TABLES

Glossary

Certification Specifications	Minimum requirements set by the **E**uropean Union **A**viation **S**afety **A**gency (EASA) that aircraft have to comply for the certification.
Flight Director	Visual attitude indicator on the Primary Flight Display which helps Pilots to execute a desired flight path.
Glide slope	Part of the instrument landing system needed for vertical guidance.
Localizer	Part of the instrument landing system needed for horizontal guidance.
Normal Law	Normal Flight Control Law.
Online Stress Assessment Tool	Novel method designed during this thesis in order to measure stress in real time and signal synchronization.

GLOSSARY

List of Abbreviations

AGL	**A**bove **G**round **L**evel
ANN	**A**rtificial **N**eural **N**etwork
ANS	**A**utonomous **N**ervous **S**ystem
ATC	**A**ir **T**raffic **C**ontrol(ler)
ATP	**A**denosine **t**riphos**p**hate
AVES	**A**ir **Ve**hicle **S**imulator
CNN	**C**onvolutional **N**eural **N**etwork
CNS	**C**entral **N**ervous **S**ystem
CS	**C**ertification **S**pecifications
CSV	**C**omma-**S**eparated **V**alues
CWC	**C**ross **W**ind **C**omponent
DLR	**D**eutsches Zentrum für **L**uft- und **R**aumfahrt
EASA	**E**uropean Union **A**viation **S**afety **A**gency
ECAM	**E**lectronic **C**entralized **A**ircraft **M**onitoring
ECG	**E**lectro**c**ardio**g**ram
EEG	**E**lectro**e**ncephalo**g**ram
EMG	**E**lectro**m**yo**g**ram
EU	**E**uropean **U**nion
FD	**F**light **D**irector
FFT	**F**ast **F**ourier **T**ransformation
fNIRS	**F**unctional **n**ear-**i**nfra**r**ed **s**pectroscopy
GRU	**G**ated **R**ecurrent **U**nit
GSR	**G**alvanic **S**kin **R**esistance
HPE	**H**uman **P**erformance **E**nvelope
HRV	**H**eart **R**ate **V**ariability
IBI	**I**nter-**B**eat-**I**nterval
ILS	**I**nstrument **L**anding **S**ystem
ISA	**I**nstantaneous **S**elf-**A**ssessment
kNN	**k**-**N**earest **N**eighbour
LoHP	**L**imits **o**f **H**uman **P**erformance
LSL	**L**ab **S**treaming **L**ayer
LSTM	**L**ong-**S**hort **T**erm **M**odel
METAR	**Met**eorological **A**erodrome **R**eport
MSE	**M**ean **S**quarred **E**rror

NASA	**N**ational **A**eronautics and **S**pace **A**dministration
NATS	**N**ational **A**ir **T**raffic **S**ervices
ND	**N**avigation **D**isplay
OSAT	**O**nline **S**tress **A**ssessment **T**ool
PFD	**P**rimary **F**light **D**isplay
PNS	**P**eripheral **N**ervous **S**ystem
RA	**R**adio **A**ltimeter
ReLu	**Re**ctified **L**inear **u**nit
RNN	**R**ecurrent **N**eural **N**etwork
SGD	**S**toachstic **G**radient **D**escent
SPO	**S**ingle **P**ilot **O**peration
SVG	**S**calable **V**ector **G**raphics
TLX	**T**ask **L**oad Inde**x**
TMP	**T**rans**m**embrane **P**otential
TWC	**T**ail **W**ind **C**omponent
UTC	**C**oordinated **U**niversal **T**ime
VAS	**V**isual **a**nalogue **s**cale
VP	**V**ersuchs**p**erson (*German for 'Test Person'*)
WFDB	**W**ave**f**orm **D**ata**b**ase
XML	E**x**tensible **M**arkup **L**anguage

Symbols and Units

Symbol	Description	Unit
b	Bias	-
l_r	Learning Rate for Adam Optimizer	-
y_{incr}	Slider Increment in OSAT	-
Δt_{indi}	Time until Update Indication in OSAT	s
$\Delta t_{y,incr}$	OSAT Time step for each Slider Increment	s
w	Weight	-
w_d	L2 Regularizarion Weight Decay	-
x	Input Vector	-
y	Label / Ground Truth	-
$\hat{y}$	Prediction	-

SYMBOLS AND UNITS

Part I

Preamble

1 Introduction

Automation has become an essential technology in aviation, with modern aircraft being highly protected and automated. However, there is still much potential for further developments and applications that can bring additional benefits through automation. One of the most important effects of automation is the transformation of the role of the flight crew from an active operator to a more passive observer and control position. As the degree of automation continues to increase, the human-machine interaction becomes even more crucial, as does the evaluation of the resources, capabilities, and limitations of flight crews. In the past, flight crews used to evaluate the systems, but with full automation, the system will evaluate the flight crew instead.
Advancing digitization and automation in aircraft systems changed the tasks that flight crews need to perform. Historically, aircraft have evolved from being mostly mechanical to becoming more computerized and automated. Today, pilots of commercial aircraft are more system managers than doing craftsmanship. The future automation will further change the tasks of flight crews, and their role and perspective will shift from being a system manager to someone who is being supported to complete a mission. Flight crews could become a subsystem of all aircraft systems, managed as a resource, such as modern flight computers or smart sensors.

1.1 Motivation

Over the course of time automation has become an established technology in aviation. Although modern aircraft are already highly automated and protected, there is still potential for further developments, applications and hence benefits through automation. Until today the job of flight crews already transformed from an active operator role to a more passive observing and managing position. While the flight crews currently evaluate the systems, with full automation it will be the other way around, with the system evaluating the available resources of pilots. Therefore, pilots could become a subsystem of all aircraft systems that are managed as a resource the same as modern flight computers or smart sensors.
As a result of further increasing the degree of automation, obtaining an objective measurement of the available resources of flight crews as a subsystem of the aircraft system becomes crucial. Automation or the adaptation of the quality of service in aircraft systems must at least know what the users' available capabilities are. In highly automated systems, having a clear picture of human performance limitations is essential, especially when considering automation that can adapt the system depth to the current situation. Without knowledge of these factors, the aircraft system would be blind to the degree of automation which would suit the flight crew best.
There are several approaches to assess stress, workload, and performance limitations in humans. A strong connection exists between the autonomous nervous system and the cardio-vascular system, making an analysis of the cardiac electric response a promising approach. This approach can provide valuable insights into human performance limitations.

1.2 Proceedings

When faced with the task of identifying stress in flight crews, the first step was to search for existing data and previous studies. Especially when applying a deep learning algorithm having sufficient and reliable data is important. Online databases like *kaggle.com* or *PhysioNet.org* contain data sets that were obtained in scientific studies and are made available for further research. Those do contain ECG records of individuals under stress. PhysioNet contains one data set of ECG recordings from physical exercise. [1] A more interesting data set is from stress detection in automobile drivers by Healey and Picard. This data set contains 17 records which include ECG, **E**lectro**m**yo**g**ram (EMG), **G**alvanic **S**kin **R**esistance (GSR) and respiration with durations of $65\,\mathrm{min}$ to $93\,\mathrm{min}$. The data from Healey and Picard was derived from a study by the same authors. In that study participants drove a car while physiological data and a video recording were taken. Besides a recording during resting, different driving scenarios were conducted in the city as well as on the highway.
In their study Healey and Picard achieved a $97\,\%$ accuracy with $5\,\mathrm{min}$ non-overlapping data interval for three different stress classes of 'low', 'medium', and 'high' stress. From all signals recorded during this study 22 features were extracted from the original signal and then input into a classification stream. This classification stream applies a Fisher projection and linear discrimination. [3] Even **N**ational **A**eronautics and **S**pace **A**dministration (NASA) investigated the question regarding real time stress detection. [4] They used the NASA Task Battery to perform three different levels of task demand while recording **E**lectro**e**ncephalo**g**ram (EEG), ECG, respiration and electrooculography. This data was then used to train a not further specified **A**rtificial **N**eural **N**etwork (ANN). Their results showed that they could determine the right amount of stress and then adapt the task load to recover mental resources. In a general —not related to transportation—, several studies have been done concerning the detection of stress based on different physiological parameter.

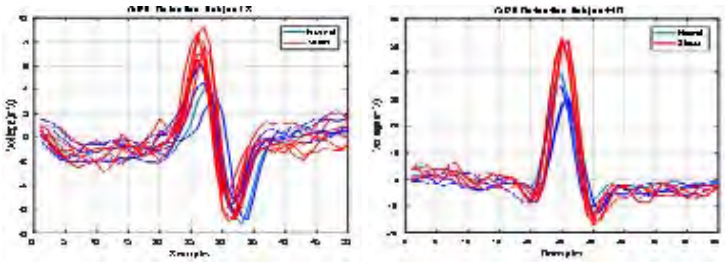


Figure 1.1: Original figures from Shahrudin, Sidek and Jusoh which show changes of ECG curves under stress (red) compared to normal (blue). Left plot shows the ECG of a male and the right a female participant.

Through investigating several physiological parameters Sharma and Gedeon found through empirical ranking that **H**eart **R**ate **V**ariability (HRV) is best suited for measuring stress. [6] Moreover, Shahrudin, Sidek and Jusoh showed in their study that the ECG morphology changes with increasing stress as well. [5] This opens up the path to consider a whole ECG signal instead of extracting a few features from it. Therefore, this thesis will try to accommodate the whole ECG for the deep learning application and investigate prediction accuracy.
Although all those studies investigated stress based on physiological signals none of those considered the flight deck environment. The only data sets available were from automobile drivers and laboratory research with limited tasks. In order to build a deep learning system to classify stress those existing data sets might be sufficient. On the other side, flight crews' response to critical and stressful situation might be more subtle than in other settings. This can be argued with the different environment and tasks flight crews are confronted with compared to cars. Therefore, the need to collect and measure physiological data from flight crews during

realistic flight scenarios arose. The idea was to create scenarios that would cause stress in pilots due to the increased task- and workload caused by the failure of various systems. In order to get appropriate stress evaluations flight crews shall give a self-evaluated feedback on how intense stress is for them. A good approach to collect such data is by using flight simulators. This not only allows to safely and reproducibly confront pilots with critical situations, but also to easily collect flight performance data.
After collecting and processing the measured data into a structured data set it can be used for machine learning applications. Then, the next step was finding a suitable machine learning algorithm for this question. For the moment a supervised learning classification task was chosen in order to classify the corresponding stress level for each ECG signal. Other approaches could be chosen by comparing a baseline ECG with the ones recorded during stressful periods. An unsupervised learning e.g. with **k**-**N**earest **N**eighbour (kNN), could then be used to discriminate between those different recordings.

This thesis starts with a theoretical introduction on relevant topics such as human physiology and machine learning. Different selected aspects are explained in order to help understand underlying principles and concepts and decisions used later. In the next chapter, several tools, concepts, and methods are explained that have been used throughout both flight simulator studies. The first study regarding Single Pilot Operations by Mr. C. Booms (Psychology Major at University Ulm) was used to gather physiological data in the background. On the contrary, the second study was conducted specifically for this thesis and in order to measure the stress feedback more precisely. In order to do so, a new online stress assessment tool was developed which is explained in chapter 5. Before that, as last part of the theoretical section, chapter 4 derives system requirements for the stress detection system. Those have been collected in cooperation with the Department of System Ergonomics at the Institute of Flight Guidance at **D**eutsches Zentrum für **L**uft- und **R**aumfahrt (DLR) Brunswick, Germany. Furthermore, this chapter lays out the fundamental design theories which the later deep learning architecture is build upon. Chapter 6 continues with the processing and merging of the data collected during the before mentioned simulator studies. Software tools have been developed in order to speed up the process and for exporting data into databases. Moreover, a closer look into the data that has been acquired is taken in order to get an understanding of how the data is structured. The following chapter, chapter 7, explains the deep learning architecture found as well as the training and revision process. Lastly, the results of all steps are concluded, discussed, and given a perspective on how further work on the stress detection system could look like.

Part II

Theory

Fundamentals 2

This thesis focuses on two major aspects: human physiological reactions and analyzing this data by applying machine learning algorithms. In order to better understand these physiological reactions and why specific measurement equipment and evaluation methods were chosen, some introduction to the human anatomy and physiology is provided. Concluding the section of organ-related topics this chapter moves on to investigate the second relevant aspect: machine learning. The basic principles and theorems of machine learning in general are covered first, then more detailed information on one type of machine learning—**L**ong-**S**hort **T**erm **M**odel (LSTM) systems— is provided.

2.1 Basic Anatomy and Physiology

When it comes to measuring physiological data one is confronted with the complexity of the human body. To get a more thorough understanding for the measuring methods this chapter will lay out fundamental aspects of anatomy and physiology.

2.1.1 Nervous System

The highly complex nervous system can be divided into a central and a **P**eripheral **N**ervous **S**ystem (PNS). While the **C**entral **N**ervous **S**ystem (CNS) consists of the brain, spinal cord as well as the twelve cranial nerves ('main nerves'), the PNS consists of the somatic and autonomous nervous system. The somatic nervous system consists of all aspects responsible for feeling, motoric innervation and communicating signals between the body and brain. Hence, it is under voluntary control. On the other side, the **A**utonomous **N**ervous **S**ystem (ANS) cannot be willingly controlled and may also be referred to as *vegetative nervous system*. Its main function is to regulate organic functions such as heart rate or digestive activity. Furthermore, within the ANS a distinction can be made between the sympathetic nervous system as well as the

parasympathetic nervous system. These two are working in opposing directions. While the sympathetic nervous system is mainly responsible for increasing body activity the parasympathetic system reduces alertness and inhibits digestion. They are often compared to 'fight and flight' and 'digest and rest', respectively. For the sake of completeness, the enteric nervous system, comprising the whole nervous system inside the digestive organs, is also part of the ANS. The most dominant nervous cord related to the parasympathetic system is the vagus nerve. This nerve cord extends through the digestive tract, lungs, heart, and many more areas across the body. Hence, it is also a very important part in regulating the heart rate. The influence can be so significant that a massive stimulus of the vagus nerve can lead to cardiac arrest. The mechanism to influence these body reactions is by using neurotransmitter molecules. These neurotransmitters are released from the sympathetic and parasympathetic nerves around the heart directly into the neuromuscular junction – the coupling between nerve ends and muscle cells. When the brain triggers the sympathetic nervous system predominantly epinephrine and nor-epinephrine are released from the nerve ends. These cause the heart to increase heart rate, blood pressure and hence blood flow, as well as reducing all non-necessary body functions. On the other side the Parasympathetic system is predominantly inhibited by the neurotransmitter acetylcholine. These neurotransmitters cause relatively fast responses within a few seconds. Which makes relevant adaptions of the body similar fast.

2.1.2 Principles of the Circulatory System

The vascular or circulatory system can be separated into a systemic and pulmonary circulation. The systemic circulation consists of all blood vessels which are distributed all over the body. The pulmonary system includes those blood vessels which carry blood to the lungs for oxygenation and then back to the heart. In both systems, the main types of blood vessels are arteries and veins. Contrary to common belief, the distinction between the two does not lie in the oxygenation status of the blood carried by them, but the transport direction relative to the heart:

1. Arteries direct blood away from the heart.

2. Veins direct blood to the heart.

Additionally, blood vessels closer to the heart are bigger in diameter and size. The biggest and most prominent blood vessels are the aorta as well as both superior and inferior vena cava. From the aorta on arteries are getting smaller and less elastic. The reason for the aorta's elasticity is that it acts as a kind of balloon, continuously contracting during the hearts relaxation phase providing a positive pressure within the circulatory system.

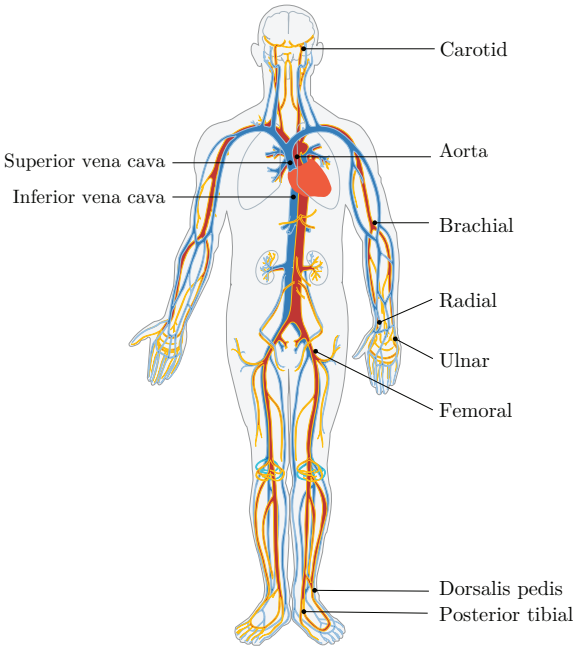


Figure 2.1: Schematic of the human circulatory system.[7]

2.1.3 The Human Heart

The heart is a hollow muscular organ located in the middle of the inside of the chest cavity. The main purpose of this specialized muscle is to pump blood through the vascular system consisting of veins, arteries, and capillaries. It has usually the size of a clenched fist of the carrier and is slightly tilted front-right with the apex pointing frontwards. It is to two thirds located on the left side of the imaginary middle line and only to one third to the right side. The heart has a septum in the middle to separate oxygen low from oxygenated blood dividing it into a left and right half respectively. Each of these halves can be further subdivided into two chambers: an atrium and ventricle. The atria collect blood from the feeding vessels and inject it into the ventricles. From here blood is pumped into the circulatory system of the body. Hence, the ventricles are not only larger but also stronger in order to produce a larger force. This is needed for two things: On the one hand for overcoming the persisting pressure inside the vascular system. On the other hand, and more importantly, to produce a large enough pressure to provide all organs with sufficient blood flow.

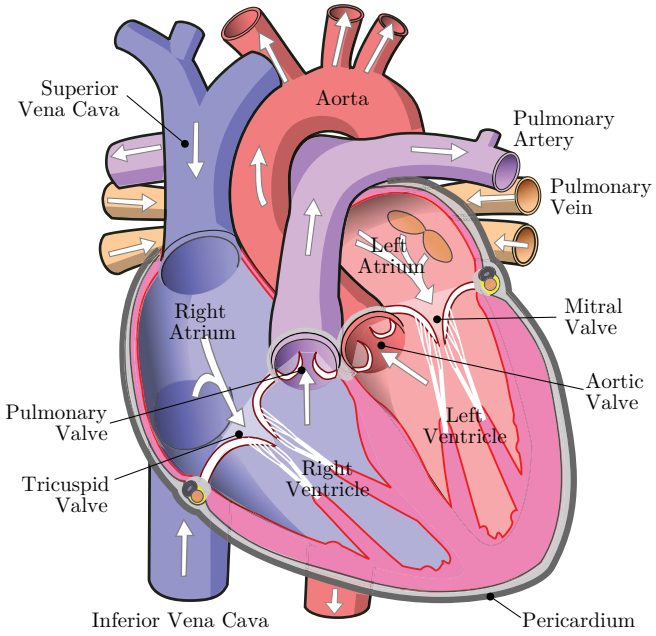


Figure 2.2: Schematic of the human heart. The direction of the blood flow is indicated by white arrows. [8] The heart can be divided into two halves each with two chambers, one atrium and one ventricle. Four valves prevent blood from flowing back into the heart.

As with every muscle, the heart is innervated by electric impulses. Mainly responsible for this clock generation is a small bundle of nervous cells located inside the left atrium called *sino-atrial node*. Therefore, sometimes a normal ECG curve is referred to as *sine rhythm* and will be explained in more detail later in section 2.2.

While the sino-atrial node is the primary pacemaker, the whole heart is permeated by nerves of the autonomous nervous system. These nerves have a significant influence on the pace, contraction strength and the so-called 'refraction time'.

2.2 Introduction to Electrocardiogram

The electrocardiography derives voltage differences from the body surface during cardiac actions and plots it over time, the ECG. Changes in height, form, and direction of the curve are a direct representation of the myocardial strength as well as direction of excitation propagation. Moreover, it represents the sum of millions of single cardiac muscle current flows. With the ECG, several examinations and information gains are possible e.g. heart rhythm and frequency, disturbances in impulse formation, propagation, and direction, malfunctions within the electrolytes distribution as well as drug influences (e.g. digoxin, a heart medication used for treating arrhythmia).

2.2.1 Cardiac Action Potential

Various ions are dissolved within cells and the blood, mainly sodium (Na^+), potassium (K^+), calcium (Ca^{2+}), and chloride (Cl^-). Over the cell membrane several passages for ions are located which allow those ions to flow either in or out. As a result of these ions flowing an electric potential between inside and outside the cell is created, the **T**rans**m**embrane **P**otential (TMP). The different proteins responsible for this transport are collectively called ion channels. The sodium-potassium pump uses **A**denosine **t**riphos**p**hate (ATP) to move three Na^+ out and two K^+ ions into the cell. Different types of ion channels in various parts of the heart result in different action potential characteristics as shown in the left part of 2.4. Matching the ECG with the underlying physiological reactions, five different phases can be distinguished. [9]

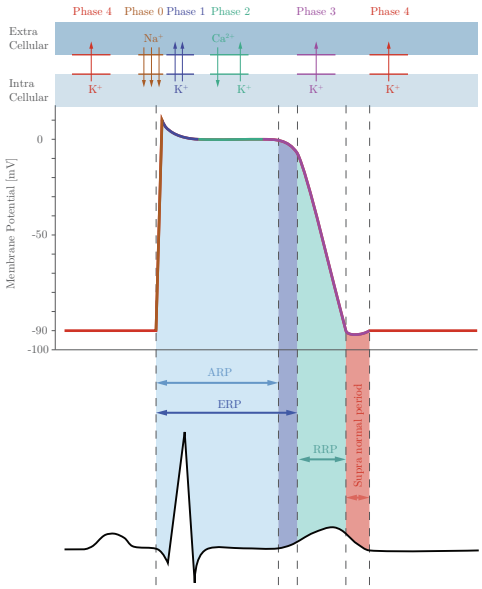


Figure 2.3: Cardiac action potential with refractory phases and correlation to an ECG. [9] Abbreviations: ARP: absolute refractory phase; ERP: effective refractory phase; RRP: relative refractory phase.

Phase 0: Rapid Depolarization Upon receiving the stimulus sodium passageways open and Na^+ rapidly flows into the cell resulting in a quick positive shift of the membrane potential. This rapid depolarization is central for the propagation of impulses through the heart muscle. [10]

Phase 1: Rapid Repolarization Upon reaching a certain potential the sodium channels close and potassium and chloride channels open resulting in this short quick drop in transmembrane potential.

Phase 2: Plateau Phase After this initial drop in membrane potential it is followed by an influx of Ca^{2+} ions. With the continuing K^+ leaking outwards the transmembrane potential is stabilized and kept just below $0\,\mathrm{mV}$ through the plateau phase.

Phase 3: Rapid Depolarization Eventually Ca^{2+} channels begin to gradually close. When all Ca^{2+} channels are closed only K^+ ion channels remain open. After a small undershoot, the supra normal period, resting potential is restored.

Phase 4: Resting Potential In normally working myocardial cells the resting potential is stable at $\approx -90\,\mathrm{mV}$. A continuous outflow of K^+ ions causes the transmembrane potential to stay stable.
With the persistent in- and outflow of those ions being perfectly balanced the membrane potential returns to the resting potential of $-90\,\mathrm{mV}$. During the polarization of cells they are resistant against new or other impulses. This time frame is referred to as refractory period with different degrees. Within the heart muscle this serves several purposes. Most dominantly to allow the ventricle to not only contract but also completely empty itself before contracting again. [9] These phases require a correctly timed stimulus for a coordinated excitation causing a normal contraction.

Absolute refractory period (ARP):	cells are completely unexcitable, no matter the stimulus
Effective refractory period (ERP):	stimulus may cause cells to depolarize, but no action potential is propagated
Relative refractory period (RRP):	greater than normal stimulus will cause cells to depolarize and will propagate an action potential
Supranormal period:	cells are hypersensitive and a smaller than usual stimulus will depolarize cells, cells are particularly susceptible to arrhythmias

Table 2.2: Definitions of different refractory periods

2.2.2 Cardiac Conduction

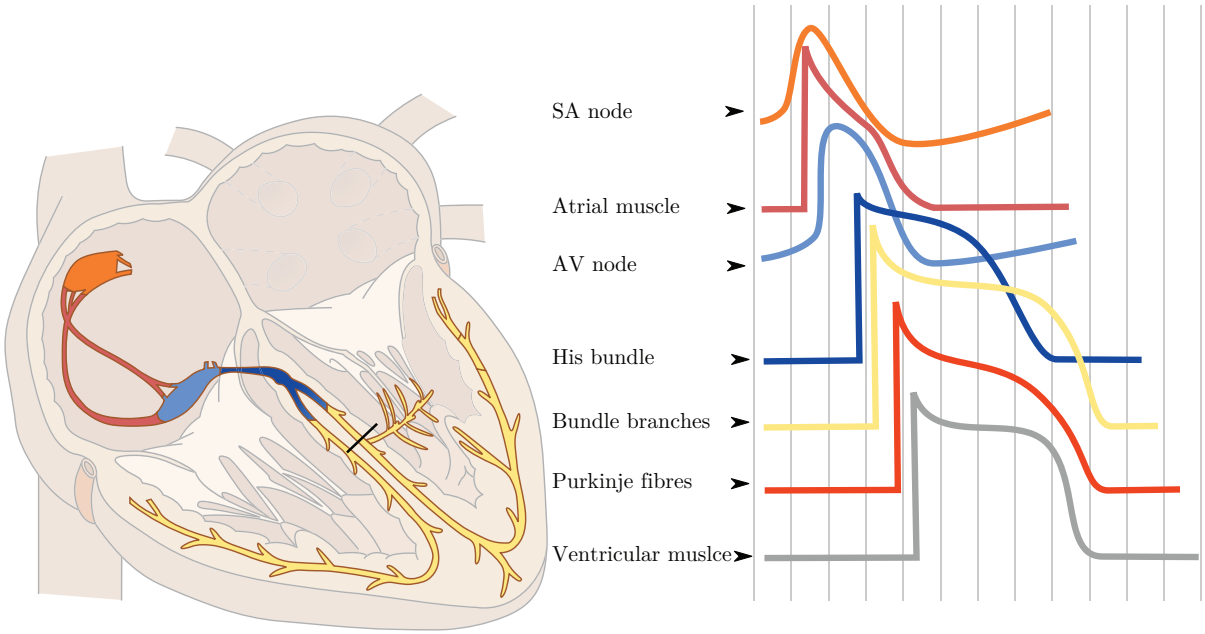


Figure 2.4: Heart conduction system. On the left, a coloured schematic of the heart is shown, on the right a diagram with time-resolved signals of different parts of the heart. Starting with the sino-atrial node in orange, the signal is further propagated to the atrial muscle in red, then the atrio-ventricular node in light blue, the His bundle in dark blue and the bundle branches in yellow. The Purkinje fibres are depicted with the dark orange stroke as they connect to the ventricular muscles which are not highlighted in the schematic. [11]

Sinus Node The sinus node is located in the right atrium and with just $10\,\mathrm{mm}$ to $20\,\mathrm{mm}$ in length and $2\,\mathrm{mm}$ to $3\,\mathrm{mm}$ width it consists of various specialized heart muscle cells. The predominant cell type is the pacemaker cell. These cells are able to autonomously excite and discharge synchronously due to a mutual entrainment. This makes the sinus node the primary pacemaker for the heart giving a pace of $60\,\mathrm{min}^{-1}$ to $80\,\mathrm{min}^{-1}$.

Atrium Coming from the sinus node the electric impulse is conducted through the atrium leading to the AV node. Research indicates that due to their anatomical structure three main pathways lead the impulse through the atrium. This ensures that all areas of the atrium get innervated.

Atrio-ventricular (AV) Node Lying in the right atrial myocardium the AV node connects partly to the bundle of His. Mainly connecting the atrium with the conduction system of the ventricle it is also responsible for imposing a slight delay of the signal. Therefore, shielding the ventricles from the atrium makes it the gatekeeper for conducting electrical impulses. Moreover, the AV node can also work as a pacemaker e.g. if the sinus node does not work anymore. This frequency is just about $40\,\mathrm{min}^{-1}$.

Bundle of His This bundle starts from the AV node and penetrates through the ventile level down to the ventricles connecting to both bundle branches.

Bundle Branches The bundle branches start right after the bundle of His moving straight down along the heart septum up to the tip of the heart. From here two major branches separate to the left and right ventricle.

Purkinje Fibres and Ventricle Purkinje fibres further separate into a fine network throughout the ventricle starting from the bundle branches. Purkinje fibres are the tertiary impulse generator with a very low frequency of about $20\,\mathrm{min}^{-1}$ to $35\,\mathrm{min}^{-1}$. Hence, the sinus node is the uncontested primary pacemaker.

2.2.3 Electrocardiogram

2.2.3.0.1 Normal Rhythm Modern ECG measurements have been developed and improved significantly compared to first discoveries by Eindhoven, see Figure 2.6 a. Nowadays, small sticky plastic electrodes are used that have a small amount of conductive gel on them. The aim is to reduce impedance and resistance for a better derivation of the electric activity on the skin.

In principle, it does not matter from which muscle electrical activity is measured. Depending on where electrodes are positioned, and if necessary filtered, the electric activity of any muscle can be measured. When performing such measurement on any skeletal muscle this would be referred to as EMG. Where the bigger the muscle is the better that electric activity can be recorded as more muscle tissue will lead to more electric activity.

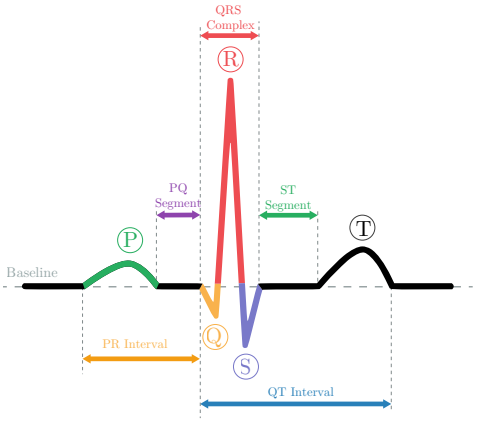


Figure 2.5: Sinus rhythm with waves and segments highlighted. [12]

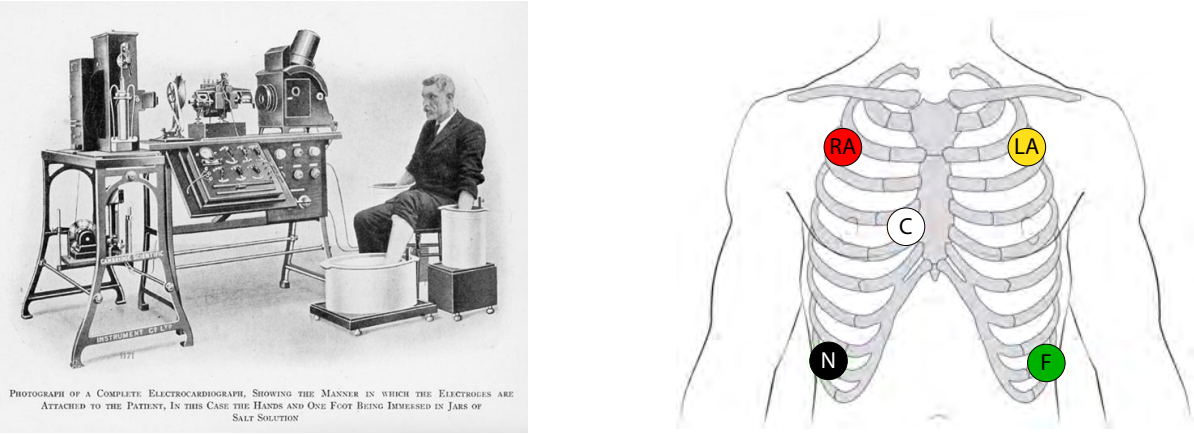
A positive polarization towards the positive lead will be seen as positive deflection and vice versa. Thus, vectors of the depolarizing and repolarizing dipole cells are recorded, consisting each of a pair of charges -+ and +-, respectively. These dipoles only exist during depolarizing and repolarizing as the charge progresses through the cells but not during rest.

A normal rhythm consists of a unique pattern with several peaks and waves which are labeled from P to T or sometimes U/J in alphabetic order. These peaks and waves can be attributed to physiological processes:

P-Wave:	Depolarization of the atrium
PQ-Time:	Complete polarization of the atrium and translation by the AV node
Q-Peak:	Depolarization of a small part of the upper ventricle septum
QRS-Complex:	Complex of three consecutive dipole vectors of the ventricle
R-Peak:	First large positive peak
S-Peak:	First negative peak after first positive peak
T-Wave:	Repolarization of the ventricle
T-P Time:	Time in which the heart muscle is relaxed

The distance from one to the next R-Peak is used to calculate the heart rate and is also called RR interval or **I**nter-**B**eat-**I**nterval (IBI). This allows to calculate a momentary heart rate compared to an averaged heart rate, e.g. by counting pulse beats for a certain time period. Moreover, RR intervals are highly responsive to the vegetative nervous system, precisely to the sympathetic and parasympathetic systems.

2.2.3.1 Electrodes



PHOTOGRAPH OF A COMPLETE ELECTROCARDIOGRAPH, SHOWING THE MANNER IN WHICH THE ELECTRODES ARE ATTACHED TO THE PATIENT, IN THIS CASE THE HANDS AND ONE FOOT BEING IMMERSED IN JARS OF SALT SOLUTION

(a) Historic ECG Machine (1911)

(b) 5-Lead ECG electrode positioning

Figure 2.6: Development of ECG machines. a: Historic ECG Machine built in 1911 by the Cambridge Scientific Instrument Company to measure human electrocardiogram according to the standards developed by Einthoven. [13] b: 5-Lead ECG electrode positioning. Colors are adapted to the ECG Systems color code.[14]

For medical diagnosis this practice has been developed since late 1800's. When taking an electrocardiogram at least two electrodes are used to conduct the electric impulse from skin to a wire. Einthoven proposed first putting both arms in a bucket with saline water connected with wires and measured the difference between these two. This is now known as Lead I or Einthoven 1. Moreover, a lead is a pair of two electrodes which are connected either as positive or negative electrode.
For Einthoven 1 this is typically expressed as:

$$I = LA - RA \tag{2.1}$$

Here, 'LA' and 'RA' stand for 'left arm' and 'right arm', respectively.
Einthoven continued his research and did the same with the left leg and the right arm to create Lead II and with the left leg and the left arm to create Lead III. [15]

$$II = LL - RA \tag{2.2}$$

$$III = LL - LA \tag{2.3}$$

Einthoven continued his research and did the same procedure with the left leg to create Lead II and Lead III with the right arm and left arm respectively. [15] These three leads are referred to as limb leads. Subsequently, he also proposed the Einthoven's triangle which hold's up its value until today.
Moreover, the placement of the electrode is significant for conducting impulses. Hence, further projections and leads have been developed.

2.2.3.1.1 Augmented Limb Leads Goldberger used the same electrode positions, but combined inputs. By taking a combination of two electrodes an artificial or augmented electrode is created. This augmented lead is then subtracted from the positive electrode. Considering the augmented vector right (aVR) the positive electrode is placed on the right arm (RA) and the average of the left arm (LA) and left leg (LL) electrode is subtracted.

$$aVL = RA - \frac{1}{2}(LA + LL) \tag{2.4}$$

$$aVR = LA - \frac{1}{2}(RA + LL) \tag{2.5}$$

$$aVF = LL - \frac{1}{2}(RA + LA) \tag{2.6}$$

2.2.3.1.2 Precordial Leads Further projections of the heart can be obtained with precordial leads. Barnes, Pardee, White et al. proposed six electrodes on the front chest at specific positions. These work as positive electrode while the Wilson central lead is used as negative electrode. The Wilson central lead is calculated by the average of all three limb electrodes RA, LA, and LL.

$$V_W = \frac{1}{3}(RA + LA + LL) \tag{2.7}$$

The benefit of those additional leads is to get more projections of different angles of the heart. As the actual electric activity of the heart is a three-dimensional vector all leads are just two-dimensional projections. With some leads specific areas of the heart can or cannot be seen which is why more leads and projects help analysing more up to the whole heart. For instance when only deriving an Einthoven II lead the electric activity of the cardiac posterior wall is not represented at all.

2.3 Stress and Workload

While these two terms indeed go hand in hand, they are often confused, mixed up or understood differently. In order to lay out a common ground, this section looks into different aspects of these terms, pointing out their definitions and where they are independent of each other.

2.3.0.1 Stress

The first time the term stress was used in a medical relation to the human body was by Hans Selye who discribed it as:

'nonspecific response of the body to any demand' [18]

Stress nowadays is the general term for reactions of the human organism to psychological or physical demands. [19] However, this term is relatively vague as it either refers to a stressor or a stress reaction. This duality is taken into consideration when assuming that stress is the disturbance of an equilibrium between the environmental requirements and individual response capabilities. Clinically, it is observable that stressors imposing stress activate the sympathetic system.
There are several ways to classify and specify stress. The first and most obvious distinctions would be either by duration or intensity. Short term stress is short in duration and has clearly defined beginning and ending points. Long term stress is a continuous persisting stimulus which is long-lasting and can become chronic. [20]
In terms of intensity, a division can be made between micro and macro stress. While micro stresses are rather smaller conflicts macro stress is more fundamental. For instance the sudden loss of work place, or a decease of a person could be perceived as macro stress. [20]
Furthermore, the quality of stress can be either perceived as positive (eustress) or negative (distress). Distress usually results from a certain amount of stress that the individual cannot cope or handle. Moreover, this can also occur when persons have no coping strategies for certain stressors. This is the unhealthy form of stress which can lead to psychiatric disorders in the long term. On the contrary, eustress is a rather short demand that can be handled. This kind of stress may also result in an increase in performance. [20]
Finally, stress can also be differentiated according to the extent to which it affects people. Considering that it can be experienced by either one individual or as a collective. For instance, experience of violence might just affect one individual, thus be categorized as individual stress. On the other hand, the threat of a crashing airplane might be experienced collectively by all individuals on board. [20]

2.3.0.2 Workload

Workload can be understood as the amount of work a person has to perform in a certain task. [21] While taskload simply represents the number of tasks to fulfill, workload is more a consequence of those tasks. Anyhow, this definition oversimplifies the term and does not take into consideration internal demands which also draw from operator's resources. In terms of human factors, putting the operator and their resources into focus, a more appropriate definition would be: "Workload is the demand placed on an operator's mental resources used for attention, perception, reasonable decision-making and action." [22]

Furthermore, the level needed for a certain task might exceed the available human resources as they are not disposable infinitely. Therefore, workload can be considered as the ratio of required task resources to available resources. Moreover, workload is not only perceived individually, but also it can vary over time. [22] Hence, stress has an influence on the resources available to an operator. This might be in either a positive way, enhancing and improving their performance or negative way further reducing available resources. Despite the close connection between those two aspects, workload considers a different and broader aspect of human capabilities.

2.3.1 Distinction between Stress and Workload

Stress and workload are closely related, but they have different causes and effects. Stress is triggered by a perceived threat or challenge, which can be anything from a difficult work project to a personal relationship problem. Stress activates the body's 'fight or flight' response, caused by the sympathetic and parasympathetic nervous systems which release hormones such as epinephrine and cortisol. These hormones cause physical symptoms such as an increased heart rate, muscle tension, and higher tissue perfusion. Stress can also affect a person's emotional well-being, leading to feelings of anxiety, depression, or irritability.

Workload, on the other hand, refers to the amount of work a person has to complete. Factors that can contribute to a high workload include having too many tasks to complete in a short period of time, having to work long hours, or having to multitask frequently. A high workload can make a person feel overwhelmed, but it is not yet the same as stress. A person may have a high workload and not feel stressed, or a low workload and still feel stressed.
A high workload can contribute to feelings of stress, and stress can make it more difficult to manage a high workload. For example, a person who is feeling stressed may have trouble focusing on their work, which can make it more difficult to complete their tasks. Additionally, stress can cause physical symptoms such as fatigue, which can make it even harder to work long hours. However, it is important to remember that stress and workload are two distinct concepts.
While in theory it is easy to distinguish stress from workload, during aircraft operation it is rather difficult. Therefore, for the course of this thesis the physiological reaction is what shall be detected.

2.3.2 Human Performance Envelope Model

The **H**uman **P**erformance **E**nvelope (HPE) is a concept created by a Future Sky Safety **E**uropean **U**nion (EU) project by Silvagni, Napoletano, Graziani et al. Within this project, the limits of human performance were defined within the context of aviation, specifically for pilots and flight crews. This multidimensional concept encompasses various aspects of human performance, such as cognitive abilities, physiological factors, and environmental factors. It is used to identify factors that contribute to human error on the flight deck and to assess the impact of new technologies and procedures on pilot and flight crew performance. These aspects are typically measured using a combination of objective and subjective methods, such as physiological markers, performance tests, and self-report measures.
Over the course of the Future Sky Safety project it was investigated which of those are the most impactful for human performance. Nine major factors were found:

1. Attention
2. Situation Awareness
3. Vigilance
4. Teamwork
5. Workload
6. Communication
7. Trust
8. Fatigue
9. Stress

More broadly the HPE can be divided into four categories: physical, cognitive, organizational, and environmental. Physical factors include things like fatigue, injury, and illness, which could impact pilots' ability to fly. Cognitive factors include things like attention, memory, and decision-making, which are crucial for pilots to make quick and accurate decisions during flight. Organizational factors include things like communication, teamwork, and training, which are vital for the safety and efficiency of flight operations. Environmental factors include things like lighting, noise, and temperature, which could affect pilots' perception and behavior.
One common way to measure cognitive abilities within the HPE is through the use of cognitive tests and assessments. These can include measures of attention, memory, and decision-making, as well as more complex tasks such as situation awareness and workload. These tests are often administered before and after a flight or simulated flight to assess changes in cognitive performance over time.
Physiological factors within the HPE are typically measured using objective markers such as heart rate, blood pressure, and brain activity. These markers can provide information about an individual's level of stress, fatigue, or workload. Additionally, self-report measures such as questionnaires and interviews can be used to assess an individual's subjective experience of these factors.

Figure 2.7: Spiderweb representation of each HPE factor. In this representation the optimal value is shown with the green line whereas a degradation is represented in red. [23]

Environmental factors within the HPE include factors such as temperature, humidity, noise, and vibration. These factors can be measured directly using measurement equipment for these physical parameters such as thermometers, accelerometers, or microphones.
In terms of the scale, the measurement values can vary depending on the specific aspect of performance being assessed. For example, cognitive performance may be measured on a scale from poor to excellent, while physiological markers may be measured in numerical values (e.g. heart rate in beats per minute).
One of the benefits of using the HPE is that it can help researchers and organizations to identify and mitigate potential sources of human error in the cockpit. By understanding the factors that contribute to flight crew's performance, manufacturers can develop strategies to improve upon these factors. This can include things like developing better training programs, improving communication systems, or designing flight decks to be more user focused.
However, it can be difficult to quantify the impact of different factors on flight crew performance. Additionally, the HPE is not always applicable to all situations or tasks. Therefore, specific measurements need to be considered whether they are appropriate for their specific use case. Overall, the HPE is a useful framework for understanding and improving pilot and flight crew performance.
In summary, the HPE encompasses various aspects of human performance, including cognitive abilities, physiological factors, and environmental factors. These aspects are typically measured using a combination of objective and subjective methods, such as physiological markers, performance tests, and self-report measures. The measurement values can vary depending on the specific aspect of performance being assessed. The HPE is a dynamic concept and the measurement values may vary depending on the context and specific situation.

2.4 Introduction to Machine Learning and Deep Learning

Machine learning is a subfield of artificial intelligence that deals with the development of algorithms and statistical models that enable computers to improve their performance on a specific task through experience. Instead of writing explicit instructions for the computer to perform a task, the program is trained on large amounts of relevant data, where it automatically learns to perform the task through its own experience. The goal is to develop models that can make accurate predictions on unseen and new data.
Machine learning algorithms can be broadly categorized into three types: supervised, unsupervised, and reinforcement learning. Supervised learning involves learning the mapping between input features and output labels based on a labeled training dataset, with the aim of making accurate predictions on unseen data. Unsupervised learning, on the other hand, involves learning patterns in data without labeled outputs, with techniques such as clustering and dimensionality reduction being popular in this field. Reinforcement learning involves learning through trial and error by taking actions in an environment to maximize a reward signal, and it is used in areas such as robotics and video game AI.
Deep learning is a subfield of machine learning that involves training artificial neural networks with multiple hidden layers. This type of learning has achieved state-of-the-art results in many

tasks, including image and speech recognition, natural language processing, and game playing. **C**onvolutional **N**eural **N**etwork (CNN) and **R**ecurrent **N**eural **N**etwork (RNN) are two popular types among deep neural networks.

CNNs are specifically designed for image classification and object detection tasks. They consist of multiple layers, including convolutional layers, activation layers, and pooling layers. In a convolutional layer, the input image is convolved with filters to extract features from the image. The activation layer then applies a non-linear function, such as a **Re**ctified **L**inear **u**nit (ReLu), to introduce non-linearity into the network. The pooling layer down-samples the output from the activation layer to reduce the dimensionality of the feature maps and reduce computation. The features extracted by the convolutional layers are then passed through fully connected layers, where the final prediction is made. The weights in the network are learned during training using a labeled dataset.

RNN are designed to process sequences of data, such as time series data or natural language sentences. Unlike feedforward neural networks, which only take into account the current input, RNN are able to model temporal dependencies by allowing information to persist across time steps.

RNNs consist of a hidden state that is updated at each time step based on the current input and the previous hidden state. This allows the network to maintain information about the sequence over time, which is crucial for tasks such as predicting future values in a time-series or recognizing patterns in speech signals.

The hidden state can be thought of as a summary of all previous inputs up to the current time step. In practice, this hidden state is represented by a vector, which is updated at each time step using a set of weights and biases, known as the RNN parameters. These parameters are learned through training with a labeled dataset, where the objective is to minimize the difference between the predicted outputs and the actual labels.

There are different variations of RNNs, such as LSTM networks and **G**ated **R**ecurrent **U**nit (GRU)s. These variations introduce gating mechanisms that allow the network to better preserve important information over time, resulting in improved performance on a variety of sequential processing tasks.

2.5 Activation Functions

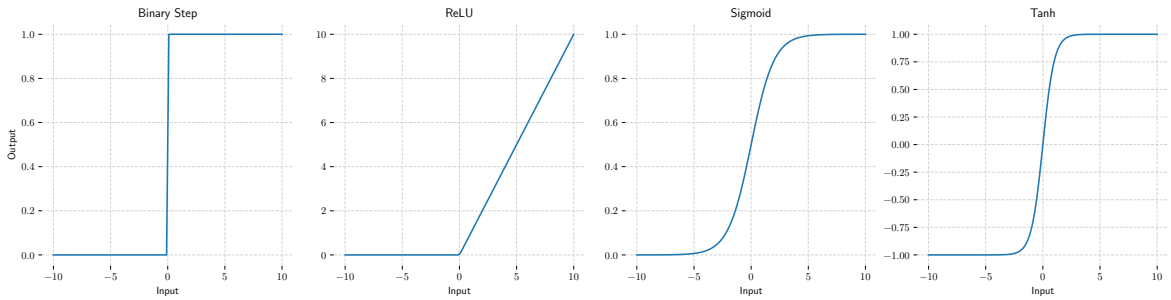


Figure 2.8: Examples of various activation functions.

The core of every neural network are the so-called neurons. They try to mimic the behaviour of a nerve cell by taking several inputs, applying some sort of activation to it and then transporting a common signal to the next nervous cell. In machine learning this activation is done by mathematical functions of different variations. They are used inside each hidden layer cell in order to apply a non-linear complexity. [24, p. 84] The, probably, simplest one would be a binary step function where the input value is compared to a threshold value and is then categorized on either side of it.

Binary Step

$$f(x) = \begin{cases} 0 & \text{for} \quad x < 0 \\ 1 & \text{for} \quad x \geq 0 \end{cases} \tag{2.8}$$

The case that includes $x = 0$ can be chosen freely.
A quite common activation function is ReLu. The ReLu function is a popular activation function that returns the input if it is positive, and 0 otherwise. ReLu is a simple and computationally efficient function that has been found to work well in many deep learning applications. [24] Moreover, ReLu uses a linear function which is the cut-off below 0.

Rectified Linear Unit

$$z = w \cdot x + b \tag{2.9}$$

$$f(x) = \text{ReLU}(x) = \max(0, z) \tag{2.10}$$

Where w is usually referred to as weight and b as bias.
The sigmoid function is a commonly used activation function that maps the input to a value between 0 and 1. It has a characteristic S-shaped curve (see Figure 2.8), which makes it useful for problems where the output needs to be a probability or a value between 0 and 1. The sigmoid function is often used in binary classification problems, where the output of the model needs to be a probability that a given example belongs to one of two classes.

Sigmoid

$$f(x) = \sigma(x) = \frac{1}{1 + e^{-x}} \tag{2.11}$$

The hyperbolic tangent function (tanh) is similar to the sigmoid function, but maps the input to a value between -1 and 1.

Hyperbolic Tangens

$$f(x) = \tanh(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \tag{2.12}$$

It has a similar S-shaped curve as the sigmoid function, but its output is centered at 0. tanh is commonly used as an activation function in the hidden layers of neural networks. However, tanh can be useful in some cases where the output of the model needs to be a value between -1 and 1, or when the input to the function is zero-centered.

2.6 Optimizers for Machine Learining

In deep learning, optimization algorithms are used to update the weights and biases of neural networks during training. The goal of these algorithms is to minimize the difference between the predicted outputs of the neural network and the actual outputs.
One commonly used optimizer is **S**toachstic **G**radient **D**escent (SGD). SGD works by computing the gradient of the loss function with respect to the model parameters (weights and biases) for a randomly selected subset of the training data. It then updates the model parameters by taking a small step in the direction of the negative gradient. [24, p. 11f] The size of the step is controlled by a hyperparameter called the learning rate l_r.

SGD is effective for optimizing many neural network architectures, but it has some limitations. For example, it can get stuck in local minima and can be slow to converge to the global minimum of the loss function. [24, p. 13] To address these issues, researchers have developed more advanced optimization algorithms, including Adam.
Adam stands for adaptive moment estimation, and it is an extension of stochastic gradient descent. Like SGD, Adam computes the gradients of the loss function with respect to the model parameters. However, Adam also keeps track of past gradients and past squared gradients, which allows it to adapt the learning rate for each parameter individually based on the historical behaviour of the gradients. Adam is known to be an efficient optimizer that can converge quickly to a good solution. It is also less sensitive to hyperparameter tuning than other optimizers like SGD. [25]

2.7 Description of Long-Short-Term-Memory Systems

LSTM networks are a type of RNN, which is a class of artificial neural network used to process sequential data. A sequence of data is simply a set of data points that have an order, such as time-series data (e.g. stock prices over time), or sequences of words in a sentence. LSTMs are different from other types of machine learning algorithms because they have a type of memory built into them that allows them to remember information from long ago, as well as more recent information. Therefore, allowing the network to better process long sequences of data and maintain long-term dependencies. [26]
An LSTM network consists of memory cells, input gates, forget gates, and output gates. The memory cells are the core component of the network and are responsible for retaining information over long periods of time. The input, forget, and output gates control the flow of information into and out of the memory cells.
At each time step in a sequence, the input to the LSTM network is processed and used to update the values of the memory cells and gates. The input is first transformed into three different vectors, which are used to update the values of the input, forget, and output gates. These vectors are computed using weight matrices and biases, which are learned during the training process. [26]

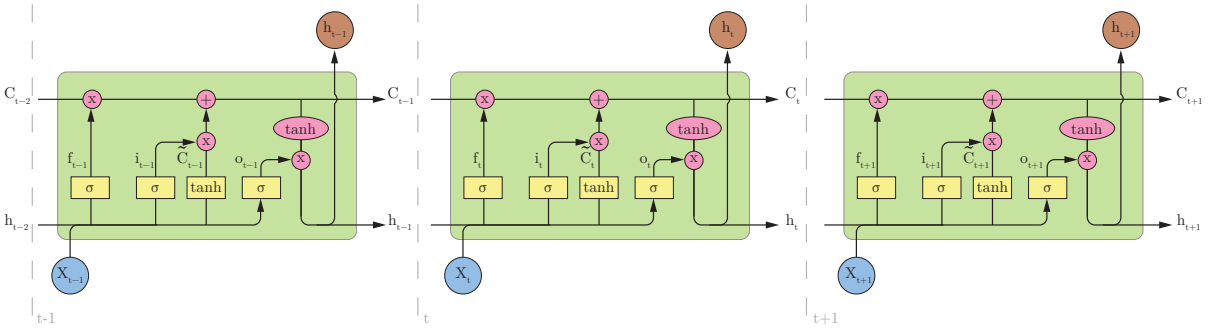


Figure 2.9: General structure of a Long Short-Term Memory cell over three time steps, $t-1$, t and $t+1$. [26] The following description applies to the state in the middle, at time step t. The cell is represented by a green box and takes as inputs the previous cell state C_{t-1}, the previous hidden state h_{t-1} and the input value x_t (blue circle). The computation of the current cell state C_t and the current hidden state h_t (brown circle) takes into account the result of the forget gate layer f_t, the input gate layer i_t, a candidate value $\tilde{C}_t$ and the output gate o_t.

As first step it is determined which information to keep from the previous memory cell state C_{t-1}. This is done by a sigmoid layer which takes the previous hidden state h_{t-1} and the current

input x_t as input. This sigmoid layer outputs a value between 0 and 1, where 1 represents 'completely keep' while 0 means 'completely ignore'. [26]

Forget Gate Layer

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{2.13}$$

Next it is computed what information from the current input should be added to the memory cell state. Fist of all another sigmoid layer called 'input gate layer' decides which values will be updated. In a next step using a $\tanh$ layer new candidate values that shall be added to the cell state is calculated. [26]

Candidate Values

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{2.14}$$

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \tag{2.15}$$

Then, the old cell state C_{t-1} is updated into the new cell state C_t. [26] This is done by combining the previous memory cell state with the new candidate memory cell state, weighted by the input and forget gates:

Cell State Update

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \tag{2.16}$$

Thereafter it is determined which information should be output for the current time step. As a first step it is decided which parts of the hidden state and input should be output. This is called the output gate. [26] The output gate is computed using a sigmoid activation function:

Output Gate

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{2.17}$$

The updated cell state is then pushed through a $\tanh$ layer to set those values between -1 and 1. Then it is multiplied with the previously calculated output gate to create the new hidden state. [26]

Hidden Layer

$$h_t = o_t \cdot \tanh(C_t) \tag{2.18}$$

The LSTM network uses the hidden state at each time step to make predictions. During model training this is then compared to a ground truth to calculate the error. This is done by choosing an appropriate loss function. The selection of an appropriate loss function heavily depends on both the type of output created as well as original labels type. Based on this, weights and biases in the network are updated to minimize the predicted error. This is done by choosing an appropriate optimizer.
LSTMs have been used in several applications such as the robot arm Dexterity by OpenAI or for training DeepMinds AlphaStar to play computer games. [27] [28] Also for stress detection based on ECG's this type of algorithm has been used and shown some good results.

Methodology 3

Within this chapter a comprehensive description of and justification for all relevant research design choices is provided. It is supposed to give an insight and deeper understanding for how choices were made under the given circumstances. The rational foundation, type of research as well as how data was collected, analyzed, and prepared for further usage is discussed. Two flight simulator studies were conducted. The data generated during these studies was then used for the deep learning application. The first study was focused on single pilot operation and physiological data was collected in the background. Some limitations during this study complicated certain applications of e.g. an online stress evaluation. But it demonstrated how physiological data can be collected and what stressful scenarios could be applied. The experiences with the first study were used for designing a second study. In this study the focus was on recording physiological data as well as monitoring stress levels more closely. In order to meet these demands, a new online stress assessment tool was developed based on existing workload/stress rating tools and scales.

3.1 Common Methodology

Some methods and tools haven been used in both studies, the **S**ingle **P**ilot **O**peration (SPO) and the **L**imits **o**f **H**uman **P**erformance (LoHP) study. Hence, these are explained and summarized in this section while pointing out differences and changes through these two studies.

Figure 3.1: AVES Flight Simulator at German Aerospace Center in Brunswick, Germany, Credit Photos: DLR, CC BY-NC-ND 3.0

3.1.1 AVES Simulation Environment

Located at the German Aerospace Center (DLR) in Brunswick, Germany is the **A**ir **Ve**hicle **S**imulator (AVES). It consists of an A320 and EC135 flight deck which can be either mounted on a fixed or a motion platform. Besides external studies the AVES can be used to test new flight dynamics and several other characteristics. For both studies the A320 flight deck was used. As the A320 is the most common aircraft in the world recruiting pilots would be significantly easier as for e.g. long-haul aircraft types. The AVES simulation environment allowed to modify weather as well as simulating system failures.

3.1.1.1 Simulator Data

The AVES can record flight data and save it into a comma-separated ASCII file in a format similar to **C**omma-**S**eparated **V**alues (CSV) called 'reca'. This file contains a header explaining each of the about 130 parameters. Data to be recorded can be customized based on the parameters stored inside the simulation model. For both studies the same base set of parameters had been used. For the second study those parameters had been expanded by a few to cover relevant failures e.g. engine oil temperature.

3.1.2 Data Collection Methods

Different sources recorded data during each flight simulator study. These data used different formats and had to be adapted for further processing. Some of applied methods were used for both simulator studies and are summarized in this section.

3.1.2.1 Data Analysis Methods/Techniques

For data analysis several Python libraries were used. Mainly Numpy and Pandas were used for any relevant data loading, calculations, or modification. Additionally, the **W**ave**f**orm **D**ata**b**ase (WFDB) software package for Python includes methods for analyzing ECG, EEG, and EMG waveform data. This package is capable of detecting R-peaks, calculating a momentary heart rate from these peaks, and other useful tools. [29] This toolkit allowed easy and reliable analysis of the recorded ECG signals. However, these libraries alone did not run any analysis, for each

analytical question individual Python scripts had to be programmed. Moreover, for the analysis of the LoHP study's data a small GUI tool was developed for faster data processing.

3.1.2.2 Questionnaires

During the studies several questionnaires were used to get feedback from participants. LimeSurvey was used as online platform to host these questionnaires, which can be found here `https://www.limesurvey.org/`.
For the sake of comprehending the participants background several questions related to demographics were asked with a questionnaire. These included questions regarding age, gender and nationality. Moreover, questions regarding flight hours and positions were asked.
During the SPO study questionnaires were also used to conduct the NASA **T**ask **L**oad Inde**x** (TLX), a demographic survey, as well as a paper feedback system for perceived stress.

3.1.2.3 Data Privacy

During both studies personal and physiological data have been collected. In compliance with German law all participants filled out and signed a data privacy agreement with DLR before participating in the simulator study. Data obtained during flight simulator, ECG, stress level, video and voice recordings, flight data as well as sociodemographic data were stored on DLR servers anonymously. Organizational data is only accessible to the DLR administration and the respective experiment manager. Data will be made available anonymously for further dissemination and usage. Moreover, all participants gave written consent to participate in those two studies.

3.2 Single Pilot Operation Study April 2022

In April 2022 a simulator study was conducted concerning single pilot operation. Within this study the possibility was given to measure physiological data of pilots flying. Several compromises had to be made to accommodate all needs e.g. the pilots should not be disturbed by anything that could draw their attention from the main tasks. Thus, limiting the feedback from the flight crew during flight. The complete design of the study and the scenarios has been done by Christian Booms, a Masters student in psychology from University Ulm, and is part of his Master's thesis.

3.2.1 Research Design

The main focus of this study was to analyze the capability of pilots to operate an A320 with no other pilot present. This single pilot operation was then compared to a normal flight crew with two persons. A total of 24 A320 First Officers participated in this study, where 14 flew in single pilot and 10 in dual pilot operation. The scenarios were randomized in order to mitigate fatigue effects statistically over all the participants. An overview of all participants and their order can be found in Table A.1.
To ensure comparability over the course of time no adjustments nor modifications were possible throughout the study. Moreover, a main principle was to not disturb the pilots with anything that may draw their attention from the scenario. Anyhow, in order to classify the ECG curves later, as well as building a 'ground truth' database, a feedback of the perceived stress levels of the operators was necessary.

3.2.2 Subjects

In this study 24 persons participated, separated into 14 single and 10 dual pilot operations. The whole group consisted of 22 male and only two female persons, aged 26 year to 52 year (mean:

32 year, standard deviation: 4.96 year). Every person was trained to the degree of First Officer on a A320 with an average of 2 791 h total flight hours, 2 234 h on A320, and 275.4 h flight hours within the last 12 months. All of them were native German speakers.

3.2.3 Flight Scenarios

This study was designed to research behaviour of pilots in single pilot operation. Five different scenarios were created, each a different approach to either Stuttgart or Frankfurt Airport. Moreover, for every but the first scenario, system and other failures were included to test pilots response and behaviour due to the changed cockpit situation. Additionally, most support and automation systems were not allowed to be used. In detail, the flight director and auto-thrust were switched off in order for pilots to fly so called raw-data. It was assumed that the failures in scenarios 2 to 5 imposed a relevant amount of workload to trigger stress in the participants. Hence, this study was well-suited to record physiological data during the scenarios.

3.2.3.1 Scenario 1

This scenario serves as a baseline scenario without disturbances or other issues. The pilot approaches Stuttgart Airport on runway 25 from 5 000 ft (1 524 m) and positioned north of the arrival track. They are cleared by the ATC for a visual approach and are required to fly manually without **F**light **D**irector (FD) or any other assistance. The Scenario ends after rolling out on the runway. The progress as well as exploratory data from four participants is shown in Figure 3.2. Graphics have been created using the GPS data from AVES and plotting them on OpenStreetMaps data with QGIS software.

3.2.3.2 Scenario 2

Weather 2nd Scenario

Wind: 10 kt **C**ross **W**ind **C**omponent (CWC)
METAR EDDS xx1000Z 33010KT 2500 BKN005 NSC
18/5 Q1004 NOSIG, RWY in USE 25, LOC25

The initial Position is the same as in scenario 1. The pilot is requested to fly a non precision approach to Stuttgart Airport on runway 25. In addition, there is medium wind coming from north at 10 Kts from heading 330 deg. The first failure is a cargo fire at about 2 500 ft **A**bove **G**round **L**evel (AGL) (762 m), secondly the ATC made several requests on short final, and finally on short final the PFD and ND flickered for a few seconds.

3.2.3.3 Scenario 3

Weather 3rd Scenario

Wind: 20 kt CWC and **T**ail **W**ind **C**omponent (TWC)
EDDS xx1000Z 35020KT 2500 BKN005 NSC
18/5 Q1004 NOSIG, RWY in USE 25, ILS25

The initial Position is the same as in scenario 1. The pilot is requested to fly an **I**nstrument **L**anding **S**ystem (ILS) approach to Stuttgart Airport on runway 25. In contrast, the wind was increased to be at 20 Kts from heading 350. In this scenario at 2 500 ft AGL (762 m) engine 1 stalled as first disturbance. At 1 000 ft AGL (304.8 m) it was simulated that the cabin is not yet

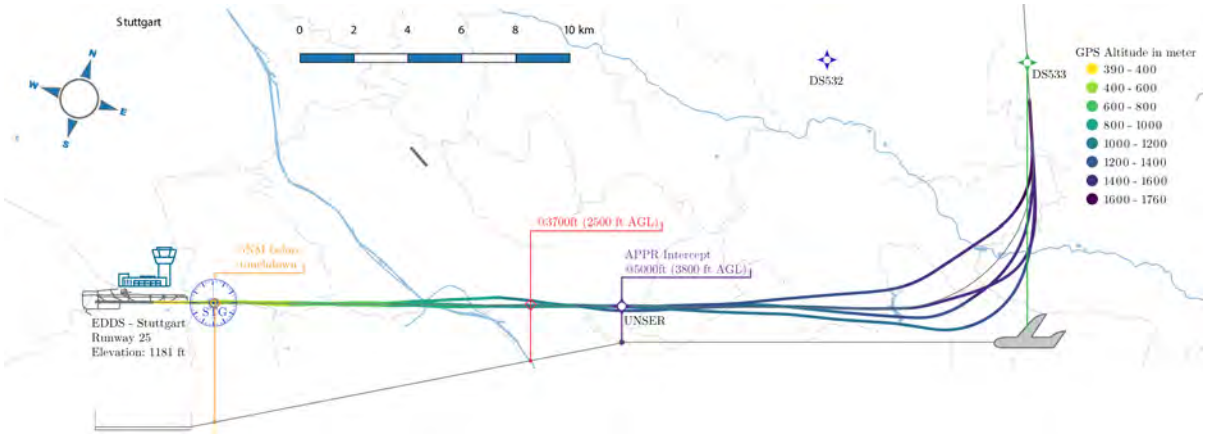


Figure 3.2: Scenario 1 is the baseline scenario without any disturbances or other issues.

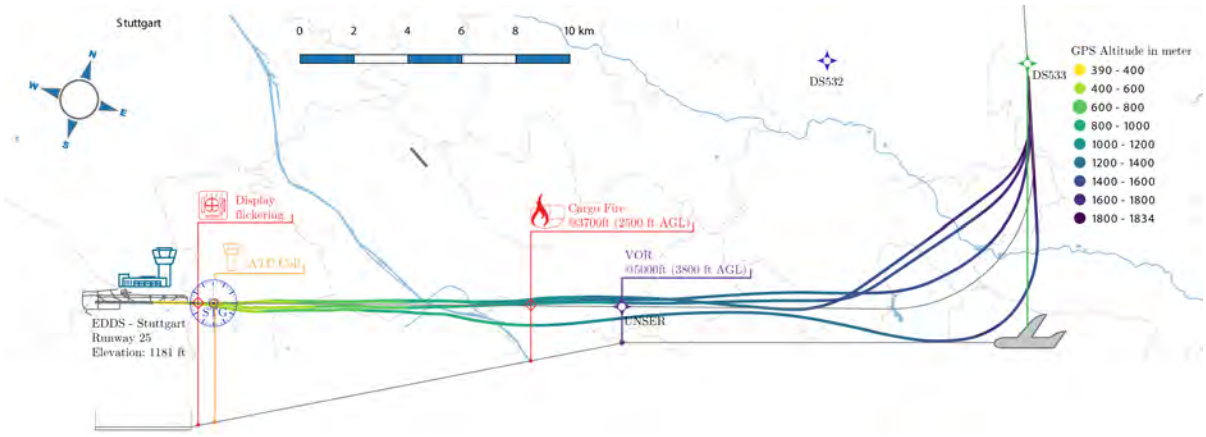


Figure 3.3: Scenario 2 starts the same as scenario 1, but introduces wind, a cargo fire, **A**ir **T**raffic **C**ontrol(ler) (ATC) requests on short final and flickering of **P**rimary **F**light **D**isplay (PFD) and **N**avigation **D**isplay (ND) on short final.

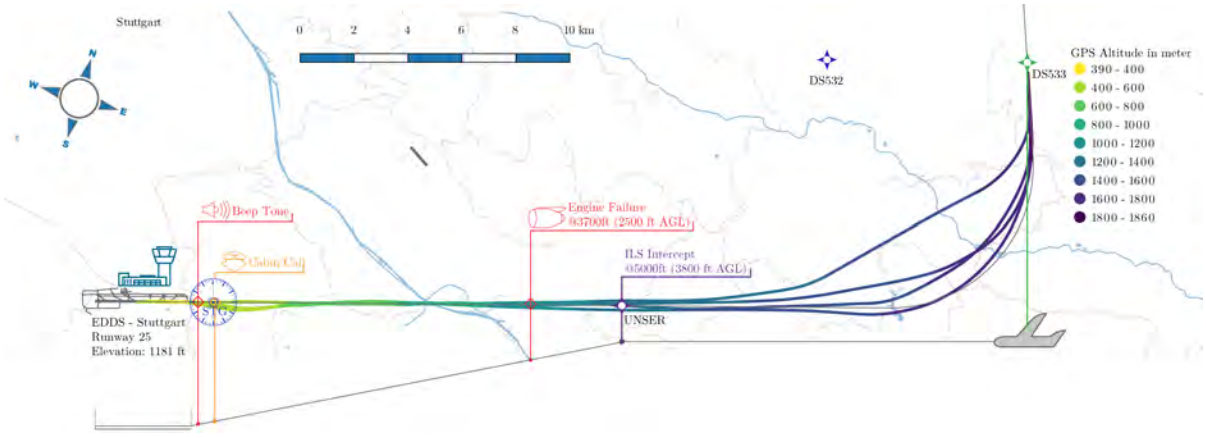


Figure 3.4: Scenario 3 starts the same as scenario 1, but introduces strong winds, an engine stall, a cabin not ready issue and a Localizer interference.

ready due to a standing passenger followed by an 'okay' from the cabin shortly after. After this a Localizer interference was simulated on the ND.

3.2.3.4 Scenario 4

Weather 4th Scenario

Wind: 30 kt CWC
EDDS xx1000Z 16030KT 2500 BKN005 NSC
18/5 Q1004 NOSIG, RWY in USE 25, LOC25

The initial Position is the same as in scenario 1. The pilot is requested to fly a non precision approach to Stuttgart Airport on runway 25. In this scenario at 2 500 ft AGL (762 m) both RA failed, which results in a rapid system degradation to Normal Law when the landing gear is extended. Flying in Normal Law is not frequently trained and changes the aircraft behaviour significantly as now the side sticks do not command attitude change, but rather require direct control with no further protections.

3.2.3.5 Scenario 5

Weather 5th Scenario

Wind: 20 kt CWC
EDDS xx1000Z 15025KT 2000 BKN005 NSC
18/5 Q1023 NOSIG, RWY in USE 25, LOC25

The initial position was about 6 miles east of Frankfurt am Main. The simulated weather was quite windy with gusts at 10 Kts in runway heading. Shortly after the go-around ATC instructed the crew to level off at 3 000 ft (914.4 m). When this altitude was reached the left engine failed. Immediately afterwards ATC requested the pilot to climb further to 4 000 ft (1 219.2 m) due to another incoming emergency aircraft. When 4 000 ft (1 219.2 m) were reached the scenario was over.

3.2.4 Data Collection Methods and Systems

During this study several physiological parameters have been measured beside the simulator data at the same time. Each part of data was recorded and stored by a different system. The simulator transmits UDP messages to synchronize with most of these systems.

3.2.4.1 Eye Tracking

Eye tracking was required by the Master's thesis of Mr. Booms researching single pilot behaviour. Although that data was not further used in this thesis it was a method applied during the simulation flights and hence be described shortly.
The eye tracking system consists of six cameras with software by DTrack2 and glasses with reflective sensor spheres by iViewETG. Software for recording and creating areas of interest was develop by the Institute of Flight Guidance's System Ergonomics department (FL-SEG) within DLR. Before recording any participants eye movement areas of interest had to be defined with a reflective sphere attached to a wooden stick. Camera alignment and glasses position need intense calibration not only beforehand but also for each participant. As each human has a different anatomical structure resulting in different eye and glasses positions in a room, calibration for each person is necessary. This was performed once when the participant was

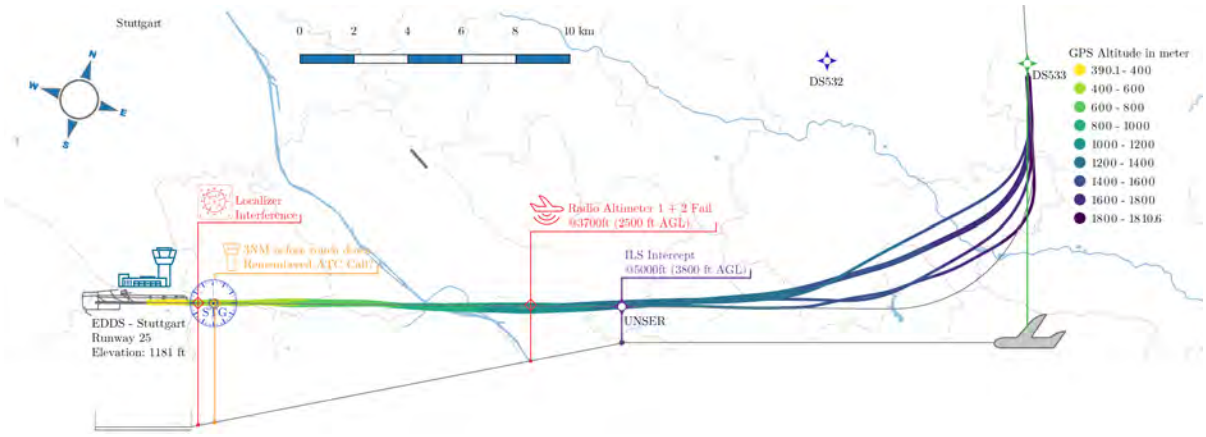


Figure 3.5: Scenario 4 starts the same as scenario 1, but both **R**adio **A**ltimeter (RA) fail on final approach. Shortly before touchdown the navigation displays flicker as an additional irritation.

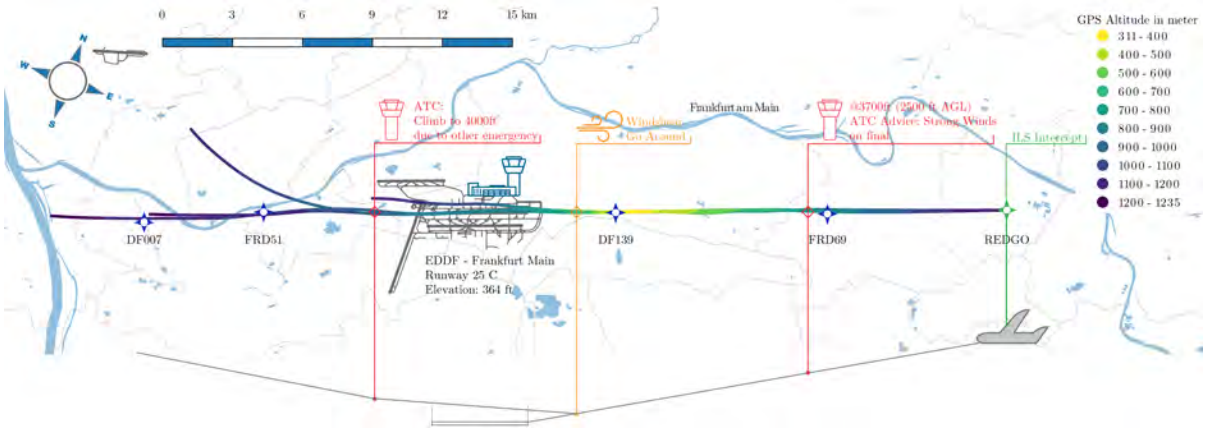


Figure 3.6: Scenario 5 starts east of Frankfurt am Main and confronts the pilots with wind, a go-around, an ATC request for levelling off, an engine failure and an ATC request for climbing.

sitting inside the simulator. During the flight participants were asked to not touch or move the glasses as this might misalign the calibration. All data were recorded on a separate laptop only used for the eye tracking system.

3.2.4.2 Capacity Heart Rate

A capacitive mat manufactured and provided by Capical GmbH was placed on the right pilot seat within the AVES. Using capacitive measurement equipment this mat recorded electric activities through clothes. The biggest benefit of this mat is the recording of ECG under otherwise barely practical circumstances, e.g. truck drivers. In a cooperation between Capical GmbH and DLR this equipment was tested for its applicability in flight deck environments. The wires of the mat were connected to a Raspberry Pi which also had been programmed by Capical GmbH. After each recording a USB drive was inserted, the Raspberry Pi restarted, and upon restart recorded files were uploaded onto the USB drive. At the same time data of the last recorded RR-interval was transmitted using a **L**ab **S**treaming **L**ayer (LSL). From outside the simulator this LSL signal was then recorded with a software originally made for EEGs. Moreover, from this position the signal was also output to a console allowing the person starting and recording the signal to see whether something is recording or not.

3.2.4.3 ECG Recording

For this study a single lead ECG system was provided. This consisted of the EG01000 ECG module manufactured by MedLab GmbH. For this single lead ECG three electrodes were placed on the chest of the participant. Namely one each on the right and left shoulder as well as one on the left upper abdomen. This allowed to derive a Lead I ECG by Einthoven. The ECG module was capable of changing recording properties. For this study it was set to have a sampling rate of $300\,\mathrm{Hz}$ and a resolution of 75 steps per $1\,\mathrm{mV}$. [30] This module was connected via USB to the same Raspberry Pi as the capacitive module. The Raspbery Pi was programmed to record all data and was automatically started and stopped by a UDP message sent by the simulator.

3.2.4.4 Stress Track

An important requirement for this study was to not disturb the pilots during flight in any way. They should be able to purely focus on the flight deck tasks. For the course of this thesis research it was important to get an as good as possible feedback of the perceived stress of the pilots. For the circumstances of this study a DIN A4 page for each scenario was designed reflecting the timeline of the scenario. After each scenario participants were handed out the respective page. They were asked to either mark with crosses or draw a continuous line of their perceived stress over the course of the scenario.

3.2.4.4.1 Data Digitization As the raw stress track data was noted on paper it needed to be digitalized. This was done simply by scanning those documents with a Sharp MX-3071 printing station. After transferring the PDF files to a computer they were manually digitalized using a vector graphics programme (Adobe Illustrator). As the templates had been created with the same software the PDF scans were overlaid and adjusted to the original lines. Then, points were created manually at regular intervals along the course. Care was taken to ensure that the points were not too far apart. The distance between the points was chosen so that a linear interpolation was a good representation. Then the file was reduced to only consist of the stress track, lines for the inner, middle and outer marker, as well as points for the boundary points of the coordinate system. This reduced file was saved as **S**calable **V**ector **G**raphics (SVG) file, which is a subclass of the E**x**tensible **M**arkup **L**anguage (XML) filetype for vector graphics,

allowing it to be read by Python scripts for merging it with the ECG curve. This process needed to be done for each of the 24 participant and each of the 5 scenarios manually.

3.3 Limits of Human Performance Study October 2022

As a result of the data evaluation of the SPO study the need to expand the data collection was identified. Therefore, the aim of this study was to not only measure physiological data such as ECG and respiration in higher quality, but also to track the momentary perceived stress of the participants more closely. Moreover, it appeared within the data that participants had experienced a higher workload but in no way reached a significant level of stress nor reached their performance limits. Hence, the scenarios needed to be modified in a way to not only increase workload but also induce stress.
For the sake of data collection the best case scenario would be flight crews under massive stress losing the required performance to maintain a safe flight. This might sound harsh or demanding for the participants, but was taken care of not to impose them to any traumatizing situations. Neither should those scenarios be so unrealistic that flight crews just lose motivation and concentration during the simulation. Another aspect which limited the design of the scenarios was the capabilities of the AVES. As this simulator is made for scientific research it is not an exact replication of an A320. With these outlines this study was further planned and defined.

3.3.1 Research Design

The aim of this study was to monitor stress in flight crews during critical and stressed situations. Scenario planning as well as construction of errors and failures were build upon conclusions of both, the Future Sky Safety Study and the Single Pilot Operation Study. Within the document 'Concept for Human Performance Envelope' [23] eleven critical situation examples were discussed. Those examples were sorted after certain HPE factors as well as their contributing situation factors. Those are for instance multiple system failures, landing in bad weather under heavy workload, and more. As a guidance these examples build the fundamental ideas to further plan scenarios upon. During the Single Pilot Operation Study five different landing scenarios were constructed by Mr. Booms. After each flight the pilot flying answered a NASA TLX questionnaire. Using this data from all flight crews helped to select suitable failures and situations in order to impose stress. In order to get a better understanding of the TLX feedback given by the flight crews violin plots for each scenario as well as operation mode were created as shown in Figure 3.7. This plot shows that the mental demand was increased in scenarios 2, 3 and 4 of the SPO Study. Another factor that might be interesting is the overall effort. An increased effort can be recognized in scenario 3 more than during the others. Therefore, the failures of these scenarios were further considered for the construction of the new scenario. During scenario 2 a cargo fire broke out, in scenario 3 an engine failed, and in scenario 4 both RA failed. A cargo fire would require an immediate landing and hence end the scenario. The goal is to have one scenario in which workload is increased over time to have sufficient time to build up a physiological reaction, because stress and the corresponding physiological reactions due to a single stressor take some time to develop and vanish again. [31]
With these inputs one scenario was created for all participants. This scenario starts on the apron at Brunswick Airport and is headed to Frankfurt am Main Airport. Flight crews were instructed that they are requested to do a ferry flight of an A320 for maintenance. Flight crews found the aircraft already started up with engines running, shortly after push back with towing equipment removed and ready for taxi. After establishing themselves on the flight deck flight crews requested taxi clearance and taxied to the runway. Shortly after lift-off the left engine, engine 1 (ENG 1), had an engine flame out. Upon solving and restarting the engine, crews were advised by the operator to continue to Frankfurt. Over the course of the flight weather

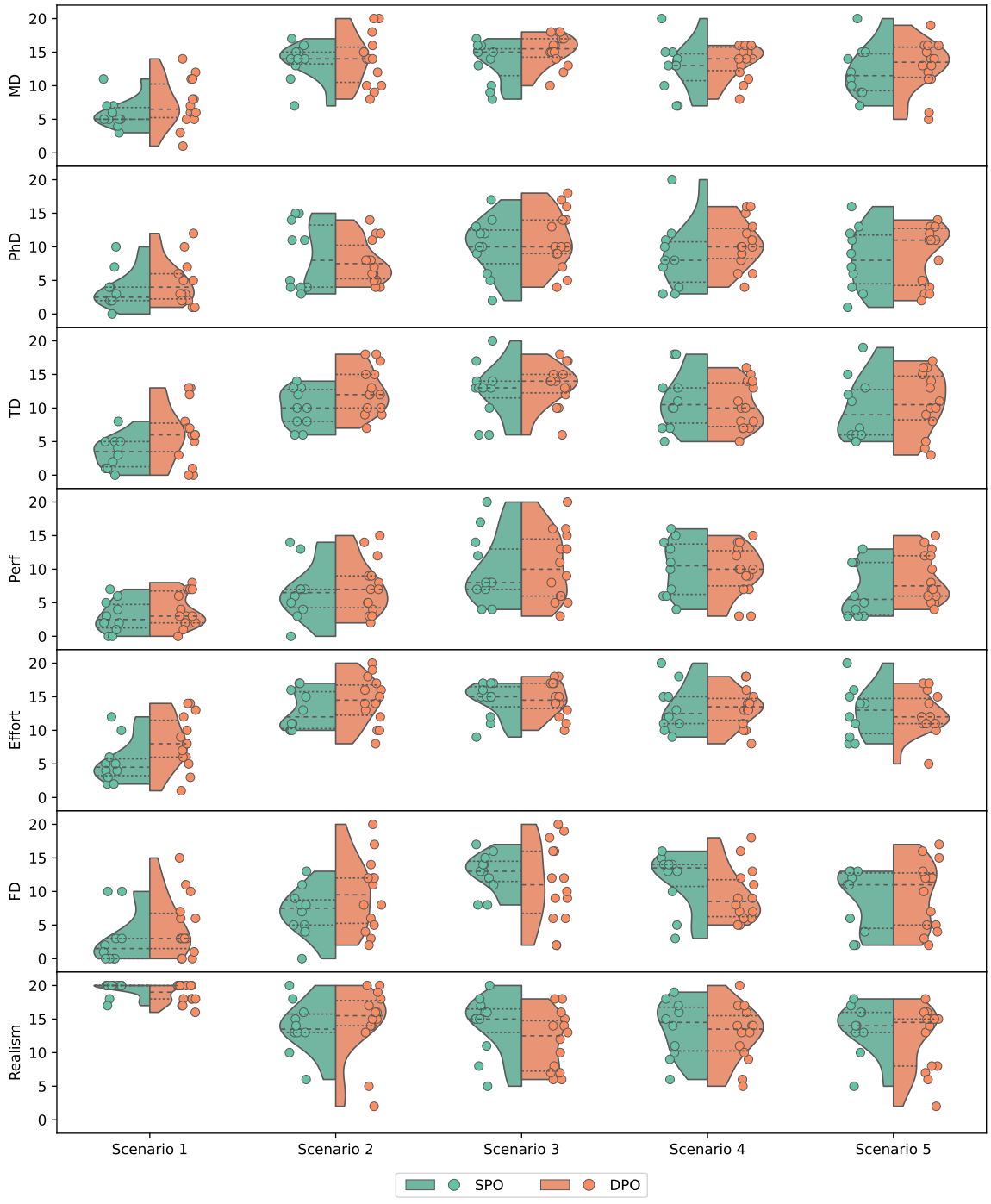


Figure 3.7: NASA TLX evaluation of the single pilot study from April 2022. The evaluation criteria of the NASA TLX as well as the evaluation on realism separated into single pilot (SPO) and dual pilot operation (DPO) are plotted for each scenario. The scattered dots represent the single values of the participants. The violin plots represent the distribution among all participants to give a quantitative overview. The dashed lines represent the median as well as the quartiles.

had been adjusted to minimal visibility as well as severe winds and gusts. Moreover, via speakers a thunderstorm was simulated to increase the noise level on the flight deck. The flight to Frankfurt continued normally with simulated ATC traffic. Shortly after leaving the top of descent the next failure was already triggered. The oil temperature of engine 2 (ENG 2) started increasing and would become critical when the engine is used in high power setting for a certain period of time. Although this failure is triggered now, it only becomes critical later in the scenario. The approach to Frankfurt Airport was a KERAX standard approach without any further failures. During the final approach on runway 25L ATC asked the flight crews to perform a swing over to runway 25C. This was done when the runway was in sight to follow the standard procedure for this and therefore increase the likelihood of the flight crew to affirm this request. At about 50 ft above ground level a wind shear and go-around warning were simulated. This forced the flight crews to perform a go-around. Shortly after gaining a positive climb rate a lightning strike was simulated which was followed by an engine stall on engine 1. Moreover, this engine stall triggered a fuel leak on the left side wing tanks which eventually lead to a loss of engine 1 later on. Before the flight crew could be instructed to do any other turns in order to realign for another landing attempt to Frankfurt they were informed that Frankfurt Airport had been closed due to bad weather. The crew was then forced to fly to the alternate which was determined to be Cologne/Bonn Airport. During reevaluating the situation flight crews also discussed other airports but weere neglected due to various reasons by ATC. At about this time the high oil temperature warning triggered as engine 1 was in idle and engine 2 in either TO/GA or Max Continuous Thrust. Closer to Cologne/Bonn Airport engine 1 shut down completely due to the fuel leak. The flight crews were directed by the ATC to the final approach to Cologne/Bonn. On final approach when the landing gear was lowered the last failure was triggered. In this case both RA failed which degrades the aircraft system mode from Normal Law all the way down to Normal Law. This is a mode that is not commonly trained by flight crews and hence is not only a surprising failure but also requires quick and fast adaptation. Upon touchdown the scenario was over and the recording was stopped. On top of these failures flight crews were asked to fly this whole flight manually without autopilot. The only assistance they had on the first part of the scenario was the flight director. But after going around in Frankfurt this was turned off as the readings became incorrect due to the simulator.

In addition to this scenario, each flight crew performed a baseline flight from Brunswick to Hamburg Airport beforehand. During this flight no failures or stressors were simulated in order to allow pilots to familiarize themselves with the simulator. Hamburg was chosen as arrival to reduce flight time as well as limiting the training to the AVES. Anyhow, data recordings were the same in both scenarios.

3.3.2 Flight Scenarios

Both scenarios started on the apron of Brunswick Airport. The aircraft was situated right after push back, with both engines running and all towing equipment removed. The flight crew still needed to get taxi and any further clearances from the tower. Using the NASA TLX data from the SPO Study as well as the results from the Future Sky Safety Human Performance Envelope Critical Situation Examples these scenarios were constructed. The aim was to create a state of maximum stress at the end of the scenario, but also making it still possible for the flight crew to land the aircraft. For both scenarios flight crews had to maintain communication with an air traffic controller which was simulated by an aerospace engineering student and private pilot.

3.3.2.1 Baseline Scenario

Flight Route Baseline Scenario

EDVE DIRBO8T **DIRBO** T902 **RARUP** RARUP2P **EDDH**

Alternate: EDDW
Cruise: FL160

Weather Baseline Scenario

Meteorological **A**erodrome **R**eport (METAR) EDVE Information K:
EDVE 081600Z 30012KT 5SM Q1004

METAR EDDH Information Q:
EDDH 081650Z AUTO 26010KT 6000 SHRA SCT037 FEW080CB 28/08 Q998 TEMPO NOSIG

METAR EDDW Information H:
EDDW 081630Z 29015KT 5SM Q1002 NOSIG

The flight plan for this scenario was to start from Brunswick Airport and heading to Hamburg Airport. One of the main purposes of this flight was to get used to the AVES simulation environment. Therefore, no failures or issues were simulated and the crew was briefed so both pilots could fly the aircraft at least once. Furthermore, no further ATC traffic was simulated during this flight. Upon request by the flight crew also directs were given by the ATC.

3.3.2.2 Stress Scenario

Flight Route Stress Scenario

EDVE NORTA8T T154 **ROBAR** T152 **KERAX** KERAX8T **EDDF**

Alternate: EDDK
Cruise: FL220

Weather Stress Scenario

METAR EDVE Information K:
EDVE 081720Z 30012KT 6000 +TSRA SCT010 SCT050CB BKN080 100/5 Q1009

METAR EDDF Information S:
EDDF 081820Z 32013G25KT 090V260 7000 +TSRA FEW010 SCT050CB BKN050 Q1016 WS ALL RWY

METAR EDDK Information Q:
EDDK 081820Z AUTO 33010G25KT 7000 -TSRA Q1010 BECMG 30014KT

Again starting at Brunswick Airport this flight was headed to Frankfurt Airport. The initial situation was the same as in the first scenario. The flight crew was briefed that this shall be a ferry flight of a faulty aircraft to maintenance and hence should not return to Brunswick

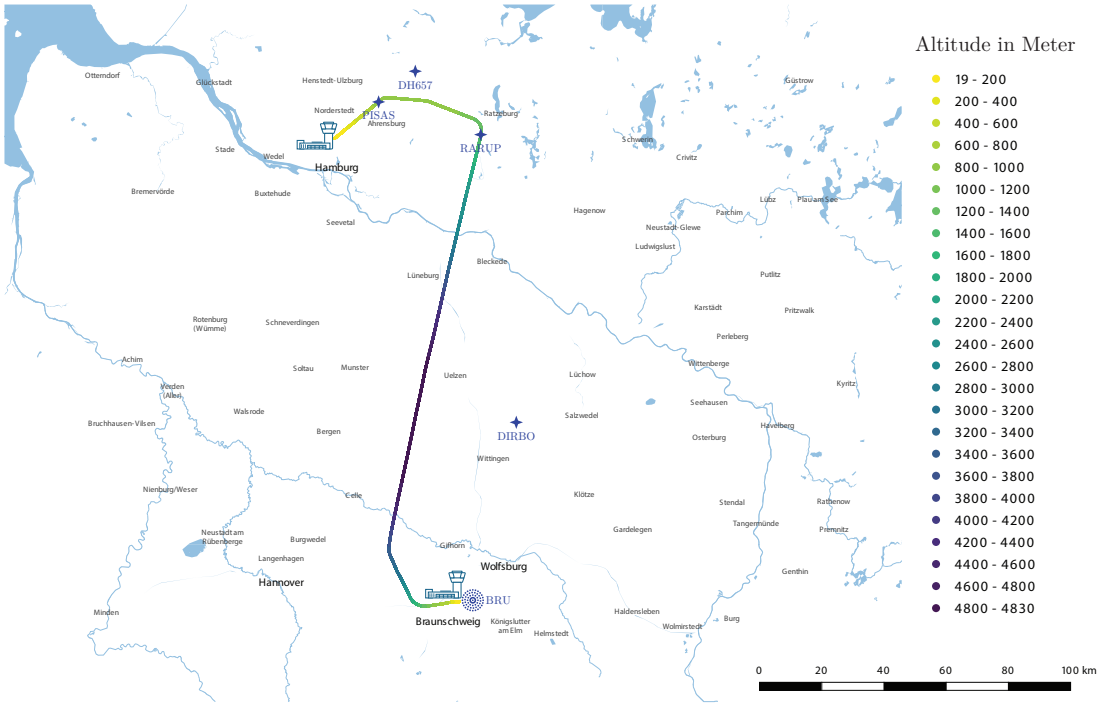


Figure 3.8: Flight progress of the baseline scenario on real world map with waypoints and an actual track from a simulator flight.

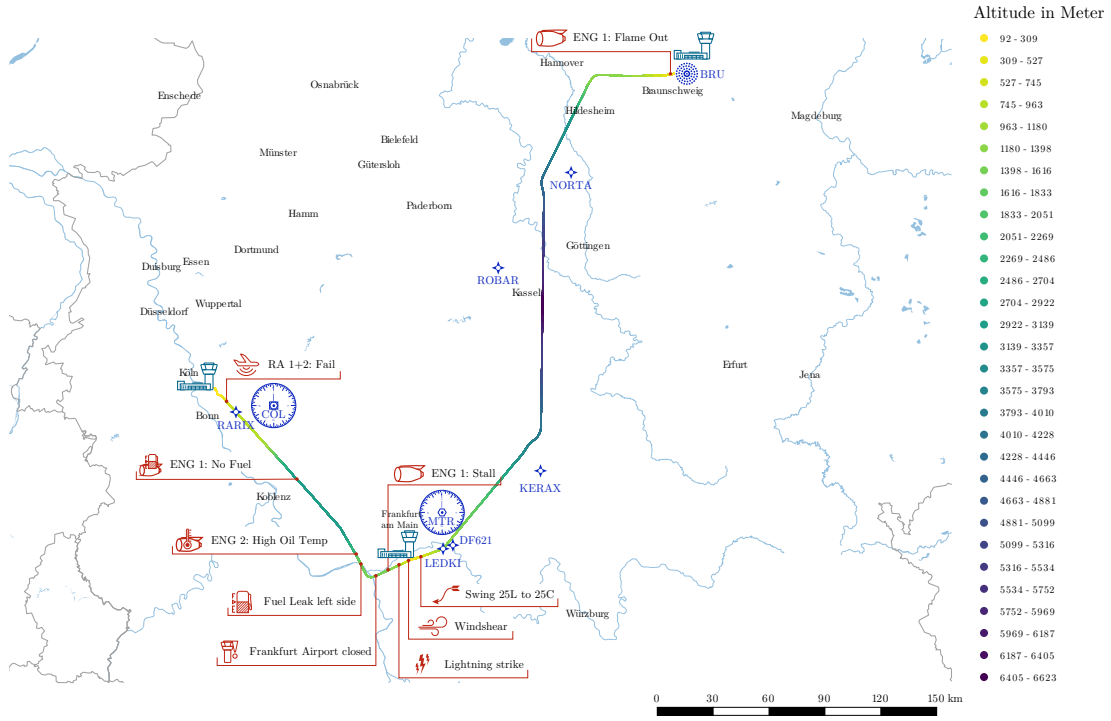


Figure 3.9: Flight progress and simulated failures of the LoHP study's stress scenario shown on a real world map.

if possible. This meant no passengers nor cargo was on board, only the two pilots plus two engineers monitoring systems. After requesting taxi clearance and during taxiing other traffic was simulated via recorded voices. Everything was planned to be normal until shortly after rotating the aircraft during the take-off. At the moment where the simulator was flying stable, an engine flame out on engine 1 was triggered. Due to the simulators tendency to go into hard bank attitude during or shortly after take-off the signal for the engine flame out was given manually by one of the operators. Moreover, due to the simulator the **E**lectronic **C**entralized **A**ircraft **M**onitoring (ECAM) display showed wrong values for the engine as well as the procedures. The flight crew was then instructed by the operators to perform the right checklists and procedures. The aim was to increase the workload but give the crew the opportunity to resolve the situation. Hence, when the engine relight procedures were followed the engine was recovered. After this the flight crews continued to proceed on track to Frankfurt. When passing $10\,000\,\mathrm{ft}$ ($\approx 3\,048\,\mathrm{m}$) and having the engines idled on the right engine (ENG 2) oil temperature started to rise. This needed to be triggered now as it takes a some time for the temperature to become hot enough to become relevant. Thus, it will stay unnoticed and become more relevant later in the scenario. During the approach on runway 25L another aircraft reported ready for take-off on that same runway. Which induced the ATC to ask whether the flight crew were able to perform a swing over. It was left open to the crew if they wanted to this and was not further enforced when they decided not to. On short final, about $50\,\mathrm{ft}$ ($\approx 15\,\mathrm{m}$) AGL, a wind shear was simulated including the airbus audio warning system. This enforced them to perform a go-around. During the go-around a lightning strike was simulated with sounds as well as a flickering on the ND and PFD. Shortly after that engine 1 was triggered to have a stall and the fuel leak on this side was started. This made the crew reduce thrust on the left side to idle and keep the right engine running on high power. Continuing on runway track the flight crew was informed by ATC that Frankfurt Airport had to be closed and were asked to divert to their alternate. This was set out to be Cologne/Bonn Airport. Directed with radar vectors the crews were navigated to the final approach point. After extending the landing gear both RA were simulated to have failed, similar to scenario 3 from the SPO Study. The combination of strong winds and the simulator behaviour tended to make it nearly impossible to handle the aircraft. The goal was to have the flight crews land safely at Cologne/Bonn Airport. Most crews managed to land the aircraft, for a few crews the simulation had to be stopped beforehand as the simulator got uncontrollable.

3.3.3 Subjects

For this study nine flight crews consisting of five Captains and 13 First Officers participated. All of which were male, aged $27\,\mathrm{year}$ to $64\,\mathrm{year}$ (mean: $38\,\mathrm{year}$, standard deviation: $8.7\,\mathrm{year}$) and had averaged 378 flight hours within the last 12 months on average. Participants were asked to not consume any caffeine at least 24 hours prior to the scheduled appointment. They provided formal written consent for participation and were to receive compensation with regard to German law. All of them fulfilled the following inclusion criteria: They were active and trained A320 pilots, no known heart disease or cardiac background. Due to technical issues the in-flight ECGs of the first and second crews were erroneous and had to be discarded.

3.3.4 Data Collection Method/System

Besides simulator data and the stress track an ECG was recorded during the scenarios. The ECG was the main aspect to perform this study at all. Not only did the recording system change compared to the SPO Study, but also the ECG board itself. During the SPO Study only ECG lead I was recorded. The new system was capable of recording 7 channels as well as respiration simultaneously.

3.3.4.1 ECG Recording

Figure 3.10: ECG System used for the October study. Left: All components used in this system. Right: Boards that are hidden within the box on the left image.

During this study a 7-channel ECG as well as a respiration curve was recorded using an EG05000 with respiration board by MedLab GmbH. [32] The board was connected to a SparkFun 5V FTDI for serial communication to a USB Port as shown in Figure 3.10. For recording and monitoring the ECG a software provided by the manufacturer was used see Figure 3.11. A total of five solid gel Ag/AgCl sensor electrodes were placed on the chest of the participants. Positions and corresponding colors are shown in Figure 2.6b.
After informing participants and collecting written consent a resting ECG was taken for 15 min. Participants were asked to sit completely still, not drink, speak, nor eat. The purpose of this was to create a baseline measurement of each participant in a completely relaxed environment. This was intentionally done before any scenario had been flown to mitigate any effects of arousal and excitement after the experiments.

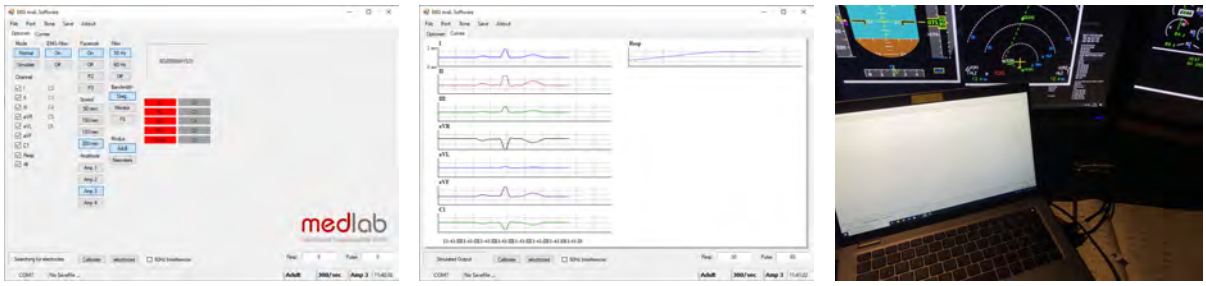


Figure 3.11: Software for recording ECG and respiration signals.

During the scenarios a live monitoring was necessary to detect any abnormalities allowing to intervene immediately. This indeed was mandatory three times as an electrode loosened and was replaced with a new one.

3.4 Debriefing Software

In order to get a feedback from the pilots after their participation in both scenarios a short debriefing was conducted. Moreover, a software tool was developed using Plotly and Dash to show the collected data in an interactive plot as shown in Figure 3.12.
During the debriefing participants were first allowed to give any kind of feedback that was on their mind. They were then asked what they liked and did not like about the scenarios and the study. After discussing those points the debriefing tool was used to look into the data and ask participants about relevant situations.

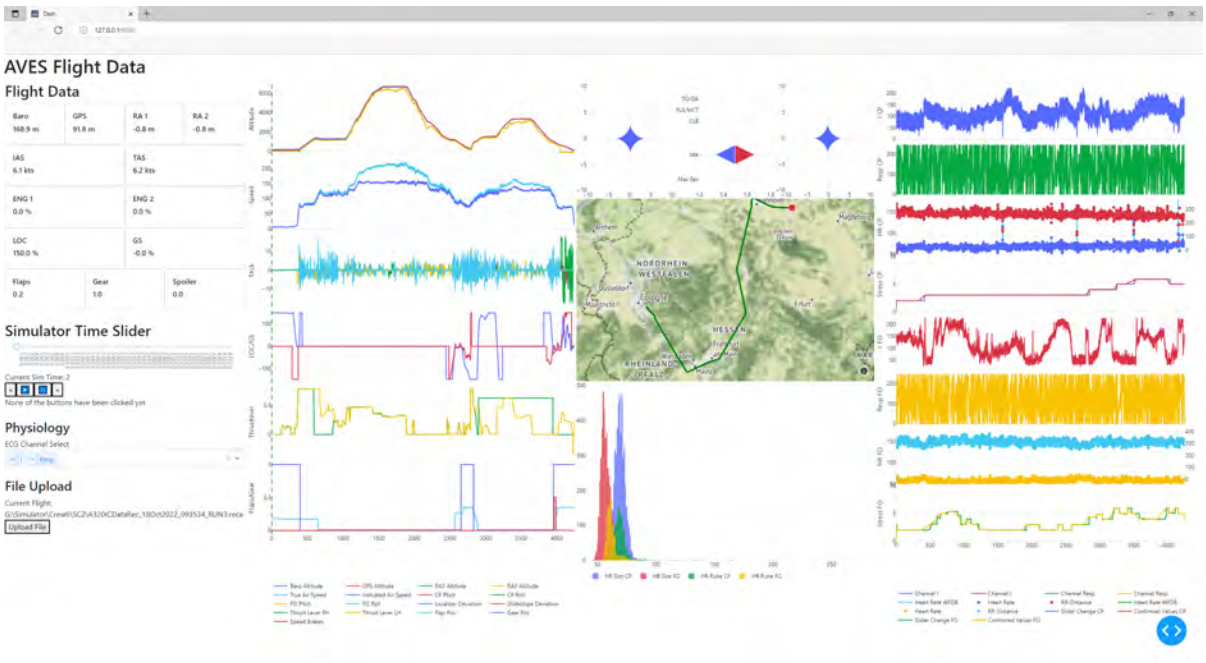


Figure 3.12: Software developed for debriefing flight simulator, ECG, and stress level data which was collected during each scenario. The screen is divided into four sections, on the left side are flight data readouts of the momentary slider position to control the progress. Both middle sections contain plots of the flight data, with an exception below the map which shows a histogram of heart rates derived from the ECG recording. On the right side are both ECG data, heart rates and stress levels. The ECG related plots can be selected on the left control panel in order to select other leads.

3.5 Online Stress Assessment System

Most of the currently existing assessment methods have been developed for workload or to analyze specific tasks. More precisely, to evaluate system complexity during design and evaluation. [33] Anyhow, the goal is to measure physiological reactions of pilots in situations of stress degrading their performance. In these situations pilots are trained to focus on three main things in the order of importance: aviate, navigate, communicate. This would mean in high stress situations pilots focus purely on flying the aircraft instead of doing anything else apart. These areas are of great interest for the present work. Hence, applying a feedback tool requires to be easy to handle and fast to complete.

3.5.1 Evaluation of Existing Feedback Methods

Throughout the development of human factors engineering several methods have been developed to track or measure task and workload. Most commonly known are either the Bedford-Workload Scale, NASA TLX, or **I**nstantaneous **S**elf-**A**ssessment (ISA). The NASA TLX method was applied during the SPO Study after each scenario to get an evaluation of the scenario. Unfortunately, most common methods either quantify workload or situation awareness but not explicitly stress. Although there are feedback methods for general or chronic stress, there is no established method for an unambiguous, fast and simple measurement. For instance the Dundee Stress State Questionnaire differentiates 11 state factors for task-induced stress. [34] Considering asking pilots to answer more than one question besides flying the aircraft would most likely result in a attention shift. This made it necessary to develop a method and tool to closely track perceived task induced stress in flight crews. Therefore, a stress slider was developed to meet the need to track stress and integrate the feedback into the flight deck work

flow.

3.5.1.1 NASA TLX

Variable	End Points	Description
Mental Demand	Low - High	How much mental and perceptual (e.g. thinking, deciding, calculating, remembering, looking, searching, etc)? Was the task easy or demanding, simple or complex, exacting or forgiving?
Physical Demand	Low - High	How much physical activity was required (e.g. pushing, pulling, turning, controlling, activity, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?
Temporal Demand	Low - High	How much time pressure did you feel due to the rate or pace at which the task or task elements occurred? Was the pace slow and leisurely or rapid and frantic?
Performance	Good - Poor	How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?
Effort	Low - High	How hard did you have to work (mentally and physically) to accomplish your level of performance?
Frustration	Low - High	How insecure, discouraged, irritated, stressed, and annoyed versus secure, gratified, content, relaxed, and complacent did you feel during the task?

Table 3.2: Descriptions of the NASA TLX variables. [35].

The NASA published the first version of the TLX in 1986 after being developed over many years. [36] Since then it has proven to be a good method for workload measurement. The NASA TLX is a multidimensional scale for obtaining workload estimates from operators while they perform a certain task or shortly afterwards. The different criteria represent somewhat independent variables: Mental Demand, Physical Demand, Temporal Demand, Frustration, Effort, Performance. The assumption is that some combinations are likely to represent the task complexity. [37] The description and range of each variable can be looked up in Table 3.2 for more details.
The participant is asked after or during a specific task to answer these six questions on a scale of 0 to 20. After the task had been completed participants were asked to further rank each variable against each other. This shall allow to pay respect to different types of tasks and their respective demand. [37]

3.5.1.2 Instantaneous Self-Assessment

ISA is a self-rating scale for participants to rate their workload on a scale of 1 to 5. This technique was developed by **N**ational **A**ir **T**raffic **S**ervices (NATS), a UK provider for air traffic control services, for assessing the design of future ATM systems. Operators are presented with a screen and a keypad with the corresponding numbers. The scale is additionally colored to mark the different levels. After 2 min of inactivity, or no update, a sound and visual feedback appears. This shall remind the operator to press a corresponding number on the keypad.

Level	Workload	Spare Capacity	Description
5	Excessive	None	Behind on tasks, losing track of the full picture
4	High	Very Little	Non-essential tasks suffering. Could not work at this level very long.
3	Comfortable Busy Pace	Some	All tasks well in hand. Busy but stimulating pace. Could keep going continuously at this level.
2	Relaxed	Ample	More than enough time for all tasks. Active on ATC task less than 50 % of the time.
1	Under-Utilised	Very Much	Nothing to do. Rather boring.

Table 3.4: ISA Workload Scale with explanations of each level. [38]

3.5.2 Online Stress Assessment Tool

For the LoHP study a new method was developed in order to measure stress during flight more accurately and time-synchronized to other signals. The definition of this tool is discussed in Chapter 5.1.

Requirements for System development

4

Before starting to create a deep learning system, setting requirements as well as formulating the question that an algorithm should solve are essential. Not only do these determine which preprocessing might be required or possible but also what kind of algorithm and structure suits best for the question at hand. Within this chapter those aspects are discussed and requirements for the deep learning system will be formulated. Moreover, those requirements imply and direct design choices which will be discussed as well.

4.1 General Requirements and Expectations on System Performance

Before starting to generate a reasonable system architecture it is important to state criteria that the system shall comply with. This not only helps to focus on necessary items but also to not under- or overdevelop and retain the system's intention. Moreover, laying out a specific research question that a deep learning algorithm shall answer. Even details that seem to be small at the beginning might have a significant impact later on. Therefore, it is important to be as clear and as precise as possible within this step.
These requirements shall not define a system that is ready to be implemented and used in a commercial aircraft. Rather it can be viewed as unit within such system to analyse incoming physiological signals for stress. Any system that shall inform flight crew must comply with all relevant specifications of international and national aviation authorities (e.g. EASA **C**ertification **S**pecifications (CS)-25). Even though such a system might be more of an assistant without an impact on aircraft systems, at least for now, it still needs a certain reliability. Considering a situation where flight crews are in a critical flight phase (e.g. landing or take-off) and a false-positive alarm goes off might result in distraction. There already have been multiple aircraft losses due to small distractions that resulted in large confusion and loss of control. Meaning that this alarm could end in such a big distraction worsening the situation or in the worst case go from a controlled situation into a fatal loss. Therefore, any system giving notices to flight crew then needs a certain reliability within its prediction in order to be useful. If this system became a basis for aircraft flight decisions even more strict requirements needed to be considered. Moreover, many Certification Specifications need to be complied with such as e.g. CS 25.1322 'Warning, caution, and advisory lights' as well as associated Acceptable Means of Compliance. Additionally, it might be required by aviation authorities to further proof reliability, effects of failure and more factors before being accepted.
Requirements set here are more fundamental in order to build up a stress detection unit for such a potential aircraft system first. They shall guide development for a stress classification

system and prepare for a possible implementation into a larger system. These requirements have been derived from group meetings and discussions at DLR FL-SEG.

First of all it must be determined when the analysis should take place and in which time steps a stress level is output. This determines not only whether a classification is done during signal recording or later but also which time performance the deep learning system needs to comply with. It is clear that an analysis after flight would be nice for maybe a personal improvement but is not the actual goal to produce a warning system for flight crews. Therefore, a near real-time analysis is required and hence a system that is capable of such performance. The next step then was to consider whether the system should take chunks of input signal or an individual datum. This could either mean collecting e.g. 5 s of ECG signal and then predicting a stress level based on this. Another path could be using a sliding window that forwards the last e.g. 5 s to the system for classification. Alternatively the signal could be forwarded once a new one is measured by the ECG module. This sampling rate establishes a run-time requirement for the deep learning system.

REQ 1 Classification and Analysis in Real Time

> **Real Time Requirement Details**
>
> The cycle time for classification and analysis needs to be lower than the time between two consecutive signals.
> This shall allow the system to classify each input signal accordingly.

This requirement heavily depends on the sampling rate of the signal providing system. During the flight simulator studies a sampling rate of 300 Hz was used for the ECG Signal. Considering this sampling rate a single evaluation cycle would need to be about 3 ms. Such a short cycle time might require to develop software more close to the hardware driver in order to comply with such requirement.

For the time being this requirement is kept in mind when designing the deep learning model but it is not a metric evaluated during model training. After the final deep learning model has been created the runtime for one cycle will be evaluated and compared to this set cycle time. If this takes longer time than acquiring the input signal segmentation might be an option. This would mean that for instance 1 s of input signal is gathered first and then forwarded to the classification unit. That deep learning model would than produce a stress level output for that second.

REQ 2 The system shall consider the special circumstances within aviation

> **Flight Crews**
>
> Flight Crews have a unique working environment that comes with special circumstances.

Not only is the stress experienced by flight crew different from e.g. in car drivers but they also undergo specific training and preparation. For instance, pilots have the obligation to have a certain amount of flight hours, simulator trainings and so on. Additionally, persons recruited by airlines for flight crews follow a specific psychological pattern. All these factors have an influence on how their stress is perceived individually. Moreover, other non-flight related aspects might trigger stress in that person.

4.2 Stress Classification System

With those general requirements established the next step is to further define stress classification aspects. Moreover, a clear research question the deep learning algorithm should solve must be formulated. Then, suitable deep learning architectures to solve this and answer this question must be explored.

REQ 3 This unit shall be responsible to classify an input signal and output a stress level value.

REQ 4 Input signals are physiological data from flight crews.

> **Input Signal**
>
> The basis for the classification are physiological signals that are measured e.g. during flight.
> The selected physiological parameters are ECG leads.

When it comes to detecting stress several physiological parameters are eligible to build the base. For the course of this project the focus shall be on ECG leads. In future concepts those might be changed to or extended with other parameters such as respiration, EMG, or even EEG/**F**unctional **n**ear-**i**nfra**r**ed **s**pectroscopy (fNIRS).

REQ 5 Based on the input signal a stress level shall be computed.

> **Research Question**
>
> With the ECG signal as input the system shall output a corresponding continuous stress level between $\mathbb{R} \in [0, 1]$.
> 0 - Absolutely no stress (resting)
>
> 1 - Maximum perceivable stress

In other studies the usual approach was to either classify between 'no stress' and 'stress' or to add one or two additional steps for 'low stress' or 'low', 'medium' and 'high' respectively. This is an abstraction that simplifies the task at hand by assuming that a human has only these binary or discrete steps in stress. In most cases, it is either hard to distinguish between those steps without a clear guidance of what exactly each level means or to evaluate by the person themselves. Moreover, the goal is to find a correlation between ECG recordings with stress and without stress. Hence, the individual needs to have both: the awareness that they are stressed right now as well as a measurable physiological reaction. Therefore, the approach here was to only set boundaries at maximum and no stress and participants could freely range themselves between those two boundaries. Additionally, before flight they were briefed on this scale including references for certain areas correlating to the ISA scale. The idea is to more accurately mimic the actual physiological reaction of the body on stress. Moreover, each individual might have a different stress experience. This scale tries to adapt to those personal and individual stress reactions by setting the maximum value to maximum stress that they can imagine. An established similar method is the **V**isual **a**nalogue **s**cale (VAS) where patients are asked to quantify their pain or itching. They can give responses either on a scale of 1 to 10 or a continuously colored scale which is then associated with the same 1 to 10 scale. [39] During both simulator studies that scale was extended to 0 to 10 and was then normalized to the above-mentioned range.

4.3 Conclusions for Model Architecture

With these requirements, formalized first assumptions and design decisions in regard to the stress classification system were possible. Moreover, those do not only set out certain boundaries for design but also narrow down reasonable algorithms. Due to not every machine learning and deep learning algorithm being suited to solve any question, that question must be well-defined. Therefore, having that question defined now, it is possible to foster through already developed algorithms for suited ones.
The input data are time series with dependencies not only on the previous but also on the next state. That means that in a healthy human once the sinus node of the heart sent its signal all other heart muscle cells will follow. Hence, this time series dependency would benefit from an algorithm that somehow keeps track of the past steps over time. One that is well suited for such a data type is a LSTM network. It is in their nature to store information over the cell state matrices. Moreover, the runtime requirement 1 limits the size of the LSTM architecture.

Part III

Application

Online Stress Assessment Tool 5

As discussed in subsection 3.5.2 a new method, or rather a modified version of existing tools, has been developed over the course of this thesis: the **O**nline **S**tress **A**ssessment **T**ool (OSAT). During the October flight simulator study this tool was used to evaluate stress from both pilots of the flight crew. They were then asked after the flight to provide written feedback on this tool. Within this chapter both the application as well as the evaluation will be discussed in more detail. Finally, based on this feedback and the data collected proposals for further improvements will be provided.

5.1 Definition of the Online Stress Assessment Tool

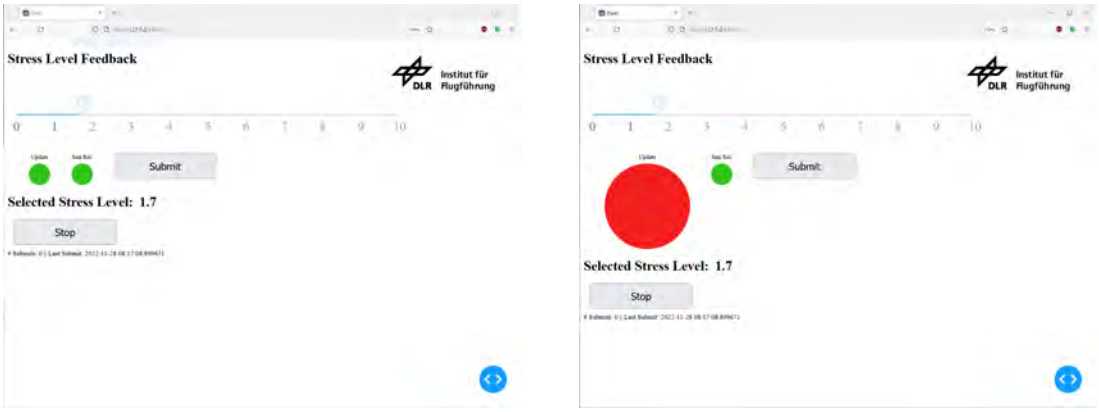
Stress is a gradual process which is difficult to be separated into precise steps. Furthermore, the goal was to get as accurate as possible feedback from the subject about the current stress level. In this sense, accurate not only refers to the extended scale, but also a higher frequency. Moreover, with the ISA method no feedback about the previously selected value was visually provided to the operator. Such kind of feedback might improve the operator's capability to compare their perceived stress level over time improving overall quality of the feedback given by the operator. On the other side, one could argue that operators then just compare the current situation to the previous. This risk shall be mitigated by briefing them on the meaning of different levels. The most suitable scale would be the ISA scale but also that would not yet cover all concerns.

A continuous slider for the stress level is proposed. Subjects can freely select their stress level replacing the ISA keypad. The slider was set out to be on a scale from 0 to 10, representing the later scale of the machine learning system. 0 is not represented by the ISA scale. It is supposed to represent a completely relaxed state as during the resting ECG. The scale from 1 to 10 is modeled after the ISA scale retaining the boundary positions, which is shown in Table 5.2. Technically, the scale could have been set up to only match full integers, but the step size was set to be 0.1 to allow automatic increases of this step size (see below). This also helped pilots during operation to have a more gradual feeling than being limited in selecting the right level. When the right value was selected a submit button had to be pressed to confirm the value. As this was an

ISA Level	Workload	Slider Level
5	Excessive	9 - 10
4	High	7 - 8
3	Busy Pace	5 - 6
2	Relaxed	2 - 4
1	Under-Utilised	1

Table 5.2: Modeled ISA workload scale adopted to larger scale.

additional task, which could be forgotten under stress, this also serves as an indicator for the actual stress of the individual.



(a) Normal Screen. (b) After 1 min of no updated submit value.

Figure 5.1: Online Stress Assessment Tool in action. Stress levels must be selected via a slider and then submitted. After 1 min without updates a green dot grew in size, became red and started flashing to remind the operator to submit a new value. After yet another minute without an interaction the slider value would slowly increase by 0.1 every 2 s.

Moreover, four parameters can be defined with this method. The first one being the time after which the participant is remembered to give a stress level submit called Δt_{indi}. The other parameters are related to the mechanism of changing the slider value itself. The first one is y_{incr}, the increment by which the stress level is increased each idle time step of length $\Delta t_{y,incr}$. This is done continuously after a specific time Δt_{incr} has passed. This means that after $\Delta t_{indi} = 1\,\text{min}$ of inactivity a previously green dot turned red, grew in size, and started flashing reminding the pilot to submit a new value. When the operator had been inactive for $\Delta t_{incr} = 2\,\text{min}$, the slider increased its value by $y_{incr} = 0.1$ every $\Delta t_{y,incr} = 2\,\text{s}$. This time increment shall reflect that the operator is so busy that they neither remembered to give an update nor recognized the flashing red light for 1 min.
The chosen parameters were first set by an educated guess rather than scientifically obtained values. Therefore, if this method is validated those parameters should be investigated as variable. For instance, a shorter or longer time period until an update is requested might influence the feedback's accuracy. The same applies to the selected stress level increment of 0.1 each 2 s. For the usage during the flight simulator study those parameter settings and exact slider values accurate to the decimal place were not decisive, but rather used to get a general evaluation of the pilots' stress. Meaning that it is not interesting whether the slider is at 5.22 or 5.03 but rather in the range of ≈6 or ≈5.

5.2 Application in Flight Simulator

Within the AVES A320 flight deck two tablets were installed on each side for Captain and First Officer. This tool was developed as a software script to provide interactive features. Utilizing the Dash framework, the software script runs on an external computer hosting a Flask web server. Via any web browser the webpages for each side can be accessed with a local IP address. When everything was ready to go for the scenario the simulator operator gave verbal 'go'. Upon that notice flight crews were asked to press the start button to manually start recording. The latency between each record start is expected to be within 2 s to 3 s, which is more than acceptable. The first attempt was to synchronize the start and stop with an AVES simulator signal. This was unfortunately not feasible as the UDP signal send from the simulator had

Figure 5.2: Online Stress Assessment Tool on tablets within the AVES. Tablets were placed on both sides to be accessed by each crew member.

a delay within 10 s to 30 s. That delay was most likely caused by the network infrastructure. This lead to the decision to manually start and stop the record as a short term solution. During the simulator flight each person was asked to adjust and submit their momentary stress level continuously. All flight crews were briefed that upon a perceived change of stress the slider shall be adjusted to match their perception. With the submit button either the changed value or the current value shall be confirmed. To remind pilots to update and submit their current stress level the update indicator turned big, red and started flashing. After 2 min of no submitted value the stress slider started to increase its value by 0.1 every 2 s. The purpose was to implement a change on the slider after a significant time. This point is further discussed in detail within the evaluation and discussion in Section 5.5.
The software script was programmed to not only record the submitted values but also any change on the slider and whether changes came from a manual submit or an automatic increase. These two data sets were saved after stopping the recording manually with a button. The data was stored on a central server hosting all other study data. Through this server data was available immediately after the end of the scenarios.

5.3 Evaluation of Data Quality

The first huge advantage over the SPO study's stress feedback is that no analog-to-digital data transformation is required. Moreover, data was available immediately after each flight and was used for discussion during debriefing. This gave the study manager the opportunity to get feedback from the flight crew about their values. Not only could pilots reflect on their performance and feedback but also provide the study management with their interpretation and reasoning.
A major factor to consider is the discrepancy between the actual stress and the stress level which was fed back. There are several factors that can influence the way stress is perceived, assessed, and reported back. [40], [41] The feedback provided by participants may also be influenced additionally by other factors e.g. experience, personal constitution and many more. Moreover, participants might consciously or unconsciously impose a certain bias on those stress levels. For instance subject bias might impact the measured values as participants are aware of what is measured. Therefore, participants could consciously or unconsciously act in a certain manner in order to achieve a certain outcome.
Additionally, stress levels provided still represent timely discreet values while stress is a continuous process. Therefore, data used from this tool needs to be further analysed and post-processed to fit more accurately. A debriefing with the participant also helps to get an

understanding what happened during data recording.

5.4 Usage Analysis

There are two sides when analysing the usage of this method. On the one hand the feedback of the users and on the other hand the resulting data. Moreover, at this point it shall be mentioned that human feedback tools always heavily depend on the capability of the user being able to self evaluate correctly. There are many psychological effects that influence any human and especially participants in a study e.g. the effect of experimenter bias. [42]
To further improve the interaction and make it more intuitive, user observation and evaluation are crucial. Although verbal feedback was already provided during debriefing formal feedback was given via an online questionnaire hosted on LimeSurvey. Participants were sent this survey one week after the study ended. In that they were asked:

'How did you like the operation of the slider scale? What could be improved?'

Out of the 18 participants six rated the slider as 'good', two as 'okay', and four stated it was easy to operate. Two participants said they had difficulties interacting with the tablet. Moreover, they had to tap twice or more for a change or to get the value submitted. Nine participants also gave further feedback on improvement. But there was no consistency among those. Any further feedback was given by only one participant each. One suggested increasing slider size, one to operate by voice, another said the markers should have corresponding explanations below the slider. Something noteworthy is that one participant stated that they found it 'difficult to assess ones own stress level'.
The most prevalent feedback received was difficulties with interacting and using the slider. This could indicate that the response time and sensitivity of the tablet device need to be improved. Besides this there was no consistent feedback but statements that indicate dissatisfaction.

5.5 Proposals for Further Improvement

It appears that usage and interactivity were the most prevalent issues. Looking at the responses by the participants the first thing to consider changing would be the tablets for input. Making the user interface more responsive and accessible might already solve some problems. Also, some participants indicated that it took too long to give their input which might be an indication that the tablet did not respond quick enough. Moreover, some users criticized the necessity to confirm the values. Nevertheless, there still might be some benefit from having this additional task, as annoying as it might be. This step might allow a further insight into the real workload and momentary capabilities of that person. As the 'Submit' button is another task to complete, missing out this step might be a reflection of a degraded human performance. Due to the risk that the button was not pressed correctly and hence did not trigger the buttons actions it is difficult to distinguish whether it was a lapse or some other factor that lead to missing this task step. Besides this issue, having hardware that is not responsive enough complicates to identify a reason for that miss.
Besides improving the hardware there is space for further improvement. Not only increasing the size of the slider but also associating colors with the levels might be options to consider.
As the goal should remain to not disturb pilots more than necessary from their flight deck tasks limiting the influence of the stress slider could reduce the impact on not only the pilots' performance but also the actual stress levels. Therefore, a fine balance between measuring accuracy and disturbance on the flight crew needs to be figured out. Not only the duration of the timeframe but also the slider increment needs to be researched further in order to find the best parameters. For now these values are only educated guessed with no reliable background.

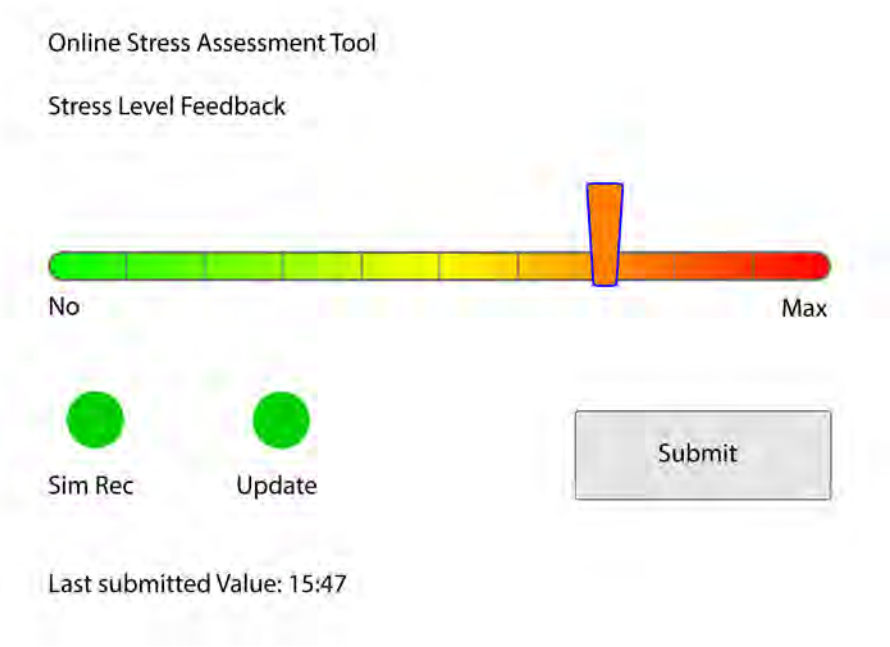


Figure 5.3: Redesign of the Online Stress Assessment Tool based on the feedback provided by participants. Slider values were replaced by a scale from 'No' to 'Max'. The slider is thicker and continuously coloured from green to yellow to red.

Especially the 1 min timeframe until a new update is requested might be too long to grasp the actual changes in hectic and stressful situation. Anyhow, this initial proposal already provided useful data despite these factors that could be improved.

Database Construction 6

Starting out with the gathered data from both simulator studies this chapter will explain how a structured database was constructed. This database will then be used in further steps and research to analyse and feed into a deep learning algorithm. The database files shall consist of all relevant signals with same length and in an easy to access format (e.g. one CSV file). The benefit of those database files in comparison to the individual signal files is that all signals are fitted together and cut to the same length. Moreover, it allows those data to be published and shared with others without having them perform a similar preprocessing.
The last thing before training the deep learning model is to get an understanding of how the data is like. This process is commonly referred to as 'data exploration'. The goal is to gain an impression on what the data consists of e.g. if there is a good distribution of the labels. This helps in further steps to separate data into a training and validation data set. Moreover, correlations between stress labels and derived parameters such as the heart rate can be illustrated in this way.

6.1 Processing of the Data Collection

In order to transform the collected data into usable databases several steps were necessary. Simple software tools were specially developed from scratch to help speeding up this process and to illustrate data during the process. Both studies were quite different and therefore data processing was approached quite differently. That is why both studies will be discussed individually.
At this point, the importance of high quality and reliable data shall be highlighted. This means that the success or failure of any deep learning application relies on the data that is used to train and validate it. For classification tasks especially the quality of those databases are even more crucial. This is because any pattern or structure represented in the data will eventually be learned by a deep learning network. Poor data quality and label distribution might lead to recognizing unwanted noise or other biases. Moreover, collecting and labeling as well as preparing data is essential for creating a solid database and thus for any good machine learning application. With regards to machine learning 'reliable data' means that both input data and labels have a true correlation. Those labels are used during model training as 'ground truth', a given truth that is used to optimize trainable parameters of the neural network. When considering image classification this might be quite obvious but when considering ECGs and stress levels it can become very abstract. In some studies features were extracted from the original ECG signal. Those can be as simple as an IBI and can become quite complex with analysis in frequency domain. [43], [44] This is usually done in machine learning to focus

on the essential information and discard the rest. In relation to the ECG input signal using only HRV features is a reduction of the signal to the heart beats instead of using the whole ECG. The approach of this thesis was to include those ECG signal information and let the machine learning algorithm figure out which parts are more important than others. This was done under the assumption that not only the heart beat plays a significant role in the cardiac representation of stress. This is supported by the study of Shahrudin, Sidek and Jusoh which showed the changes of QRS-morphology under stress. When it comes to modifying or augmenting the stress levels provided by the participants this can have an even more significant impact. Although a certain fed back stress level might not fit the expectation or observation, the individual just might be experiencing it that way. Hence, no modification to the submitted stress levels was done besides a linear interpolation to fit the ECG's sampling rate.

Next, a data format in which the collected data is converted and merged to needs to be selected. There is a plethora of common data formats. Looking for a suitable data format turned out two valid options. First, the recommended format used by the largest collection of physiological data, physioNet.org, is WFDB. Second, a rather widespread and more user accessible format: CSV. There is one clearly specified CSV format defined by RFC 4180 [45], but implementors often deviate from the standard, e.g. by changing the separator symbol. Nonetheless, the latter was used as its input and output usage is straight forward and quite easy to implement using Python. At the end the merged data file was created for further use.

In general, the transformation and merging process can be described as follows: The first step to start off with is to look into each record and at least plot this data once. With this graphical representation a first scan for any abnormalities or signal noises was conducted. For some simulator runs some artefacts were detected as well as files with no data recorded at all. All files with no recording and heavy noise over the whole signal length were discarded from further processing and analysis. In a next step, the stress level and ECG signals were matched and sampled to the same length. Finally, that data was output to a CSV file.

6.1.1 Data from Single Pilot Study April 2022

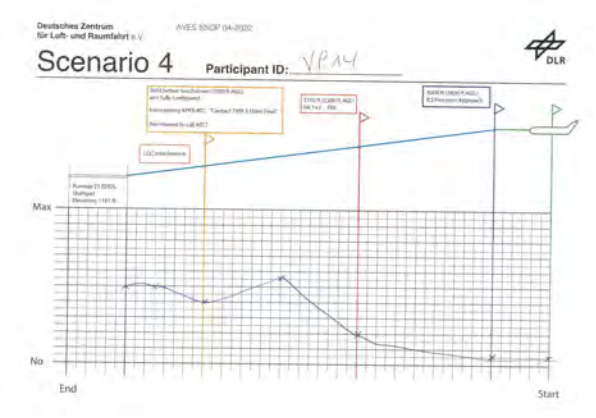
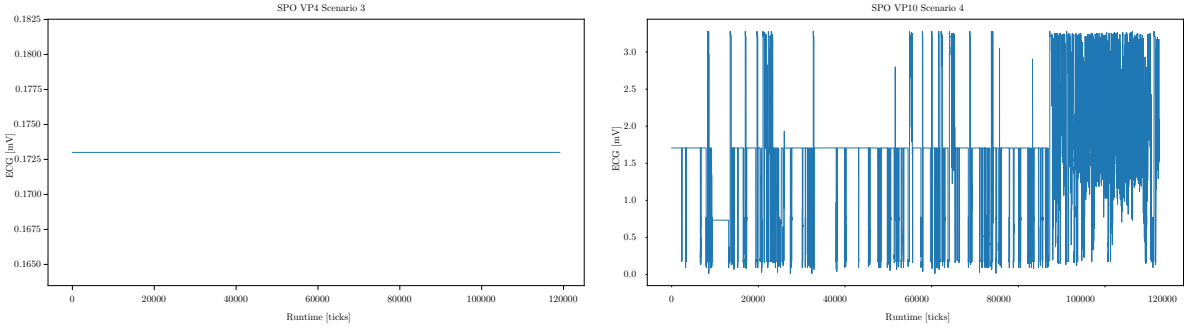
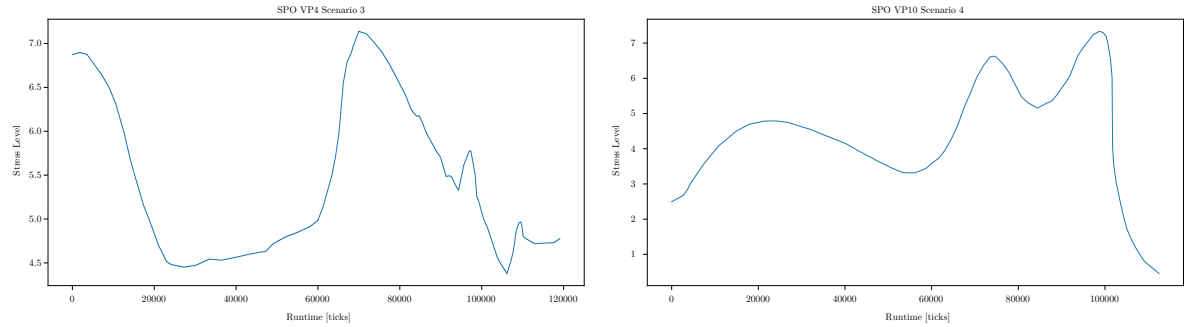


Figure 6.1: Example feedback page from person 14 for scenario 4 of the Single Pilot Operations Study April 2022 as it came out the scanner before any further processing steps were made. The coloured flags represent guidance for the pilots for filling out the form and represent certain events during the scenario.

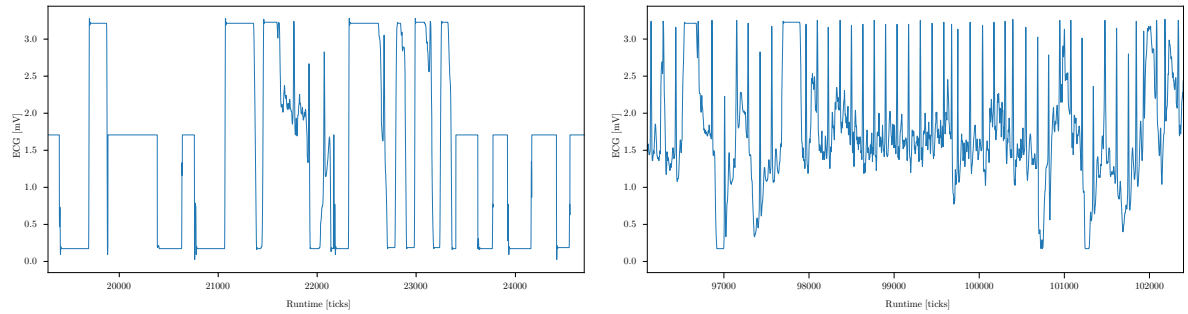
The challenge at hand is that there is no synchronization between the ECG and stress data. Starting and stopping the recording of the ECG module was triggered by the simulator. When the operator manually gave a signal that the simulation is running the recording of the ECG module started. Unfortunately, the simulator data recording started a significant time (about 10 s to 60 s) before the actual recording began. This resulted in huge differences in signal length between simulator and ECG signals. Having no correlation between the simulator and ECG data signals makes it impossible to match those two in any way. For the course of this project this is not extremely important, because the correlation of interest is between the ECG signal and the stress level.



(a) ECG recordings.



(b) Stress level tracks.



(c) Details of **V**ersuchs**p**erson (*German for 'Test Person'*) (VP)10's ECG record.

Figure 6.2: Examples of erroneous data recording from SPO Study. a: ECG data which was recorded from VP4 Scenario 3 and VP10 Scenario 4 on the left and right, respectively, b shows the corresponding stress level tracks. c: Zoomed in area of the ECG from VP10 shows some contact issues in comparison to VP4 where only a constant flat line was recorded.

6.1.1.1 Data Merging

The ECG signal was saved in a GEN4-File, a proprietary file format developed by Capical GmbH. A Python software script was provided for reading this binary format and make it accessible. Stress levels were filled out by pilots using a pen and a sheet of paper with the flight track. Those paper documents were scanned with a printing station and saved as PDF files. In order to digitize those stress level tracks a vector graphics program was used, Adobe Illustrator 2021. Each PDF file was loaded into this software and was then scaled and rotated to fit with the original template of the respective scenario. It was not always possible to align the scan to fit all four corners at the same time, which indicated a distortion from the scanning process. This distortion was not significant so that further mitigation was not needed, but it is a small uncertainty that was added. When the scan was aligned as good as possible on the drawn stress track points were added manually. Those were put in a regular distance to match the original line as closely as possible. When the stress track was covered completely only those points, including two corner points for coordinate system restriction, were exported in the SVG format. An SVG file has the benefit that all points are stored in an XML-like file structure. Therefore, this file can simply be read by any XML library to extract those points and coordinate system. From this point on all further modification and analysis was automatable using software. The SVG files were loaded in Jupyter Labs, the data extracted and then used further.
With the help of the bottom left and top right points a coordinate system for each file could be created in which the stress track was located. Using the maximum x and y values the stress track was scaled to have a scale from 0 to 1. From this point on the stress track was scaled to the length of the ECG signal by simple multiplication. Now the discreet points of the stress track were up-sampled by applying a 1-D interpolation between each point. During the digitizing process it was ensured that the distance between two consecutive points is so small that a linear interpolation is sensible. Now both, ECG and stress signal, have the same length and same amount of data points. More precisely, for each ECG data value there is a corresponding stress value. In the last step, this data was saved as an CSV file for later use. The complete process is also displayed as a schematic in Figure 6.3.

6.1.1.2 Evaluation of Data Quality

Overall the recording was successful for most participants. In total, 105 files were created, 21 files for each scenario with a total runtime of around $666\,\mathrm{min}\ 19\,\mathrm{s}$ (11.993.700 Rows with $300\,\mathrm{Hz}$ Sampling rate). For every scenario, the mean run time was between $6\,\mathrm{min}$ and $7\,\mathrm{min}$ and the longest overall scenario was a scenario 3 with about $7:33\,\mathrm{s}$. Recordings from three participants had to be discarded due to no recording or measuring errors. All of them were during single pilot operation. In Figure 6.2 is an example of a recording that was so superimposed with artefacts that it would not make sense to further use them for training the model. Although the stress recognition system must be robust for such artefacts or disruptions in the input signal such large number of measuring noise in a such small overall sample might distort model recognition. Moreover, files with a flat-line signal, mostly due to an unattached electrode, were also excluded for model training to prevent a similar overfitting to those erroneous data. Within other recordings only a few motion artefacts could be found which is seen as acceptable as realistic use cases will likely contain these artefacts as well. Training the model with such artefacts increases the robustness of the prediction. Another software unit could be used to detect whether the signal is correct or not before feeding it to the stress detection unit. Besides that, there had been no further issues in terms of recording quality within the data.

6.1.2 Data from Limits of Human Performance Study 2022

For this study, all relevant data were recorded digitally, so no analog-to-digital conversion was required. In order to visually control and cut both signals to one synchronized signal file a

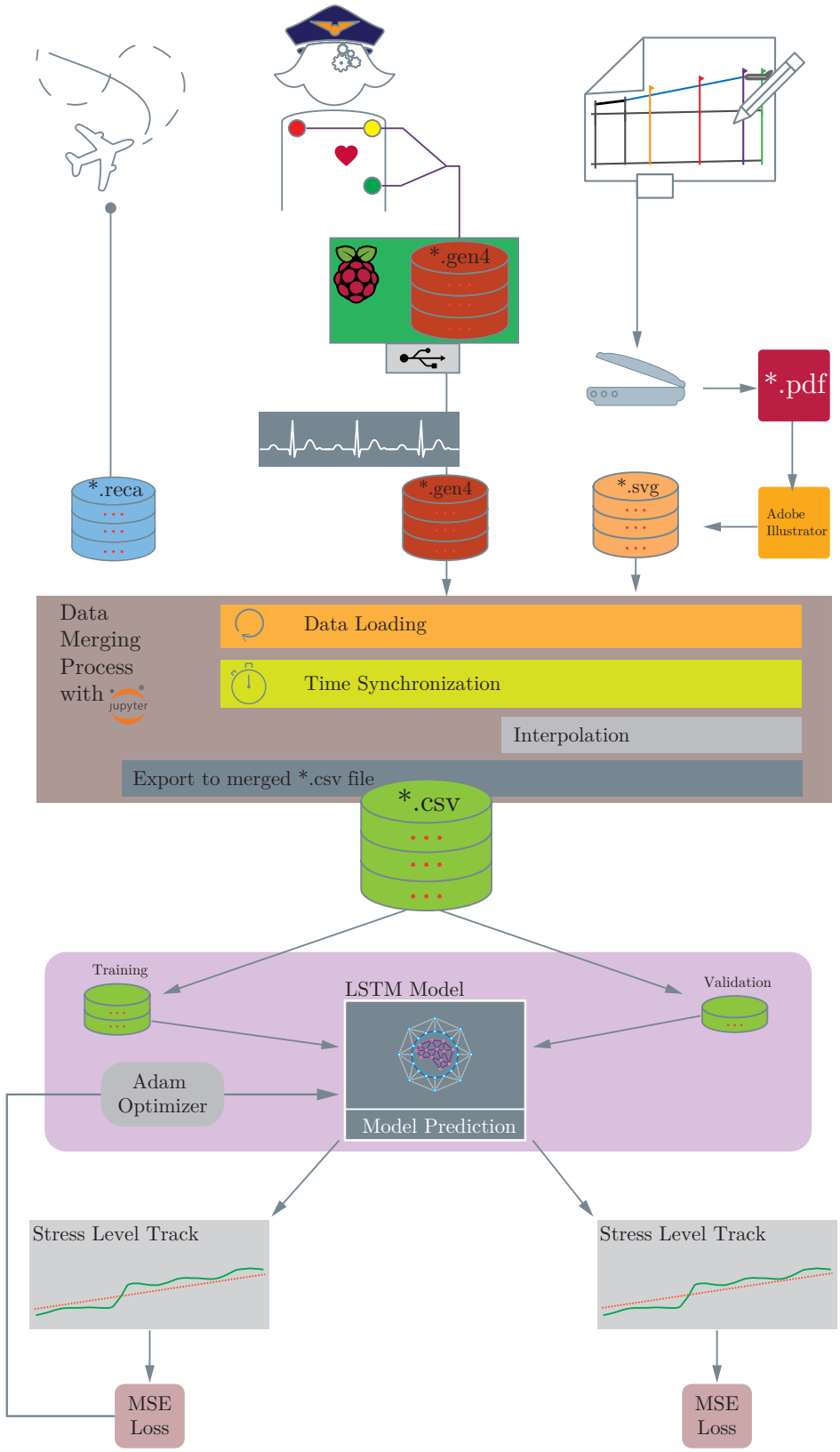


Figure 6.3: Data collecting and processing flow of data from the SPO study. Starting on top with the data collection leading the way down to data processing and finally the usage within the LSTM architecture for training and validation.

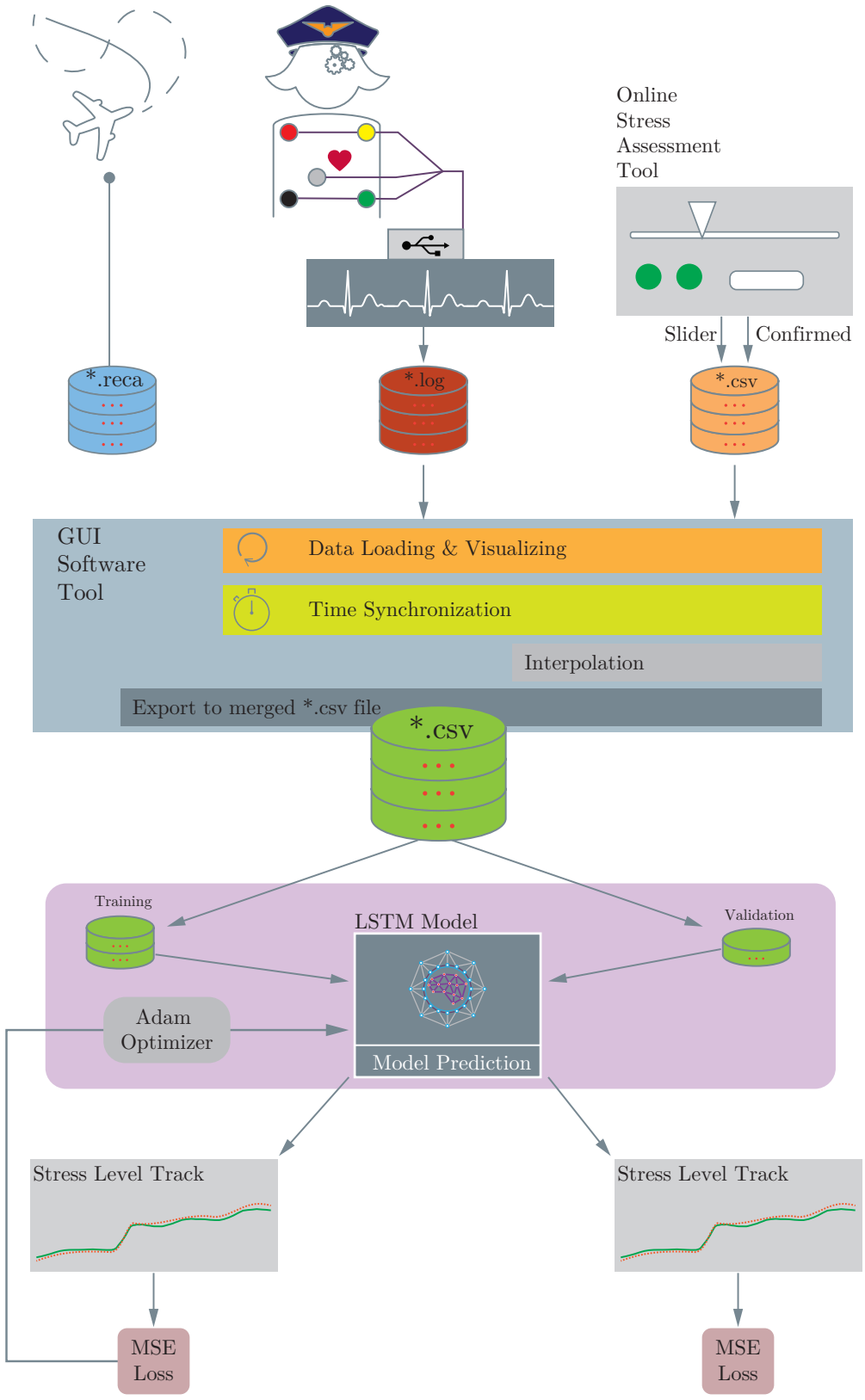


Figure 6.4: Data collecting and processing flow of data from the LoHP study. Starting on top with the data collection leading the way down to the data merging process amd finally the usage within the LSTM architecture for training and validation. The LSTM section is the same as in Figure 6.3.

small GUI tool was developed. Within this tool both, ECG and stress track files, were loaded and plotted with **C**oordinated **U**niversal **T**ime (UTC) time on the x-axis. On the left side of the tool sections of the data could be selected and analysed for statistic parameters of the ECG lead. Those parameters are commonly used when it comes to analyse the heart rate variability. In later steps, this tool could be used to select areas where stress is expected and then analyse only this section. Anyhow, the most usable overlapping signal excerpt was selected and then exported as a CSV file containing ECG signals, stress track, and UTC time. This not only helped to visualize and investigate artefacts within the two signals but also enabled a quick processing of the data. Overall, a small offset could still remain because the two computers on which ECG and stress levels were recorded were not time synchronized. As long as this offset is within the range of a few milliseconds to microseconds this is assumed to be negligible for this case. In future projects, fixing this would not only improve correlating data but also further decrease the time required for post-flight processing.

6.1.2.1 Data Merging with GUI Software

This tool has three major areas (Figure 6.5): the left side user panel, the middle graph section, and statistic read-outs on the right side. The top two graphs consist of the ECG channel II and the respiration curve. Other channels can be selected and displayed by selecting the corresponding signal in the drop-down menu on the left side user panel. Using the XQRS method within WFDB toolkit R-peaks were detected and plotted as grey lines in the top row plot. In the bottom row both confirmed and slider stress tracks were plotted. IBI in ms and instantaneous heart rate in min^{-1} were calculated from the ECG signal and then also displayed in the middle. These values should just be a more accessible representation of the ECG signal during editing. But those calculated values must be taken with consideration as the heart rate might not be the only thing changing under stress.

In order to merge ECG and stress levels both signals need to have some type of time stamp. The files of the Online Stress Assessment Tool contain the UTC time alongside each submitted and slider values. Furthermore, ECG files meta-data contain the last modification made to that file which represents the moment recording was stopped in the software. As we got the time for the last row transmitted by the ECG module as well as the sampling frequency of $300\,\mathrm{Hz}$ a precise UTC time for each signal row can be calculated. Due to the fact that each recording, both ECG and stress level, were started and stopped manually, offsets are present in each and every recording pair. Using the time stamps a good synchronization between those two signals was realized. Moreover, during the export process the stress level tracks were linearly interpolated to match the sampling frequency of the ECG signal.

To indicate the starting and stopping points for ECG and stress they were indicated with a red line on the respective other plot. So, the red lines on the ECG plot represent the first and last signal of the stress track and vice versa. In Figure 6.5 the red lines for the starting points are shown as example. Next, the selection slider can be used to adjust the selection which will be used for export later. Therefore, it can be assured that there are not only always two signals present but also that those correspond to about the same moment in time. In a next step those two signals were exported into one CSV file containing all 7 ECG leads, respiration curve, UTC time, slider as well as confirmed stress level values. The software tool was programmed to compute a 1-D and Akima 1-D interpolation for the slider and confirmed values respectively. In a final step, the exported file can be loaded into this tool as well in order to check that nothing unexpected happened. Especially the interpolation might have been too extreme and had to be changed to a simple linear interpolation.

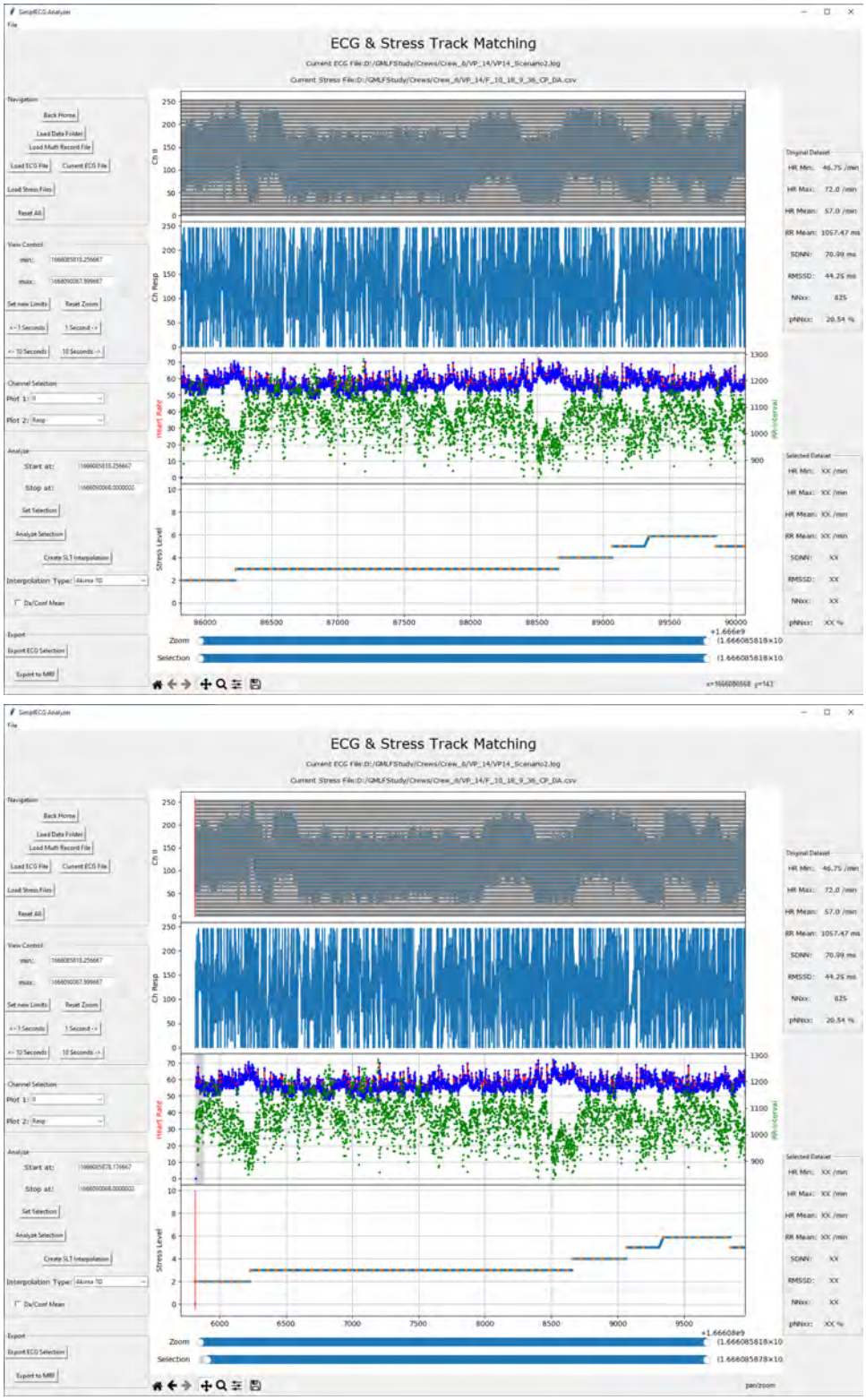
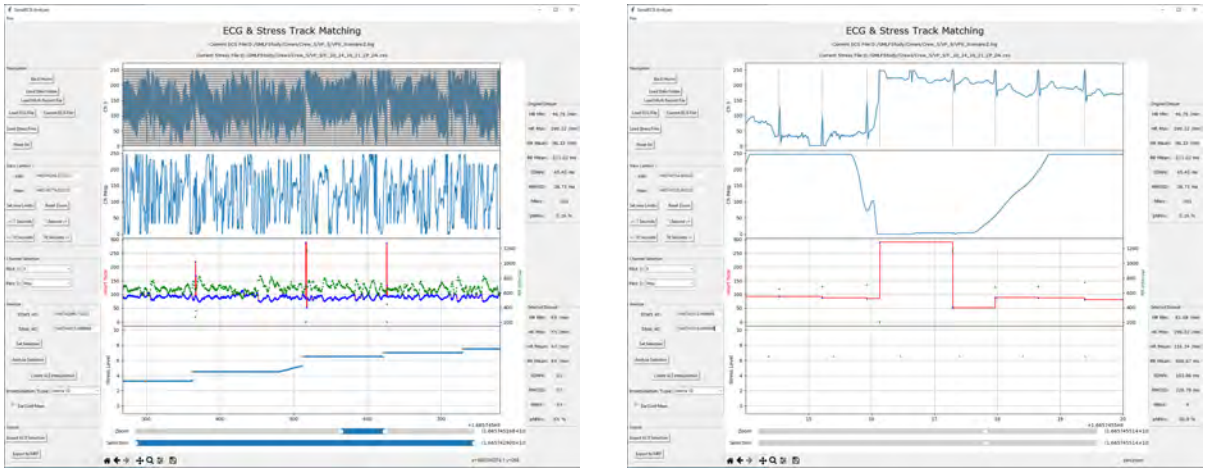


Figure 6.5: Tool developed to process and merge data from Limits of Human Performance Study October 2022. Red lines indicate the start and end of the active interval. are visible on the respective other axis. In the bottom screenshot only one red line (start) is visible, because the data is cropped on the right.

6.1.2.2 Evaluation of Data Quality

During most flight simulator runs data was recorded without any bigger issues. From this study a total of 1 670 min of ECG (30 598 289 rows at 300 Hz sampling rate) and 1 452 min in simulator recordings were obtained from 18 individuals, 534 min and 918 min of the baseline scenario and the stress scenario, respectively. The average recording time of the baseline scenario is about 33 min and 70 min for the stress scenario, the longest record was 97 min during a stress scenario. Only over both scenarios of the very first crew the data recording system had a total loss which caused no recordings from those flights. Moreover, during the stress scenarios of VP3, VP5, and VP6 the recording system had to be restarted resulting in two files and an interruption of about 1 min to 2 min. Those three files needed to be merged and cut manually. Unfortunately, this took quite some time and model training was started before finishing these mergers. This resulted in an exclusion from the training and validation set for this thesis but could be considered in later research. During all other files no larger distortions, noise, or artefacts could be detected. Two to three participants each had an electrode disconnect during flight or artefacts from movement. Those superimposed signals should not influence the detection performance of the deep learning network noticeably. Moreover, those are more realistic dropouts and interferences which the model should build up a tolerance against.

6.2 Data Exploration



(a) Zoomed in Section with artifacts.

(b) Further zoomed in on the middle artifact.

Figure 6.6: Examples of detected artefacts within the ECG signal. a: Zoomed in section of VP9's stress scenario. The three red spikes in the heart rate on the third plot are motion artifacts falsely detected as R-peaks. b: The same artefact as in a, but zoomed in. The top plot shows how the ECG signal behaves. The following R-peak, which is cut off by the maximum signal range, is only partially visible and was not detected.

After a single record CSV file had been created the data was ready to be fed into the machine learning system. At this point, a small analysis was conducted, not only to get a good split for training and validation, but also to get an impression on how ECG and stress level data fit together. This analysis shall help to get a better understanding of how the data is structured and what it consists of e.g. in terms of distribution. Therefore, the stress levels were plotted as histograms to show those labels' distribution. Another approach could be correlating heart rates calculated from the ECG with the labels at the corresponding time step. This would assume that there must be a correlation between the heart rate and stress level. Therefore, this method should be taken with care as the representation of stress in the ECG signal is impacted

by other factors too, e.g. changes within the QRS-complex morphology.
Moreover, as stress has several influences and is perceived differently by each individual editing or augmenting the stress data could falsify the actual truth. Although the personal feedback might be distorted and biased, any changes could make the actual correlation between physiological reaction and perceived stress even worse. Additionally, the question of how to modify these values not only by quality but also by quantity would arise, so whether to increase or decrease the stress level and if so by which amount. It could be argued that any lapse or miss by the flight crew is already so critical that the stress level should be an immediate maximum. But on the other hand, the goal is to match a physiological reaction. Although a pilot makes lapses or mistakes because of a workload overflow, this does not mean that this person is really under stress or more precisely is not physiologically experiencing a higher degree of stress. For all these reasons no further modification of the data was done. But this is something to keep in mind when not only looking at the data but also training the model and evaluating its performance. Instead, complete files may be excluded from the training or validation data set. This might be the case if a certain participant's labels deviate to either side of the distribution.

6.2.1 Data from Single Pilot Study 2022

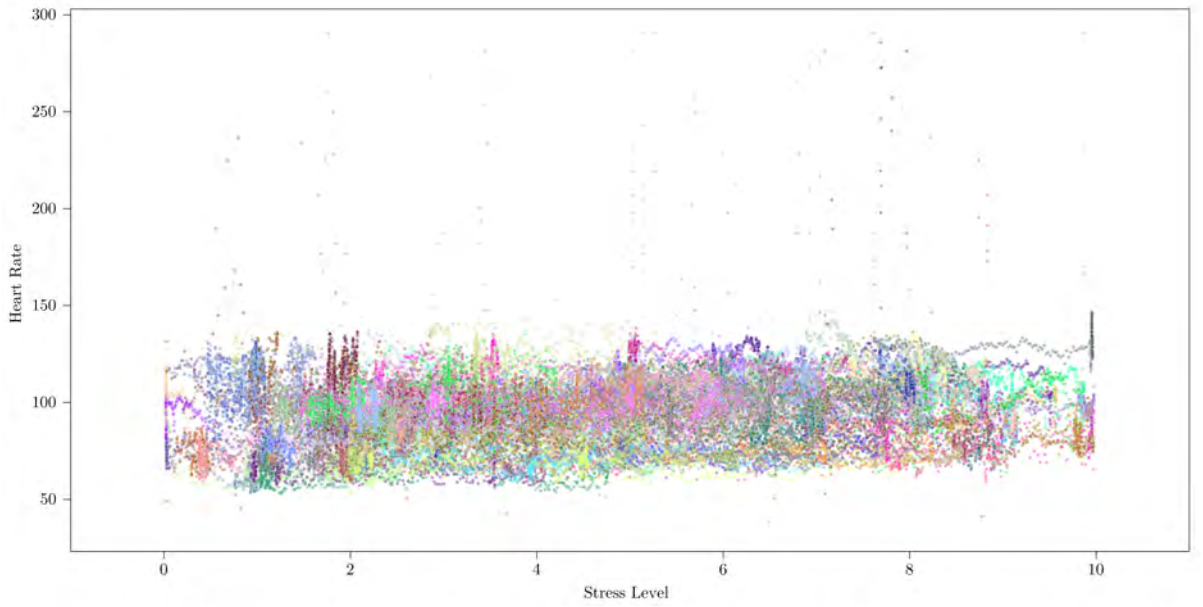
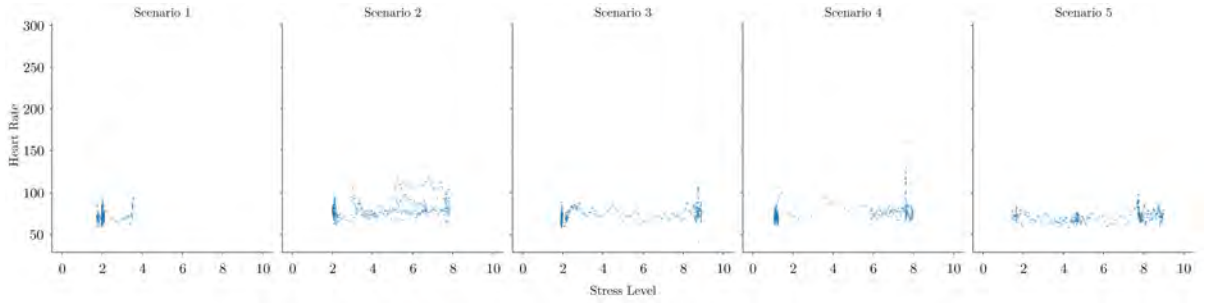
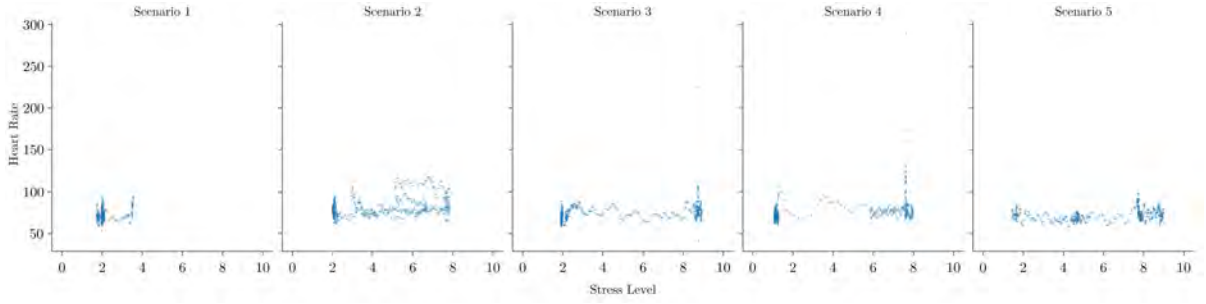


Figure 6.7: Correlation between stress level and heart rate over all scenarios. Each color represents a unique combination of participant and scenario. This plot serves a purely illustrative purpose. In the print version it might be difficult to distinguish between some points in certain areas. The digital version contains easier-to-distinguish vector graphics and the data are available digitally as well.

For this data, only one ECG lead and the digitalized stress track were available. As a first step, in Figure 6.9 for each scenario a histogram of the stress levels was plotted. The purpose of this plot is to get an understanding on how those labels are distributed. Therefore, in this step the separation into single and dual pilot operation during the simulator study was neglected. As a next step the distribution of calculated heart rate against the momentary stress levels was plotted.
Whereas the first plots in Figure 6.7 took all stress levels of each file Figure 6.9 contains only the stress level at the moment of the R-peak. So these only contain a fraction of the original amount of stress level labels. Anyhow, this chart here will give an overview of the distribution from both heart rate and corresponding stress level. The heart rate was calculated using WFDB's



(a) Without jitter.



(b) With jitter.

Figure 6.8: Example from SPO study of correlating stress levels and heart rates from VP09 for each of the five scenarios. a: shows the data points without. b: has some jitter added to the data in order to see more densely populated areas.

processing toolkit for Python. By detecting the R-peak of the ECG the IBI can be calculated and hence the momentary heart rate. It is subject to the assumption that there is a tight, in some cases even just linear, correlation between stress and heart rate. In reality this might not or not always be the case. In chapter 2.3 this is elaborated more thoroughly. For the sake of understanding the data this is a good abstraction and approximation for a first look. But it should not be used for any further interpretation in terms of correlation. Anyhow, it shows that there might be a few artefacts within the ECG signal. As a heart rate above $200\,\mathrm{min}^{-1}$ is very unlikely and above $250\,\mathrm{min}^{-1}$ would be considered ventricular fibrillation and results in unconsciousness and can lead to sudden cardiac death. Those values most likely come from a spike artefact within the ECG signal which was interpreted as R-Peak by the WFDB algorithm as shown in Figure 6.6. Most values are situated between $60\,\mathrm{min}^{-1}$ to $120\,\mathrm{min}^{-1}$ with a significant spike around $95\,\mathrm{min}^{-1}$ to $105\,\mathrm{min}^{-1}$. There is slightly higher amount of values to the left of that distribution than to the right. The stress level distribution on the other hand contains several spikes. The most dominating ones are at about 2 and 5. The overall amount seems to be decreasing with increasing stress level and then dropping around stress level 9 only to then peak again at maximum stress. Anyhow, all stress levels seem to be represented over all scenarios although not equally. In an ideal case for training the deep learning model the labels would be distributed equally among all stress levels. Nonetheless, there should be a sufficient amount of all labels throughout the data.

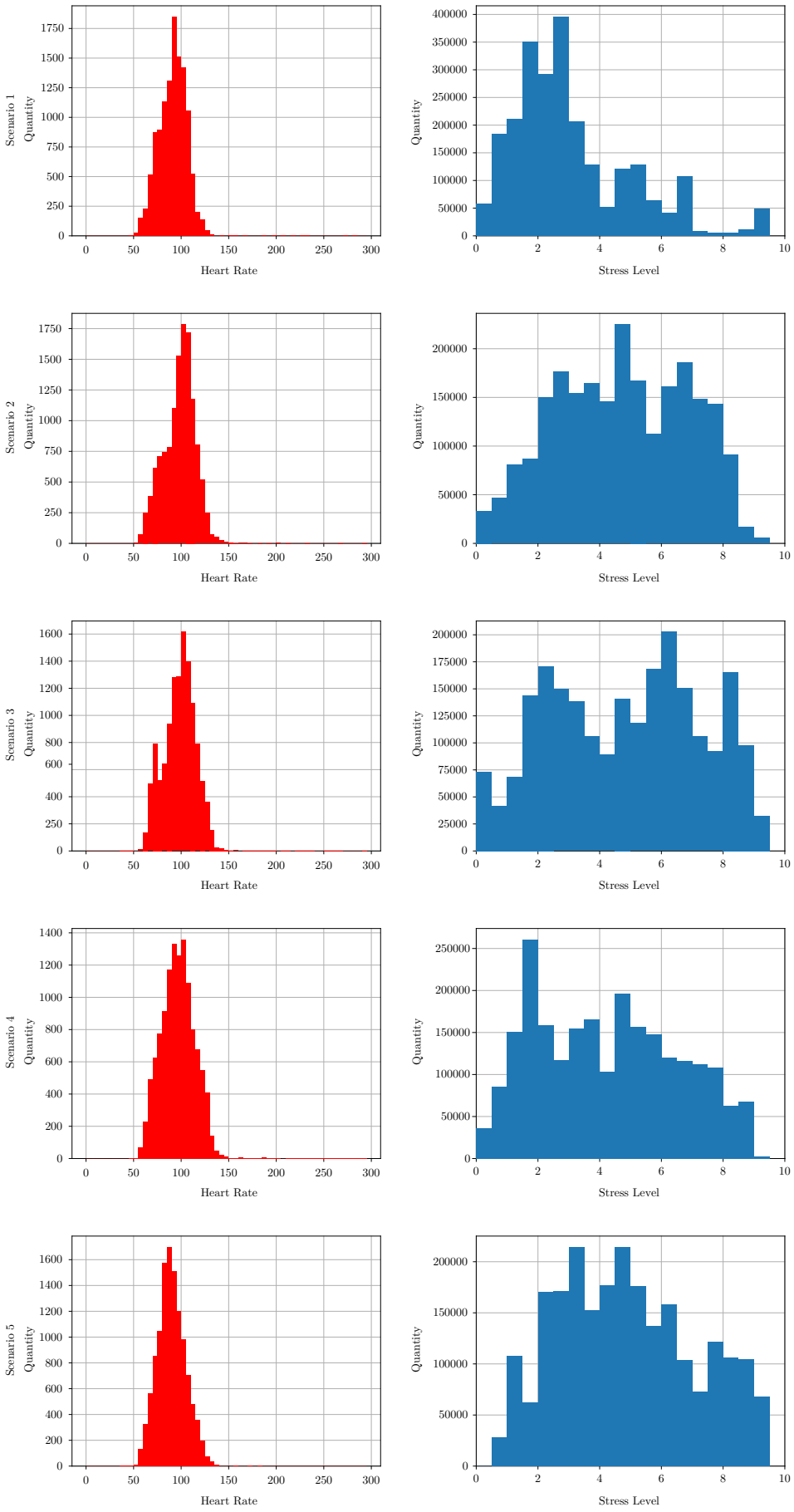


Figure 6.9: Distributions of all stress levels and heart rates separated by scenarios of the SPO Study.

The final step is getting an impression on how the momentary calculated heart rate correlates with the stress level feedback. Therefore, every heart rate was plotted against the corresponding interpolated stress level for every participant. Two versions were created: one with only one color for all files and one where each color represents an individual file. The first one shows the overall distribution among all participants and scenarios. Whereas the second gives more details about the individual contribution to that distribution. Although this looks cluttered it gives a good overview of the overall correlation. But to further investigate how this distribution is composed each individual contribution was plotted in a grid plot. Each column representing a scenario and each row another participant. An example row is provided in Figure 6.8 and the whole distribution is given in Appendix A.
Overall it can be noticed that at least for those scenarios that should trigger stress a slight increase in stress levels is visible. But there is quite a significant skew in the data: higher stress level values occur only rarely. Therefore, further cutting and excluding sections with average stress levels could change the overall distribution but also decreases the overall amount of data available for training. Analysing how model detection performance changes when excluding certain areas or complete files is an interesting question for further research. Considering the overall already low amount of data no further modifications or exclusions of the original files were made. Moreover, it might also be discussed further whether those labels represent actual stress of each participant correctly, see section 10.1.

6.2.2 Data from Limits of Human Performance Study 2022

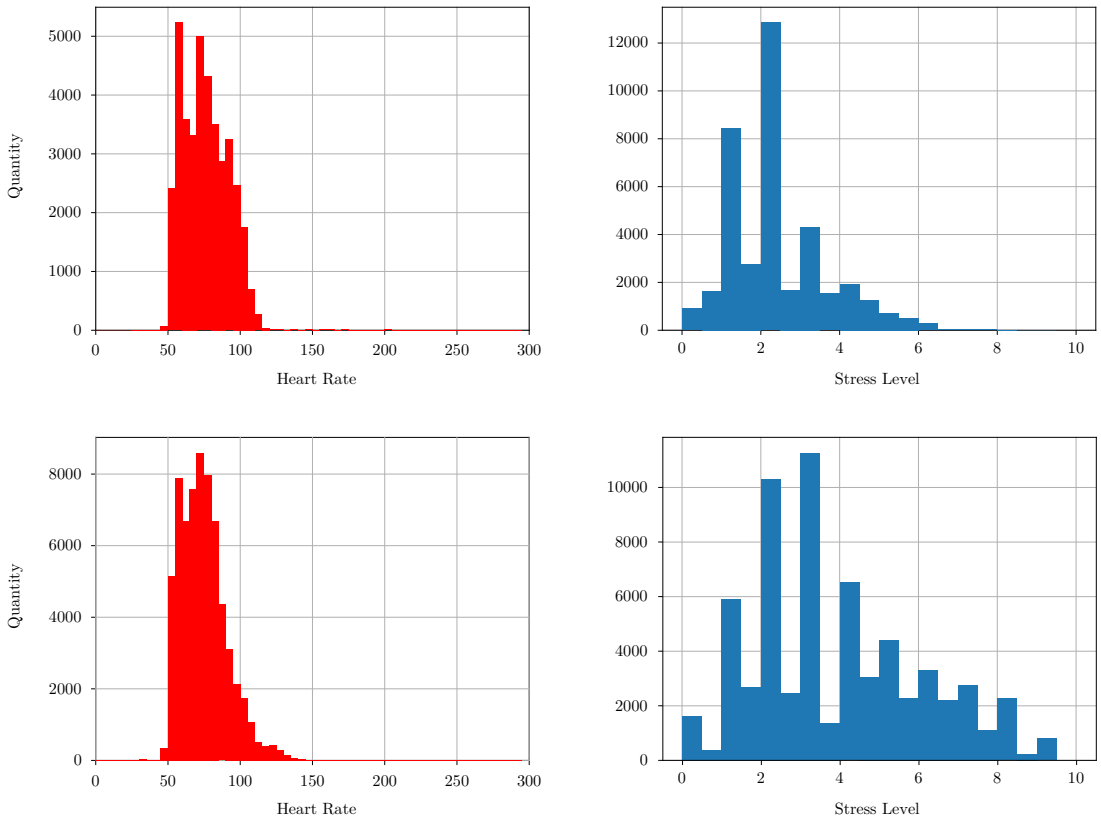


Figure 6.10: Distributions of all stress levels and heart rates separated by scenarios. The left column shows histograms of the stress levels, the right column histograms of the heart rates. The top row corresponds to the baseline scenario, the bottom row to the stress scenario.

In this study seven ECG leads were available, so one had to be chosen for further analysis and

model training. Generally, a good overall representation of the heart is obtained with channel II and was therefore taken. To explore this data set the same approach was used as with SPO study's data set. Moreover, stress level feedback collected two metrics, one value representing any change on the stress level slider and one when the confirm button was pushed. In general, those confirmed values will always be represented in the slider values too. As there had been several cases of exceeding the 2-minute window, the slider stress values were considered for this analysis as well as for model training later. There are other different approaches possible which will be further discussed in section 9.2.

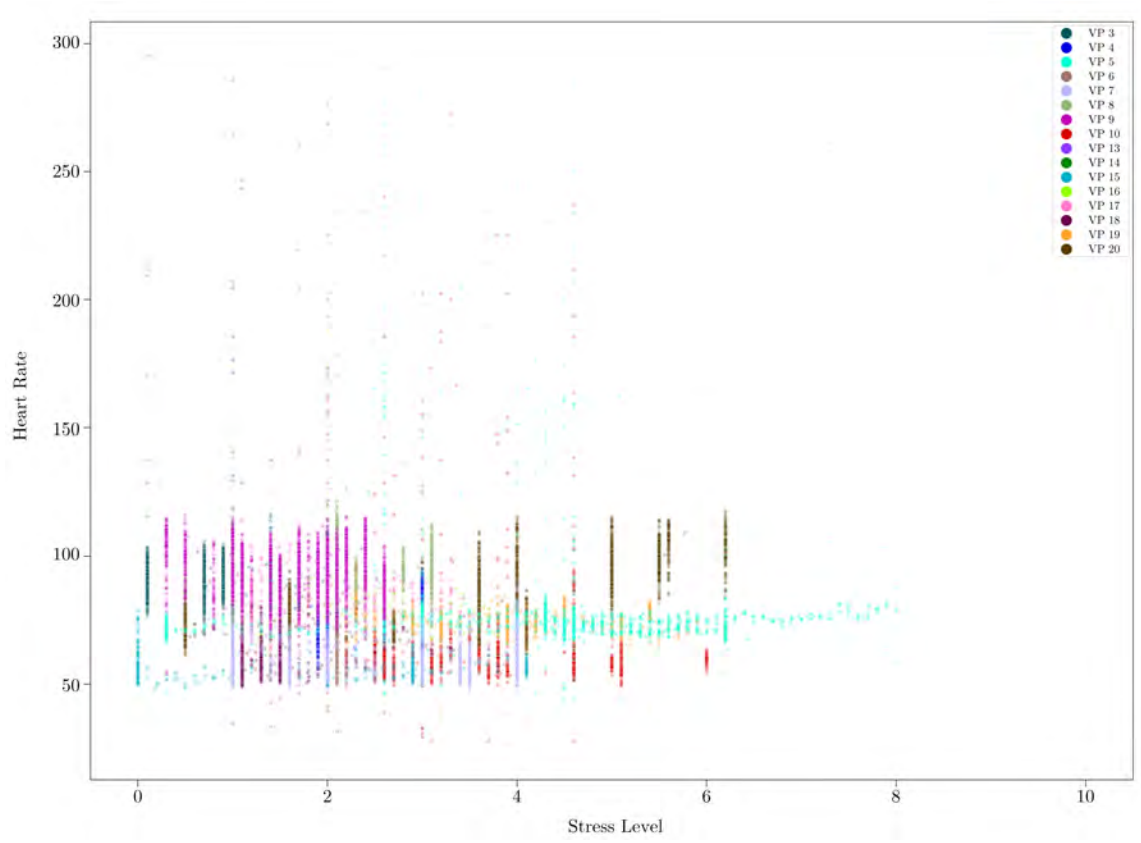
In a first step histograms were created showing the distribution of stress levels and heart rates derived from ECG, see Figure 6.10. The data has been separated into both scenarios for better distinction. For the baseline scenario the highest amount of stress levels are located around one and two with another small peak at three. Only a few are around four and even less with higher stress levels. For the heart rate, all are distributed around a normal frequency of $50\,\mathrm{min}^{-1}$ to $100\,\mathrm{min}^{-1}$.

In the stress scenario stress levels are also more distributed in the higher levels. With the stress level increasing the frequency of those labels drastically decreases. There are four distinct peaks at stress levels one, two and three and four. From above four the distribution seems to linearly decrease until the maximum stress level. At least there are a few maximum stress levels represented, in comparison to baseline scenario where there are, unsurprisingly, none. On the contrary, the heart rates are almost indistinguishable to those of the baseline scenario. There are just a few more heart rates above $100\,\mathrm{min}^{-1}$ up to about $145\,\mathrm{min}^{-1}$.

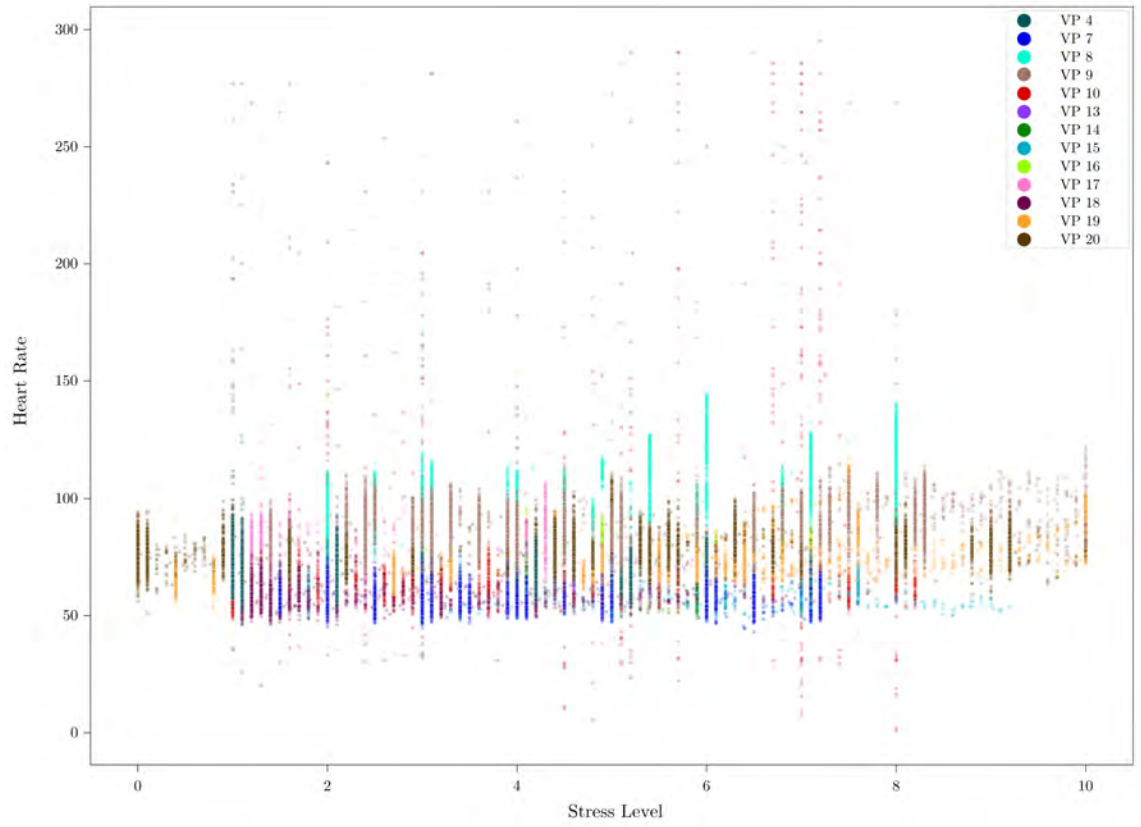
This rose the interesting question how those two metrics correlate and which participant contributed in which way. Therefore, a scatter plot for both scenarios was created plotting stress levels against the corresponding heart rates, which is shown in Figure 6.11. This shows that only five participants overall said that their stress level was at 8 or above. Only three participants had a maximum stress level during this scenario. Any heart rates above $250\,\mathrm{min}^{-1}$ can be ignored as those are most likely movement artefacts. Interestingly, it seems there is hardly any correlation between the stress level feedback and heart rates at all. A clear pattern cannot be seen from those plots. Additionally, most stress levels seem to be independent of heart rates as there is a broad band between $50\,\mathrm{min}^{-1}$ to $100\,\mathrm{min}^{-1}$ no matter the stress level. The same applies for baseline scenario, although higher stress levels are not that represented there is a huge cluster between stress levels one and four. The same spectrum of heart rates can be found as in the stress scenario. This at least allows the conclusion that there is no clear and simple correlation found within those two parameters.

In order to get a better overview of each individual's contribution, a scatter plot of stress level against heart rate for each participant was created. An example of VP15 is shown in Figure 6.12, plots from all participants can be found in Appendix A. Unfortunately, this does not disclose much more information besides which participant had a maximum of stress. Again, no clear correlation could be found in neither plot.

Overall, it can be observed that between the baseline scenario and the stress scenario the stress levels are shifted towards higher stress levels. This indicates that there was some kind of stress present during the stress scenario. Only a few participants reflected to be on maximum stress level whereas the majority is in a broad span between two and eight. Most stress levels are distributed within a broad band of heart rates, e.g. VP15 has heart rates between $55\,\mathrm{min}^{-1}$ to $70\,\mathrm{min}^{-1}$ at stress level 6. From these plots a clear correlation between heart rate and stress levels could not be deduced.



(a) Baseline Scenario.



(b) Stress Scenario.

Figure 6.11: Distributions of stress level against heart rate separated by scenarios from the LoHP Study. Each color represents an individual participant. The top diagram represents the baseline scenario, the bottom diagram represents the stress scenario.

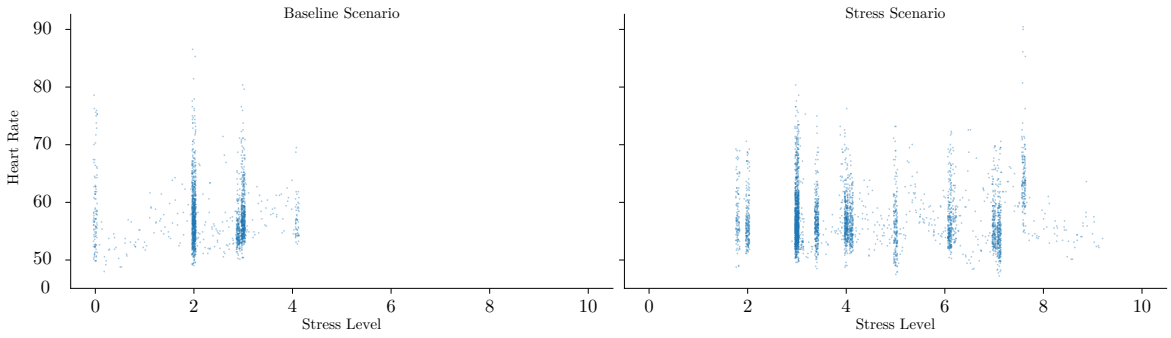


Figure 6.12: Example from the LoHP Study of VP15 of correlating stress levels and heart rates. The data has been jittered in order to improve visibility of more densely populated areas.

Stress Classification Stream 7

With the databases created almost everything is ready for training a classification algorithm to detect stress. A suitable algorithm for time series predictions are Long Short-Term Memory networks. As a part of RNN it benefits from carrying information from one time step to the next. This makes these networks promising for analyzing a time series where the next state depends on the previous. When looking at the nature of ECG signals these have a high dependency on their previous state as well. Over the course of this chapter the search for a neural network architecture as well as the model training process will be discussed in more detail. There are several parameters which determine not only the shape of the network but also model prediction performance.

7.1 LSTM Model Architecture

When it comes to designing artificial neural networks there are no deterministic formulae which direct the shape of it. A common approach is to start with a very simple classification stream, evaluate the performance of it, then improve it based on performance results. [46, p. 29] With growing network and database sizes this becomes a very tedious and time-consuming process. Therefore, a first simple LSTM architecture was created based on simple ideas. First, the ECG signal will be fed as input with one single datum at a time. Another approach to this would be to feed the network a short time period of that signal at once e.g. splitting the signal into chunks of 10 s. [47] Additionally, another option would be using a sliding window of a certain time length. With that, a certain timed window is continuously progressed through the whole length of data. Anyhow, these are all viable options and one had to be chosen for this first attempt. With the aspect in mind to keep it quite simple for the first try, hence the single signal was chosen.

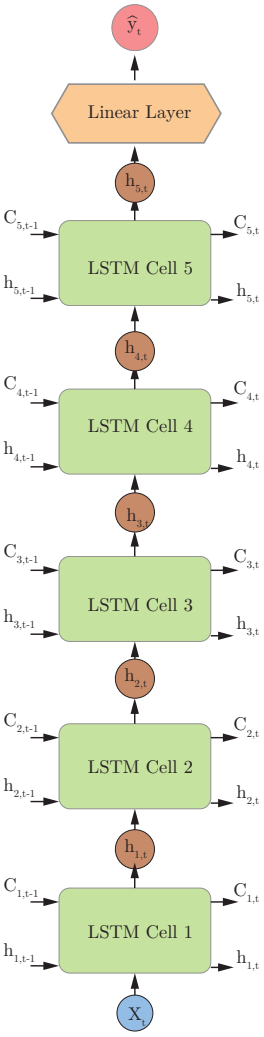


Figure 7.1: Schematic structure of the LSTM architecture with hidden Layer as defined in Section 7.1.

Modules	Parameters			
	Hidden		Cell State	
	w	b	w	b
LSTM 1	262 144	1 024	1 024	1 024
LSTM 2	65 536	512	131 072	512
LSTM 3	16 384	256	32 768	256
LSTM 4	4 069	128	8 192	128
LSTM 5	1 024	64	2 048	64
Linear	16	1		
Total				528 273

Table 7.1: Overview of trainable parameters of the used LSTM architecture. Trainable parameter are separated into weights (w) and biases (b).

First Model Architecture

LSTM Cells: 5

LSTM Cell 1:	Hidden Layers	=	256
LSTM Cell 2:	Hidden Layers	=	128
LSTM Cell 3	Hidden Layers	=	64
LSTM Cell 4:	Hidden Layers	=	32
LSTM Cell 5:	Hidden Layers	=	16
Linear Layer:	In size: 16	Out size:	1
Optimizer:	Adam		
	learning rate:	$l_r = 1 \cdot 10^{-3}$	
	weight decay:	$w_d = 0$	
Loss function:	Mean Squared Error		

The first model architecture consisted of five stacked LSTM cells. Each of these cells, besides the first, takes as input the hidden layer of the previous cell. The first cell has a single value, the ECG signal, as input and consists of 256 hidden layers. Each following cell has half as many hidden layers as the previous cell. Lastly, after the fifth LSTM cell a linear layer is applied. This layer merges down the hidden layer vector of the last LSTM cell into a single output value and returns it as prediction (see 7.1). This architecture has a total of 528 273 trainable parameters of weights and biases, the exact parameter distribution can be found in Table 7.1.

Linear Layer

$$y = xA^T + b \tag{7.1}$$

x: Input Vector
A: Weight Matrix, learnable weights with shape [Out features x In features], initialized with $\mathcal{U}\left(-\sqrt{k}, \sqrt{k}\right)$, where $k = \frac{1}{infeatures}$
b: Bias, learnable bias of shape [Out features]

Source: PyTorch Documentation [48]

Stacking these LSTM cells shall allow a better abstraction from the ECG signal to the associated

stress level. This allows more hidden neuron cells to maintain a certain knowledge of the input data. In the next steps, both could be modified, the amount of stacked LSTM cells as well as the number of hidden layers. This can be done either individually, e.g. only changing the amount of hidden layers while maintaining the number of LSTM cells, or with both at the same time.

Another common practice is to preprocess the raw ECG signal by detecting the R-Peak. [43] Usually certain features such as the heart rate or the frequency distribution are extracted and then used as input for deep learning algorithms. This was opted out from because of two reasons: First, although **F**ast **F**ourier **T**ransformation (FFT) and other methods for extracting features are quite fast it is at least questionable and needs to be proven that this still retains the original soft real-time requirement 1. Second, an interesting question at hand is also if the deep neural network is capable of learning such interpretation on its own. As those deep neural networks are known for their good capability to abstract and extract knowledge even from very noisy or distorted data. [49] Therefore, no specific features were extracted from the original ECG signal.

7.1.1 Selection of Loss Function

When looking at the options available for loss functions a simple and applicable loss function is **M**ean **S**quarred **E**rror (MSE) also known as quadratic loss or L2 loss. So for this application the stress level is seen as the ground truth and hence y_i, and the prediction made by the LSTM model is $\hat{y}_i$.

Mean Square Error Equation

$$MSE = \frac{1}{N}\sum_{i}^{N}(y_i - \hat{y}_i)^2 \tag{7.2}$$

This is then subtracted from each other and squared. All squared errors of the signal are then summed up and divided by the signal length. That gives the loss for each file input into the model. For the training data this is then forwarded to the optimizer while the validation loss is only output and saved.

7.1.2 Selection of Optimizer

The task of an optimizer is to find minima, ideally a global minimum[1]. A common stochastic optimizer is adaptive momentum estimation also known as 'Adam'. This is an established optimizer which is known for being efficient in computing and straight forward to implement. [50] When implementing this optimizer with PyTorch in the first model architecture only a learning rate with an initial value of $l_r = 1 \cdot 10^{-3}$ was specified, no weight decay. The first few iterations with the first architecture revealed that the model might overfit to the training data. Therefore, a smaller learning rate was chosen in order to reduce suspected overfitting. This could be an issue as each file is presented to the model individually due to host computer's RAM capacity. Therefore, the learning rate was adapted to $1 \cdot 10^{-5}$ instead of $1 \cdot 10^{-3}$. To further help reduce overfitting an L2 penalty was implemented with a weight decay of $1 \cdot 10^{-8}$. The losses with these hyperparameters seem to have improved but still showed some overfitting behaviour after a few iterations. Hence, in a third attempt the learning rate as well as weight decay were further reduced to $l_r = 1 \cdot 10^{-6}$ and $w_d = 1 \cdot 10^{-8}$.

[1] Depending on the research question to solve this could also be used to find a global maximum, e.g. when looking for maximum revenue.

7.1.3 Training and Validation Data Split

When training a classification algorithm it is advised to separate the existing data into two sets. One data set is used for training and the other for validating the model performance. In this sense training means letting the model predict classifications for a data set, calculate the loss and then use the prediction and loss to change the models parameters to try to reduce the loss. Validation means that the same first two steps are made, but the model is not tuned or modified based on the results. This is why sometimes it is said that validation data is not 'known' by the model.

Mode	VP	SC1	SC2	SC3	SC4	SC5
Single Pilot	VP1	T	T	V	T	V
	VP2	T	T	T	T	T
	VP3	nV	nV	nV	nV	nV
	VP4	nV	nV	nV	nV	nV
	VP5	T	V	T	T	T
	VP6	V	T	T	T	T
	VP7	T	T	V	T	T
	VP8	T	T	T	V	T
	VP9	T	T	T	T	V
	VP10	nV	nV	nV	nV	nV
	VP11	T	T	T	T	T
	VP12	T	V	T	T	T
	VP13	T	T	V	T	T
	VP14	T	T	T	V	T
	VP15	V	T	T	T	T
Dual Pilot	VP16	T	T	V	T	T
	VP17	T	T	T	T	V
	VP18	T	T	T	V	T
	VP19	T	T	T	T	V
	VP20	T	V	T	T	T
	VP21	T	T	T	V	T
	VP22	V	T	T	T	T
	VP23	V	T	T	T	T
	VP24	T	V	T	T	T
	n	21	21	21	21	21
	80%	16.80	16.80	16.80	16.80	16.80
	Rounded	17	17	17	17	17
	Val. Split	4	4	4	4	4

Table 7.2: Data Split for Single Pilot Operation Study. VP: *Versuchsperson* (German for 'test subject'); SC1 to SC5: Scenarios 1 to 5; T: Training data set; V: Validation data set; nV: invalid data set; Val. Split: Validation split.

Furthermore, the model state is not saved or stored in any way after doing predictions on validation data. Overall a good rule of thumb for train-to-test separation is around 80:20, respectively. So if there are 100 equally usable files 80 would be used for training the model and 20 for validation purpose.

A more general discussion is whether whole participants should be used exclusively for training and/or validating. It can be argued that this could be beneficial as well as disadvantageous for the overall model performance. One factor that might require this split is if the model should be able to detect stress without knowing that individual at all. But on the contrary, it would probably improve the overall detection performance if the model already had been trained on some data of that individual. One good reason for this is that each person has a different ECG pattern. Actually, it is so individual that it has already been shown that persons can be distinguished and recognized only by their ECG signal. [51] Hence, if the model is already familiar with at least a baseline it can be assumed that this improves performance for further detection. But this still has to been proven in further research. Anyhow, it can be assumed that if such system would ever be applied in a real world environment that for each flight crew member some kind of baseline record was taken and the model trained with it. This would result in the system knowing at least some resting signal before analysing for stress. For the sake of this work a good overall performance was more desirable than being capable of analysing unknown persons. Therefore, the splits made did not separate whole participant data sets into validation.

7.1.3.1 Split for Single Pilot Study Data

From the SPO study 105 data files with a total of 11 851 200 rows which is equivalent to 658.4 min have been acquired. Their lengths vary between 93 000 (scenario 3) and 142 500 (scenario 1) rows of data with an overall average of ≈6.27 min. During this study, five different scenarios where flown as well as two different operation modes of flight crews. In order to pay tribute to these different factors the data split should be made with each of these groups. So, every single pilot mode scenario 1 files will be separated accordingly. There are eleven valid sets from single pilot and ten from dual pilot operation. Each of those had all five scenarios covered, which made selecting a split easier. For each single and dual pilot an online randomizer tool was used to create some random selections. These have been slightly adjusted to result in the split chosen in Table 7.2.

7.1.3.2 Split for Limits of Human Performance Data

For this study, the data split was done in a similar fashion as with the previous data set. The only thing that was looked after is that from each participant no more than one set of either resting, baseline or stress was within the validation set. This was done in order to reduce bias towards individual responses and feedbacks. From all 18 participants a resting ECG was successfully recorded and obtained. Due to unsuccessfully recordings two data sets of the baseline scenario as well as five of the stress scenario had to be discarded. This resulted in 18 resting ECG files, 16 files from the baseline scenario and 13 from the stress scenario. A total of 47 files with 1 699.90 min recording time from the LoHP study have been acquired. Lengths range from 66 546 and 321 252 rows (mean: 13.73 min), 530 425 and 682 050 rows (mean: 33.39 min), and 884 256 and 1 758 042 rows (mean: 70.65 min) for resting, the baseline scenario and the stress scenario, respectively. The mean overall scenarios and resting data is 39.25 min.

To maintain an about 80:20 split four data sets of each resting and the baseline scenario as well as three of the stress scenario were separated for validation. Again, an online random number generator was used to generate four numbers between 1 and 18. The resulting split is shown in Table 7.3. Due to memory limitations on the host computer, files longer than 200 000 rows were split into separate files. As the signal is fed into the system one at a time this should make no difference during training.

Subject	Resting	Baseline	Stress
VP1	V	nV	nV
VP2	T	nV	nV
VP3	T	T	nV
VP4	T	V	T
VP5	T	T	nV
VP6	T	V	nV
VP7	V	T	T
VP8	T	V	T
VP9	T	T	T
VP10	T	T	V
VP11	nE	nE	nE
VP12	nE	nE	nE
VP13	T	T	T
VP14	V	T	T
VP15	T	T	V
VP16	T	T	T
VP17	V	T	T
VP18	T	T	V
VP19	T	V	T
VP20	T	T	T
n	18	16	13
80%	14.4	12.8	10.4
Rounded	14	12	10
Split	4	4	3

Table 7.3: Data Split for Limits of Human Performance Study. VP: *Versuchsperson* (German for 'test subject'); T: Training data set; V: Validation data set; nE: non-existant data set (participant cancelled on short notice); nV: invalid data set.

7.2 LSTM Model Training and Revisions

The software script was written in Python using the PyTorch framework and was executed on a host computer which was equipped with an NVIDIA RTX A5000 with 24 GB RAM GPU. Although very capable for this task, the GPU RAM limited not only the amount of data that could be loaded at the same time but also the model's architecture. This lead to a reduction in LSTM network hidden-layer size as well as the aforementioned file slicing. Slicing or batching the data into manageable chunks is a common practice and does not change the model's performance. What significantly influences performance is the size of the hidden layers. Additionally, the amount and length of the data sets determined runtime significantly. As for the data obtained from the SPO study one iteration over all files took about $2\,\mathrm{h}$ including validation. While for the data from the LoHP study one iteration took about $6\,\mathrm{h}$ for the same process. The increase can be traced to the increased amount of files and data. There might still be room for improvement as memory allocation could be more streamlined.

During the course of training resulting losses for each file were exported into an CSV file per iteration. In order to keep an overview over the changes those losses were plotted frequently. When plotting those losses it was observed how the performance changed over those iterations. When it became obvious that the model is either over- or underfitting or some stagnation, the training process was aborted. In this case, first the learning rate was adapted to be larger or smaller.

7.2.1 First Training Approach

For the first model, hyperparameters such as the learning rate were left at default values provided by PyTorch. Therefore, a learning rate of $l_r = 0.001$ was set with the mentioned architecture. At this time, only data of the SPO study was available. Hence, model training and improving was only based on this data set's performance. This setup ran initially for ten iterations and then the losses were evaluated to see whether it appears worthwhile to continue. From each file's MSE the mean loss, median, and sum of losses across all files is shown in Figure 7.2.

In this graph, it is clearly visible that the error initially increases significantly. Interestingly, the validation set median continues to decrease up to the third iteration and then climbs again not reaching the initial value again. All other values increase, then decrease again to level off over the course of the next iterations, but not decreasing below the initial loss. This indicates that the model could change the values within the hidden layers too drastically. Each file within the training set is presented to the model individually and its hidden layers are updated after each file. A too large learning rate and too quick adaptation to the training data might result in a too fast adaptation to the first files presented. These two factors both indicated reducing the learning rate and implementing an L2 regularization.

7.2.2 Second Model Training

Now the learning rate was decreased to $l_r = 1 \cdot 10^{-5}$. Furthermore, another hyperparameter was implemented to help prevent overfitting and too fast weight changes during training. Weight decay for Adam optimizers is an additional regularization of the hidden layer weights. This L2 regularization helps in this way to prevent overfitting to the training data. This is done by adding a term to the loss as shown in Equation 7.3. The weight decay was initially set to $w_d = 1 \cdot 10^{-6}$. With these parameters set a new learning approach was started. The same loss results are shown in Figure 7.3.

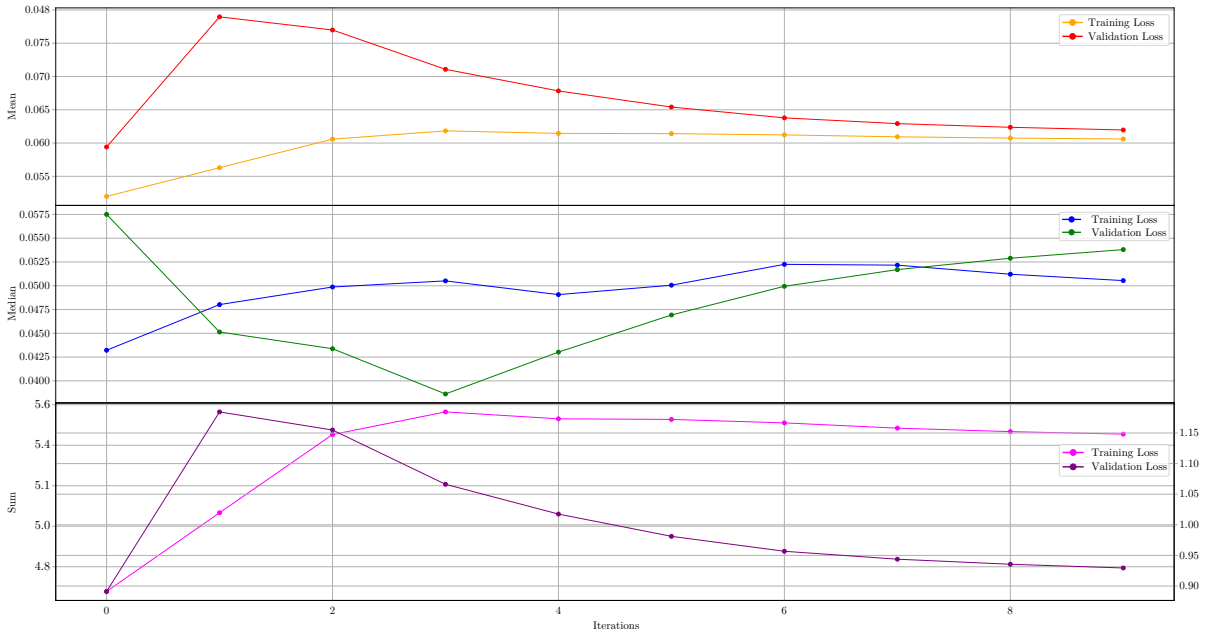


Figure 7.2: MSE loss of the first model architecture. $lr = 0.001$, $w_d = 0$. The top plot shows the mean MSE loss of the training data (yellow) and validation data (red). In the middle is the median loss in blue and green, as well as the sum of losses on the bottom in magenta and violet of the training and validation data, respectively.

Equation of L2 Regularization

$$final\ loss = loss + w_d * \sum \frac{all\ weights^2}{2} \tag{7.3}$$

Overall, the model seems to have converged after the 9th to 10th iteration. Over the first three initial iterations the loss continuously decreases. The validation median stagnates at iteration 4 just to increase again and level off at about the same value as initially started. Moreover, the mean and the sum continue to decrease but then also increase again but not so significantly. After that they level off and do not change over further iterations. This behaviour could also indicate that the optimizer tries to approach a certain minimum but overshoots. In that case an actual minimum cannot be reached closer, which indicates that smaller changes within the hidden layer parameters are needed then currently possible. Therefore, it was tried to further reduce hyperparameters, namely $l_r = 1 \cdot 10^{-6}$ and $w_d = 1 \cdot 10^{-8}$.

7.2.3 Final Model Architecture

With those settings the model training was restarted for 100 iterations on both data sets. The MSE losses from each file were tracked over each iteration. For each iteration, the mean of all losses, the median of all losses and the sum of both were computed and plotted as shown in Figure 7.4. Compared to the previous run, shown in Figure 7.3, the overall evolution along iterations appears to be similar. Especially the overshoot in the validation median seems to be almost the same, but at a finer scale. The losses of the last iteration on the validation set show that this version finds a lower mean loss compared to the previous one. The sums of losses from the training data set converge to a value of about 5.16 after dropping sharply from an initial loss of about 9.9. The validation loss starts to slightly increase again after the 70th iteration and continues to raise, but more slowly. This is, again, an indication for overfitting of the model. Further decreasing hyperparameters would not make any sense with these results.

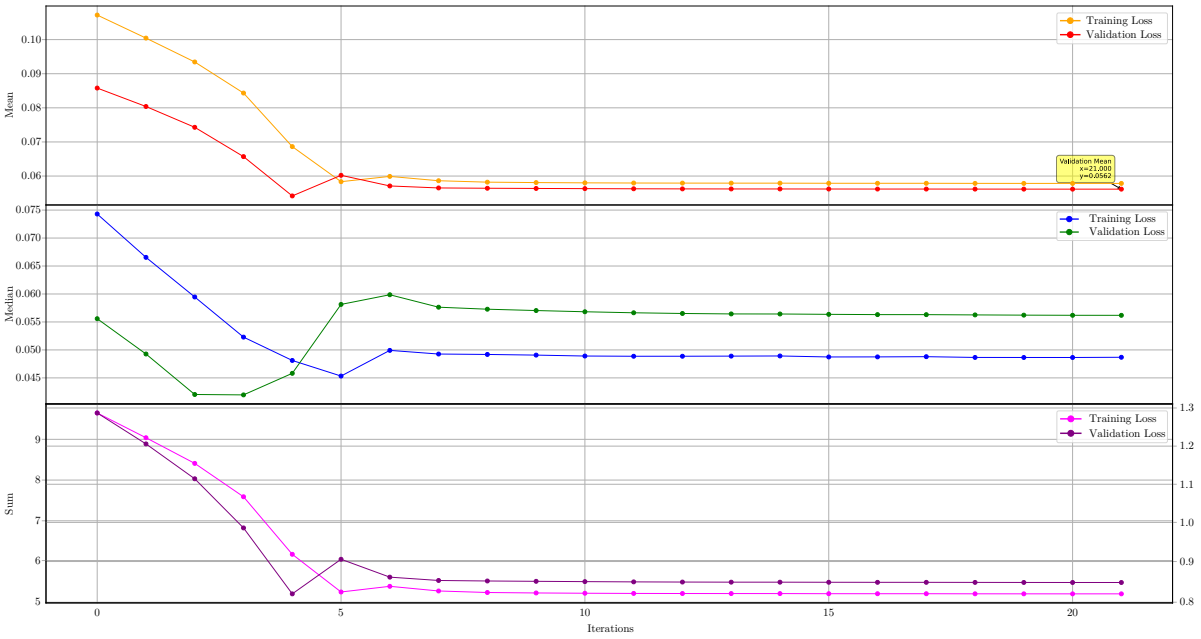


Figure 7.3: MSE loss of the second model architecture. $lr = 1 \cdot 10^{-5}$, $w_d = 1 \cdot 10^{-6}$. The colors and their meanings in this plot are the same as in Figure 7.2

Doing so might result in an even finer graduation but find the same hidden and cell state values eventually. One way to prevent overfitting now would be reconsidering the initial data split into training and validation sets. As it seems that there are some information within the current validation set that the model is unable to approach.
The LoHP study's data were next. The data were used to train and optimize the current model with the found hyperparameters as described in Section 7.2.3. Also, 100 iteration were done on this data and the same metrics captured as with the previous data. This is shown in Figure 7.5 and appears quite differently than with the other data. Looking at the development of the mean loss shows that both, training and validation loss, decrease quickly first and then level off. Contrary to the previous data set the validation mean does not increase again. A similar behaviour is found with the sum of all losses. Doing more iterations could reveal whether there will be an overfitting or not. What is very interesting though is the course of the mean validation loss value. The median is decreasing until the 20th iteration and then starts to increase at about the same rate again. This pattern repeats two more times and ends up with a higher middle value than during initial iteration.

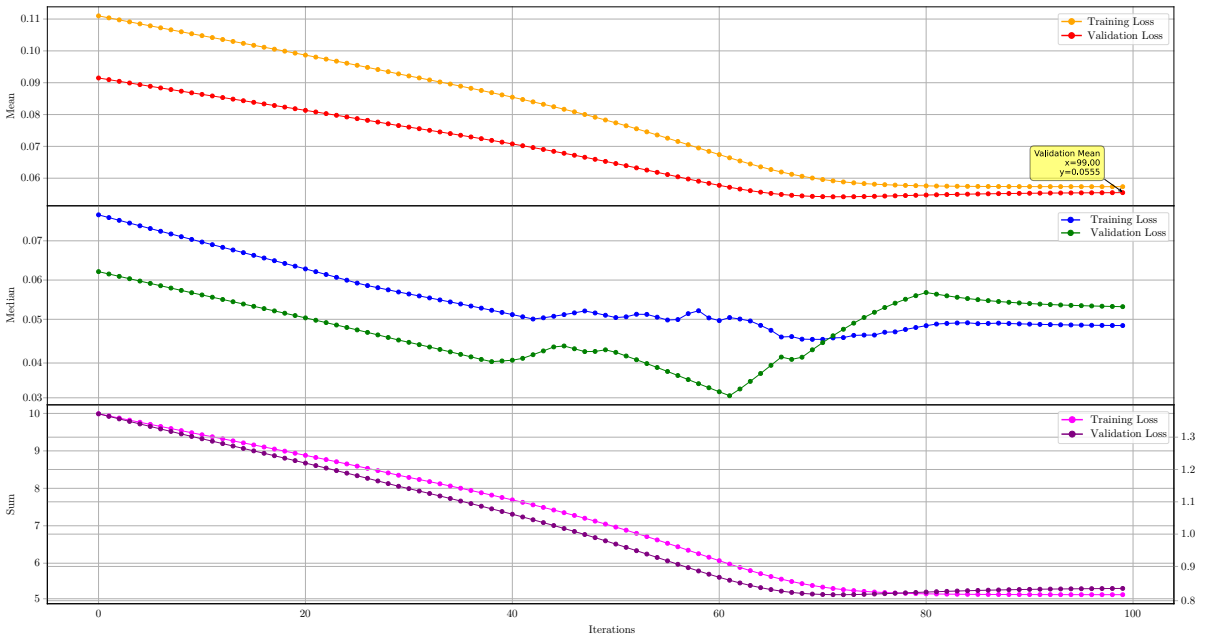


Figure 7.4: MSE loss of the final model architecture and SPO data. Plots of mean loss, median and sum of all training and validation files of the SPO study. $lr = 1 \cdot 10^{-6}$, $w_d = 1 \cdot 10^{-8}$. The colors and their meanings in this plot are the same as in Figure 7.2

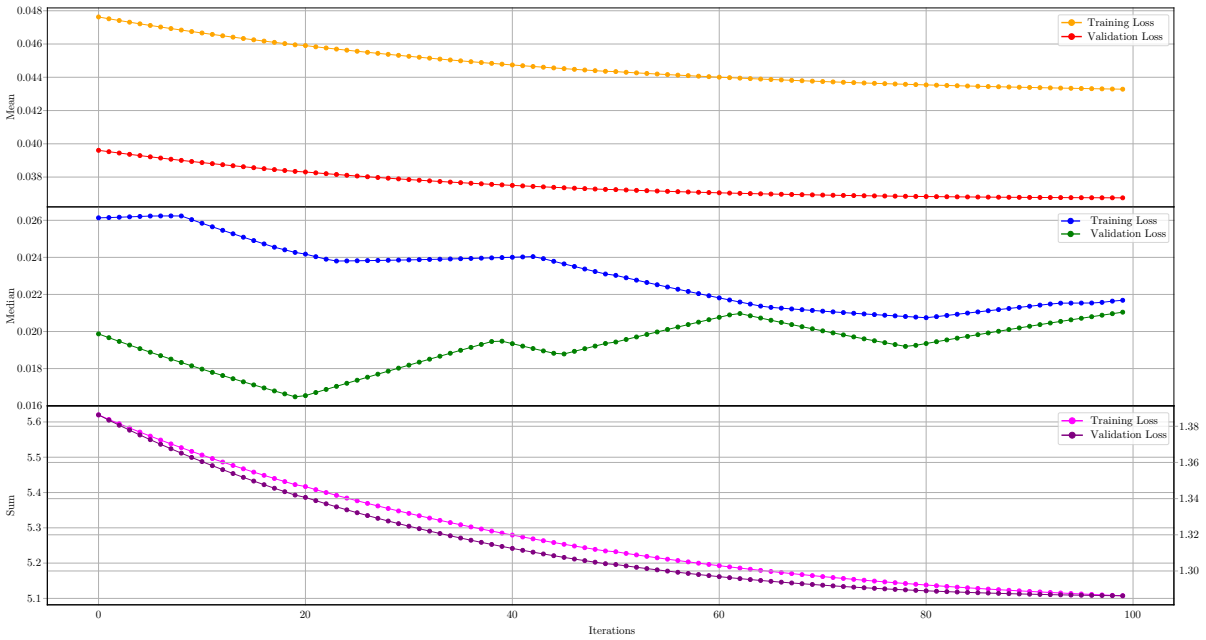


Figure 7.5: MSE loss of the final model architecture and LoHP data. Plots of mean loss, median and sum of all training and validation files of the LoHP study. $lr = 1 \cdot 10^{-6}$, $w_d = 1 \cdot 10^{-8}$. The colors and their meanings in this plot are the same as in Figure 7.2

Part IV

Epilogue

Methodology Conclusions 8

Physiological data were successfully collected during two flight simulator studies. This data will not only be used during this thesis for training a deep learning network but is also meant to be disseminated. This might help other researchers test their approaches on detecting stress with this data. Many tools were used in order to achieve this goal. Those tools come along with certain advantages as well as disadvantages. For both studies specific scenarios were created in order to not only trigger stress in flight crews, but also to be accommodated into the AVES. Furthermore, two different methods were used for getting an as good as possible stress level feedback after and during flight. The following chapter will elaborate on what limitations and disadvantages have been discovered throughout the two flight simulator studies. Additionally, conclusions will be drawn from the methods applied as well as recommendations given for future simulator studies.

8.1 Methodology Limitations

Over the course of this project multifarious limitations have resulted in minor and major difficulties which will be discussed within this chapter. Those start with the capabilities of the simulator to implement scenarios properly and go to the limitations imposed by the subjects themselves. Moreover, there are fundamental physiological and psychological effects that influence the outcome. They will be discussed in order to find improvements for future projects with similar research questions.
In an ideal world, pilots could have flown in an actual aircraft performing high demand maneuvers while measuring their physiological parameters. Due to various safety factors, this is not possible and hence requires an as good as possible representation in a lab settings. Limitations and aspects of the flight simulator will be discussed in chapter 8.2 in more detail.
Although a lot of research had been done in this field up until today, yet there is no reliable and objective measurement for stress either with or without deep learning systems. In this project, a discriminative approach was chosen and therefore a 'ground truth' for the corresponding labels needed to be established. In this case, it means an as precise as possible evaluation of each individual's current stress levels is required. This feedback required precision in two dimensions: quality (correct stress level evaluation to a corresponding physiological measurement) and of time-correlation. This means that when a participant says they experienced a higher stress this needs to be time-synchronous to the recorded physiological reaction. Such kind of feedback system also heavily depends on the individuals' capabilities to self-reflect and correctly self-evaluate their stress perception and reaction. This will always bring a certain bias to the data as it is at least doubtful that everyone is always capable of evaluating their stress

or other physiological reactions correctly. There are known participant biases such as subject bias or demand characteristics where a participant consciously or unconsciously acts the way they think the researcher wants them to. [52], [53] At least for the conscious part, mitigating measures in order to reduce this bias were taken e.g. raising no expectation on the participants and randomized scenarios for the SPO study.
Another factor to consider is the fact that pilots are not only chosen by their technical skill sets, but also by psychological and personal traits. During selection before pilot training those people are psychologically examined and only when certain aspects fit allowed to further proceed their career. This creates the special situation that there are certain personality traits being more and less represented in pilots. Therefore, a certain base resilience for stress is present in pilots making it more difficult to trigger certain stress levels. Additionally, pilots undergo intense training to prepare them for emergency cases and train them to maintain certain professionalism even when confronted with dire situations. This was especially reflected by the feedback provided by the pilots. During the final approach to Cologne/Bonn Airport in the stress scenario of the LoHP study the aircraft status was very critical and close to be uncontrollable. Even in this situation not many pilots said they were in the range of maximum stress. This can have of course multifarious factors that might even be different from one individual to another. All these biases increase uncertainty for the collection of a precise stress feedback.

8.1.1 Gender Data Gap

During both studies, most participants identified as male and only two during the SPO and none during LoHP study participated as female. 'Gender data gap' refers to the lack of information and representation of individuals based on their gender in data sets and research. This gap is an issue for scientific data, because it can lead to the production of biased and incomplete findings, as well as perpetuate gender-based inequalities. This is especially important when it comes to physiological data, because physiological differences between men and women have a significant impact on health, well-being, and stress. For example, women and men have different hormonal profiles, cardiovascular systems, and responses to medication, which can influence the onset, progression, and treatment of various medical conditions. If physiological data only includes men, it could lead to an incomplete understanding of the physiological differences between men and women and the development of an analysis that is not effective for women. Addressing this gap by including women in medical research and collecting gender-specific data is critical for improving the systems capabilities of detecting stress. Therefore, future studies should definitely include more women although female pilots are still underrepresented in airlines.

8.2 Flight Simulator Studies

Both studies took place in the AVES in Brunswick, Germany. This simulator was designed to be a flexible platform for several simulation environments and to test different systems and controllers. The A320 section is not certified by the manufacturer or in any other way. A reason for that is to keep the simulator flexible for research. While the SPO study took place on the fixed platform the LoHP study used the motion platform. After theoretical planning which errors could impose a significant amount of stress on the crew, it needed to be verified that such failures are possible to implement in the AVES. As this simulator is not a certified A320 simulator not all systems are simulated and their functions might be different. That means that not all failures possible in the A320 could be simulated inside AVES. This not only limited the possible errors but also shaped significantly the final scenarios.
Several pilots gave the feedback that controlling and maneuvering the simulator was way different from what they were used to. For instance, the rudder inputs were extremely sensitive

and needed to be used with a lot of caution. As this seems to be good as it increases stress in pilots, it mostly disturbed their usual workflow. This breaks simulator immersion by reminding them that they are in a simulator and hence decreasing their stress levels. This disrupted the overall simulation experience and reduced the overall realism for participants. Therefore, giving pilots no need to be stressed as they were made aware that they are inside a simulator.

8.2.0.0.1 Video Recording The AVES flight simulator has two fixed installed cameras that can be used for recording during the scenarios. This could help during data processing to get a better understanding on what happened during the simulator flight. Unfortunately, due to a technical error from both studies no reliable video recording was obtained. This not severely impacted the data evaluation and creation of the database but might have helped in some cases during post-processing the data.

8.2.1 Conclusion of Flight Simulator Studies

Despite all limitations, both studies successfully collected physiological data as well as flight data from the simulator. After finding scenarios that can be realized in AVES those mostly worked out well and did trigger some amount of stress. Unfortunately though, not to a degree that was expected or required for the deep learning training. As data exploration showed, only a few participants said they reached the maximum stress level. From a deep learning data perspective, it would be good to have a more equally distributed data set.
Simulator realism might be one factor that could have influenced stress in pilots to be lower than it would be in an actual real situation. Additionally, some pilots gave the feedback that they were so distracted by some simulator issues that they just relaxed, because it was nothing they could do something about. This shows that for such experiments and research questions simulator realism is very important.
Time and data recording synchronization between AVES and other systems could not be established. This resulted in a time and work intense post-processing in order to synchronize simulator data and e.g. ECG data. Although time synchronization was possible in general, several issues could not have been resolved in time. During the LoHP study preparation it turned out that there was a time delay within the synchronization signal of the SPO study's data. Due to a lack of time in the simulator before the LoHP study, those issues could not be resolved. This resulted in the usage of different tools and methods than actually planned. Ideal would be a central data acquisition system which acquires all relevant data from all systems. Such systems not only take care of time synchronization but also merge those different sources together into one.
Due to that fact, flight and ECG data have not yet been synchronized. This could open up another approach to analyze ECG data. First of all, stressful times can be correlated with the triggering of failures. Secondly, baseline parameters of the ECG can be computed based on the resting ECG. Now ECG data of the stressful times with the resting ECG could show deviations.

8.2.2 Prospects for Future Flight Simulator Studies

Building a prospect from those conclusions gives aspects on what to consider when designing similar flight simulator studies in the future. The biggest criticism throughout both studies and all pilots was simulator realism. Participating pilots need to almost forget that they are in a flight simulator. This also needs to be considered when designing flight scenarios. Those scenarios need to be designed in collaboration with not only experienced training pilots, but also simulator operators. Additionally, using a certified simulator might help overcome those simulator related issues. But acquiring data from those simulators might be slightly more difficult as they are usually not made for data recording.

Having a centralized data acquisition system will also reduce the time necessary for data post-processing. This will improve signal synchronization and hence allow better correlation between two signals.
Another point to consider is the design of the scenarios. Although the LoHP study faced pilots with quite severe failures and aircraft conditions, there still might be room to induce more stress. One aspect that could be further exploited are consecutive missed approaches. Considering the stress scenario of the LoHP after the first go-around in Frankfurt another landing attempt could have been implemented. Also, after diverting to Cologne another go-around before finally landing on another runway could be implemented. Those multiple landing attempts are tiring and require the flight crews to continuously re-plan their strategy. Additionally, this can create a sense of hurry and thus stress.

8.3 Conclusions and Evaluation

The most critical method used turned out to be the flight simulator and the scenarios used. This opened the discussion on other ways to acquire physiological data from pilots during flight. One idea was to approach airlines and collaboratively ask their pilots if they would wear some ECG device during their normal operation. Although collecting physiological data allows no further correlation with their perceived stress levels or flight data. The flight data is a less critical point as they could be obtained with online websites such as `flightradar24.com`. Most airliner flights are rather quiet and errors or failures are quite rare. Therefore, even more 'baseline' data would be available and not so much data with stress. On the other hand, if pilots were instructed to report only stressful situations afterwards, at least some correlation could be possible.
Another option could be working with aircraft manufacturers and their flight test departments. Flight tests can be very stressful and demanding for pilots and can also lead to some stress reactions. Moreover, flight test aircraft have a large system on board for flight data acquisition. This would at least allow the correlation between physiological and aircraft data. Additionally, these approaches would allow to gather real world data in contrast to the simulation environment.
Over the course of both flight simulator studies about $2\,000\,\mathrm{min}$ of data was collected. This data not only contains the ECG leads of the pilots, but also their stress level feedbacks. There certainly have been limitations that reduced the overall data quality, yet quality was sufficient for further usage.
The stress level feedback provided by the pilots is prone to biases and depends on their capabilities to correctly self-reflect. The Online Stress Assessment Tool specially made for this thesis successfully collected data from those pilots during simulator flights.

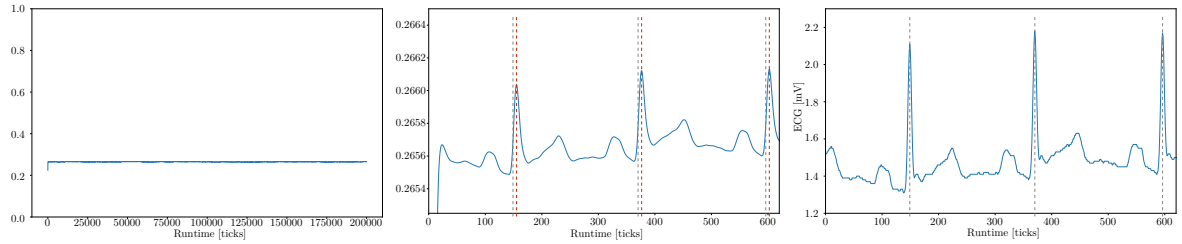
Machine Learning System Evaluation

9

After adapting and modifying the deep learning model in section 7.1 this chapter focuses on evaluating and discussing the models' results. Two things are interesting when evaluating such systems: first, how the loss changes over each iteration and secondly, what the model's actual output is. The first aspect is a more theoretical and mathematical approach to evaluating the performance of the model. This is due to the abstract metric that is calculated over training and validation data sets. It is more comprehensible when visualizing the metrics directly. As this would take a lot of memory space to save these values for each iteration it is only done once at the end. Based on these results ideas and propositions on how to improve models performance will be discussed.

9.1 Training Results

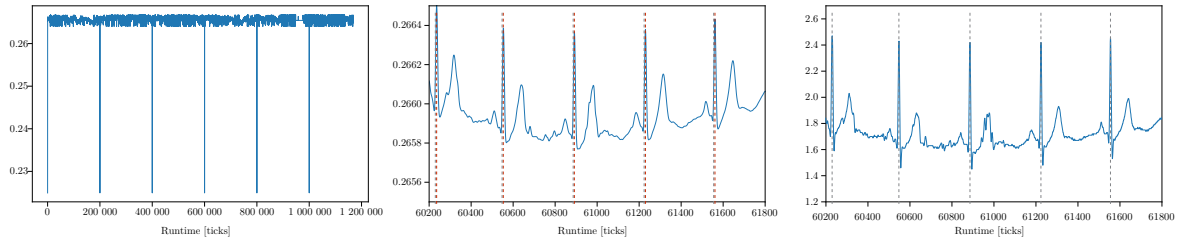
With the three iterations described in chapter 7 the final model architecture with hyperparameter was concluded. Unfortunately, not because it had the best performance in stress prediction, but rather because there was no time left. Computation for training this model with the collected data was very time-demanding. For the SPO study's data, the first 10 iterations took about one day to complete. For 100 iterations, the training started on 8th February 10:32 and lasted until 17th February 2023 at 00:16 (about 9 days). For the LoHP study, though, the first 100 iterations took about 19 days (started 18th January at 14:04, ended 6th February 2023 at 19:02).



(a) Prediction. (b) Zoomed in on start. (c) Same segment from actual ECG and Input.

Figure 9.1: Predictions made by the LSTM Model on an resting ECG from LoHP study from VP17. The y-Axis is still in LSTM output range of $\mathbb{R} \in (0, 1)$. a: Overview of LSTM System prediction. b: Enlarged the beginning of the prediction which reveals an ECG curve similar to the actual input shown in c.

As the losses of this model architecture have been discussed in section 7.2.3 this section focuses on another aspect. In order to understand what the predictions of this LSTM architecture looks like both validation data sets were fed into the respecting last model state. The predictions were saved into a CSV file and then plotted. Two representative examples from the LoHP study will be further discussed here. The first example are predictions made upon a resting ECG lead from VP17. At a first glance, Figure 9.1 shows that there is an initial sharp rise and then an almost continuous value of about 0.26. But when zooming in very closely, the track actually reveals a similar shape as an actual ECG, shown in Figure 9.1. A similar behaviour is observed with all other predictions too, Figure 9.2 shows an example of predictions on a stress scenario record.
Comparing those plots with the actual ECG leads shows that the model is trying to imitate and repeat the ECG pattern. Furthermore, giving a more damped and smoothed track than the original ECG including a delay from a few 10 ticks. This indicates that this architecture is not yet capable of performing this abstraction between ECG and stress levels.



(a) Prediction. (b) Zoomed in on middle section. (c) Same segment from actual ECG and Input.

Figure 9.2: Predictions made by the LSTM Model on an ECG from LoHP study's stress scenario from VP10 shown in a. b: Enlarged on a section of the prediction which also reveals an ECG curve resembling the actual input shown in c. Which is a similar behaviour as in Figure 9.1

The runtime required for computing those predictions was at maximum below 1 min, more within a range of 20 s to 30 s. Considering the longest runtime among validation files (VP15, stress scenario) has a record runtime of ≈67 min. This means that this type of model would comply with the real-time requirement 1 easily.

9.2 Suggestions for Performance Improvements

Considering the results discussed in section 9.1 reveals that not timely performance is an issue, but actual predictions. In order to face this issue, three main approaches can be used.

9.2.0.0.1 More Input Data One would be changing the way data is presented as input to the model. Right now, every single signal datum is input into the model individually. The alternative would be segmenting these ECG leads into an input vector of a few seconds (e.g. 2 s to 10 s) and then doing a stress level prediction on this data. This would result in expanding the dimensions of both hidden and cell state layers by the input size. This means that more data is presented to the model at once and a prediction is made based on this more information. Hence, this could improve the capability of the LSTM hidden layer and cell state layer to find a better abstraction from the input.

9.2.0.0.2 Expanding the Model Architecture Another concept is adding other types of neural layers before and/or after the LSTM block. As of right now, the LSTM seems to be good at finding some patterns within the ECG but no correlation to the stress levels. A fully connected layer with a ReLu activation could be a first approach. The layer would be added

after the output of the last LSTM cell. A good option for a layer before the LSTM cells would be CNN as they perform overall well on feature extraction from images.

9.2.0.0.3 Finding Stressful Events Finally, another overall method besides the supervised learning used in this work could be used for finding stressful events within the ECG signal. Clustering methods or even semi-supervised learning algorithms seem promising. Clustering or unsupervised learning algorithms do not need any label, but rather try to find discriminative functions to separate data. This could mean using the resting ECG and baseline scenario recordings from LoHP study's data as a baseline. Comparing this baseline to segments of the stress scenario, where stress is expected, could reveal whether stress is detected correctly or not. Unfortunately, there are no established methods for pure time series data. All classifiers rely on an extraction of features from the original data. ECGs are difficult in this respect as many metrics besides the heart rate are available as possible feature. These can include features based on the IBI as with HRV, or more exotic metrics such as different peak lengths. [54] Moreover, that leads back to the initial discussion on the issue that such feature extraction already makes some physiological assumptions, as discussed in section 2.3 and section 7.1. The representativeness of stress reactions of these assumptions is at least arguable. [55] A viable approach could be segmenting the time series ECG data into chunks of data around the R-peak and comparing them to other chunks fitted in the same way. This would be a similar approach as used in Shahrudin, Sidek and Jusoh.

9.3 Conclusions and Prospects for Deep Learning Stress Level Classification

With the results gathered, the most significant factor to work on is the overall model architecture. It could be possible that the initial output given by the model, after the first few iterations, was more prominent in amplitude as it is in the end. Meaning that the aspect which was optimized during model training could be dampening the ECG input. As discussed in section 9.2 there are some promising modifications that could be considered. Looking into those different algorithms further as well as analysing their behaviour and performance could improve overall detection capabilities. For the moment, the most significant finding is that with the final architecture the model is not creating the desired output. Instead of producing some sort of constant signal it outputs a damped and delayed ECG. A promising aspect in order to overcome this issue is by adding additional layers to help the abstraction process. This could e.g. include a CNN layer before the actual input in order to already extract information from the ECG. Other valid options could involve SoftMax and/or fully connected layers.
The model training process was very time extensive and required much computational power. Trying out other architectures will consume a considerable amount of computational time. Improving data loading and structuring could be beneficial for speeding up the overall process. Data loading time is assumed to have had the greatest impact as each file had to be loaded individually. Stacking the whole data sets into one tensor and loading it in different batches and input sizes might mitigate this issue.
There are still promising steps and modifications that can be made in order to improve the overall prediction capabilities.

Discussion & Outlook 10

In this last chapter, conclusions drawn earlier and evaluations will be further discussed and an outlook on the realization of future stress detection will be provided. Over the course of this project, several aspects have been discovered that need to be taken into account. Not only in terms of study design and measurement equipment, but also for deep neural network stress detection or classification. Moreover, the question if stress after all is the right metric for detecting potentially dangerous work-overload situations needs to be discussed. With a final outlook and personal placement, this work shall conclude.

10.1 Discussion

10.1.1 Online Stress Assessment Tool

The new online stress assessment method is a digital tool that was designed to provide real-time feedback on the stress levels of flight crews during flights. The biggest benefits of this tool is that it provides time-precise self-evaluation of stress by pilots for a certain moment. This is in contrast to the SPO feedback method, where the stress level was a vague track over the flight after each scenario. It is at least arguable whether participants were capable of correctly provide feedback of their stress in amplitude as well as precisely in time domain after those scenarios. During the LoHP study no disturbances or issues occurred during data recording. However, there were some concerns raised during the evaluation of the tool. For example, in the evaluation questionnaire, two participants said they were just submitting values in order to silence the tool. Although this might be a signal for work overload, there are few things to mitigate such behavior besides a verbal system. With that, pilots would verbally tell their momentary stress level, which is then noted by the research manager. This would increase the amount that is talked inside the cockpit. This can disrupt the work cycle of the pilots, while this new tool can be integrated into their routine. Furthermore, pilots are still in control when to give confirmed values, choosing a good moment when there is no other more important task to perform.
After the LoHP study participants were asked to give feedback to this tool. The criticism provided by the users is mainly in regard to tablet responsiveness. Most users said they needed to press several times until the slider changed or pressing the submit button was registered. Other suggestions were a larger scale, adding colors to it, and audio notification. Those suggested improvements by participants could also be investigated in coming studies to validate this method. However, the remaining question is whether this tool is good to measure online stress which could be answered by a thorough method validation.

When looking at the data that came out of this tool there are two values available: slider and confirmed. Where the confirmed values correspond to the slider value at the moment that button is pressed, the slider value tracks every change. This means that mathematically speaking the confirmed values are always included in the slider values. Assuming the submit button always worked properly (each time it was pressed by a participant a value was submitted) this allows further analysis and interpretation. This means that e.g. changing the slider to another value, but forgetting to submit could indicate a more demanding workload or already beginning performance degradation.
In summary, the new online stress assessment tool is a promising development for stress assessment for flight crews, as it provides time-precise feedback and allows a good time synchronization with ECG and simulation data. However, there are some concerns that need to be addressed, such as mitigating work overload and improving tablet responsiveness. Further studies and method validation are needed to fully assess the effectiveness of this tool in measuring online stress for flight crews.

10.1.2 Data Collection from Flight Simulator Studies

A highlight throughout this thesis is the data collected and transformed into databases. Over the course of the two flight simulator studies more than $2\,000\,\mathrm{min}$ of ECG and stress tracks were measured from 42 participants[1]. Although quite some data was gathered, deep learning applications need and rely on huge amounts of data in order to reduce the issue of overfitting. [56] This is especially true for supervised algorithms with prediction labels on which the model is optimized upon. With that said two aspects are important: first the length of data recorded from one individual and second the number of individuals of whom data is recorded. Additionally, certain reactions, that shall be detected later, need to be represented within a significant quantity. This means that not only more test participants, but also more time in which a relevant stress level is present would be appreciated. Looking at the data exploration of chapter 6.2 reveals that only a few participants said they were in a high stress regime. In order to say whether more data is needed signs of overfitting on training data would need to be an issue. One approach to even out the label distribution would be by cutting out non-stress intervals and hence increasing the overall ratio of high stress levels present in the data sets. As a matter of fact, that might create a more equally distributed data but cut out relevant data for no reason. Furthermore, good deep learning systems are usually capable of learning such difference even with an unequal distribution. Hence, removing data might not overcome the issue of the under-representation of high stress levels. Moreover, the influence of such data exclusion is at least questionable and needs to be investigated by further research.
Another aspect to further consider is whether a feedback system to measure stress (levels) is a good solution. This method heavily relies on participants to correctly self-reflect on their momentary stress level. The stress levels need to be collected in a timely and precise manner so that they can be used further. Those demands bring along several issues. It might be not only difficult for one individual to correctly self-reflect on their stress, but also that stress perception and sensation vary significantly between individuals. Introducing a finer scale of stress might make it even more difficult to differentiate between certain stress levels. On the one hand, having two to four defined levels might make it easier for participants to categorize themselves into. On the other hand, this might make it more difficult to distinguish when a critical performance degradation sets in. Going back to the original idea of detecting situations where pilot's performance is critical, the influence of stress might be not significant enough in order to be distinguished. Therefore, reformulating the question and measurement scale might be necessary.

[1] There have not been 42 different individuals participating in both studies. Some individuals participated in both studies. Unfortunately, due to the anonymized data it was not possible to find out who did.

10.1.3 Stress Detection with ECG Signals

In the quest to develop effective stress detection systems using ECG signals, there are a number of challenges that need to be managed. Firstly, it is important to note that stress might not always be reflected within the ECG, and even when it is, some stress is not severe enough to have a significant influence on the cardio-vascular system. This may make it difficult to accurately detect stress levels using ECG signals. Additionally, there are times when pilots might be in dangerous situations without being under physical stress, for example, if they have not yet realized the situation they are in. Even deep neural networks, which are commonly used in image classification tasks, can find stress detection in ECG signals challenging, since it is not always easy to label and classify the data. One issue is obtaining stress levels from participants.

Most preprocessing techniques for ECG analysis already apply some kind of assumptions, for instance, by using the instantaneous heart rate or IBI the assumption is that this is the predominant metric for stress. This is only true to a certain degree but the ECG signal comprises more information than that, e.g. the QRS morphology also changes.[5] Another assumption is that there must be a linear correlation between heart rate and stress, meaning that higher stress levels should result in higher heart rates. However, these assumptions might only hold true for certain levels of stress, under certain circumstances, and might vary between individuals. Moreover, each individual has another idea or limit when talking about maximum stress. This could mean using a 'one-size-fits-all' approach for stress detection, meaning that the same model is used for each individual, could be difficult. A better approach could be developing a model for each individual in order to consider these interpersonal differences.

Another aspect to consider is how in a real world application the ECG would be taken. Current ECG systems that produce a continuous derivation of electric activity of the heart rely on using skin pads. Those skin pads easily detach which disables taking an ECG. About 12 % of the SPO data was not usable, because of invalid recording probably due to disconnected or misplaced electrodes. During the LoHP study this was mitigated by continuous monitoring throughout the simulator scenarios. Conductive systems, as used in the SPO study, are capable of measuring electric activities of the heart but can only detect the R-peak. Hence, all analysis can only be done based on those R-peaks or more precisely the IBI. Moreover, several factors complicate this conductive measurement and can result in no possible detection. Those include but are not limited to thickness of clothes, distance from upper body to measuring mat, movements and vibrations, or individual positioning in respect to the mat. In theory, assuming that a reliable stress detection is possible based on R-peaks, this would solve the issue of continuous heart monitoring but in practice it has several aspects to further improve. A viable solution could be integrating such measurement equipment into textiles that are in direct skin contact as demonstrated by Nigusse, Mengistie, Malengier et al. or Tsukada, Tokita, Murata et al.

Moreover, the baseline assumption that stress will always result in a cardiac response might not hold true. During the LoHP study it was seen that sometimes the area of interest, work-overload of pilots resulting in performance reduction, might be very subtle. The goal of such system should be detecting how much remaining resources the flight crew have and stress might be just one metric within many to determine those resources. Other studies have shown that acute mental stress can be detected by EEG as well but still has issues with measuring and evaluating those signals.

10.1.4 Aspects in the LSTM System and the Deep Learning System

The model architecture discussed here has shown to reduce the overall loss on the training and validation data during model training. Yet, the output produced by the LSTM cells is rather a damped and delayed mimic of the original ECG. Although quite interesting, it does not seem to have found a correlation or abstraction from the input towards the output. Having this model

run more iterations would probably not have changed the output further as losses have slightly converged already. In chapter 9.2 appropriate measures have been discussed in order to face those issues. The most promising one seems to be a combination of changing the input format and adding different types of layers before or after the LSTM block. As a first attempt the input length should be extended to include at least 1 s and could be experimented to increase this up to 10 s to 30 s. Moreover, adding fully connected layers after the LSTM blocks could improve the abstraction capability.
Another completely new approach could be using different algorithms and layers for the stress classification task. As LSTM tries to mimic the shape of an ECG they could be used for an ECG prediction. This could take the previous n-seconds as input data and try to predict the ECG curve for the next m-seconds[2]. The stress classification could then analyse this ECG prediction for stress. Such a system might be capable of detecting stress even before it becomes an issue.
Another aspect that could be considered is whether training a model for each individual participant would result in a better stress prediction. This could take into account the fact that each person has such a different ECG curve that it is almost like a fingerprint. The ECG lead can be so different from one person to another that even an identification via ECG is possible. [59] Moreover, every person has a different tolerance against stress and will react differently in certain situations.

10.2 Conclusions and Outlook

In this thesis, it was successfully demonstrated that with mentioned resources and a comparatively limited amount of time it was feasible to

1. build a working ECG from prefabricated parts,
2. to write software tools for acquiring and visualizing ECG data,
3. to design simulator studies for stressing pilots,
4. to design a novel way to measure stress during flights, OSAT,
5. to use an A320 simulator, said ECG, and said stress measurement tool for recording pilots physiological data during flights,
6. to write software tools for viewing, editing, and processing time-synchronous ECG and stress data, and
7. to design an LSTM model, train, and evaluate it with the acquired data.

Most issues identified during the practical realization were either mitigated right away or in post-processing. Some of them still remain, but solutions are on the horizon for virtually all of them. If the issues related to the prediction algorithm are sorted out it would remain a sub-system for a yet to build aircraft service system. Therefore, the best approach would be shaping the usage definition of a stress detecting system even further. Such system definition would include, but not be limited to, a failure mode analysis and considering CS verification related aspects. Hence, the use case definition of such larger system might also reshape the research question that deep learning algorithm has to solve.
It can be said that with the mentioned modifications in section 9.2 this setup could turn into a viable detection system. The simulator scenarios could also be tuned further in order to increase the perceived stress in pilots. All those discussed aspects pave the way well for further research in this field.
Any relevant software, data, and other things are publicly available in a GitHub repository under `https://github.com/0Patrice/StressDetectionInFlightCrews`. [60]

[2] n and m are meant to be placeholders for a more specific time still to be defined.

Part V

Appendix

Bibliography

[1] P. Albrecht, *The mit-bih st change database*, 1992. DOI: 10.13026/C2ZW2H. [Online]. Available: https://physionet.org/content/stdb/ (cit. on p. 4).

[2] J. A. Healey and R. W. Picard, *Stress recognition in automobile drivers*, 2008. DOI: 10.13026/C2SG6B. [Online]. Available: https://physionet.org/content/drivedb/ (cit. on p. 4).

[3] J. Healey and R. Picard, 'Detecting stress during real-world driving tasks using physiological sensors', *IEEE Transactions on Intelligent Transportation Systems*, vol. 6, no. 2, pp. 156–166, Jun. 2005. DOI: 10.1109/tits.2005.848368. [Online]. Available: https://doi.org/10.1109/tits.2005.848368 (cit. on p. 4).

[4] G. F. Wilson, J. D. Lambert and C. A. Russell, 'Performance enhancement with real-time physiologically controlled adaptive aiding', *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 44, no. 13, pp. 61–64, Jul. 2000. DOI: 10.1177/154193120004401316. [Online]. Available: https://doi.org/10.1177/154193120004401316 (cit. on p. 4).

[5] N. S. N. Shahrudin, K. A. Sidek and A. Z. Jusoh, 'Electrocardiogram (ECG) based stress recognition integrated with different classification of age and gender', *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 15, no. 1, p. 199, Jul. 2019. DOI: 10.11591/ijeecs.v15.i1.pp199-210. [Online]. Available: https://doi.org/10.11591/ijeecs.v15.i1.pp199-210 (cit. on pp. 4, 56, 89, 93).

[6] N. Sharma and T. Gedeon, 'Objective measures, sensors and computational techniques for stress recognition and classification: A survey', *Computer Methods and Programs in Biomedicine*, vol. 108, no. 3, pp. 1287–1301, Dec. 2012. DOI: 10.1016/j.cmpb.2012.07.003. [Online]. Available: https://doi.org/10.1016/j.cmpb.2012.07.003 (cit. on p. 4).

[7] U. Drj. 'Circulatory system'. License: CC BY-NC-SA 3.0 ; Edits by Author: Font, Lines and Points changed, Tags 'Popliteal' removed, further markers added. (2012), [Online]. Available: https://www.textbookofcardiology.org/wiki/File:Figure_2.svg (visited on 28/10/2022) (cit. on p. 10).

[8] U. Wapcaplet. 'Diagram of the human heart'. License: CC BY-SA 3.0 unported ; Edits by Author: Font, Lines and Points changed. (2006), [Online]. Available: https://commons.wikimedia.org/w/index.php?title=File:Diagram_of_the_human_heart_(cropped).svg (visited on 28/10/2022) (cit. on p. 11).

[9] J. M. Nerbonne and R. S. Kass, 'Molecular physiology of cardiac repolarization', *Physiological Reviews*, vol. 85, no. 4, pp. 1205–1253, Oct. 2005. DOI: 10.1152/physrev.00002.2005. [Online]. Available: https://doi.org/10.1152/physrev.00002.2005 (cit. on p. 12).

[10] A. O. Grant, 'Cardiac ion channels', *Circulation: Arrhythmia and Electrophysiology*, vol. 2, no. 2, pp. 185–194, Apr. 2009. DOI: 10.1161/circep.108.789081. [Online]. Available: https://doi.org/10.1161/circep.108.789081 (cit. on p. 12).

[11] U. Spjkrul. 'Conductionsystem'. License: CC BY-NC-SA 3.0 ; Edits by Author: Font, ECG curve and Logo on bottom removed. (2011), [Online]. Available: https://www.textbookofcardiology.org/wiki/File:Conductionsystem.svg (visited on 28/10/2022) (cit. on p. 13).

[12] U. A. A. Atkielski). 'Schematic diagram of normal sinus rhythm for a human heart as seen on ecg (with english labels).' License: Public Domain. (2007), [Online]. Available: https://commons.wikimedia.org/wiki/File:SinusRhythmLabels.svg (visited on 03/11/2022) (cit. on p. 14).

[13] S. L. Barron, 'Development of the electrocardiograph in great britain', *BMJ*, vol. 1, no. 4655, pp. 720–725, Mar. 1950. DOI: 10.1136/bmj.1.4655.720. [Online]. Available: https://doi.org/10.1136/bmj.1.4655.720 (cit. on p. 15).

[14] M. Cadogan. '5-lead electrode placement'. License: CC BY-NC-SA 4.0 ; Edits by Author: Colors and Names changed. (2022), [Online]. Available: https://litfl.com/ecg-lead-positioning (visited on 28/10/2022) (cit. on p. 15).

[15] W. Einthoven, 'Ueber die form des menschlichen electrocardiogramms', *Pflüger, Archiv für die Gesammte Physiologie des Menschen und der Thiere*, vol. 60, no. 3-4, pp. 101–123, Mar. 1895. DOI: 10.1007/bf01662582. [Online]. Available: https://doi.org/10.1007/bf01662582 (cit. on pp. 15, 16).

[16] E. Goldberger, 'A simple, indifferent, electrocardiographic electrode of zero potential and a technique of obtaining augmented, unipolar, extremity leads', *American Heart Journal*, vol. 23, no. 4, pp. 483–492, 1942 (cit. on p. 16).

[17] A. R. Barnes, H. E. Pardee, P. D. White, F. N. Wilson and C. C. Wolferth, 'Second supplementary report by the committee of the american heart association for the standardization of precordial leads2', *Journal of the American Medical Association*, vol. 121, no. 17, pp. 1349–1351, 1943 (cit. on p. 16).

[18] S. Tan and A Yip, 'Hans selye (1907–1982): Founder of the stress theory', *Singapore Medical Journal*, vol. 59, no. 4, pp. 170–171, Apr. 2018. DOI: 10.11622/smedj.2018043. [Online]. Available: https://doi.org/10.11622/smedj.2018043 (cit. on p. 16).

[19] H. H. Publishing. '"stress"'. (14th Feb. 2023), [Online]. Available: https://www.health.harvard.edu/q-through-z#S-terms (cit. on p. 17).

[20] V. Keim. 'Stress'. (2016), [Online]. Available: https://www.pschyrembel.de/Stress/K0LQG (visited on 02/11/2022) (cit. on p. 17).

[21] S. M. Jex, 'Stress and job performance: Theory, research, and implications for managerial practice', 1998 (cit. on p. 17).

[22] Skybrary.aero. 'Workload (OGHFA BN)'. (2022), [Online]. Available: https://www.skybrary.aero/articles/workload-oghfa-bn (visited on 02/11/2022) (cit. on p. 17).

[23] S. Silvagni, L. Napoletano, I. Graziani, P. Le Blaye and L. Rognin, 'Concept for human performance envelope', Future Sky Safety, Tech. Rep., 2015 (cit. on pp. 18, 19, 33).

[24] A. Ng, *Cs229 lecture notes*, 2022. [Online]. Available: https://cs229.stanford.edu/notes2022fall/main_notes.pdf (cit. on pp. 20–22).

[25] D. P. Kingma and J. Ba, *Adam: A method for stochastic optimization*, 2014. DOI: 10.48550/ARXIV.1412.6980. [Online]. Available: https://arxiv.org/abs/1412.6980 (cit. on p. 22).

[26] C. Olah. 'Understanding LSTM Networks'. (2023), [Online]. Available: https://colah.github.io/posts/2015-08-Understanding-LSTMs/ (visited on 30/01/2023) (cit. on pp. 22, 23).

[27] L. Schiavo, J. Clark, G. Brockman, I. Sutskever, D. Yoon and B. Barry. 'Learning dexterity'. (16th Feb. 2023), [Online]. Available: https://openai.com/blog/learning-dexterity/ (cit. on p. 23).

[28] S. Stanford. 'Deepmind's ai, alphastar showcases significant progress towards agi'. (16th Feb. 2023), [Online]. Available: https://medium.com/@towardai/deepminds-ai-alphastar-showcases-significant-progress-towards-agi-93810c94fbe9 (cit. on p. 23).

[29] C. Xie, L. McCullum, A. Johnson, T. Pollard, B. Gow and B. Moody, *Waveform database software package (wfdb) for python*, 2022. DOI: 10.13026/MMPM-2V55. [Online]. Available: https://physionet.org/content/wfdb-python/4.0.0/ (cit. on p. 26).

[30] M. medizinische Diagnosegeräte GmbH, *Technical manuel eg01000*, Version 2.5, https://www.medlab-gmbh.de/english/downloads/eg01000_25.pdf, 2012 (cit. on p. 32).

[31] J. Koolhaas, P. Meerlo, S. de Boer, J. Strubbe and B. Bohus, 'The temporal dynamics of the stress response', *Neuroscience & Biobehavioral Reviews*, vol. 21, no. 6, pp. 775–782, Nov. 1997. DOI: 10.1016/s0149-7634(96)00057-7. [Online]. Available: https://doi.org/10.1016/s0149-7634(96)00057-7 (cit. on p. 33).

[32] M. medizinische Diagnosegeräte GmbH, *Technical manuel eg05000*, Version 1.12, https://www.medlab-gmbh.de/english/downloads/eg05000_112.pdf, 2018 (cit. on p. 39).

[33] Jenkins, D. P, N. A. Stanton, L. A. Rafferty, P. M. Salmon, G. H. Walker and C. Baber, *Human factors methods*, en, 2nd ed. London, England: Ashgate Publishing, Jul. 2013 (cit. on p. 40).

[34] G. Matthews, L. Joyner, K. Gilliland, S. Campbell, S. Falconer and J. Huggins, *Dundee stress state questionnaire*, 1999. DOI: 10.1037/t27031-000. [Online]. Available: https://doi.org/10.1037/t27031-000 (cit. on p. 40).

[35] Human Performance Research Group, NASA Ames Research Center, 'Nasa task load index (tlx) paper and pencil package v.1.0', Ames Reserch Center, Moffett Field, California: NASA, 1988 (cit. on p. 41).

[36] S. G. Hart, 'Nasa-task load index (NASA-TLX); 20 years later', *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 50, no. 9, pp. 904–908, Oct. 2006. DOI: 10.1177/154193120605000909. [Online]. Available: https://doi.org/10.1177/154193120605000909 (cit. on p. 41).

[37] 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, pp. 139–183. DOI: 10.1016/s0166-4115(08)62386-9. [Online]. Available: https://doi.org/10.1016/s0166-4115(08)62386-9 (cit. on p. 41).

[38] B. Kirwan, A. Evans, L. Donohoe, A. Kilner, T. Lamoureux, T. Atkinson and H. MacKendrick, 'Human factors in the atm system design life cycle', ser. FAA Eurocontrol ATM R&D Seminar, 1997 (cit. on p. 42).

[39] M. D. Miller and D. G. Ferris, 'Measurement of subjective phenomena in primary care research: The visual analogue scale', en, *Fam. Pract. Res. J.*, vol. 13, no. 1, pp. 15–24, Mar. 1993 (cit. on p. 45).

[40] L. Guillet, D. Hermand and E. Mullet, 'Cognitive processes involved in the appraisal of stress', *Stress and Health*, vol. 18, no. 2, pp. 91–102, 2002. DOI: 10.1002/smi.927. [Online]. Available: https://doi.org/10.1002/smi.927 (cit. on p. 51).

[41] H. Hinds and W. J. Burroughs, 'How you know when you're stressed: Self-evaluations of stress', *The Journal of General Psychology*, vol. 124, no. 1, pp. 105–111, Jan. 1997. DOI: 10.1080/00221309709595510. [Online]. Available: https://doi.org/10.1080/00221309709595510 (cit. on p. 51).

[42] R. Rosenthal and K. L. Fode, 'The effect of experimenter bias on the performance of the albino rat', *Behavioral Science*, vol. 8, no. 3, pp. 183–189, Jan. 2007. DOI: 10.1002/bs.3830080302. [Online]. Available: https://doi.org/10.1002/bs.3830080302 (cit. on p. 52).

[43] H.-G. Kim, E.-J. Cheon, D.-S. Bai, Y. H. Lee and B.-H. Koo, 'Stress and heart rate variability: A meta-analysis and review of the literature', *Psychiatry Investigation*, vol. 15, no. 3, pp. 235–245, Mar. 2018. DOI: 10.30773/pi.2017.08.17. [Online]. Available: https://doi.org/10.30773/pi.2017.08.17 (cit. on pp. 55, 73).

[44] S. Pourmohammadi and A. Maleki, 'Stress detection using ECG and EMG signals: A comprehensive study', *Computer Methods and Programs in Biomedicine*, vol. 193, p. 105 482, Sep. 2020. DOI: 10.1016/j.cmpb.2020.105482. [Online]. Available: https://doi.org/10.1016/j.cmpb.2020.105482 (cit. on p. 55).

[45] Y. Shafranovich, *Common Format and MIME Type for Comma-Separated Values (CSV) Files*, RFC 4180, Sep. 1981. DOI: 10.17487/RFC4180. [Online]. Available: https://www.rfc-editor.org/rfc/rfc4180 (cit. on p. 56).

[46] A. Ng, 'Machine learning yearning. deeplearning.ai', *URL: https://www. deeplearning. ai*, 2018 (cit. on p. 71).

[47] J. He, K. Li, X. Liao, P. Zhang and N. Jiang, 'Real-time detection of acute cognitive stress using a convolutional neural network from electrocardiographic signal', *IEEE Access*, vol. 7, pp. 42 710–42 717, 2019. DOI: 10.1109/access.2019.2907076. [Online]. Available: https://doi.org/10.1109/access.2019.2907076 (cit. on p. 71).

[48] P. Contributors. 'Linear Layer'. (2023), [Online]. Available: https://pytorch.org/docs/stable/generated/torch.nn.Linear.html#torch.nn.Linear (visited on 30/01/2023) (cit. on p. 72).

[49] M. Koziarski and B. Cyganek, 'Examination of the deep neural networks in classification of distorted signals', in *Artificial Intelligence and Soft Computing*, Springer International Publishing, 2016, pp. 680–688. DOI: 10.1007/978-3-319-39384-1_60. [Online]. Available: https://doi.org/10.1007/978-3-319-39384-1_60 (cit. on p. 73).

[50] D. P. Kingma and J. Ba, *Adam: A method for stochastic optimization*, 2014. DOI: 10.48550/ARXIV.1412.6980. [Online]. Available: https://arxiv.org/abs/1412.6980 (cit. on p. 73).

[51] K.-K. Tseng, D. Lee and C. B. Chen, 'Ecg identification system using neural network with global and local features.', *International Association for Development of the Information Society*, 2016 (cit. on p. 74).

[52] A. P. Association. 'Subject bias'. (3rd Mar. 2013), [Online]. Available: https://dictionary.apa.org/subject-bias (cit. on p. 84).

[53] J. McCambridge, M. de Bruin and J. Witton, 'The effects of demand characteristics on research participant behaviours in non-laboratory settings: A systematic review', *PLoS ONE*, vol. 7, no. 6, H. R. Baradaran, Ed., e39116, Jun. 2012. DOI: 10.1371/journal.pone.0039116. [Online]. Available: https://doi.org/10.1371/journal.pone.0039116 (cit. on p. 84).

[54] M. Vaglio, L. Isola, G Gates and F. Badilini, 'Use of ecg quality metrics in clinical trials', *2010 Computing in Cardiology*, pp. 505–508, 2010 (cit. on p. 89).

[55] G. E. Billman, 'The LF/HF ratio does not accurately measure cardiac sympatho-vagal balance', *Frontiers in Physiology*, vol. 4, 2013. DOI: 10.3389/fphys.2013.00026. [Online]. Available: https://doi.org/10.3389/fphys.2013.00026 (cit. on p. 89).

[56] M. A. Bansal, D. R. Sharma and D. M. Kathuria, 'A systematic review on data scarcity problem in deep learning: Solution and applications', *ACM Computing Surveys*, vol. 54, no. 10s, pp. 1–29, Jan. 2022. DOI: 10.1145/3502287. [Online]. Available: https://doi.org/10.1145/3502287 (cit. on p. 92).

[57] A. B. Nigusse, D. A. Mengistie, B. Malengier, G. B. Tseghai and L. Van Langenhove, 'Wearable smart textiles for long-term electrocardiography monitoring—a review', *Sensors*, vol. 21, no. 12, p. 4174, Jun. 2021. DOI: 10.3390/s21124174. [Online]. Available: https://doi.org/10.3390/s21124174 (cit. on p. 93).

[58] Y. T. Tsukada, M. Tokita, H. Murata, Y. Hirasawa, K. Yodogawam, Y.-k. Iwasaki, K. Asai, W. Shimizu, N. Kasai, H. Nakashima and S. Tsukada, 'Validation of wearable textile electrodes for ECG monitoring', *Heart and Vessels*, vol. 34, no. 7, pp. 1203–1211, Jan. 2019. DOI: 10.1007/s00380-019-01347-8. [Online]. Available: https://doi.org/10.1007/s00380-019-01347-8 (cit. on p. 93).

[59] A. Fratini, M. Sansone, P. Bifulco and M. Cesarelli, 'Individual identification via electrocardiogram analysis', *BioMedical Engineering OnLine*, vol. 14, no. 1, Aug. 2015. DOI: 10.1186/s12938-015-0072-y. [Online]. Available: https://doi.org/10.1186/s12938-015-0072-y (cit. on p. 94).

[60] P. Lorrig, *Stressdetectioninflightcrews*, https://github.com/0Patrice/StressDetectionInFlig 2023 (cit. on p. 94).

Appendix for SPO Study

A

A.1 Randomization of Scenarios for Participants from SPO Study

ID	Scenario Order	Operation
VP 1	1 - 2 - 3 - 4 - 5	Single
VP 2	2 - 1 - 3 - 4 - 5	Single
VP 3	3 - 2 - 4 - 1 - 5	Single
VP 4	2 - 4 - 3 - 1 - 5	Single
VP 5	3 - 1 - 4 - 2 - 5	Single
VP 6	3 - 1 - 2 - 4 - 5	Single
VP 7	4 - 3 - 2 - 1 - 5	Single
VP 8	2 - 3 - 1 - 4 - 5	Single
VP 9	2 - 4 - 1 - 3 - 5	Single
VP 10	2 - 1 - 4 - 3 - 5	Single
VP 11	1 - 3 - 4 - 2 - 5	Single
VP 12	3 - 4 - 2 - 1 - 5	Single
VP 13	4 - 1 - 2 - 1 - 5	Single
VP 14	4 - 2 - 1 - 3 - 5	Single
VP 15	1 - 2 - 3 - 4 - 5	Dual
VP 16	4 - 3 - 2 - 1 - 5	Dual
VP 17	3 - 2 - 4 - 1 - 5	Dual
VP 18	2 - 4 - 3 - 1 - 5	Dual
VP 19	3 - 1 - 4 - 2 - 5	Dual
VP 20	3 - 1 - 2 - 4 - 5	Dual
VP 21	2 - 1 - 3 - 4 - 5	Dual
VP 22	1 - 3 - 4 - 2 - 5	Dual
VP 23	4 - 2 - 1 - 3 - 5	Dual
VP 24	2 - 1 - 4 - 3 - 5	Dual

Table A.1: Randomization of Scenarios for each Participant for the Single Pilot Operations Study early 2022

A.2 Demographic Questionaire Results from SPO Study's Participants

ID	Age	Gender	Nationality	Position	Current aircraft type(s)	Previous aircraft type(s)	Number of flight hours within last 12 months	Total number of flight hours	Flight hours on A320 Family aircraft	Scenario Order	Operation Mode
VP1	31	Male	German	FO	A320		200	1800	250	1 - 2 - 3 - 4 - 5	Single
VP2	52	Male	Deutsch	FO	A320	A320/330	700	11000	10000	2 - 1 - 3 - 4 - 5	Single
VP3	30	Female	German	FO	A320	Crj9	400	3000	1200	3 - 2 - 4 - 1 - 5	Single
VP4	27	Male	German	FO	A320		72	604	503	2 - 4 - 3 - 1 - 5	Single
VP5	27	Male	German	FO	A320	A319, A320	130	735	617	3 - 1 - 4 - 2 - 5	Single
VP6	26	Male	German	FO	A320		500	1500	1500	3 - 1 - 2 - 4 - 5	Single
VP7	30	Male	German	FO	A320		300	3000	3000	4 - 3 - 2 - 1 - 5	Single
VP8	29	Male	German	FO	Airbus A320 Family	Embraer 195, Cessna Citation CJ1+	438	2200	870	2 - 3 - 1 - 4 - 5	Single
VP9	34	Male	German	FO	A32X		200	3000	2500	2 - 4 - 1 - 3 - 5	Single
VP10	28	Male	German	FO	A320		55	830	730	2 - 1 - 4 - 3 - 5	Single
VP11	33	Male	German	FO	A320 Family	C152, BE33, PA28, DA42	620	2940	2720	1 - 3 - 4 - 2 - 5	Single
VP12	28	Male	Austria	FO	A320		60	960	800	3 - 4 - 2 - 1 - 5	Single
VP13	32	Male	German	FO	A320Fam	DH4	639	2900	1800	4 - 1 - 2 - 1 - 5	Single
VP14	38	Male	German	FO	Airbus A320	BAE Avro RJ100	500	7800	6800	4 - 2 - 1 - 3 - 5	Single
VP15	34	Male	German	FO	A320		350	1600	1600	1 - 2 - 3 - 4 - 5	Dual

Continued on next page

Single Pilot Operations Study Demographics

Table A.2 – continued from previous page

ID	Age	Gender	Nationality	Position	Current aircraft type(s)	Previous aircraft type(s)	Number of flight hours within last 12 months	Total number of flight hours	Flight hours on A320 Family aircraft	Scenario Order	Operation Mode
VP16	31	Male	German	FO	A320	None	300	2000	2000	4 - 3 - 2 - 1 - 5	Dual
VP17	30	Male	German	FO	A320		200	1700	1300	3 - 2 - 4 - 1 - 5	Dual
VP18	32	Male	German	FO			263	1600	1600	2 - 4 - 3 - 1 - 5	Dual
VP19	32	Male	German	FO	A320 Fam		4	2800	2800	3 - 1 - 4 - 2 - 5	Dual
VP20	36	Male	German	FO	A320 Family	–	400	4000-5000	4000-5000	3 - 1 - 2 - 4 - 5	Dual
VP21	34	Female	German	FO	A320		0	1800	1524	2 - 1 - 3 - 4 - 5	Dual
VP22	31	Male	German	FO	A220, A320		200	2450	520	1 - 3 - 4 - 2 - 5	Dual
VP23	34	Male	German	FO	A320		4	3600	3400	4 - 2 - 1 - 3 - 5	Dual
VP24	30	Male	German	FO	A320 Family		150	3000	2800	2 - 1 - 4 - 3 - 5	Dual

Table A.2: Single Pilot Operations Study Demographics

A.3 AVES Simulator Data Header from SPO Study

Column	Internal Name	Unit	Description
000000	system-time	[-]	
000001	del.navLocDeviation	[-]	
000002	fms.navLocDeviation	[-]	
000003	model.navGSDeviation	[-]	
000004	fms.navGSDeviation	[-]	
000005	wcls.cp.PosPit	[deg]	sidestick, capt., pitch, position
000006	wcls.cp.PosRol	[deg]	sidestick, capt., roll, position
000007	wcls.fo.PosPit	[deg]	sidestick, FO, pitch, position
000008	wcls.fo.PosRol	[deg]	sidestick, FO, roll, position
000009	wcls.cp.PosOffPit	[Deg]	sidestick, capt., pitch, position offset
000010	wcls.cp.PosOffRol	[Deg]	sidestick, capt., roll , position offset
000011	wcls.fo.PosOffPit	[Deg]	sidestick, FO, pitch, position offset
000012	wcls.fo.PosOffRol	[Deg]	sidestick, FO, roll , position offset
000013	model.simIAS	[-]	Real indicated airspeed used for trimming
000014	model.simTAS	[-]	Real true airspeed used for trimming
000015	model.navBaroAltADIRU1	[-]	
000016	model.navBaroAltADIRU2	[-]	
000017	model.navBaroAltADIRU3	[-]	
000018	model.navGPSAltitudeGPS1	[-]	
000019	model.navGPSAltitudeGPS2	[-]	
000020	model.navRadioAltRA1	[-]	Radio altitude value measured by radio altimeter 1
000021	model.navRadioAltRA2	[-]	Radio altitude value measured by radio altimeter 2
000022	model.simPilotAlt	[-]	WGS84 altitude of pilot seat (from equations of motion)
000023	model.simAltitude	[-]	Real geometric altitude used for trimming
000024	model.navIASADIRU1	[-]	
000025	model.navIASADIRU2	[-]	
000026	model.navIASADIRU3	[-]	
000027	model.navTASADIRU1	[-]	True Air Speed (TAS)

Continued on next page

Table A.3 – continued from previous page

Column	Internal Name	Unit	Description
000028	model.navTASADIRU2	[-]	True Air Speed (TAS)
000029	model.navTASADIRU3	[-]	True Air Speed (TAS)
000030	model.fctlFlapsDeflec	[-]	(0 ... 40) flap actual position
000031	model.fctlFlapSelectedPosition	[-]	Selected flap config
000032	model.simTrimFlaps	[-]	Flap angle used as trim input
000033	model.gearGearPositionMainLH	[-]	0: Gear up, 1: Gear fully extracted
000034	model.gearGearPositionMainRH	[-]	0: Gear up, 1: Gear fully extracted
000035	model.gearGearPositionNose	[-]	0: Gear up, 1: Gear fully extracted
000036	model.gearGearFullyExtendedNose	[-]	0: Gear not fully extended, 1: Gear fully extended
000037	model.gearGearFullyExtendedMainRH	[-]	0: Gear not fully extended, 1: Gear fully extended
000038	model.gearGearFullyExtendedMainLH	[-]	0: Gear not fully extended, 1: Gear fully extended
000039	model.gearGearFullyRetractedNose	[-]	0: Gear not fully retracted, 1: Gear fully retracted
000040	model.gearGearFullyRetractedMainRH	[-]	0: Gear not fully retracted, 1: Gear fully retracted
000041	model.gearGearFullyRetractedMainLH	[-]	0: Gear not fully retracted, 1: Gear fully retracted
000042	model.gearBrakeGearCmdLH	[-]	0.0: Brake off 1.0: Brake Full
000043	model.gearBrakeGearCmdRH	[-]	0.0: Brake off 1.0: Brake Full
000044	model.gearExtractRetractGearCmd	[-]	0: up 1:down
000045	model.simTrimGearBrakeLH	[-]	LH brake signal used as trim input
000046	model.simTrimGearBrakeRH	[-]	LH brake signal used as trim input
000047	model.simTrimGearExtension	[-]	Gear extension signal used as trim input
000048	model.gearGearUplockedNose	[-]	1: Nose gear uplocked 0: Nose gear not uplocked
000049	model.gearGearUplockedMainLH	[-]	1: Main LH gear uplocked 0:Main LH not uplocked
000050	model.gearGearUplockedMainRH	[-]	1: Main RH gear uplocked 0:Main RH gear not uplocked
000051	model.gearGearDownlockedNose	[-]	1: Nose gear downlocked 0:Nose gear not downlocked

Continued on next page

AVES Sim Data Header from SPO Study

Table A.3 – continued from previous page

Column	Internal Name	Unit	Description
000052	model.gearGearDownlockedMainLH	[-]	1: Main LH gear downlocked 0: Main LH gear not downlocked
000053	model.gearGearDownlockedMainRH	[-]	1: Main RH gear downlocked 0: Main RH gear not downlocked
000054	model.gearGearRetActMainRHState	[-]	State of the main RH gear actuator
000055	model.gearGearRetActMainLHState	[-]	State of the main LH gear actuator
000056	model.gearGearRetActNoseState	[-]	State of the nose gear actuator
000057	model.simEngThrustLH	[-]	LH Engine Thrust
000058	model.simEngThrustRH	[-]	RH Engine Thrust
000059	model.fctlThrustIdleCmdLH	[-]	0: Eng cmd not Idle, 1: Engine cmd Idle
000060	model.fctlThrustIdleCmdRH	[-]	0: Eng cmd not Idle, 1: Engine cmd Idle
000061	model.engEPRCmdAThrLH	[-]	EPR commanded by A/THR
000062	model.engEPRCmdAThrRH	[-]	EPR commanded by A/THR
000063	model.engThrustInIdleLH	[-]	0: Thrust is not in idle 1:Thrust is in idle
000064	model.engThrustInIdleRH	[-]	0: Thrust is not in idle 1:Thrust is in idle
000065	model.simTotalThrust	[-]	Sum of both engine thrust forces used for trimming
000066	model.simDeltaThrust	[-]	Delta of both engine thrust forces used for trimming
000067	model.afltAThrModeEngaged	[-]	0: Disengaged 1: Engaged Autothrust Mode
000068	model.afltAThrThrEngaged	[-]	0: Disengaged 1: Engaged Autothrust Mode
000069	model.afltAThrApprThrEngaged	[-]	0: Disengaged 1: Engaged Autothrust Mode
000070	model.afltAThrApprSpeedEngaged	[-]	0: Disengaged 1: Engaged Autothrust Mode
000071	model.afltAThrApprMachEngaged	[-]	0: Disengaged 1: Engaged Autothrust Mode
000072	model.afltAThrAlphaFloorEngaged	[-]	0: Disengaged 1: Engaged Autothrust Mode
000073	model.afltAThrSpdModeEngaged	[-]	0: Disengaged 1: Engaged Autothrust Mode
000074	model.afltAThrMachModeEngaged	[-]	0: Disengaged 1: Engaged Autothrust Mode
000075	model.afltAThrRetardModeEngaged	[-]	0: Disengaged 1: Engaged Autothrust Mode

Continued on next page

AVES Sim Data Header from SPO Study

Table A.3 – continued from previous page

Column	Internal Name	Unit	Description
000076	model.afltModeChangeAThr	[-]	0: No Change 1: Change Mode Change Flag for controller blending purpose
000077	model.fctlThrLeverAngleLH	[-]	Thrust Lever angle from full reverse (-20 deg) to TOGA (+45 deg)
000078	model.fctlThrLeverAngleRH	[-]	Thrust Lever angle from full reverse (-20 deg) to TOGA (+45 deg)
000079	model.afltAThrThrustLocked	[-]	0: Thrust not locked; 1: Thrust locked
000080	model.navTunedILSThresholdLat	[-]	Latitude of threshold of runway for tuned ILS
000081	model.navTunedILSThresholdLon	[-]	Longitude of threshold of runway for tuned ILS
000082	model.simTrimThrottleLH	[-]	LH throttle setting used as trim input
000083	model.simTrimThrottleRH	[-]	RH throttle setting used as trim input
000084	fcu_heading	[-]	temp. heading offset fpr fcu
000085	fcu_speed	[-]	temp. speed offset fpr fcu
000086	fcu_flightpath	[-]	temp. flightpath offset fpr fcu.
000087	fcu_altitude	[-]	temp. altitude offset fpr fcu
000088	model.navWindSpeedADIRU1	[-]	
000089	model.navWindSpeedADIRU2	[-]	
000090	model.navWindSpeedADIRU3	[-]	
000091	model.navGroundSpdADIRU1	[-]	
000092	model.navVertSpdADIRU1	[-]	
000093	model.navGroundSpdADIRU2	[-]	
000094	model.navVertSpdADIRU2	[-]	
000095	model.navGroundSpdADIRU3	[-]	
000096	model.navVertSpdADIRU3	[-]	
000097	model.navGPSGroundSpdGPS1	[-]	
000098	model.navGPSGroundSpdGPS2	[-]	
000099	model.gearWheelSpdMGL	[-]	Wheel speed of left main gear
000100	model.simTrimWheelSpdCmd	[-]	Command for the trim wheel speed CAN input.
000101	model.afltVertSpdSetpoint	[-]	Actual setpoint for the Autoflight Controller
000102	model.afltVertSpdModeEngaged	[-]	0: Disengaged 1: Engaged Vertical Mode

Continued on next page

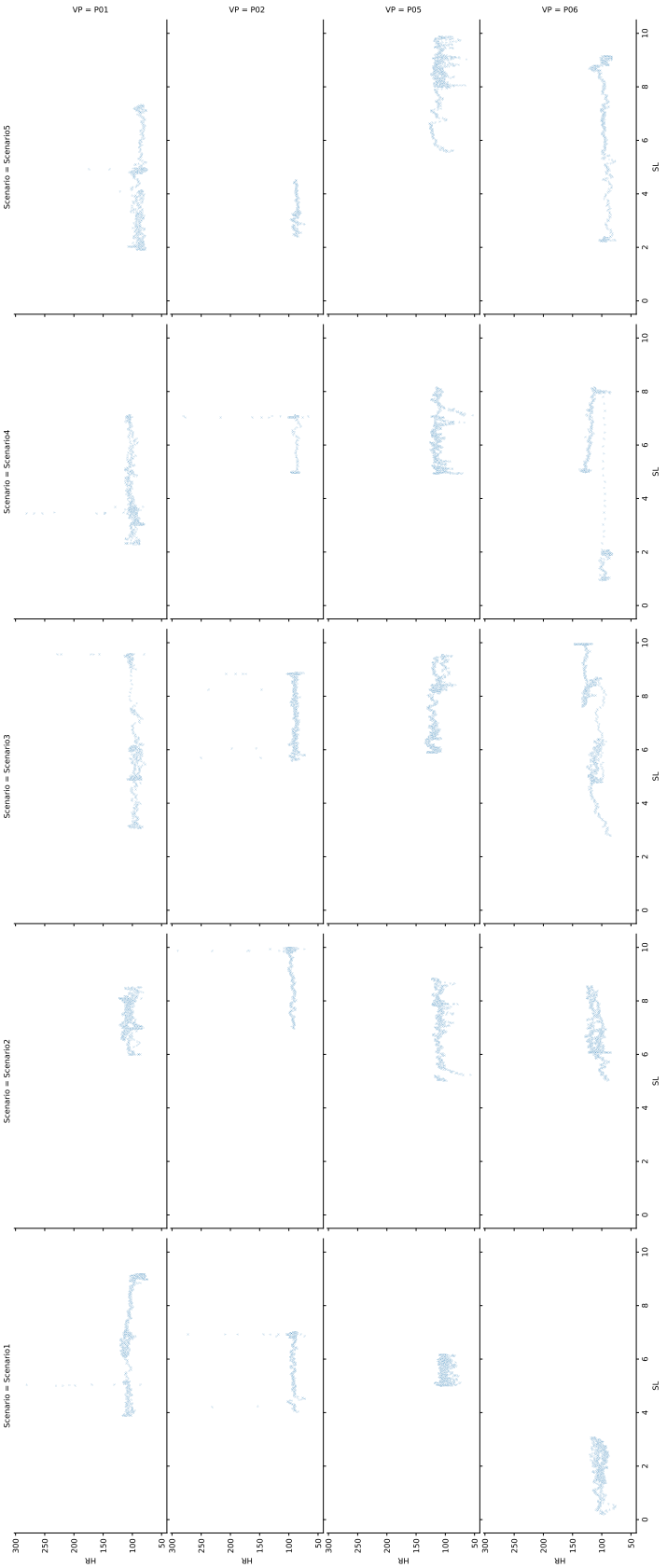
AVES Sim Data Header from SPO Study

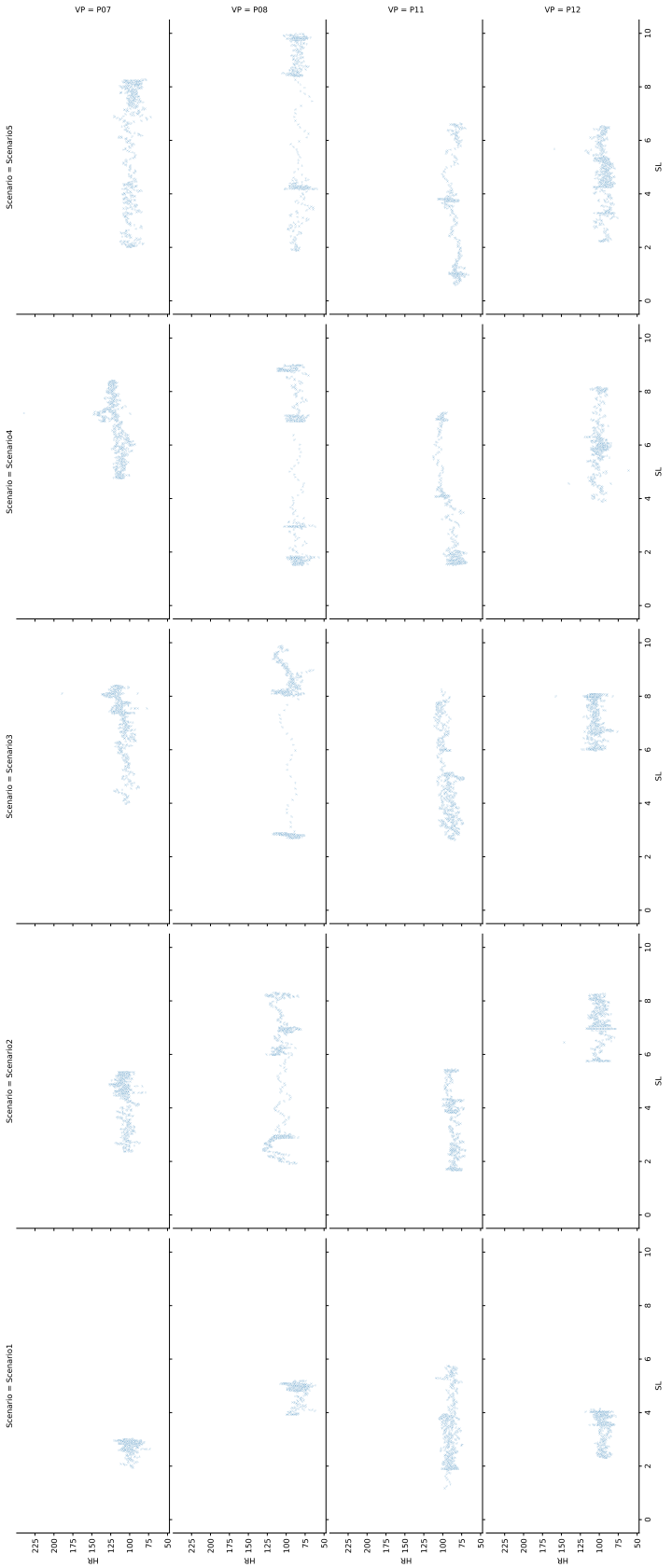
Table A.3 – continued from previous page

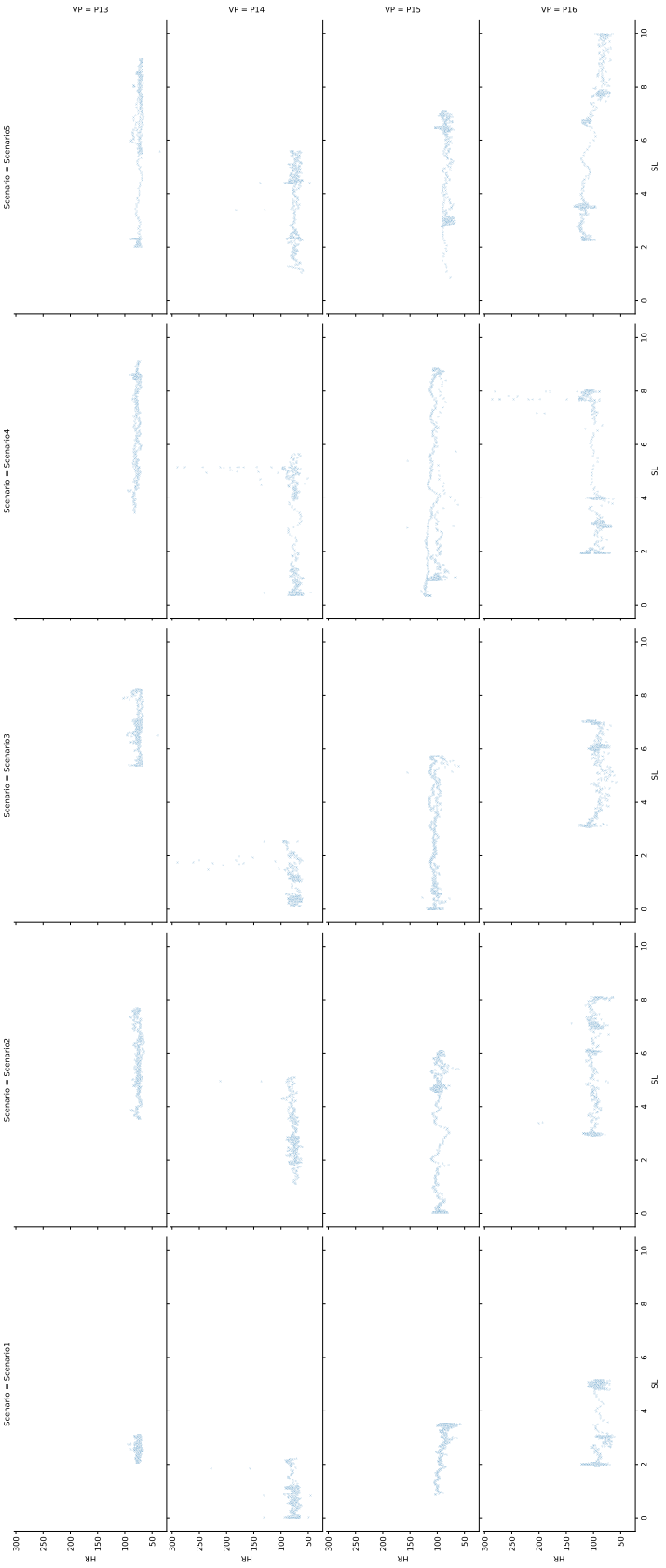
Column	Internal Name	Unit	Description
000103	model.afltAThrSpdModeEngaged	[-]	0: Disengaged 1: Engaged Autothrust Mode
000104	model.afltSpdModeEngaged	[-]	0: Disengaged 1: Engaged Speed Mode
000105	model.fctlSpdBrakeCmd	[-]	0.0: Full Retracted-1.0: Full Deflected
000106	model.afltLowerSpdLimit	[-]	Lower speed limit for the FMGC speed command
000107	model.simXPosNED	[-]	Aircraft x-position in local NED-coordinates
000108	model.simYPosNED	[-]	Aircraft y-position in local NED-coordinates
000109	model.simZPosNED	[-]	Aircraft z-position in local NED-coordinates
000110	model.simCGLat	[-]	WGS84 latitude of center of gravity (from equations of motion)
000111	model.simPilotLat	[-]	WGS84 latitude of pilot seat (from equations of motion)
000112	model.simCGLon	[-]	WGS84 longitude of center of gravity (from equations of motion)
000113	model.simPilotLon	[-]	WGS84 longitude of pilot seat (from equations of motion)
000114	model.simCGAlt	[-]	WGS84 altitude of center of gravity (from equations of motion)

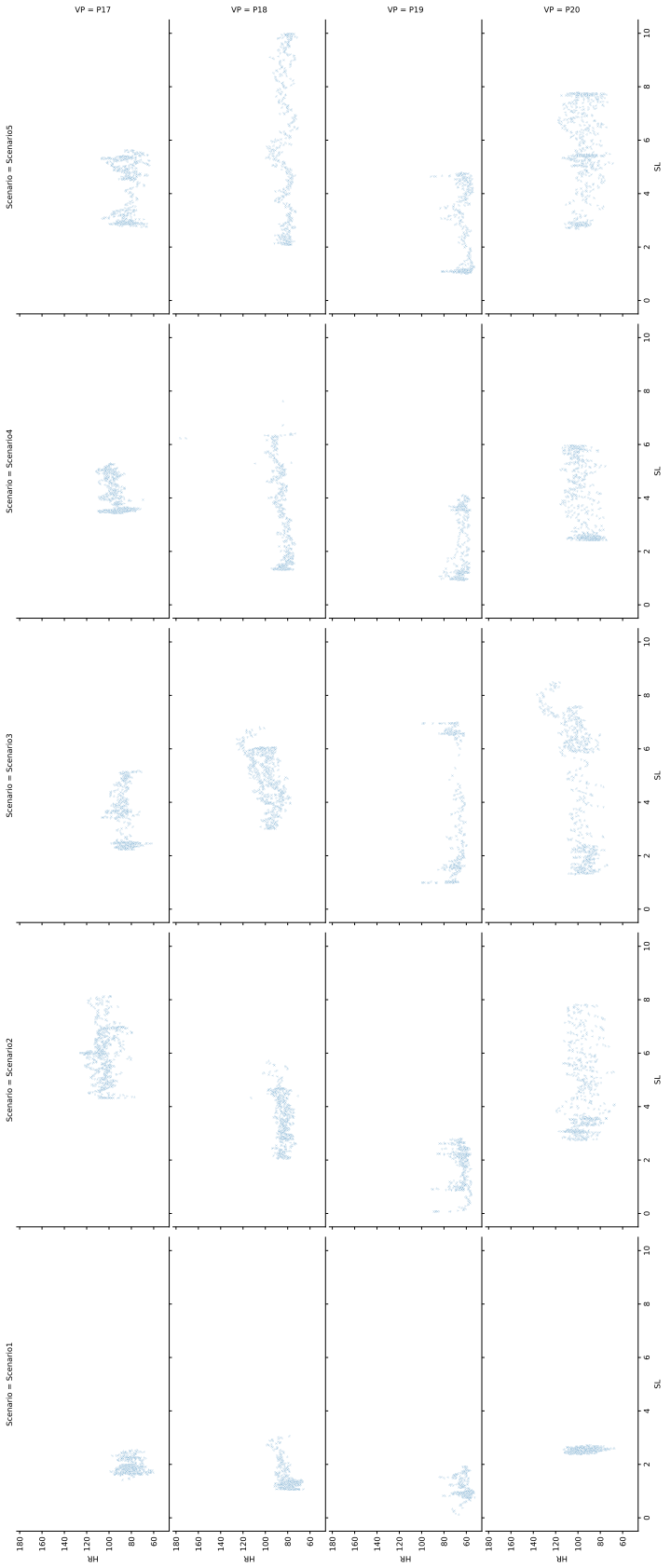
Table A.3: AVES Sim Data Header from SPO Study

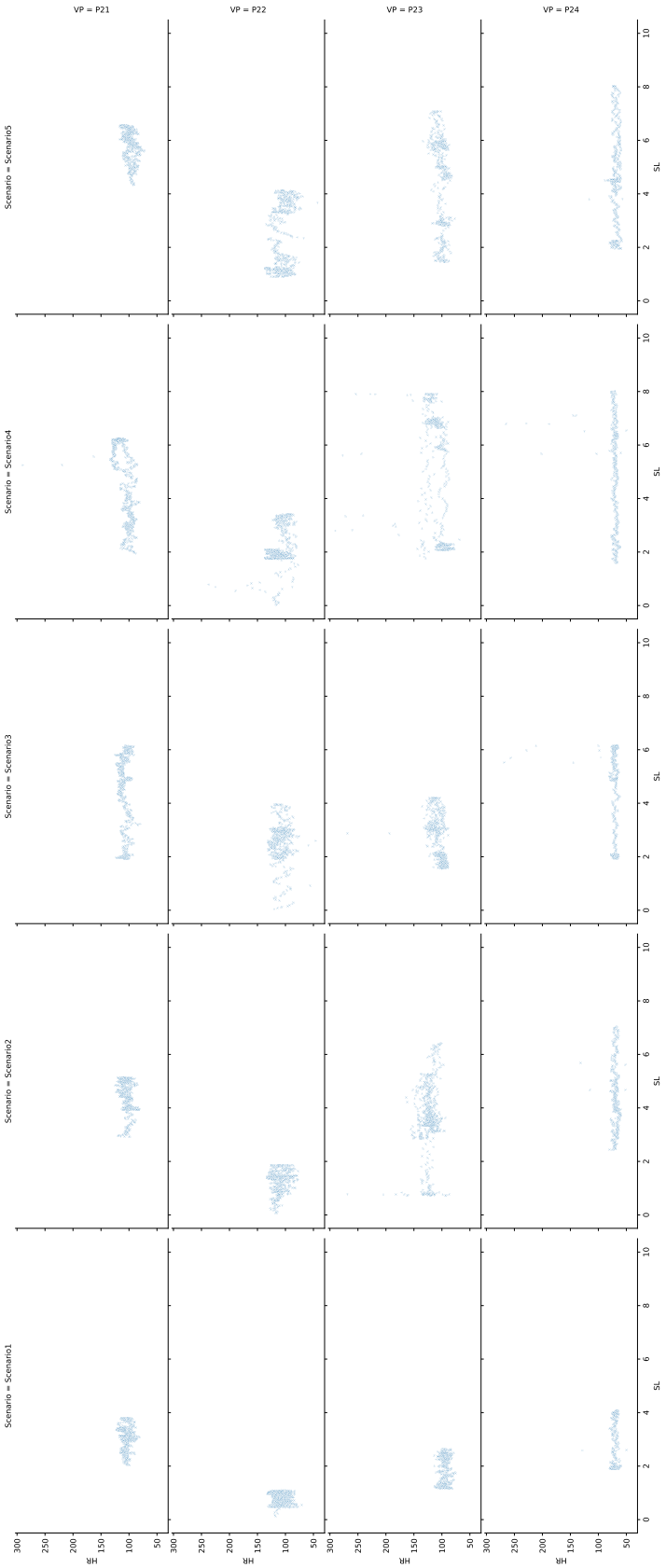
A.4 Data Distribution from SPO Study


VP = P01
VP = P02
VP = P05
VP = P06
Scenario = Scenario5
Scenario = Scenario4
Scenario = Scenario3
Scenario = Scenario2
Scenario = Scenario1
HR
SL


VP = P07
VP = P08
VP = P11
VP = P12
Scenario = Scenario5
Scenario = Scenario4
Scenario = Scenario3
Scenario = Scenario2
Scenario = Scenario1
HR
SL


VP = P13
VP = P14
VP = P15
VP = P16
Scenario = Scenario5
Scenario = Scenario4
Scenario = Scenario3
Scenario = Scenario2
Scenario = Scenario1
HR
SL


VP = P17
VP = P18
VP = P19
VP = P20
Scenario = Scenario5
Scenario = Scenario4
Scenario = Scenario3
Scenario = Scenario2
Scenario = Scenario1
HR
SL


VP = P21
VP = P22
VP = P23
VP = P24
Scenario = Scenario5
Scenario = Scenario4
Scenario = Scenario3
Scenario = Scenario2
Scenario = Scenario1
HR
SL

Appendix for LoHP Study

B

B.1 Demographic Questionaire Results from LoHP Study's Participants

ID	Age	Gender	Nationality	Position	Current aircraft type(s)	Previous aircraft type(s)	Number of flight hours within last 12 months	Total number of flight hours	Flight hours on A320 Family aircraft	Realism 2nd Scenario
VP1	31	Male	germany	First Officer	Airbus A320 Family	Cessna Cj1+	500	3000	3000	7
VP2	37	Male	German	First Officer	A320 family	none	500	3400	3400	2
VP3	41	Male	Germany	Captain	A320FAM	Training A/C like PA42, Beech Bonanza and private Ultralight Aircraft	4	9200	8500	2
VP4	30	Male	German	First Officer	A320	Sailplanes	350	1700	1200	2
VP5	64	Male	german	Captain	A380	B737, DC10, A320, A330, A340	0	18000	12000	2
VP6	30	Male	german	First Officer	A320	none	20	3000	2900	3
VP7	39	Male	German	First Officer	B747	Airbus 320	700	11000	7000	3
VP8	39	Male	Austria	First Officer	A350	A320, MD11, A340family, A330	500	9000	4000	4
VP9	33	Male	German	First Officer	A320 Family	A320 Family	450	6700	6500	2
VP10	35	Male	German	First Officer	Airbus A320	Airbus A320	200	3000	2800	4
VP13	48	Male	German	Captain	A320fam	B737 B747-200	650	16800	12000	7

Continued on next page

Limits of Human Performance Study Demographics. Realism 1 = not realistic, 10 = very realistic.

Table B.1 – continued from previous page

ID	Age	Gender	Nationality	Position	Current aircraft type(s)	Previous aircraft type(s)	Number of flight hours within last 12 months	Total number of flight hours	Flight hours on A320 Family aircraft	Realism 2nd Scenario
VP14	34	Male	german	First Officer	A320,319,321	B737	700	4000	2000	3
VP15	33	Male	German	First Officer	A320 Fam	none	0	3200	2900	6
VP16	36	Male	GERMAN	First Officer	A320	B737	192	5436	2718	4
VP17	49	Male	german	Captain	Airbus A319, A320, A321	Avro RJ85, Airbus A300/A310, Airbus A330/A340	500	13500	500	8
VP18	33	Male	German	First Officer	A320 Family	SEP	660	3400	3150	6
VP19	46	Male	German	Captain	A320	B737, MD11,A330/340	550	17000	2000	4
VP20	27	Male	German	First Officer	Airbus A320 family	A319, A320	328	1199	1082	1

Table B.1: Limits of Human Performance Study Demographics. Realism 1 = not realistic, 10 = very realistic.

B.2 AVES Simulator Data Header from LoHP Study

Column	Internal Name	Unit	Description
000000	system-time	[-]	
000001	model.simModelTime	[-]	Header output for a/c model containing the simulation time
000002	model.navLocDeviation	[-]	
000003	fms.navLocDeviation	[-]	
000004	model.navGSDeviation	[-]	
000005	fms.navGSDeviation	[-]	
000006	wcls.cp.PosPit	[deg]	sidestick, capt., pitch, position
000007	wcls.cp.PosRol	[deg]	sidestick, capt., roll, position
000008	wcls.fo.PosPit	[deg]	sidestick, FO, pitch, position
000009	wcls.fo.PosRol	[deg]	sidestick, FO, roll, position
000010	wcls.cp.PosOffPit	[Deg]	sidestick, capt., pitch, position offset
000011	wcls.cp.PosOffRol	[Deg]	sidestick, capt., roll , position offset
000012	wcls.fo.PosOffPit	[Deg]	sidestick, FO, pitch, position offset
000013	wcls.fo.PosOffRol	[Deg]	sidestick, FO, roll , position offset
000014	model.simIAS	[-]	Real indicated airspeed used for trimming
000015	model.simTAS	[-]	Real true airspeed used for trimming
000016	model.navBaroAltADIRU1	[-]	
000017	model.navBaroAltADIRU2	[-]	
000018	model.navBaroAltADIRU3	[-]	
000019	model.navGPSAltitudeGPS1	[-]	
000020	model.navGPSAltitudeGPS2	[-]	
000021	model.navRadioAltRA1	[-]	Radio altitude value measured by radio altimeter 1
000022	model.navRadioAltRA2	[-]	Radio altitude value measured by radio altimeter 2
000023	model.simPilotAlt	[-]	WGS84 altitude of pilot seat (from equations of motion)
000024	model.simAltitude	[-]	Real geometric altitude used for trimming
000025	model.navIASADIRU1	[-]	

Continued on next page

AVES Sim Data Header from LoHP Study

Table B.2 – continued from previous page

Column	Internal Name	Unit	Description
000026	model.navIASADIRU2	[-]	
000027	model.navIASADIRU3	[-]	
000028	model.navTASADIRU1	[-]	True Air Speed (TAS)
000029	model.navTASADIRU2	[-]	True Air Speed (TAS)
000030	model.navTASADIRU3	[-]	True Air Speed (TAS)
000031	model.fctlFlapsDeflec	[-]	(0 ... 40) flap actual position
000032	model.fctlFlapSelectedPosition	[-]	Selected flap config
000033	model.simTrimFlaps	[-]	Flap angle used as trim input
000034	model.gearGearPositionMainLH	[-]	0: Gear up, 1: Gear fully extracted
000035	model.gearGearPositionMainRH	[-]	0: Gear up, 1: Gear fully extracted
000036	model.gearGearPositionNose	[-]	0: Gear up, 1: Gear fully extracted
000037	model.gearGearFullyExtendedNose	[-]	0: Gear not fully extended, 1: Gear fully extended
000038	model.gearGearFullyExtendedMainRH	[-]	0: Gear not fully extended, 1: Gear fully extended
000039	model.gearGearFullyExtendedMainLH	[-]	0: Gear not fully extended, 1: Gear fully extended
000040	model.gearGearFullyRetractedNose	[-]	0: Gear not fully retracted, 1: Gear fully retracted
000041	model.gearGearFullyRetractedMainRH	[-]	0: Gear not fully retracted, 1: Gear fully retracted
000042	model.gearGearFullyRetractedMainLH	[-]	0: Gear not fully retracted, 1: Gear fully retracted
000043	model.gearBrakeGearCmdLH	[-]	0.0: Brake off 1.0: Brake Full
000044	model.gearBrakeGearCmdRH	[-]	0.0: Brake off 1.0: Brake Full
000045	model.gearExtractRetractGearCmd	[-]	0: up 1:down
000046	model.simTrimGearBrakeLH	[-]	LH brake signal used as trim input
000047	model.simTrimGearBrakeRH	[-]	LH brake signal used as trim input
000048	model.simTrimGearExtension	[-]	Gear extension signal used as trim input
000049	model.gearGearUplockedNose	[-]	1: Nose gear uplocked 0: Nose gear not uplocked
000050	model.gearGearUplockedMainLH	[-]	1: Main LH gear uplocked 0:Main LH not uplocked
000051	model.gearGearUplockedMainRH	[-]	1: Main RH gear uplocked 0:Main RH gear not uplocked

Continued on next page

AVES Sim Data Header from LoHP Study

Table B.2 – continued from previous page

Column	Internal Name	Unit	Description
000052	model.gearGearDownlockedNose	[-]	1: Nose gear downlocked 0:Nose gear not downlocked
000053	model.gearGearDownlockedMainLH	[-]	1: Main LH gear downlocked 0: Main LH gear not downlocked
000054	model.gearGearDownlockedMainRH	[-]	1: Main RH gear downlocked 0: Main RH gear not downlocked
000055	model.gearGearRetActMainRHState	[-]	State of the main RH gear actuator
000056	model.gearGearRetActMainLHState	[-]	State of the main LH gear actuator
000057	model.gearGearRetActNoseState	[-]	State of the nose gear actuator
000058	model.simEngThrustLH	[-]	LH Engine Thrust
000059	model.simEngThrustRH	[-]	RH Engine Thrust
000060	model.fctlThrustIdleCmdLH	[-]	0: Eng cmd not Idle, 1: Engine cmd Idle
000061	model.fctlThrustIdleCmdRH	[-]	0: Eng cmd not Idle, 1: Engine cmd Idle
000062	model.engEPRCmdAThrLH	[-]	EPR commanded by A/THR
000063	model.engEPRCmdAThrRH	[-]	EPR commanded by A/THR
000064	model.engThrustInIdleLH	[-]	0: Thrust is not in idle 1:Thrust is in idle
000065	model.engThrustInIdleRH	[-]	0: Thrust is not in idle 1:Thrust is in idle
000066	model.simTotalThrust	[-]	Sum of both engine thrust forces used for trimming
000067	model.simDeltaThrust	[-]	Delta of both engine thrust forces used for trimming
000068	model.afltAThrModeEngaged	[-]	0: Disengaged 1: Engaged Autothrust Mode
000069	model.afltAThrThrEngaged	[-]	0: Disengaged 1: Engaged Autothrust Mode
000070	model.afltAThrApprThrEngaged	[-]	0: Disengaged 1: Engaged Autothrust Mode
000071	model.afltAThrApprSpeedEngaged	[-]	0: Disengaged 1: Engaged Autothrust Mode
000072	model.afltAThrApprMachEngaged	[-]	0: Disengaged 1: Engaged Autothrust Mode
000073	model.afltAThrAlphaFloorEngaged	[-]	0: Disengaged 1: Engaged Autothrust Mode
000074	model.afltAThrSpdModeEngaged	[-]	0: Disengaged 1: Engaged Autothrust Mode
000075	model.afltAThrMachModeEngaged	[-]	0: Disengaged 1: Engaged Autothrust Mode

Continued on next page

AVES Sim Data Header from LoHP Study

Table B.2 – continued from previous page

Column	Internal Name	Unit	Description
000076	model.afltAThrRetardModeEngaged	[-]	0: Disengaged 1: Engaged Autothrust Mode
000077	model.afltModeChangeAThr	[-]	0: No Change 1: Change Mode Change Flag for controller blending purpose
000078	model.fctlThrLeverAngleLH	[-]	Thrust Lever angle from full reverse (-20 deg) to TOGA (+45 deg)
000079	model.fctlThrLeverAngleRH	[-]	Thrust Lever angle from full reverse (-20 deg) to TOGA (+45 deg)
000080	model.afltAThrThrustLocked	[-]	0: Thrust not locked; 1: Thrust locked
000081	model.navTunedILSThresholdLat	[-]	Latitude of threshold of runway for tuned ILS
000082	model.navTunedILSThresholdLon	[-]	Longitude of threshold of runway for tuned ILS
000083	model.simTrimThrottleLH	[-]	LH throttle setting used as trim input
000084	model.simTrimThrottleRH	[-]	RH throttle setting used as trim input
000085	fcu_heading	[-]	temp. heading offset fpr fcu
000086	fcu_speed	[-]	temp. speed offset fpr fcu
000087	fcu_flightpath	[-]	temp. flightpath offset fpr fcu.
000088	fcu_altitude	[-]	temp. altitude offset fpr fcu
000089	model.navWindSpeedADIRU1	[-]	
000090	model.navWindSpeedADIRU2	[-]	
000091	model.navWindSpeedADIRU3	[-]	
000092	model.navGroundSpdADIRU1	[-]	
000093	model.navVertSpdADIRU1	[-]	
000094	model.navGroundSpdADIRU2	[-]	
000095	model.navVertSpdADIRU2	[-]	
000096	model.navGroundSpdADIRU3	[-]	
000097	model.navVertSpdADIRU3	[-]	
000098	model.navGPSGroundSpdGPS1	[-]	
000099	model.navGPSGroundSpdGPS2	[-]	
000100	model.gearWheelSpdMGL	[-]	Wheel speed of left main gear
000101	model.simTrimWheelSpdCmd	[-]	Command for the trim wheel speed CAN input.
000102	model.afltVertSpdSetpoint	[-]	Actual setpoint for the Autoflight Controller

Continued on next page

AVES Sim Data Header from LoHP Study

Table B.2 – continued from previous page

Column	Internal Name	Unit	Description
000103	model.afltVertSpdModeEngaged	[-]	0: Disengaged 1: Engaged Vertical Mode
000104	model.afltAThrSpdModeEngaged	[-]	0: Disengaged 1: Engaged Autothrust Mode
000105	model.afltSpdModeEngaged	[-]	0: Disengaged 1: Engaged Speed Mode
000106	model.fctlSpdBrakeCmd	[-]	0.0: Full Retracted-1.0: Full Deflected
000107	model.afltLowerSpdLimit	[-]	Lower speed limit for the FMGC speed command
000108	model.simXPosNED	[-]	Aircraft x-position in local NED-coordinates
000109	model.simYPosNED	[-]	Aircraft y-position in local NED-coordinates
000110	model.simZPosNED	[-]	Aircraft z-position in local NED-coordinates
000111	model.simCGLat	[-]	WGS84 latitude of center of gravity (from equations of motion)
000112	model.simPilotLat	[-]	WGS84 latitude of pilot seat (from equations of motion)
000113	model.simCGLon	[-]	WGS84 longitude of center of gravity (from equations of motion)
000114	model.simPilotLon	[-]	WGS84 longitude of pilot seat (from equations of motion)
000115	model.simCGAlt	[-]	WGS84 altitude of center of gravity (from equations of motion)
000116	flKlf.rec.runEkgRecording	[-]	run EKG recording (0 - IC; 1 - RUN; 2 - HALT)
000117	flKlf.env.runWindshear	[-]	run windshear
000118	flKlf.env.windshearLimit	[kts]	windshear limit
000119	flKlf.env.windshearStep	[kts/s]	windshear step
000120	flKlf.disp.runNavDisplayFlickerCpt	[-]	run NAV display flicker (CPT)
000121	flKlf.disp.runNavDisplayFlickerFO	[-]	run NAV display flicker (F/O)
000122	flKlf.nav.runLocInterference	[-]	run localizer interference
000123	flKlf.nav.locInterferenceLimit	[deg]	localizer interference limit
000124	flKlf.nav.locInterferenceStep	[deg/s]	localizer interference step
000125	flKlf.eng.runEngStallLH	[-]	run engine stall (left hand)
000126	flKlf.eng.runEngStallRH	[-]	run engine stall (right hand)
000127	flKlf.eng.runEngFailureLH	[-]	run engine failure (left hand)

Continued on next page

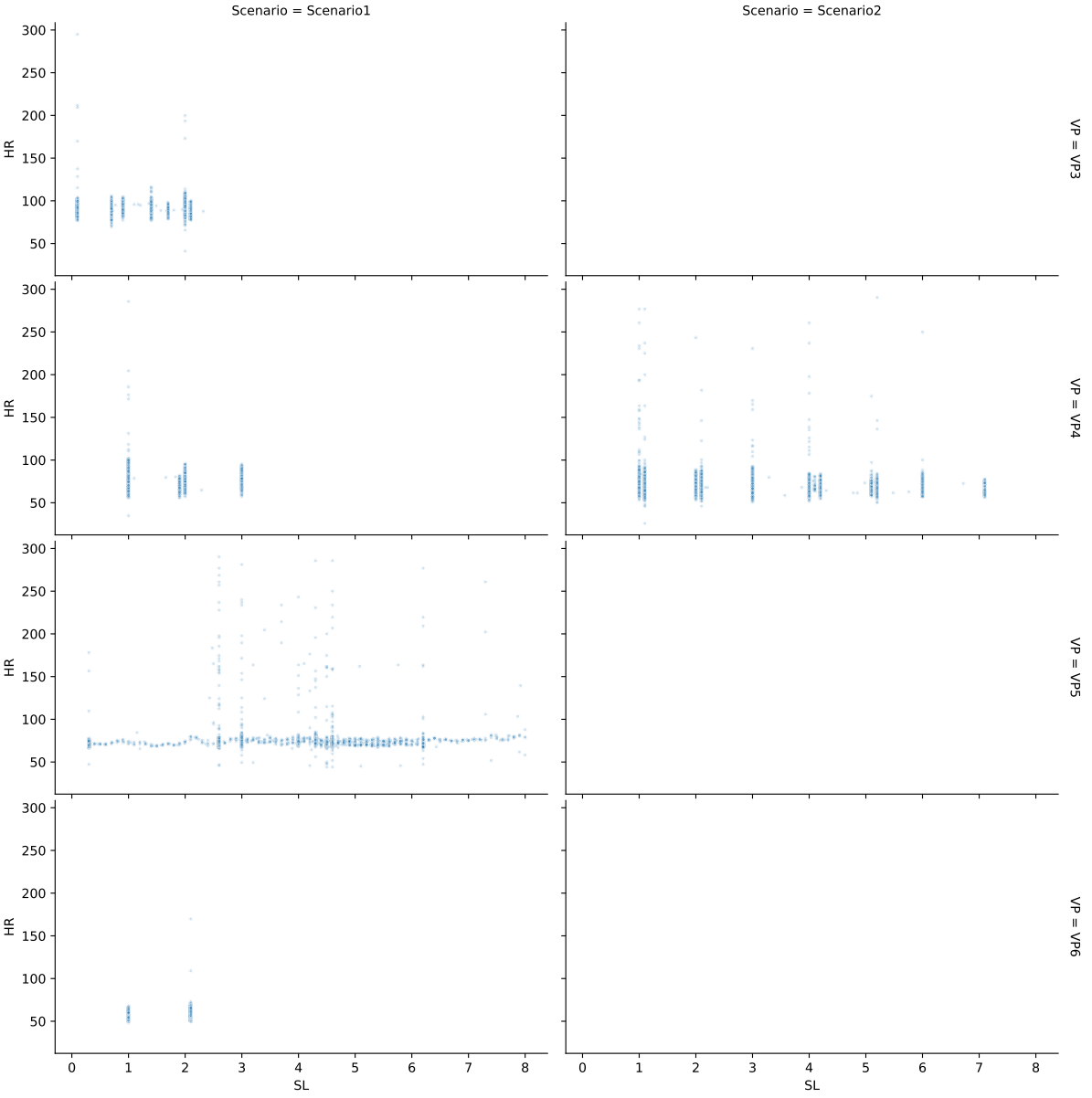
AVES Sim Data Header from LoHP Study

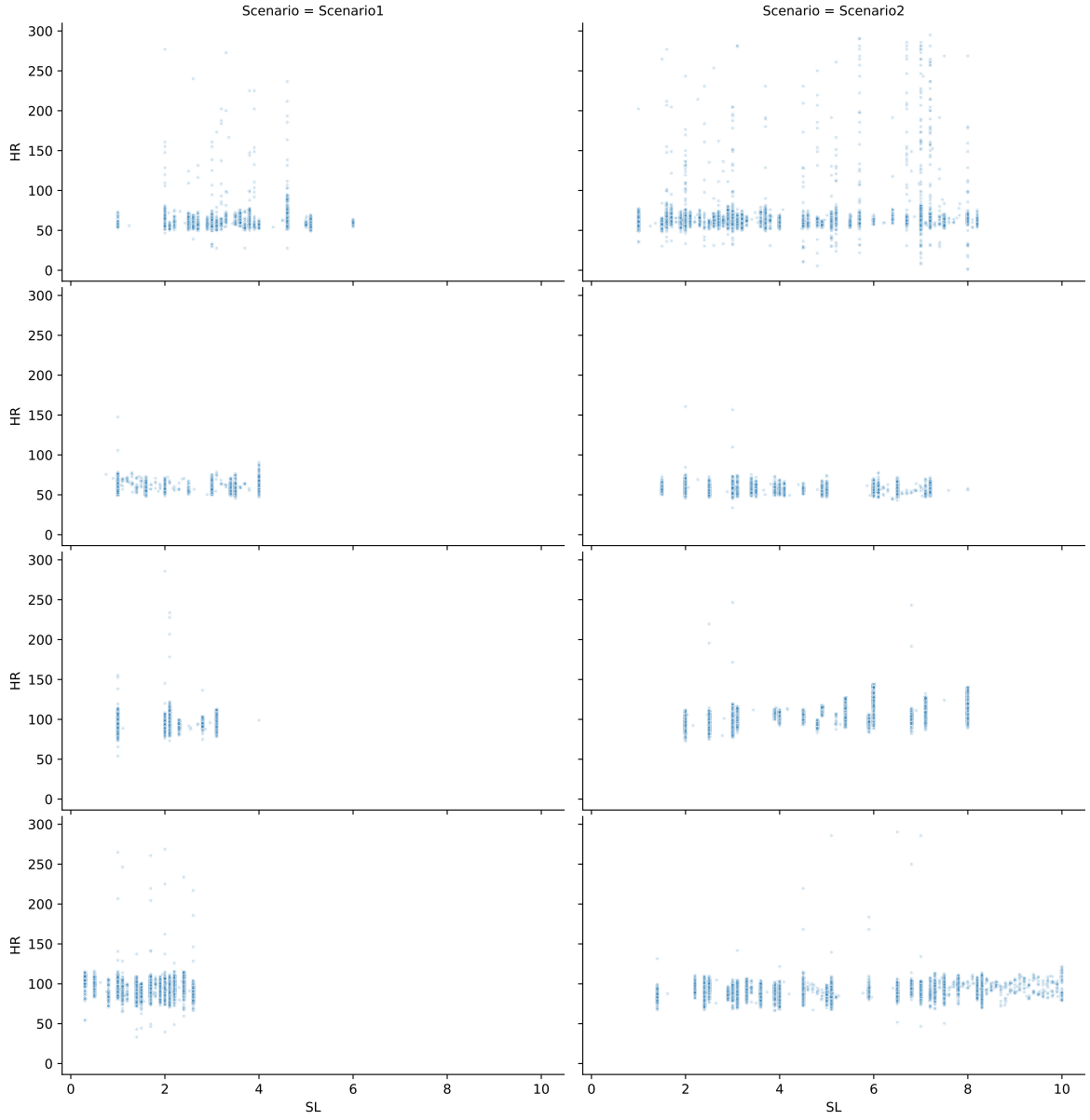
Table B.2 – continued from previous page

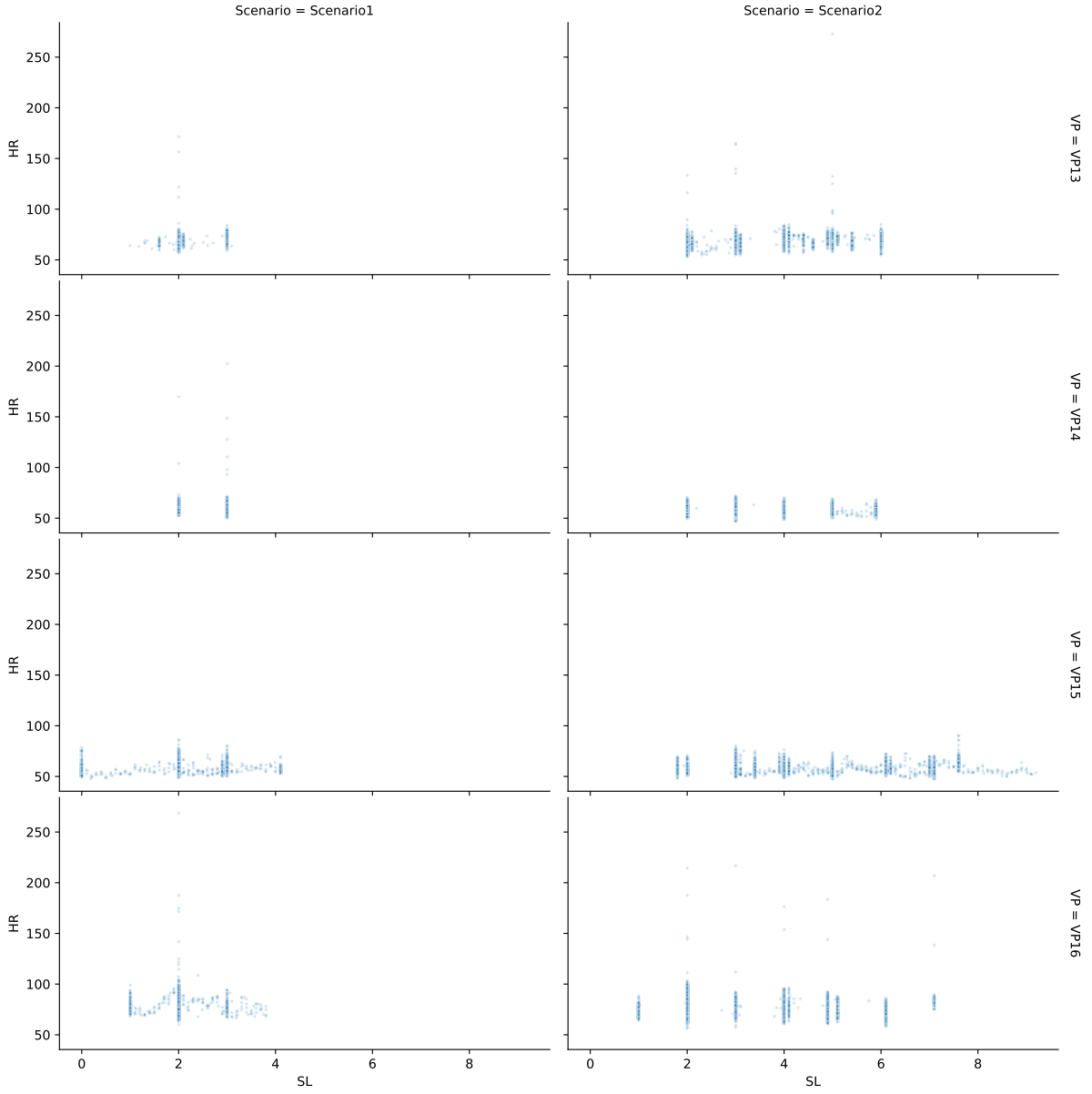
Column	Internal Name	Unit	Description
000128	flKlf.eng.runEngFailureRH	[-]	run engine failure (right hand)
000129	flKlf.eng.runEngShutdownLH	[-]	run engine shutdown (left hand)
000130	flKlf.eng.runEngShutdownRH	[-]	run engine shutdown (right hand)
000131	ecam.engOilLHTemp	[-]	engine 1 oil temperature
000132	ecam.engOilRHTemp	[-]	engine 2 oil temperature
000133	sim:i.a320SE.engOilLHTemp	[-]	engine 1 oil temperature
000134	sim:i.a320SE.engOilRHTemp	[-]	engine 2 oil temperature
000135	model.fuelOnBoard	[-]	Total amount of fuel on board
000136	model.fuelInnerLeakRateLH	[-]	Leak rate of inner LH wing tank
000137	model.fuelInnerLeakRateRH	[-]	Leak rate of inner RH wing tank

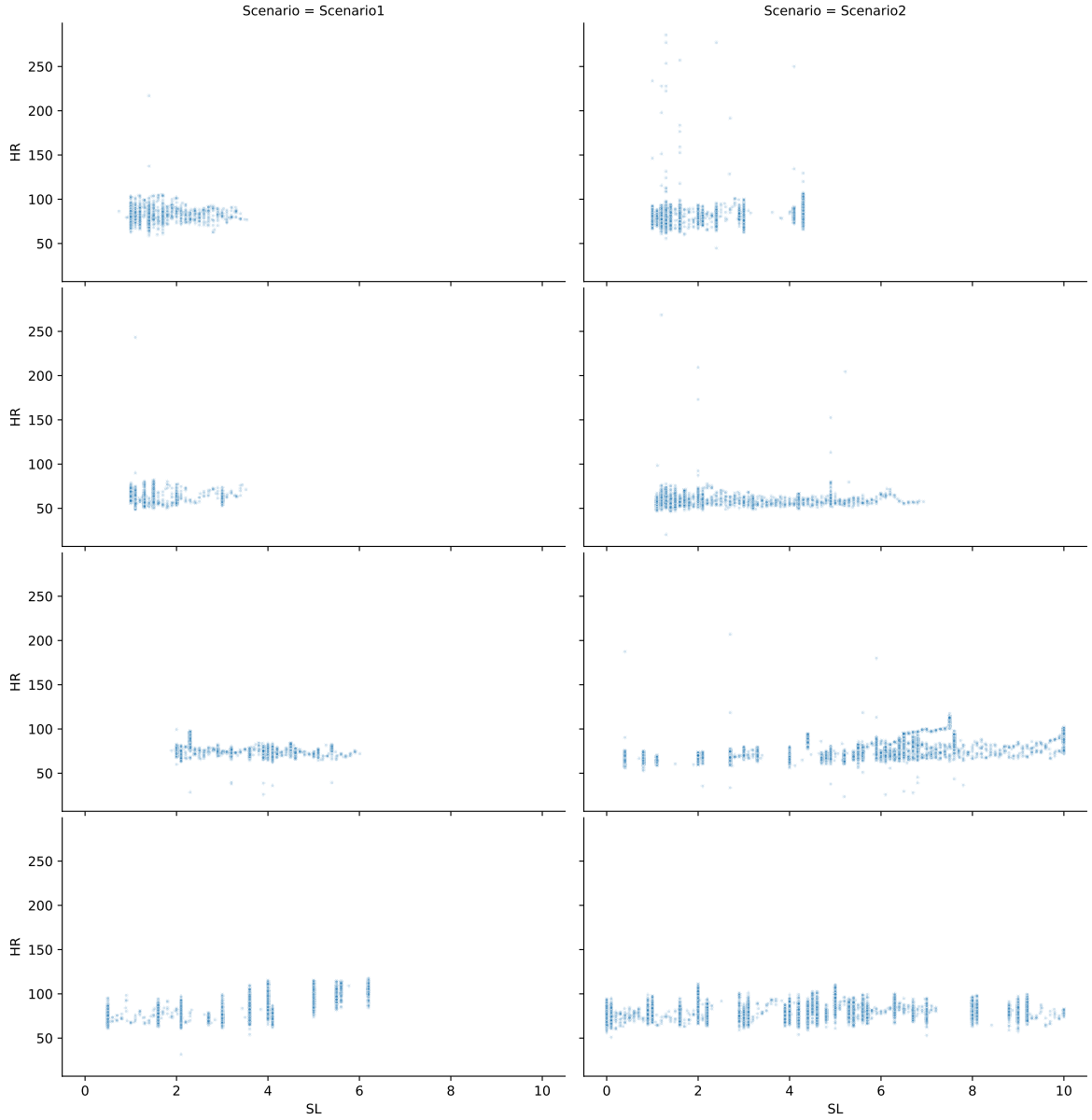
Table B.2: AVES Sim Data Header from LoHP Study

B.3 Data Distribution from LoHP Study


Scenario = Scenario1
Scenario = Scenario2
VP = VP3
VP = VP4
VP = VP5
VP = VP6
HR
SL
300
250
200
150
100
50
0
1
2
3
4
5
6
7
8


Scenario = Scenario1
Scenario = Scenario2
VP = VP10
VP = VP7
VP = VP8
VP = VP9
HR
SL
0
50
100
150
200
250
300
2
4
6
8
10


Scenario = Scenario1
Scenario = Scenario2
VP = VP13
VP = VP14
VP = VP15
VP = VP16
HR
SL
50
100
150
200
250
0
2
4
6
8


Scenario = Scenario1
Scenario = Scenario2
VP = VP17
VP = VP18
VP = VP19
VP = VP20
HR
SL
50
100
150
200
250
0
2
4
6
8
10