RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

ISAE SUPAERO Institut Supérieur de l'Aéronautique et de l'Espace

Université de Toulouse

Doctorat de l'Université de Toulouse

Délivré par l'ISAE-SUPAERO

Interaction humain/système de drones et facteurs humains : Prise en compte de l'estimation de l'état de fatigue d'un opérateur dans le design d'interactions adaptatives pour le contrôle de drones longue endurance

---

Thèse présentée et soutenue, le 26 septembre 2024 par

Marcel HINSS

École doctorale
SYS - Systèmes

Spécialité
Systèmes embarqués

Unité de recherche
ISAE-ONERA DECISIO DECIsion Supervision et Interaction pour l'Opération de systèmes complexes

Thèse dirigée par
Raphaëlle ROY

Composition du jury
M. Fabien LOTTE, Examinateur, Président, INRIA Bordeaux
Mme Anne-Marie BROUWER, Rapporteure, Radboud University Nijmegen
M. Philippe CABON, Rapporteur, Université Paris Cité,
M. Joost DE WINTER, Deflt University of Technology
M. Timothy MERRITT, Examinateur, Aalborg Universitet
Mme Raphaëlle ROY, Directrice de thèse, ISAE-SUPAERO

Membres invités
Mme Anke BROCK, ENAC

Acknowledgements

Three years of work, three years living in France, three years of cycling to work with the wind suspiciously constantly blowing in the wrong direction, except for the days you decide to take the metro. Three years of learning, exploring, testing, re-testing, writing, re-writing, rewriting.final, rewriting.final2, rewriting.final2.the.last.one, rewriting.final2.the.one.I.promise.

All of this and much more would not have been imaginable without the support of many people, many of whom I will mention here and some that I will forget. For the forgetting, I am sorry. I blame my bad memory and the summer heat robbing me of my sleep. Maybe a bit of wind would not be that bad after all.

To Raphaëlle, before and when I arrived in France for an Internship in 2020, my university's policy on foreign internships was very strict. For a long time, I was not sure whether I could do an internship. Despite not knowing me, you assisted me until we knew I could stay. During my PhD, your support was unwavering, as you allowed me to delve into my research interests and provided guidance whenever I felt lost or unsure. In our meetings (that, because of me, tended to go overtime), you always took the time to understand the ideas I brought forward. Beyond that, you supported me effortlessly in all of my personal projects. I will never be able to thank you enough for all that and much more.

Anke, it is without a doubt that I could not have finished this project without your continued support. You introduced me to a new research domain and provided me with many new ideas that shaped many aspects of our work. I always appreciated your ability to focus on the bigger picture and ask the right questions at the right time. But there is much more that you did for me. You always took the time to make sure I was doing ok and reassured me when I was uncertain. I want to thank you sincerely for all this and so much more! It was a pleasure working under your supervision.

Words may never fully capture the efforts both my supervisors made patiently listening to my often unnecessarily complicated ideas, translating my mix of French, English and German into coherent phrases and rereading texts littered with spelling mistakes and punctured with incorrect punctuation.

I want to thank both the Reviewers Anne-Marie Brouwer and Philippe Cabon as well as the Examiners Joost de Winter, Timothy Merritt and Fabien Lotte of my Defence Jury for taking the time and effort to read this manuscript, prepare and travel to the defence. Seeing the length of this manuscript, no one would hold it against you to temporarily regret your choice of being on this jury.

It was an honour to work at ISAE-SUPAERO and ENAC, to be part of two prestigious schools, and to interact with their students and staff. This project would not have been possible without the support from the DGA, which financed our project.

If you want to go fast, go alone; if you want to go far, go together. I never had the feeling of walking alone, and that is due to some of the great people I had the pleasure of working with who provided me with feedback on many aspects of my work. Thank you to Reinmar, Emilie, Victor, Vincenzo, Tien, Miguel, Xavier, Arnaud, Clément, Darlann, Kalou, Alexandre and Arie.

In my time at ISAE-SUPAERO, I had the pleasure of meeting, talking, laughing, discussing, and sometimes even crying with some of the most remarkable people I have known in my life. Frédéric Dehais, equally devoted to research and surfing, thank you for making me appreciate both time and time again. To Caro, Maëlis and Mathias, you are the people I spend most of my time with, and my only regret is not having more time with you.

Caro, I look to you when I need advice, ranging from appending a list of lists in Python to

anything in my personal life. I can count on you to tell me what you think and on your advice, being with the purest intentions at all times.

Maëlis, being around you is true bliss, and you carry fun to every place you go. You have the gift of keeping things easy and making me laugh daily. You are extraordinarily talented at taking another person's point of view and can always say the right thing that sheds everything in a new light.

Mathias, your devotion to happiness in life is infectious. Rarely have I seen someone so open to new experiences, someone who is so motivating and patient at the same time. You make everybody feel like the main character in a great adventure movie.

Thank you to all three of you! With you, bad times pass fast, and good times feel even better.

But I also express my gratitude to the rest of the CNE Lab, each of whom has supported me in one way or another over the years. Claire, the only person to bring a sweater to a trip to the sun (just in case); Sebastien, unbeaten champion in opening doors in the most violent way; Seva, the jack of all trades; Juan, the keyboard master; Big K Kalou; Pietro, the pétanque magician; Sebastien Velut, the only other remaining open space survivor; Christoph the Rock-star; Emilie, the master of organization; Clément, who fears no cold waters; Mr Open Science Quentin and Simon, Toulouse Kebab expert.

I also spend considerable time at ENAC. Jean-Paul Imbert was always there to help me with any issue or to let me fly in the newest flight simulator. Special thanks go to the entire team of LII, including Sylvain, Stéphane, Dong-Back, Célia, Laetitia, Loïc, Mathieu, Jérémie, Catherine, Nadine, Vincent, Pierre-Loïc, Céline, Cécile and Alexandre who I was happy to see and share a coffee with whenever I made the long trip from ISAE-SUPAERO to ENAC (896 meters, but consider the wind).

Doing a PhD did not stop me from meeting some people that I will hold near and dear for the rest of my life: Paula, the only person to support living with me for more than two years (besides my parents), and despite my decision to wake up at 06:00 a.m. for a period of 6 months; Samy, the fastest man on an electric scooter in all of Toulouse. Laura, the best French teacher for the worst words and Blaer, partner in crime and the greatest Mads Mikkelsen fan from all of Iceland.

I also want to thank my friends who are away from Toulouse and keep in contact with me despite my usual response time to a text message being around five days. This is no ordinary group of people, and that statement has been tested, retested, fact-checked, cross-validated, plotted, analyzed, classified, clustered, peer-reviewed and adjusted for the correct chance level. Saying that this group is extraordinary is no hypothesis; it has a DOI, Citations, an Impact Factor, an Appendix, Supplementary Materials, a link to a GitHub, multiple affiliations, no conflict of interest and is Open Access. To state it bluntly, this is as close to a scientific fact as I will ever be.

So, in no particular order, a list of amazing people who I would like to thank for being there: Thorge, Jules, Moritz, Luca, Julius, Mohit, Anna, Tobi, Emie, Chris, Lukas, Simon, Marc, Rick, Max, Juliette, Raffi, Francesco, Peter, Ami, Eva, Niklas, Vanessa, Elli, Karen, Kev, Jenny, Chris, Shirley, Ole, Tanja, Per, Helge, Susi, Jan, Till, and Amelie.

Finally, to the two persons who have seen it all. My cheerleaders, wingman and wing-woman, emergency contacts, and the two people who always backed me: my parents, Thomas and Tonia. With you on my team, I could never lose.

Abstract

Uncrewed Aerial Systems (UAS) are common in many industries and an important pillar of many modern militaries. While technology advances rapidly, military long-endurance UAS still suffer from high rates of human error-related mishaps, especially when operators experience mental fatigue.

In particular, the frequent switching between tasks required by operators presents an opportunity for improvement. Cognitive flexibility, the mental ability to switch between tasks or responses, is an important executive function. However, little research has investigated cognitive flexibility during complex tasks such as UAS control. In this thesis, we address this lack, and the possibility of mitigating the adverse effects of fatigue on cognitive flexibility was also explored. For this, visual alerts were developed to inform participants when they had to switch between tasks. To increase the effectiveness of visual alerts, their use should be restricted to periods of poor performance (e.g. due to mental fatigue), during which the positive impact of visual alerts may be maximized.

Passive Brain-Computer Interfaces (pBCI) and physiological computing systems using electroencephalography (EEG), electrocardiography (ECG) and eye-tracking may detect sub-optimal states, such as mental fatigue. Detecting mental states poses several theoretical issues. Most notably, a mental state is not necessarily a good performance indicator. Directly estimating performance, on the other hand, has proven to be more challenging for BCIs and Physiological Computing, partially because performance can be defined in many ways.

Following several interviews with operators, five experiments were designed. (i) The first experiment, named the Double Task Switching (DTS) experiment tests a novel behavioural protocol to investigate cognitive flexibility and the effects of similarity on task-switching costs. Results showed that switching between similar and dissimilar tasks significantly differs in accuracy and reaction time. This experiment is followed by (ii) a second behavioural paradigm investigating whether the mere presence of visual alerts can improve performance. It builds on the DTS protocol and is called the DTS-II experiment. The results of this study show that visual alerts without any prior explanation do not appear to improve performance when switching between tasks.

Two further experiments explore the possibility of using EEG, ECG and eye-tracking to detect mental fatigue during realistic UAS simulations. (iii) The Remot3e experiment (Remote search, 3 Es for EEG, ECG & Eye-Tracking) focuses on a search task. It constitutes a primary investigation into the differences between Time-on-Task (TOT) and performance-based mental fatigue estimation. Then, (iv) the UASOS (UAS Operator Simulation) experiment combines the DTS protocol and the Remot3e Task, adding a navigation task to recreate some of the fundamental aspects of UAS operations. Both experiments highlight the difference in using different definitions of mental fatigue for constructing and training physiological computing systems. Notably, Time-On-Task (TOT, the time a participant has been performing a task) and task performance are compared.

Finally, (v) we investigate if visual alerts can be used to adapt an adaptive interface in UAS simulations to improve performance. For this, Visual Alerts were added to the UASOS experiment. The preliminary results highlight the impact of a simple adaptation on behaviour within a complex work environment.

The work in this manuscript shows that cognitive flexibility should be considered a priority in human factor research and that simple adaptations, such as visual alerts, can improve cognitive flexibility. It also highlights the importance of mental fatigue and how different definitions

of the construct (based on performance or TOT) can result in major performance differences when researching pBCIs and Physiological Computing systems.

Résumé

L'utilisation de drones est fréquente dans de nombreuses industries et dans les activités militaires modernes. Toutefois les drones militaires à longue endurance souffrent encore d'un taux élevé d'accidents liés à des erreurs humaines, en particulier lorsque les opérateurs ressentent de la fatigue mentale (FM). En particulier, la flexibilité cognitive exigée des opérateurs offre une possibilité d'amélioration. Cependant, peu de recherches ont été menées sur celle-ci lors de tâches complexes tel que le contrôle d'un drone. Dans cette thèse, nous remédions à cette lacune, et la possibilité d'atténuer les effets négatifs de la FM sur la flexibilité cognitive a également été explorée. Pour ce faire, des alertes visuelles ont été développées pour informer les participants lorsqu'ils devaient passer d'une tâche à l'autre. Pour accroître l'efficacité des alertes visuelles, leur utilisation devrait être limitée aux périodes de faible performance (par exemple, en raison de la FM), au cours desquelles l'impact positif des alertes visuelles peut être maximisé. Les interfaces cerveau-ordinateur passives (pBCI) et les systèmes informatiques physiologiques utilisant l'électroencéphalographie (EEG), l'électrocardiographie (ECG) et l'oculométrie peuvent détecter des états sub-optimaux comme la FM. Dans ce cas, les alertes visuelles sont susceptibles d'avoir un impact positif plus important. La détection d' états mentaux pose plusieurs problèmes : Un état mental n'est pas nécessairement un bon indicateur de performance. D'autre part, l'estimation peut s'avérer difficile, notamment parce que la performance peut être définie de différentes manières. À la suite de plusieurs entretiens avec des opérateurs, cinq expériences ont été conçues. (i) La première expérience, appelée expérience Double Task Switching (DTS), teste un nouveau protocole comportemental pour étudier la flexibilité cognitive. Les résultats ont montré que le passage d'une tâche similaire à une autre impacte significativement la performance. (ii) Cette expérience est suivie d'un second paradigme visant à déterminer si la simple présence d'alertes visuelles peut améliorer les performances. Les résultats de cette étude montrent que les alertes visuelles sans explication préalable ne semblent pas améliorer les performances. Deux autres expériences explorent la possibilité d'utiliser l'EEG, l'ECG et le suivi des yeux pour détecter la FM lors de simulations réalistes de drone. (iii) La première expérience se concentre sur une tâche de recherche et constitue une première étude des différences entre l'estimation de la FM basée sur le temps passé sur la tâche (TOT) et celle basée sur la performance. (iv) L'expérience suivante combine ensuite les deux expériences précédentes en ajoutant une tâche de navigation pour recréer certains aspects de l'opération de drones. Les résultats mettent en évidence la différence dans l'utilisation de définitions différentes de la FM pour le design de systèmes informatiques physiologiques. Notamment, le TOT et la performance de la tâche sont comparés. (v) Enfin, nous étudions si les alertes visuelles peuvent être utilisées pour adapter une interface adaptative dans les simulations de drones afin d'améliorer les performances. Pour ce faire, des alertes visuelles ont été ajoutées à l'expérience précédente. Les résultats préliminaires mettent en évidence l'impact d'une simple adaptation basée sur le comportement dans un environnement de travail complexe. Les travaux présentés dans ce manuscrit montrent que la flexibilité cognitive devrait être considérée comme une priorité dans la recherche sur les facteurs humains et que des adaptations simples peuvent améliorer la performance. Ils soulignent également l'importance de la FM et la manière dont différentes définitions de ce concept peuvent entraîner des différences de performance majeures pour les pBCI et les systèmes d'informatique physiologique.

Abbreviations

ACC	Accuracy
ATC	Air Traffic Controller
AUI	Adaptive User Interface
BCI	Brain-Computer Interface
BOLD	Blood oxygen level-dependent
BLOS	Beyond Line of Sight
BPM	Beats Per Minute
ECG	Electrocardiogram
EEG	Electroencephalography
EOg	Electrooculography
ERP	Event-Related Potential
ET	Eye-Tracking
FAA	Federal Aviation Administration
fMRI	Functional Magnetic Resonance Imaging
GCS	Ground Control Station
HCI	Human-Computer Interaction
HMI	Human-Machine Interaction
HR	Heart Rate
HRV	Heart Rate Variability
IC	Intelligence Coordinator
ICAO	International Civil Aviation Organization
ISI	Inter-Stimulus Interval
KSS	Karolinska Sleepiness Scale
LOA	Level of Automation
LOS	Line of Sight
MF	Mental fatigue
MART	Malleable Attention Resource Model
NIRS	Near Infra-Red Spectroscopy
PCS	Physiological Computing System
PIC	Pilot in Command
pBCI	passive Brain-Computer Interface
QSN	St Mary Hospital Sleep Questionnaire
RMSSD	Root Mean Square of Successive Differences
RSME	Rating Scale Mental Effort
RT	Reaction time
SA	Situational Awareness
SDNN	Standard Deviation of NN intervals
SNR	Signal-to-Noise Ratio
SO	Sensor Operator
SPS	Samn-Perelli Scale
TOD	Time of Day
TOT	Time on Task
TSC	task-switching cost
UAS	Uncrewed Aerial System
UAS OPS	Uncrewed Aerial System Operations
UAV	Uncrewed Aerial Vehicle
USAF	United States Air Force
VAS	Visual Analogue Scale

Tab. 0.1 Abbreviations used in the Manuscript.

Contents

Contents 7

1 Introduction 13

I State of the Art 17

2 Machines 19
I.2.1 UAS operations . . . 19
I.2.1.1 Hardware . . . 20
I.2.1.2 Operators . . . 23
I.2.1.3 Missions . . . 23
I.2.1.4 Human Factor Issues . . . 25
I.2.1.5 Interviews . . . 27
I.2.2 Adaptive Interfaces . . . 29
I.2.2.1 Cockpits: How interfaces have developed over time . . . 29
I.2.2.2 Human-Machine Interaction and Adaptive Interfaces . . . 31
I.2.2.3 Why an Adaptive User Interface rather than an improved User Interface? . . . 34
I.2.2.4 Adaptive Interfaces for long endurance uncrewed aerial system operations . . . 37
I.2.2.5 When to adapt . . . 44
I.2.2.6 Design Motivation & Considerations . . . 45

3 Humans 49
I.3.1 Mental fatigue . . . 49
I.3.1.1 Definitions of Mental fatigue . . . 50
I.3.1.2 Theories of Mental fatigue . . . 52
I.3.1.3 Within and between subject designs for mental fatigue assessment 55
I.3.1.4 Effects of mental fatigue . . . 56
I.3.1.5 Modelling and labeling mental fatigue . . . 61
I.3.2 Cognitive flexibility . . . 65
I.3.2.1 Theories of cognitive flexibility . . . 65
I.3.2.2 Experimental paradigms of task-switching . . . 66
I.3.2.3 Similarity, Bivalency, Inter-stimulus Intervals & Task-Rule Congruency 67
I.3.2.4 Cognitive flexibility in complex operations . . . 70

4 Sensors 71
I.4.1 EEG . . . 72
I.4.1.1 Mechanisms . . . 72
I.4.1.2 The EEG signal . . . 76
I.4.1.3 Preprocessing . . . 81
I.4.1.4 Analysis . . . 83
I.4.2 ECG . . . 86
I.4.2.1 Mechanisms . . . 86
I.4.2.2 Analysis & Processing . . . 87
I.4.3 Eye Tracking . . . 90
I.4.3.1 Mechanisms . . . 90
I.4.3.2 How to measure movements . . . 91
I.4.3.3 Types of eye movements & other metrics . . . 92

5 Algorithms 95
I.5.1 Machine Learning . . . 95
I.5.1.1 Creating a supervised machine learning pipeline . . . 97
I.5.1.2 Issues in Machine Learning . . . 103
I.5.1.3 Conclusion . . . 104
I.5.2 Physiological Computing . . . 105
I.5.2.1 Physiological Computing Systems & Brain-Computer Interfaces . . . 105
I.5.2.2 Performance estimation vs Mental fatigue estimation . . . 110
I.5.2.3 Metrics for mental fatigue detection . . . 112
I.5.2.4 The variability issue and transfer learning . . . 116

II Contributions 119

6 Double Task Switching: Cognitive flexibility and mental fatigue 121
II.6.1 Motivation and Research Goals/ Questions . . . 122
II.6.2 Materials and methods . . . 123
II.6.2.1 Ethics Statement . . . 123
II.6.2.2 Design . . . 123
II.6.2.3 Sample size . . . 126
II.6.2.4 Recruitment inclusion criteria and compensation . . . 128
II.6.2.5 Procedure . . . 128
II.6.2.6 Variables . . . 129
II.6.2.7 Questionnaires . . . 130
II.6.2.8 Analysis Plan . . . 130
II.6.3 Results . . . 131
II.6.3.1 Outliers . . . 131
II.6.3.2 Reaction Time . . . 131
II.6.3.3 Accuracy . . . 133
II.6.4 Discussion . . . 135
II.6.5 Conclusion . . . 137

7 Double Task Switching II: Countermeasures for performance improvement 139
II.7.1 Motivation and Research Goals/ Questions . . . 139
II.7.2 Methods . . . 140
II.7.2.1 Ethics Statement . . . 140
II.7.2.2 Design . . . 140
II.7.2.3 Participants . . . 142
II.7.2.4 Recruitment inclusion criteria and compensation . . . 142
II.7.2.5 Procedure . . . 143
II.7.2.6 Variables . . . 143
II.7.2.7 Questionnaires . . . 144
II.7.2.8 Analysis Plan . . . 145
II.7.3 Results . . . 145
II.7.3.1 Reaction Time . . . 145
II.7.3.2 ACC . . . 147
II.7.3.3 Subjective Analysis . . . 148
II.7.4 Discussion . . . 150

8 Remot3e: Research into mental fatigue classification and labelling 151
II.8.1 Motivation and Research Goals/ Questions . . . 152
II.8.2 Material and methods . . . 152
II.8.2.1 Participants and ethical approval . . . 152
II.8.2.2 Procedure . . . 153
II.8.2.3 Materials . . . 154
II.8.2.4 Database quality assessment . . . 154
II.8.2.5 Mental fatigue estimation . . . 156
II.8.3 Results . . . 159
II.8.3.1 Subjective measures . . . 159
II.8.3.2 Behavioral measures . . . 160
II.8.3.3 Physiological measures . . . 160
II.8.3.4 Mental fatigue estimation . . . 162
II.8.4 Discussion . . . 165
II.8.4.1 Limitations . . . 167

9 UASOS: Moving towards ecological validity 169
II.9.1 Motivation and Research Goals/ Questions . . . 170
II.9.2 Task Development . . . 170
II.9.2.1 Overview . . . 170
II.9.2.2 Characteristics . . . 171
II.9.3 Methods . . . 173
II.9.3.1 Participants . . . 173
II.9.3.2 Procedure . . . 173
II.9.3.3 Materials . . . 173
II.9.3.4 Performance and Mental State classification . . . 174
II.9.3.5 Statistical Analysis . . . 175
II.9.4 Results . . . 176
II.9.4.1 Subjective Data . . . 176
II.9.4.2 Behavioral Data . . . 176

II.9.4.3 EEG . . . 178
II.9.4.4 ECG . . . 180
II.9.4.5 Classification . . . 180
II.9.5 Discussion . . . 180
II.9.6 Conclusion . . . 182

10 UASOS-II: Real-Time implementation of solutions 183
II.10.1 Motivation and Research Goals/ Questions . . . 183
II.10.2 Methods . . . 184
II.10.2.1 Participants . . . 184
II.10.2.2 Procedure Changes . . . 184
II.10.2.3 Analysis . . . 186
II.10.3 Results . . . 187
II.10.3.1 Behavioral . . . 187
II.10.3.2 ECG . . . 194
II.10.3.3 EEG . . . 194
II.10.4 Discussion . . . 196

III Discussion & Afterword 199

11 Discussion 201
III.11.1 RO1) UAS Operations & Interfaces . . . 202
III.11.1.1 RO1 a) Why do Uncrewed Aerial Systems (UAS) suffer from a high rate of human factor-related mishaps? . . . 202
III.11.1.2 RO1 b) Which modifications in interface design can improve the safety of UAS operations? . . . 203
III.11.1.3 Open questions, limitations & future research . . . 204
III.11.2 RO2) Mental fatigue . . . 205
III.11.2.1 RO2 a) What is a useful definition of mental fatigue within complex operations? . . . 205
III.11.2.2 RO2 b) How do markers of mental fatigue (physiological, behavioural & subjective) perform at predicting mental fatigue depending on different definitions of the construct? . . . 206
III.11.2.3 Open questions, limitations & future research . . . 208
III.11.3 RO3) Cognitive flexibility . . . 210
III.11.3.1 RO3 a) What do we need to understand about cognitive flexibility to move from fundamental tasks to more realistic environments? . 210
III.11.3.2 RO3 b) Are there ways to reduce the effects of task-switching? . . . 211
III.11.3.3 Open questions, limitations & future research . . . 212
III.11.4 Conclusion . . . 213

12 Afterword 215
III.12.1 Some thoughts on the ethics of military UAS operations . . . 215

IV References 221

V Scientific Output & Appendices 265

13 Scientific Output 267

V.13.1 Journal Articles . . . 267
V.13.2 Conference Articles . . . 267
V.13.3 Conference Abstracts . . . 268
V.13.4 Internship . . . 268

14 Appendices 272

V.14.1 Picture Sources . . . 272
V.14.2 Questionnaires . . . 273
V.14.2.1 Demographics . . . 273
V.14.2.2 Edinburgh Handedness Questionnaire . . . 274
V.14.2.3 Ishihara Colorblindness Test . . . 275
V.14.2.4 Karolinska Sleepiness Scale . . . 276
V.14.2.5 Samn-Perelli Scale . . . 277
V.14.2.6 Visual Analogue Scale . . . 278
V.14.2.7 Rating Scale Mental Effort . . . 279
V.14.2.8 NASA Task Load Index . . . 280
V.14.2.9 The St Mary's Hospital Sleep Questionnaire . . . 280
V.14.3 Remot3e - Saccade detection . . . 281
V.14.4 Doulbe Task Switching - Exploratory Analysis . . . 281
V.14.4.1 Methods . . . 281
V.14.4.2 Results . . . 282
V.14.4.3 Discussion . . . 285
V.14.5 UASOS - Open Source Code . . . 287
V.14.5.1 Characteristics . . . 287
V.14.5.2 Code/Software . . . 288
V.14.5.3 Usage Notes . . . 288
V.14.6 UASOS - Transfer Learning . . . 289
V.14.6.1 Motivation . . . 289
V.14.6.2 Methods . . . 289
V.14.6.3 Results . . . 290
V.14.7 COG-BCI - An Open EEG Database for passive BCIs . . . 291
V.14.7.1 Introduction . . . 291
V.14.7.2 Methods . . . 292
V.14.7.3 Validation results . . . 304
V.14.7.4 Conclusion of the dataset validation . . . 317

Chapter 1

Introduction

In 1935, the British Military used a modified and remote-controlled DH82 "Queen Bee" airplane to train anti-aircraft gunners. This is not only one of the first uses of an uncrewed aerial vehicle but also the origin of the name drone for uncrewed aerial vehicles, as a reference to male bees[1].

Today, Uncrewed Aerial Systems (UAS) are common in many industries and an important pillar of many modern militaries (Austin, 2011). While technology advances rapidly, military long-endurance UAS still suffer from high rates of human error-related mishaps (Haider, 2021). As these UAS can remain airborne for periods exceeding 40 hours, several teams of operators are required for a single mission, often controlling the UAV for up to 8 hours. Long, intense missions with little to no breaks result in operators being mentally fatigued, thus reducing their ability to direct attention, remain flexible, and maintain situational awareness (Hinss, Brock, and Roy, 2022a, see Section I.2.1.4).

The interface design of the Ground Control Stations (GCS) from which the operators pilot and control the UAS likely adds to the mishap risk (see Section I.2.1.4). In particular, the frequent switching between tasks required by operators was identified as an opportunity for improvement.

Cognitive flexibility, the mental ability to switch between tasks or responses, is an important executive function (Koch et al., 2018). Cognitive flexibility has mostly been investigated using very controlled and fundamental task-switching protocols or measures of cognitive flexibility to correlate with performance during a complex task (Koch et al., 2018, see Section I.3.2). To the best of our knowledge, very little research has tried recreating fundamental task-switching protocols in more ecological contexts. One of the research goals of this thesis was to investigate if fundamental aspects of task-switching apply to more ecological settings, particularly UAS operations (see Chapters 6 & 9).

[1] https://www.iwm.org.uk/history/a-brief-history-of-drones

Going beyond understanding cognitive flexibility in ecological settings, we also wanted to examine if there are ways to mitigate the negative impact of reduced cognitive flexibility. Giving participants more time to switch or prepare to switch to another task has improved performance in prior studies (Kiesel et al., 2010, see Section I.3.2.3). We thus hypothesised that visual alerts may similarly affect performance improvement (see Chapters 7 & 10).

However, in a realistic environment, visual alerts and increased preparation time come at the cost of longer reaction times. Even if preparation results in quicker responses, this is likely less than the added preparation time. So, any performance improvement must be defined by accuracy or other non-reaction time-related performance metrics.

As participants may, when vigilant, perform their task quickly and accurately using visual alerts could result in an overall negative impact on performance when looking at both reaction time and accuracy (see Section I.2.2.3).

To mitigate this, their use should be restricted to periods of poor performance, during which the positive impact of visual alerts may be maximized. Unfortunately, measuring performance in real time is difficult in the real world, as the ground truth is unknown.

Passive Brain-Computer Interfaces (pBCI) and Physiological Computing systems may solve this issue. These systems measure cerebral (pBCI) or peripheral (Physiological Computing) activity to detect mental states using EEG, ECG, Eye-Tracking, and other sensors (Fairclough, 2017; Lotte and Roy, 2019, see Section I.5.2). These systems often use supervised Machine Learning algorithms (see Section I.5.1) to detect mental states accurately, objectively, and automatically.

If these systems can detect suboptimal states, such as mental fatigue, in which the operator is likely to profit most from the visual alerts, then these systems can be used to trigger said alerts.

Detecting mental states poses several theoretical issues. Most notably, a mental state is not necessarily a good performance indicator. Directly estimating performance has proven to be more challenging for BCIs and Physiological Computing, partially because performance can be defined in many ways (see Section I.5.2.2). In two experiments, we investigate how we can predict performance in realistic tasks and combine this with visual alerts to create an adaptive interface that supports UAS operators during their long-endurance UAS missions.

The structure of the manuscript is as follows: The state-of-the-art part is divided into three main sections. The chapter Machines focuses on UAS operations (see Section I.2.1) and Adaptive Interfaces (see Section I.2.2). In the chapter Humans, the concepts of mental fatigue (see Section I.3.1) and cognitive flexibility (see Sections I.3.2) are explored. Finally, the chapter Algorithms provides an introduction to machine learning (see Section I.5.1) and physiological computing (see Section I.5.2).

The Contributions part encompasses five experiments, each described in a separate chapter (see Figure 1.1 for a visual representation of the contributions and their relation to the topics of this manuscript). (i) The first experiment, named the Double Task Switching Experiment (DTS, see Chapter 6), tests a novel behavioural protocol to investigate cognitive flexibility and the effects of similarity on task-switching costs (see Chapter 6). (ii) This experiment is followed by a second behavioural paradigm investigating whether the mere presence of visual alerts can improve performance. It builds on the DTS protocol and is called the DTS-II (see Chapter 7 experiment).

Two further experiments explore the possibility of using EEG, ECG and Eye-Tracking to detect mental fatigue during realistic UAS simulations (see Chapters 8 & 9). (iii) The first experiment, named Remot3e (Remote search, 3 Es for EEG, ECG & Eye-Tracking), focuses on

a search task and constitutes a primary investigation into the differences between TOT and performance-based mental fatigue estimation(see Chapter 8). (iv) The UASOS (UAS Operator Simulation) experiment then combines both the DTS protocol and the Remot3e Task, adding a navigation task to recreate some of the fundamental aspects of UAS operations(see Chapter 9).

(v) Finally, we investigate if visual alerts can be used as an adaptive interface in UAS simulations to improve performance (see Chapter 10). For this, the Visual Alerts were added to the UASOS experiment.

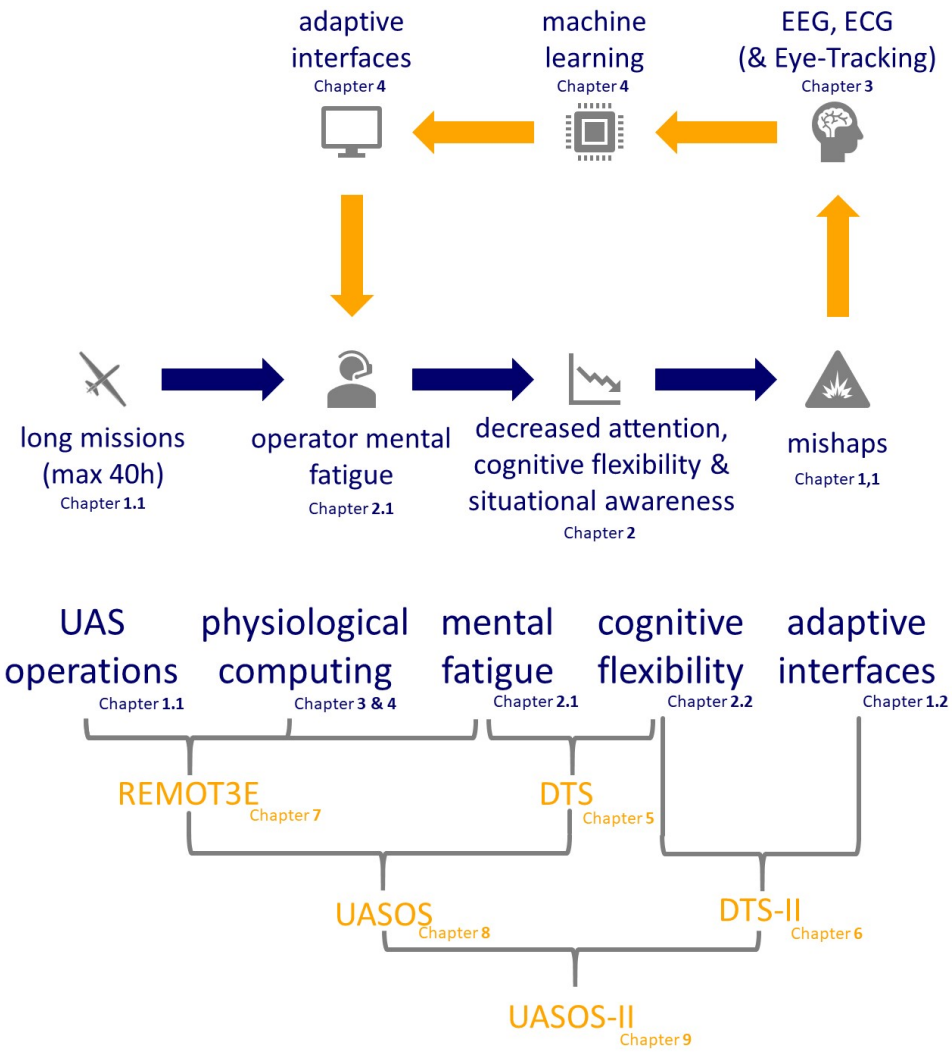

Fig. 1.1 Top: Visual representation of the work presented in this manuscript; Bottom: Visual Summary of the Contributions to the central topics of the manuscript.

Research Objectives

In line with the aforementioned issues related to UAS operation and mental fatigue detection, the following research objectives (RO) were drawn:

- **RO1)** UAS operations & interfaces
  - **a)** Why do Uncrewed Aerial Systems (UAS) suffer from a high rate of human factor-related mishaps?
  - **b)** Which modifications in the digital navigation and visual search interface design can improve the performance of UAS operations?
- **RO2)** Mental fatigue
  - **a)** What is a useful definition of mental fatigue within complex operations?
  - **b)** How do markers of mental fatigue (physiological, behavioural & subjective) perform at predicting mental fatigue depending on different definitions of the construct?
- **RO3)** Cognitive flexibility
  - **a)** What do we need to understand about cognitive flexibility to move from fundamental tasks to more realistic environments?
  - **b)** Are there ways to reduce the negative effects of task-switching?

Part I
State of the Art

Chapter 2

Machines

I.2.1 UAS operations

Central Ideas

- Uncrewed Aerial Systems (UAS) are steadily gaining popularity for military and civilian use.
- Their capacities (endurance, complexity) are increasing faster than the domain of human factors can catch up, resulting in significantly higher mishap rates as compared to crewed operations.
- UAS are not independent of humans in their operation and in the catastrophic outcomes when errors occur.

One question to ask: **How can we make UAS safer, without sacrificing the advantages these systems offer?**

Please note that many aspects of UAS operations (UAS OPS) are classified, making it difficult to find reliable sources and verify all the information. In addition, as UAS in military and civilian operations become exponentially more important, it is difficult to ascertain that certain information is still current. With that being said, the goal of this section is not to give a detailed analysis of UAS OPS but rather a general idea of what UAS OPS involve.

The following subsections will detail the different types of UAS that exist (see Section I.2.1.1), then detailing the MQ-9 Reaper (I.2.1.1) and its Ground Control Station (see Section I.2.1.1). Finally we will provide some oversight into UAS OPS (see Section I.2.1.3) and the Human Factor issues (see Section I.2.1.4) with UAS OPS.

I.2.1.1 Hardware

Terms such as Uncrewed Aerial Vehicle, Uncrewed Aircraft, Uncrewed Aerial System, Drone, and Remotely Piloted Aircraft are used interchangeably [1]. For the remainder of the manuscript, we will follow the following convention: The word drone is reserved for civilian use uncrewed aircraft that operate within *Line-of-Sight (LOS)* (Austin, 2011; Haider, 2021). *UAV* are defined as

> *"Vehicles that use aerostatic or aerodynamic lift, and overall don't generally fly on a ballistic trajectory can be categorized as an aircraft. These vehicles can be propelled by a motor (e.g. rotary or jet) to create lift and sustain flight. If these aircraft do not house a pilot within the airframe and are operated remotely using various levels of automated functions, they are considered an UA, excluding cruise missiles."* - Haider (2021)

Finally, *UAS* comprise not only the actual aircraft but also the Remote control station, payload, launch & recovery element and data link.

Each component of a UAS plays an important role. The *Remote Control Station* also referred to as a *Ground Control Station (GCS)*, *Ship Control Station* or *Airborne Control Station* (depending on location), is the human-machine interface that allows operators to interact with the UAV. For some UAS, there are two *RCSs*, one for launch and recovery (take-off and landing) and one for the mission element. The *payload* includes sensors such as video and infrared cameras, normal and synthetic aperture radars, or, in the case of Strike UAS, weapons. Some UAV, such as the Scan Eagle (see Table 2.1), cannot take off and land themselves and require catapults, arresting hooks or similar systems. Finally, a UAS needs a data link to communicate between the UAV and the RCS. This may be done in LOS cases with radio antennas, but *beyond-line-of-sight (BLOS)* operations require a satellite link (Austin, 2011; Haider, 2021).

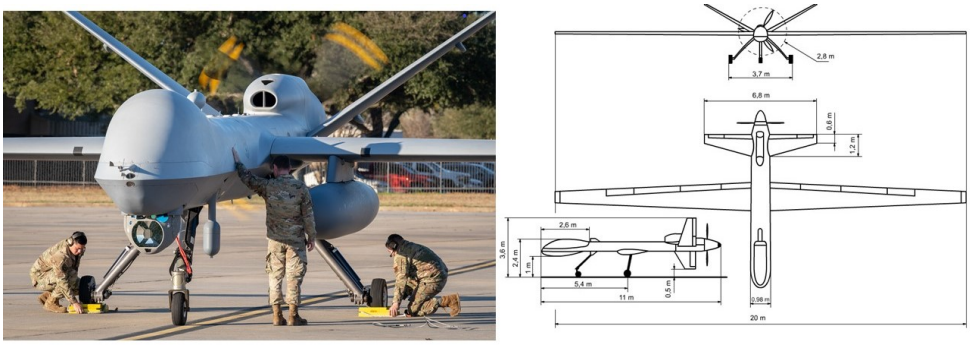

Fig. 2.1 MQ-9 Reaper from the US-Airforce, with its dimensions on the left side (Picture Source: see Appendix V.14.1).

No definitive taxonomy of UASs exists, and the development of many different systems is blurring the lines. NATO divides UAS into three classes, based mostly on maximum take-off weight. Table 2.1 provides an overview of the NATO classification (Haider, 2021). Most examples are *fixed-wing aircraft*, but *Rotary wing* UAVs exist in classes II and I. UAS can operate on different *levels of Automation (LOA)*, with several different taxonomies for assigning levels to a specific UAS (Vagia et al., 2016). We will follow the division proposed by Radovic (2019), focusing on UAS specifically. Here, six levels, 0-5, are proposed (See Figure 2.2).

[1] For the sake of inclusivity the general naming convention is changing from Unmanned to Uncrewed (Haider, 2022)



RQ-4 Global Hawk Class III - HALE | **MQ-9 Reaper** Class III - MALE | **WK450 Watchkeeper** Class II - Tactical | **ScanEagle** Class I - Small | **RQ-11 Raven** Class I - Mini | **Black Hornet** Class I - Micro

Class	Category	Altitude	Operating Radius	Example Platform
III >600 kg	Strike/Combat	max FL65	Unlimited (BLOS)	MQ-9 Reaper*
	HALE	max FL650**	Unlimited (BLOS)	RQ-4 Global Hawk
	MALE	max FL450**	Unlimited (BLOS)	MQ-9 Reaper
II 150 - 600 kg	Tactical	max FL180**	200km (LOS)	Watchkeeper
I <150 kg	Small (>15 kg)	max 5000 ft	50km (LOS)	Scan Eagle
	Mini (<15 kg)	max 3000 ft	25km (LOS)	Skylark
	Micro (<66 J***)	max 200 ft	5 km (LOS)	Black Hornet

Tab. 2.1 NATO classification of UV classes. *Only if the UV is armed; ** FL refers to Flight Level, steps of 100 feet above mean sea level, not used in aviation below 5000 feet; *** Aircraft with less than 66 Joules of energy, as these are unlikely to cause significant harm to humans

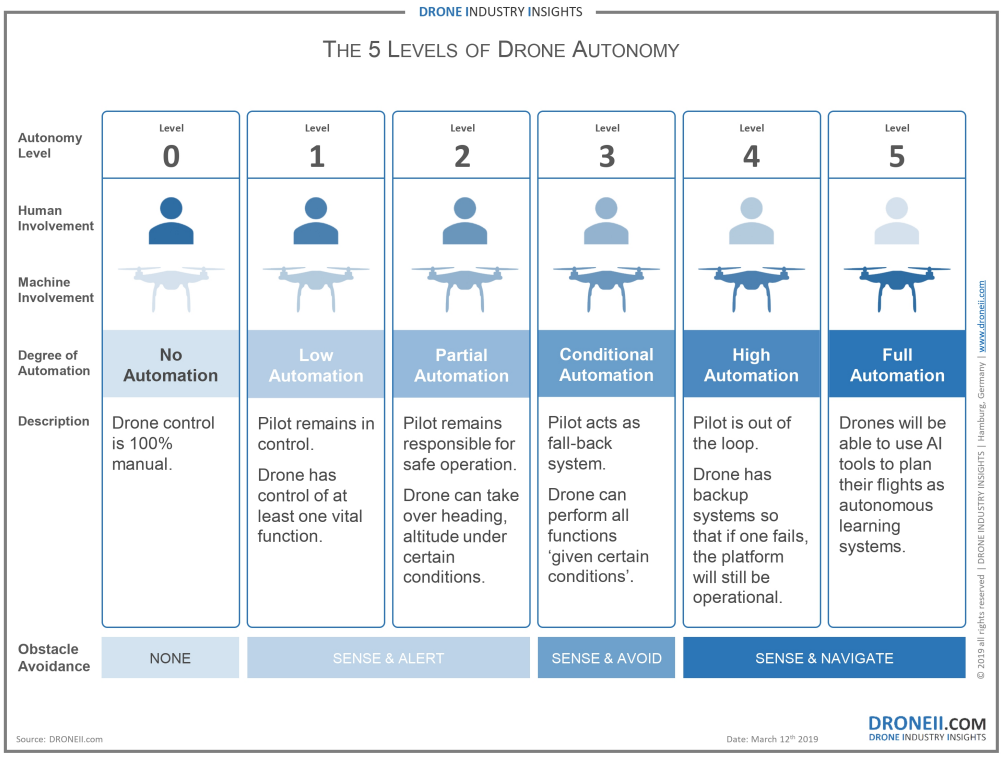

Fig. 2.2 The different levels of Automation, taken from Radovic (2019) (Picture Source: see Appendix V.14.1).

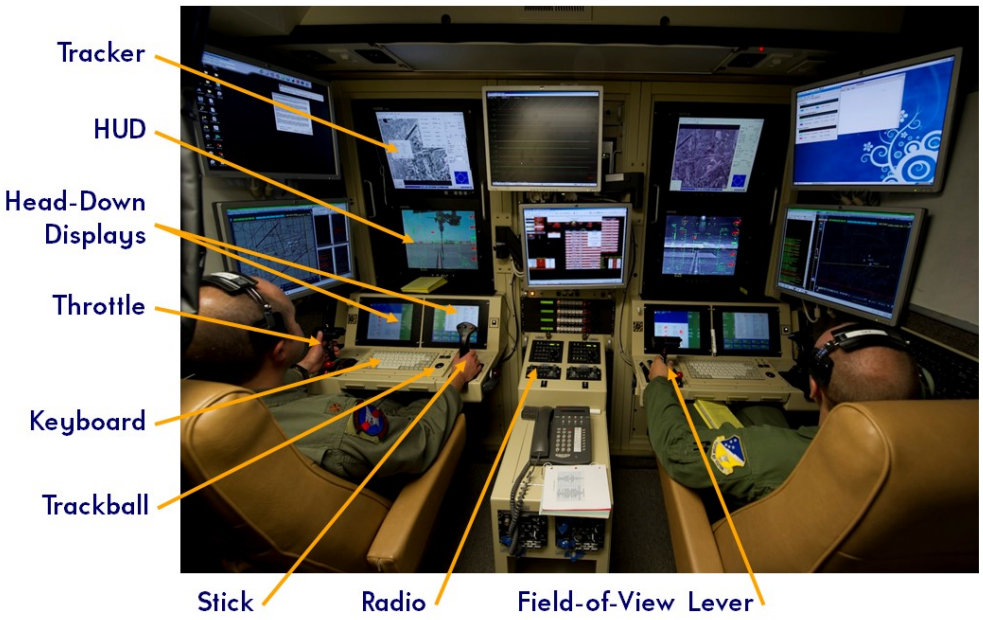

Fig. 2.3 MQ-9 Reaper Block 30 Ground Control Station (Picture Source: see Section V.14.1)

The MQ-9 Reaper

The MQ-9 Reaper (see Figure 2.1) was developed from its predecessor, the MQ-1 Predator (Chappelle et al., 2010). It is larger, faster and has more endurance, making it an asset to the militaries of the US, France, the UK, India, Italy, Japan, the Netherlands, Poland and Spain (Yerushalmy, 2023). It has a 20m wingspan and a maximum take-off weight of 4763 kg. A *turboprop engine* powers the reaper, which can fly for up to 36 Hours (Information differs between sources), with a maximum speed of 240 knots and a service ceiling of 50.000 feet. The equipment of the UAV can differ between operators and mission requirements. Options include radar, infrared, synthetic aperture radar, anti-submarine warfare equipment, signal intelligence sensors and laser designation targeting pods. MQ-9 can carry various missiles, such as air-to-ground and air-to-air, as well as guided bombs (Beno & Adamcik, 2014).

The MQ-9 Ground Control Station

The information on the MQ-9 Reaper ground control station is mostly classified (Cullen, 2011). Currently, two main types of the GCS exist: the more widespread block 30 (see Figure 2.3)[2] and the newer block 50[3]. The current block 30 GCS includes five main screens, rudder/brake pedals, a trackball sidestick, thrust lever, radio, touchscreens and a keyboard for each operator. The future block 50 GCS remains faithful to this but improves the HMI and would allow for single-seat operations (only one operator).

[2] https://www.ga-asi.com/ground-control-stations/block-30
[3] https://www.ga-asi.com/new-block-50-ground-control-station-flies-mq-9-reaper

Army & Air Force	Civillian	Naval
Reconnaissance	Aerial photography	Fleet detection & shadowing
Surveillance	Crop monitoring	Anti-submarine warfare
Search and rescue	Searching for survivors	Fleet protection
Fall of shot plotting	Powerline inspection	
Target designation	Fire surveillance	
Contamination monitoring	Traffic surveillance	
Landmine detection	Meteorology	
Electronic intelligence	Research	
Base security		

Tab. 2.2 Mission statements for UAS OPS (Austin, 2011). Please note that many mission statements overlap between branches but are not repeated in this table (e.g. search and rescue & surveillance). Furthermore, Air Force and Army missions were grouped due to a large overlap.

I.2.1.2 Operators

The composition of teams of operators for UAS changes depending on the vehicle, country, military/civilian branch, and mission[4]. The operators of the MQ-9 Reaper in the USAF typically consist of a Pilot (Pilot in Command, PIC), sensor operator (SO) and intelligence coordinator (IC) (Chappelle et al., 2010). The PIC responsibility is to, plan and prepare the flight, as well as operate the aircraft (Oncu & Yildiz, 2014). The SO employs sensors to acquire, track, and monitor airborne, maritime, and ground objects, enemy combatants, and assets. They also conduct operations following the rules of engagement and assist the pilot through all phases of flight. The IC, on the other hand, is not responsible for the operation of the UAV but assists in coordinating with ground forces, collecting information and assuring that the entire crew understand the mission objectives[5]. An advantage of UAS is their ability to quickly involve specialists while the UAV is in the air, such as a doctor for medical advice while surveying an area after an earthquake (Gabbard, 2023).

I.2.1.3 Missions

UAS operators' work may be broadly categorised as complex operations or complex tasks (Cuevas et al., 2007). This means that the work is difficult and often involves some level of automation, requiring trust in the system. Complex operations often occur in teams where several humans/operators work on one system. Rather than focusing on one aspect, complex operations that often involve several sub-tasks require situational awareness where distributed information can be integrated to form a complete picture of a system's current state (Endsley & Kaber, 1999; Endsley, 1996; Fabiani et al., 1989). Besides UAS OPS, and aviation in general, naval operations, nuclear power plants and robot control/supervision are common examples of complex operations (Chen, Barnes, & Harper-Sciarini, 2010; Cuevas et al., 2007; Lau et al., 2008).

UAS are a versatile asset that can perform many different missions. Table 2.2 highlights some types of missions in naval, ground, air force and civilian environments. Despite the

[4] https://dronewars.net/who-has-armed-drones/

[5] https://www.acc.af.mil/News/Photos/igphoto/2000633596/

substantial differences between types of missions, most of these missions profit from some of the unique capabilities of many UAS systems. UAS OPS are commonly characterized as "dull, dirty & dangerous" (Austin, 2011; Gupta et al., 2013; Oncu & Yildiz, 2014).

Dull: UAS are well-suited for surveillance and reconnaissance missions due to their extended loiter time, allowing them to remain in a certain area for extended times (Beno & Adamcik, 2014; Oncu & Yildiz, 2014). This is only possible because operators can work in shifts to operate a single UAS.

Dirty: The remote aspect also allows UAS to fly into areas with atomic, biological or chemical (ABC, A for atomic) contamination.

Dangerous: UAS can also be used in missions that would normally be highly risky to the pilot or carry a major political risk. Power line inspections, forest fires and extreme weather scenarios pose major risks to the pilot. The political risk of escalation is also lower, as aggressive actions against a UAS have been shown not necessarily to result in retaliation. For an example, please refer to the 2023 Black Sea Drone Incident [6].

A final aspect that makes UAS OPS popular is the comparatively low financial cost of operating UAS (Austin, 2011; Oncu & Yildiz, 2014). Due to the versatility of modern UAS, it is difficult to estimate what a 'typical' mission consists of. Four different phases may still be highlighted. A mission starts following the preflight preparations, such as an outside check, fuelling, and briefing. A team of operators called the Launch-and-Recovery Element controls the UAV and is responsible for taxiing to the runway and taking off. Once the UAV is airborne, a handoff procedure is performed, and the control is given to another team of operators called the mission element. These are the operators that will then perform the mission. Handoff procedures can be performed from one GCS to another GCS, from one team to another team in the same GCS, or from one crew member to another crew member (Arrabito et al., 2010). In some cases, the Launch-and-Recovery and the Mission Element are the same operators. The Mission Element may also be switched several times during a mission if its duration exceeds the operator's duty time. In these cases, operators perform a so-called hand-over procedure to transfer command from one team to another.

One frequent occurrence during missions is the so-called retask. In this case, the UAS has to change its mission from what was planned to a more urgent situation (see Section I.2.1.5).

When returning the UAV to its base, the command is again transferred to the Launch-and-Recovery Element, which then proceeds to land the UAV.

[6] https://www.reuters.com/world/europe/us-says-reaper-drone-crashes-into-black-sea-after-russian-intercept-2023-03-14/

I.2.1.4 Human Factor Issues

Key Points

- UAS OPS suffer significantly more human factors related mishaps compared to manned flight operations.
- Both fatigue and mental fatigue issues due to long operations and 24-hour scheduling are major issues for the operators.
- Poor interface design has resulted in many mishaps.
- The challenging environment with frequent task changes and mission changes may put an excessive burden on the operator's cognitive flexibility.

The following section will examine some human factor issues in UAS OPS. The statistics presented by the articles should be carefully considered for several reasons. Most of the statistics presented here are based on US military operations. However, mishap statistics may differ significantly across countries' military branches and systems. Statistics may be further skewed due to information being classified or organizational structures. For example, only mishaps of a certain severity have to be reported within the US military (Waraich et al., 2013). The US military classifies mishaps into five classes summarized in table 2.3.

UAS OPS are not only limited to military operations, but human factors issues also persist in the civilian sector, for example, for first responders using UAS in search and rescue missions (Hoang, Grøntved, et al., 2023).

Another aspect to consider is that mishaps are not the only possible negative outcome of human error in UAS OPS. Countries worldwide are now operating armed UAS, classified as Strike UAS. The risk for civilian casualties is big, especially in asymmetrical conflicts such as the Afghanistan war (Carlisle, 2022). This risk is not limited to Strike UAS; false intelligence from any sourpoce, including any UAS, can result in unwanted casualties.

The Law of Armed Conflict in the United Nations Charter Article 52 states that casualties as collateral damage should be avoided but are deemed acceptable Moreno-Ocampo (2006). Furthermore, according to international law, there is no obligation to report civilian casualties.

	Cost	Injury	Destruction
Class A	>2.500.000$	Fatality or permanent total disability	Destruction of aircraft
Class B	>600.000$	Permanent partial disability; hospitalization of three or more personnel	
Class C	>60.000$	Any injury or occupational illness that causes loss of one or more days away from work	
Class D	>25.000$	Any injury	
Class E	Mishaps not meeting any classification criteria, but still deemed important to investigate/report		

Tab. 2.3 Mishaps are classified according to the highest class where they meet any criteria. Table redacted: Space vehicles are not deemed necessary to be included (Air Force, 2024).

Recent changes in US military policy may help reduce civilian casualties and better understand the root of these fatal errors (Austin, 2023). This shift in policy came in the wake of the Kabul airstrike in which missiles launched from the US-operated MQ-9 Reapers resulted in the death of 10 civilians, among them an aid worker and seven children[7].

Unfortunately, this is not an isolated incidence. A 2021 study by the RAND think tank showed a 10 % probability of civilian victims under optimal circumstances (good weather, prior intelligence and 3 UAS working simultaneously). Under non-optimal circumstances, the probability increased to about 50 % with conditions of light clouds, no prior intelligence and only one UAS available (Tingstad et al., 2021).

Consequently, when discussing Human Factors in UAS OPS, we must remember that mishaps are only one side of the coin.

Mishap rates within the US military for human-operated aircraft average around 1 per 100.000 flight hours. For UAS, the rate is approximately 50 for every 100.000 hours (Waraich et al., 2013). Due to several fallacies, the human element is more easily overlooked in UAS OPS. There is the fallacy that uncrewed means that UAS operate without humans, that the human has been completely automated and that the only thing a human might have to do is to direct the UAS somewhere (Cooke, 2006). In fact, many mishaps are attributed to human factors (Giese et al., 2013). A 2004 study of 56 US military UAS mishaps concluded that 11% were caused only by human factors, and another 22% were partially attributed to human factors. However, this study could not attribute any cause to 30% (undetermined or awaiting determination at the time of the report) of all mishaps, likely increasing the numbers even more (Manning et al., 2004). A 2008 study found even more severe results and attributed 60.2 % of the 221 examined mishaps to human factors. In the USAF in particular, this percentage almost reached 80% (Tvaryanas et al., 2008).

Attributing a cause to these accidents is difficult. The HFACS model, used by several aviation administrations and others, identifies the human causes of an accident (Shappell & Wiegmann, 2000). The model considers mental fatigue a factor in accidents, but proving a causal relationship between an accident and mental fatigue is difficult (Durham, 2020). Still, mental fatigue cannot be dismissed in UAS OPS. A 2017 study on USAF MQ-9 Pilots and SO showed that many operators suffer from fatigue, anxiety and difficulties in shifting and focusing. During missions, operators reported constant fatigue and attributed it to mission duration rather than workload or mission type. For the operators, this resulted in slower reaction times and a lower ability to multitask and to transition between tasks (Scheiman et al., 2017).

In a review of studies investigating occupational risks of remote piloting operations, the authors found that UAS crews scored significantly higher on measures of fatigue than an external reference group of other military pilots. While 6.8 % of the Airborne warning and control system aircrew met the criteria for shift work sleep disorder (DSM IV), the percentage for predator (MQ1) crew members was as high as 53.6 % (Armour & Ross, 2017; Thompson et al., 2006). Fatigue in UAS crews seemed robust against modifications of the work shift schedule, supporting the previous point of mission time as a large contributor to fatigue in UAS crews. An experimental study of UAS simulations also showed that longer missions resulted in performance decrement and task disengagement (Wohleber et al., 2019). Fatigue issues are not limited to US operators. A study on French UAS operators showed that fatigue impacts all operators (Caid et al., 2016).

[7] https://www.defense.gov/News/News-Stories/Article/Article/2780257/dod-august-29-strike-in-kabul-tragic-mistake-kills-10-civilians/

Task-switching and cognitive flexibility present a further risk for UAS OPS. As reported in Scheiman et al. (2017), operators struggle to transition between tasks or multitask, especially when fatigued. Studies by (Squire & Parasuraman, 2010; Squire et al., 2006) further highlight this issue. For a more detailed examination of cognitive flexibility in UAS and complex operations, please refer to Section I.3.2.4.

A particular situation that has resulted in several mishaps is the handoff procedure when control of the UAS is shifted from one operator to another (Arrabito et al., 2010; Williams, 2006). In a striking example, the procedures were not followed precisely, resulting in the engine and stabilizer shut off and the UAV entering an uncommanded dive and crashing (Class A mishap, Williams, 2004).

UAS OPS also suffer from the sometimes poor design of the GCS. In 2006, the US military lost a 1.5 million $ UAV because the pilot mistook the ignition switch for the landing gear button. The buttons were closely located to each other, and the pilot shut off the engine instead of lowering the landing gear during an approach. On two separate occasions, ill-illuminated displays have resulted in class A mishaps. In one case, the pilots could not read the airspeed correctly, resulting in a fast approach and hard landing. In the other case, the pilot misread the display and, thinking the UAV had landed, turned off the engine, only to realize that the UAV was still airborne (Waraich et al., 2013).

The HMI is particularly important for UAS OPS; the previous examples highlight why. While not impossible, these accidents would have been less likely in manned aircraft. It is difficult for a pilot to mistake having landed and still being airborne. Similarly, the deceleration following an engine shutoff may have prompted a response by the pilot. (However, similar situations have occurred in manned flights as well Ghimire, 2023). Special design considerations must account for the lack of sensory input to the operation (Austin, 2011).

In the future, these issues are likely to become even more central as technology moves towards multi-drone and drone swarm operations (Hoang, van Berkel, et al., 2023).

I.2.1.5 Interviews

In order to deepen our understanding of the current situation in French military and better understand UAV operators' needs, we interviewed. We interviewed six military Uncrewed Aerial System (UAS) pilots to understand their needs better and gathered their impressions of their current interface (note: personal information about the pilots is classified). These pilots were current and former operators of the MQ-9 Reaper system (see Section I.2.1.1) in the French Air Force (Armée de l'air et de l'espace). We wanted to know how their cognitive functioning is affected by time-on-task and mental fatigue[8]. The current interface (see Section I.2.1.1) for operators allows easy screen layout customisation using a tactile device. Still, they have to reset the configuration when their shift ends. One operator specified that failure messages were not visible enough.

> *"For example, the engine parameters are displayed in a particular place, always in red. There should be a screen dedicated to engine parameters with real gauges, alarms in red when it exceeds. Having numbers in a row, like in Reaper, is not user-friendly at all."* - Quote from a pilot.

Pilots operating long-endurance UASs often experience fatigue due to shift and night work and cognitive fatigue due to sudden changes in workload. A prime example is a frequently

[8]Due to ongoing operations and operators training, we only had limited opportunities to engage with operators

occurring situation referred to as *retask*. During a retask, the mission that was initially planned is changed in real time. Maintaining and regaining situational awareness following such an event is particularly challenging.

For the operators, fatigue manifests itself in different ways. Sometimes, operators stare at elements on the screen without paying attention to them (mind wandering). Nervous fatigue following an intense phase of heightened workload generates a slackness, which does not translate into a desire to sleep but a decreased capacity. Pilots tend to be less attentive, slower, or forget certain details. With time-on-task, capacities progressively decrease compared to their initial state, notably due to boredom during monotonous tasks. To overcome this, some operators suggested that a countermeasure could be applied when they focus too much on one screen element.

I.2.2 Adaptive Interfaces

Central Ideas

- Adaptive User Interfaces (AUI) consist of three components and use information on the user to improve usability by self-adaptation.
  - Afferential Component collecting user information
  - Efferential Component handling the adaptations
  - Inferrential Component using the Afferential Component to guide the adaptation
- AUIs can be a superior aid to operators than static interfaces, even if the static interfaces are "better designed".
  - They may allow for an optimal workload for the operator.
  - AUIs may allow for a better global performance even if this comes at the cost of a performance decrease on a single.
- Although mitigating mental fatigue is difficult, AUIs can help manage the negative effects of mental fatigue rather than reducing mental fatigue directly

One question to ask: **What conditions need to be satisfied to build an AUI that can support operators suffering from mental fatigue?**

In this section we will discuss the concept of adaptive user interfaces (AUI) with specific focus on how they may facilitate complex operations. Section I.2.2.1 provides a historical account of the importance of Human-Machine Interaction in ever more complex systems, by looking at the evolution of aircraft cockpits. Section I.2.2.2 provides a broad overview of the relevant concepts, and in Section I.2.2.3 some motivations for using AUIs are detailed. Following this there is an exploration of adaptive interfaces for long endurance UAS in Section I.2.2.4, before asking when an AUI should adapt (Section I.2.2.5) and finally discussing some design considerations in Section I.2.2.6.

I.2.2.1 Cockpits: How interfaces have developed over time

Interfaces and Aviation have been connected since the early days of flight. The importance of pilots getting the necessary information quickly and accurately and being able to respond to information in time has resulted in some astonishingly thought-through designs.

Looking at the evolution of cockpits easily shows how technological developments have changed the role of the pilot and thus their requirements for a cockpit interface (see Figure 2.4, Huddlestone et al., 2017). World War I Era planes such as the British Sopwith Pup were very simple machines that relied mostly on mechanical sensors, such as an air pressure gauge for indicating altitude or a dynamic pressure indicator for measuring airspeed[9] (see Figure 2.4 all the way on the left).

[9] Mechanical sensors such as Altimeters and Air Speed Indicators still used.



Fig. 2.4 Evolution of cockpit design highlighted by typical designs of the corresponding time. From Left to right: (1) Sopwith Pup, a completely mechanical cockpit from a WWI-ear British fighter bi-plane (2) Lockheed P-38 Lightning WWII era fighter plane, mechanical cockpit with basic electronics (e.g. radio). (3) Lockheed L-1011 Tristar widebody jet-age airliner with early digital cockpit instruments and some of the first advanced autopilots (such as autoland capabilities). (4) Airbus A320-200 first completely digital cockpit and fly-by-wire cockpit. (5) General Atomics MQ-9 Reaper UAS with a digitally remotely controlled ground station. (6) Ukrainian remotely controlled UAS using VR. (Picture Source: see Section V.14.1)

A giant leap occurred between the wars and during World War 2 when pilots started using electronic interfaces for radio communication and navigation (see Figure 2.4second from the left). In the jet age, interfaces were perhaps at their most visually complex.

New technologies such as turbine engines, pressurized cabins and more complex autopilots required the pilots to move away from monitoring basic flight parameters such as attitude and altitude to overseeing systems (see Figure 2.4 third from the left).

This development was further introduced by the 'glass cockpit' (digital displays) and Fly-by-Wire technology. The Airbus A320 was the first commercial airliner to have control surfaces (yoke and flightstick) that were not mechanically linked to the actual flight surfaces (see Figure 2.4 third from the right).

Nowadays, interfaces of uncrewed vehicles have to adapt to rapid technological shifts (e.g. Virtual Reality), tackle the problem of flying a plane without physiological feedback (their bodies remain stationary) and deal with operator fatigue Lim et al., 2018, see Figure 2.4 second to the left and the leftmost).

I.2.2.2 Human-Machine Interaction and Adaptive Interfaces

Adaptive interfaces are part of the domain of Human-Computer Interaction (HCI) defined as:

> *"Human-computer interaction is a discipline concerned with the design, evaluation and implementation of interactive computing systems for human use and with the study of major phenomena surrounding them"*

by Hewett et al. (1992).

One of the main goals of HCI is to increase the so-called *usability* of a specific system. This is:

> *"the extent to which a product can be used by specified end-users to achieve specified goal with effectiveness, efficiency and satisfaction in a specified context of use"*

-"ISO 9241-110:2020 Part 110: Interaction principles" (2021)

To achieve usability, 5 components must be considered when engineering a system according to "ISO 9241-110:2020 Part 110: Interaction principles" (2021).

- **Safety and Security**: the interface must not be harmful and ought to help to reduce the user error.
- **Effectiveness**: the interface should aid and not hinder the task of the user.
- **Efficiency & Functionality**: the interface should expedite the user task.
- **Joy & Fun**: use of the interface should not be a laborious endeavour;
- **Ease of Learning & Memorization**: the learning curve should not be steep, and the procedures learned should be memorable.

AUIs aim to achieve this by adapting themselves based on some input data. As stated by Browne (2016): *"...user interfaces that adapt to a particular user's ability and knowledge of a system"*. Adaptive interfaces are different from both adaptive automation and adaptable interfaces. In Adaptive automation, some of the task load is taken over by some agents, reducing the user's workload. Adaptive interfaces change the interface but do not directly take over tasks for the operator. Adaptable interfaces differ from this, as these interfaces are controlled by the user who can change the interface at any given time (Findlater & McGrenere, 2004).

A simple example of adaptive interfaces are dynamic roadmaps that zoom in when a user slows down, and vice versa (Partala et al., 2006). It should be noted that some definitions of adaptive interfaces are more narrow, and would consider this roadmap a context aware user interface rather than an adaptive user interface, as the data that feeds the adaptation (movement speed is not user related) (Pradeep & Krishnamoorthy, 2019). Other definitions consider all input that is user-relevant, sufficient for an adapative user interface (Paterno, 2014).

Adaptive systems may be much more complex, such as the AugCog system for future soldiers. This system uses Electroencephalography (EEG) and Electrocardiography (ECG) as well as behavioral metrics, to determine the mental state of the operator and to adapt several aspects of the HMI for better usability (Dorneich et al., 2005)[10].

[10]The synopsis of this paper was part of a student project at ENAC from October 2023 - March 2024. Arnaud Lagarde, Clément Douchin and Darlann Bertolone conducted a literature review on Adaptive interfaces and performance measurement methods applied to transportation.

AUI are composed of three primary components as defined by Miraz et al. (2021), that will be highlighted with the two aforementioned examples: (1) A roadmap that dynamically adjusts its size to the users speed, and the AugCog future soldier system (see Figure 2.5, Dorneich et al., 2005; Partala et al., 2006).

- **Adativity Afferential Component**: To be considered an AUI, the system needs user-related data to personalize the interface. This information may be obtained explicitly or implicitly.
  - **Roadmap**: The sensor that measures acceleration, or the GPS that can calculate position change.
  - **AugCog**: The EEG and ECG sensors collect physiological data.
- **Adaptivity Efferential Component**: This component is responsible for the actual implementation of the adaptations and should handle the type of the adaptation and how it is implemented.
  - **Roadmap**: A system that can zoom in or out of a map.
  - **AugCog**: A system to simplify communications, navigation and task offloading
- **Adaptivity Inferential Component**: The final component can be seen as the bridge between the previous two. The component takes the information from the Afferential Component, infers the user's state and dictates the Efferential Component to activate or deactivate the adaptations.
  - **Roadmap**: An algorithm that checks for speed changes and, if detected, informs the Efferentical component whether to zoom in or out based on the change in speed.
  - **AugCog**: The system decides, based on the physiological data, if the operator needs support and, if so, if the support should come in the form of simplified communications, navigation or task-offloading.

AUIs can take on various shapes and forms. For a comprehensive overview of many different AUIs we adapted the taxonomy of adaptations for user interfaces originally proposed by Feigh et al. (2012) (Hinss, Brock, and Roy, 2022a, see Figure 2.6).

The taxonomy was adapted to better reflect the difference between adaptive interfaces and automation. While adaptive interfaces may change how information is presented or manipulated, in adaptive automation, the interface can take over certain tasks (Hinss, Brock, & Roy, 2022a). The difference is particularly important as changes in the task structure due to automation may be so large that the automated and the initial tasks no longer resemble each other (Parasuraman et al., 2000).

Definitions of AUIs vary between sources and one recurring question has been what is considered an adaptation. There can be a broad understanding of the concept that considers any change in the interface based on user-relevant data to render an interface adaptive. The more narrow understanding, on the other hand, considers only changes in an interface that alter the interaction between the human and the interface to be adaptive (Miraz et al., 2021; Paterno, 2014; Pradeep & Krishnamoorthy, 2019). This discrepancy arises particularly when considering feedback to the user. Following the broad definition, displaying the heart rate, alarms based on a fatigue index, or other similar concepts would be considered an adaptive interface,

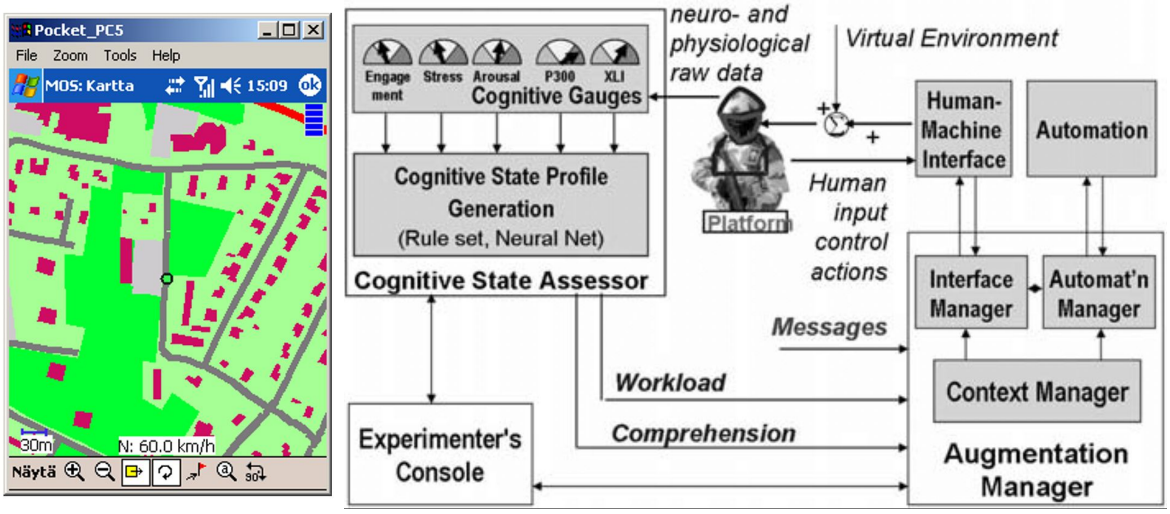

Fig. 2.5 Examples of two adaptive interfaces: (1) A map that can adjust its size based on the users speed and orientation (taken from Partala et al., 2006). (2) The system architecture of the AugCog system (Dorneich et al., 2005).

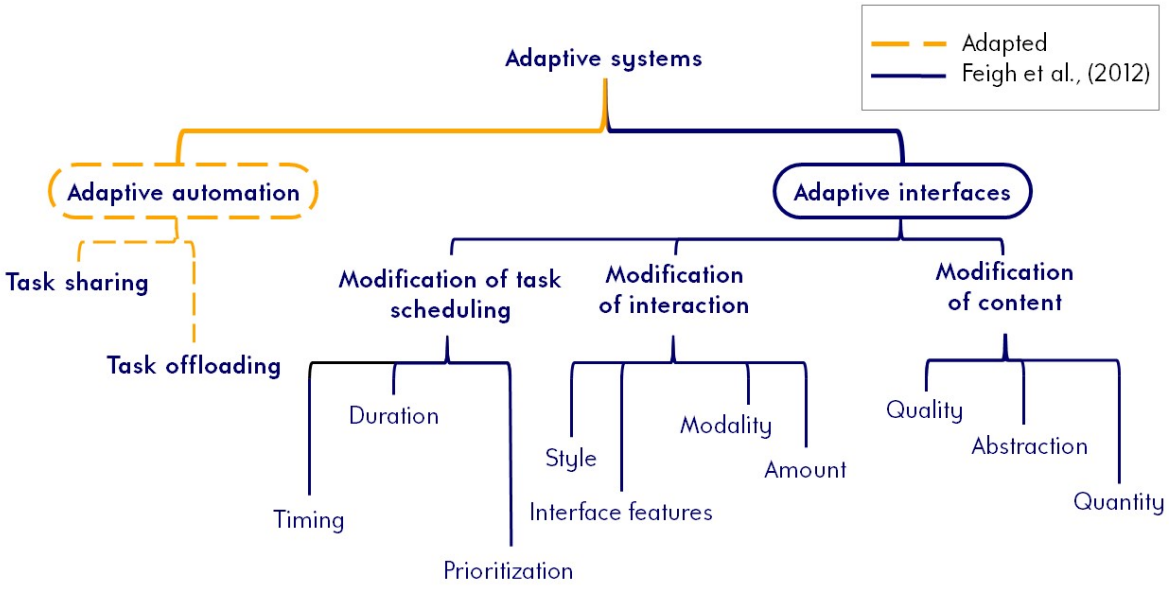

Fig. 2.6 Proposed provisional taxonomy of all types of adaptations based on Feigh et al. (2012) (Picture Source: see Section V.14.1)

as they change what is displayed on the interface based on the users' data. The narrow definition would not include this, as the interaction between the user and the interaction does not change due to the information the interface receives from the user. In this manuscript, we will follow the narrow definition, as the broad definition would allow a simple heart monitor to be considered an adaptive interface. The counterargument to the narrow definition is that that being aware of one's own state could alter the interaction between a human and a machine. However, in this case, the change in interaction is initiated by the human, not by the interface. Similarly, a time display may alter the interaction between a user and the interface as the user becomes aware of the time that they have spent using the interface. For the remainder of this manuscript, we will thus abide to the following definition of an adaptive user interface:

> *An adaptive user interface (AUI) is any kind of interface that uses relevant information about the user, context or environment to change any aspect of the actual interface by changing its interaction between the human and the interface. Interfaces that adapt upon the explicit request of a user are considered adaptable and not adaptive. Adaptations that include automations are considered adaptive automation interfaces.*[11]

I.2.2.3 Why an Adaptive User Interface rather than an improved User Interface?

Key Points

- The malleable attention resource theory (MART) states that operators should always have an optimal workload for the best performance, avoiding over and underloading.
- When evaluating AUIs, multidimensionality can show why the adaptation of adaptive interfaces may have negative consequences if used at all times.

Adaptive interfaces are generally designed to improve a system's usability. But if there is a way to improve the HMI, why not improve the interface and use the adaptation at all times?

Central to this is the malleable attention resource theory (MART, Young and Stanton, 2002a, see Figure 2.7). The theory expands on previous research positing that attention draws from a limited pool to allow humans to deal with a specific task (Kahneman, 1973). According to this, the brain has a certain upper limit for processing capacity, and it will allocate resources to meet the demands of any given task. MART expands on this by stating that the resource pool will diminish if only a few resources are demanded. In a highly automated situation where the required resources are low, the pool of resources is limited. This explains why dealing with unexpected events is difficult when the workload is very low (Young & Stanton, 2002b).

If an operator works with a non-adaptive interface, there is the risk of cognitive overload. This may be due to a change in the workload or their cognitive resources diminishing (see Section I.3.1.2). In that case, they may not be capable of dealing with the changed situation. An operator in the same situation may be capable of handling said situation with an improved interface. However, the improved interface risks the operator having a diminished pool of resources available when the situation is normal, thus resulting in a cognitive underload. An adaptable interface offers the benefits of both scenarios. During normal situations, the operator

[11] The words context or environment are included to highlight that the information need not necessarily come from the user directly but that the information needs to be user-relevant. *Its* specifies, that the change in interaction needs to originate from the AUI. Interaction is included to define that the interaction between the AUI and the human needs to change for an interface to be considered adaptive.

fully engages in their task, preventing underload. However, the interface provides support in more critical phases, and the operator can manage the situation without experiencing cognitive overload (see figure 2.7).

The argument that the MART justifies using adaptive interfaces is not without criticism. Operators could also use an improved interface and do additional tasks to avoid underload. Should resources no longer allow for that, they could drop the task.

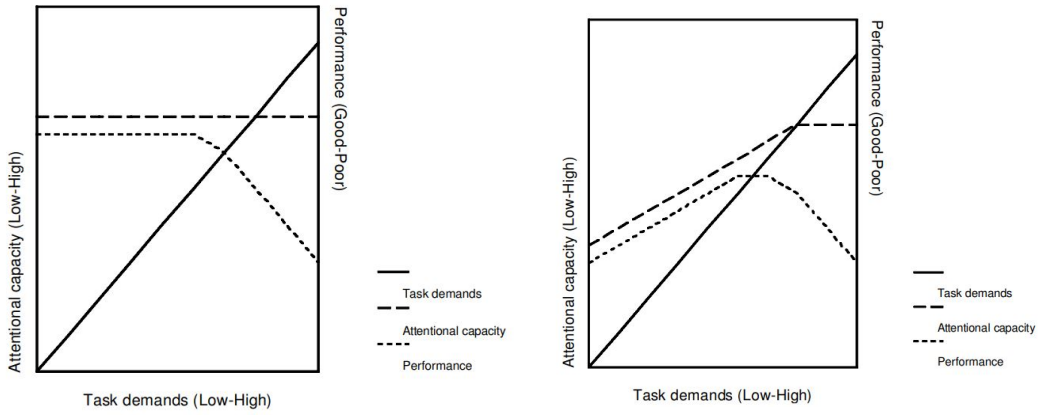

Fig. 2.7 MART: The left side shows the relation between task demands performance and attentional capacities, assuming a fixed capacity (Stokes & Wickens, 1990) (see the horizontal attentional capacity line). On the right, the recruited attentional resources increase with task demands until the maximal capacity is reached. Figures taken from (Young & Stanton, 2002a). (Picture Source: see Section V.14.1)

Another aspect to consider is that certain adaptations (such as reducing the amount of available information) might reduce the interface's functionality and, hence, it may not be usable in all types of situations and for all tasks.

This becomes apparent when evaluating an interface on multiple dimensions rather than just one. To illustrate this, we will propose an abstract example:

We have a certain task that must be done with a certain interface. We can evaluate the performance using accuracy and reaction time (see Figure 2.8 top left). Accuracy is more important than reaction time. The operator can be in two states 1 (vigilant, good performance) and 2 (fatigued, bad performance). In state 1, performance on both metrics is better. Finally, we have an adaptation that can be turned on or off. The adaptation will force the operator to sacrifice reaction time for accuracy.

We can borrow from economics the idea of constructing a Utility curve (Geanakoplos, 1989). This curve calculates the utility (in our case, the overall performance) based on the individual reaction time and accuracy scores. We can plot this 3-dimensional function into a 2-dimensional plot. The Y axis shows accuracy, and the X axis shows reaction time but is inversed; larger values equal a shorter reaction time. This simplifies the plotting (see Figure 2.8 top left).

With our utility function, we can calculate the performance for every point on this coordinate system. The outcome corresponds to the three dimensions; a higher value means better performance. We will draw indifference curves since we cannot plot a 3-dimensional space on two dimensions. All points on these lines have the same utility (performance, see Figure 2.8

top right). We can see that we need to perform the task much faster if we are making more errors to achieve the same performance.

We can plot the four scenarios on our utility curve to understand the utility of adaptive interfaces. First, the vigilant operator without the adaptive interface. Their utility (performance) is high due to a good combination of reaction times and accuracy. On the other hand, the fatigued operator's performance has degraded with regard to accuracy and reaction time (see Figure 2.8 bottom left). Now, using the adaptive interface results in a further decrement in reaction time, but the overall utility (performance) increases through the improvement in accuracy. Interestingly, if the non-fatigued operator uses the adaptive interface, their performance overall would decrease as the overall accuracy is already comparatively high (see Figure 2.8 bottom right). So, the active interface is only useful if it is adaptive.

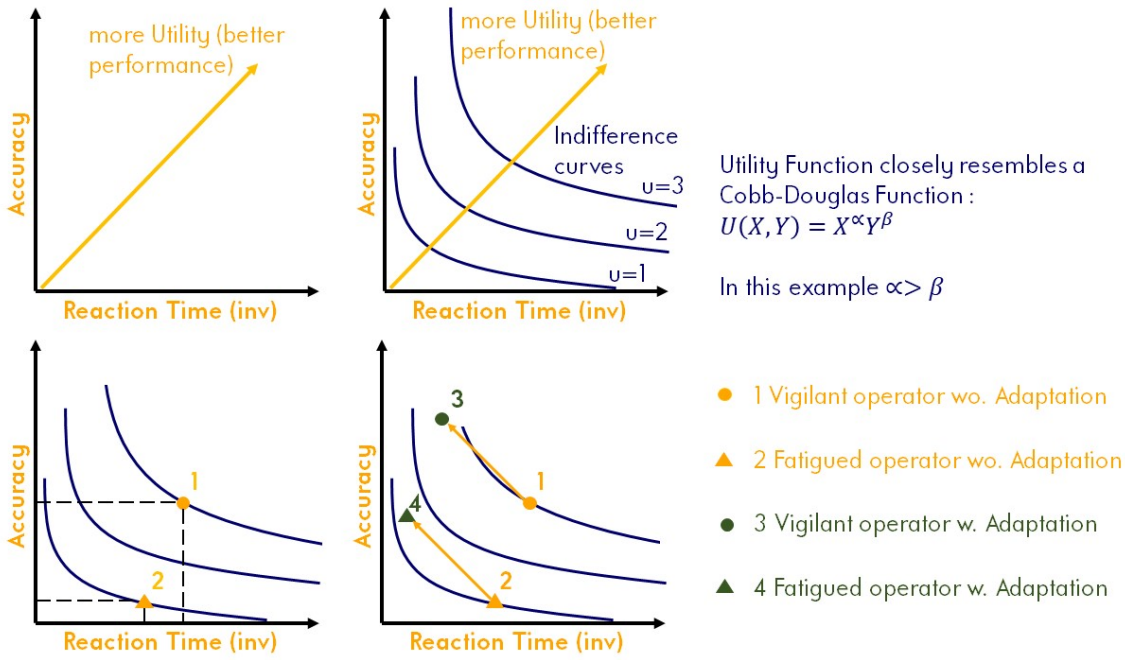

Fig. 2.8 Performance on two dimensions: Top Left: Coordinate system for accuracy and reaction time (inversed); Top Right: Plotting Indifference curves (Cobb-Douglas, approximated, Cobb and Douglas, 1928). Each point on each curve has the same utility (in our example performance, abbreviated as U in the figure). Bottom left: Adding both the operator in a vigilant and fatigued state. In the fatigued state, the operator has a lower utility (performance), and both metrics, reaction time and accuracy, have lower values (compare dotted lines). Bottom right: Effects of an adaptation. Here, the yellow vector represents the performance change (the same vector is used for both operators). For the fatigued operator, the vector results in a higher utility (point 4 is above the indifference curve of point 2). For the vigilant operator, the vector results in a lower utility (point 3 is below the indifference curve of point 1).

I.2.2.4 Adaptive Interfaces for long endurance uncrewed aerial system operations

Key Points

- To mitigate fatigue directly, operators need to rest.
- AUIs may instead mitigate the effects caused by mental fatigue, rather than reducing mental fatigue.
- Mental fatigue causes a decrement in attention, cognitive flexibility and Situational Awareness, AUI for managing these issues exist.

As identified in Section I.2.1.4, mental fatigue and interface design issues in their HMI pose major issues for Uncrewed Aerial System (UAS) operators. To accommodate the effects of prolonged use and mental fatigue may mean inhibiting their onset. Alternatively, adaptive systems may mitigate the effects of prolonged use and accommodate the operator's cognition changes. Some research has shown the effects of caffeine, stream bathing, chicken essence, and physical activity to reduce mental fatigue (Jacquet et al., 2021; Mizuno et al., 2009; Van Cutsem et al., 2018; Yamano et al., 2013). Still, these options lack practical applicability in many critical situations.

External motivation, especially as a reward, has effectively reduced mental fatigue (Hopstaken et al., 2015a). However, motivation may not be a viable solution due to two issues: It is unclear how long the effect of external motivation on an operator lasts. As suggested by Christie and Schrater (2015), it may only allow for short-term mitigation. In addition, systems that deliver a reward tend to result in operator exploitation behaviour. The sole objective becomes maximizing direct reward (the external motivation), thereby neglecting other aspects of the system (Dubey, 1986).

Despite these issues, efforts have been made to mitigate mental fatigue directly. Bright light exposure has been shown to reduce mental fatigue on a subjective level and some performance measures. Smolders and de Kort (2014) tested if bright light could reduce mental fatigue after it was previously induced with cognitively demanding tasks. Results showed that bright light exposure in participants previously reporting mental fatigue could reduce said sensations. Furthermore, participants improved some of the simpler cognitive tasks with bright light exposure. More challenging tasks showed a performance decrement. While not an adaptive interface itself, bright light exposure may be used as such in a future study (for examples of light in the automotive industry, see Figalová et al., 2022; Hassib et al., 2019).

The alternative to mental fatigue mitigation is mental fatigue management, where the effects of mental fatigue (see Section I.3.1.4) are addressed rather than mental fatigue itself. The following sections will detail different approaches for mental fatigue management based on adaptive interfaces for attention, cognitive flexibility and situational awareness. The following sections on attention, cognitive flexibility and Situational awareness will cover the Adaptivity Efferential Component of an AUI (Miraz et al., 2021). The adaptive interfaces discussed in this section are summarized in Table 2.4.

Attention

Adaptive interfaces to mitigate attention problems have generally taken one of two approaches: automation or adaptation of salience of alarms and information.

One approach for attention-based adaptive interfaces, be it for Uncrewed vehicles, robots, etc., is automating some tasks to allow the operator to focus on a reduced set of tasks. Parasuraman et al. (2005) examined if partial automation of responsibilities could help operators during Uncrewed vehicle operations. Participants had to control a UAS and an uncrewed ground vehicle. A performance-based system of adaptive automation was integrated into the task. If operators' performance on a change detection task deteriorated, then parts of their operation were automated. A significant improvement in attention was shown if operators were supported by adaptive automation, as opposed to a condition without adaptive automation.

Deza et al. (2017) proposed another noteworthy system. Their participants were supported in a visual search task by an eye-tracker-supported algorithm that determined if they had sufficiently explored a visual scene and provided recommendations to continue with another image or to explore further. In real-time tests, the system increased the search speed while maintaining accuracy. Though not used adaptively in the present study, the system may be particularly useful when participants are fatigued.

Alternatively, a set of studies has explored exploiting the benefits of bottom-up attention by adapting the salience of alarms and information, thereby decreasing the risk of misses. To accommodate the increased mental workload in air traffic controllers, Aricò et al. (2016) proposed several adaptations, of which two focused on attention. The system could highlight aircraft if they were calling the controller to facilitate locating an aircraft. In addition, if there were a mid-air collision risk, the interface would show a specific animation highlighting the emergency. This study showed that if adaptations were activated during high mental workload, these could improve performance and allow for recovery from high mental workload states. A similar adaptive interface was used to increase the attention of UAS operators (Park et al., 2021). In a preliminary study, participants had to fly a simulated UAS while attention was monitored using a mobile EEG setup. Whenever the participants' attention decreased, they were notified of their decrease with an alert. Effects were insignificant, but this might be due to the small sample size (N= 4).

Another study supporting visual attention was performed by Chew et al. (2021). Here, gaze data was used to inform an adaptive system of what visual input the operator of a remote-controlled forklift needed. Based on fixation clusters, the system would present different information in the visual system. Results showed that operators had a significantly reduced perceived workload when using the adaptive system.

Success was also seen with the air traffic control (ATC) "Red Alert" system. Here, the entire screen was draped in a red hue if there was a risk of a mid-air collision between aircraft. The increase in salience improved the behaviour during the ATC task (see Figure 2.9, Saint-Lot et al., 2020). The effectiveness of systems using alarms as adaptation is closely related to their accuracy. Inaccurate alarms should be avoided at all costs, as false alarms and nuisance alarms can harm performance (alarm fatigue; Onnasch et al., 2013; Ruskin et al., 2021; Wickens and McCarley, 2019). Therefore, several adaptive interfaces have used alarm suppression to increase operational performance. These systems avoid alarm fatigue and allow operators to focus on relevant objectives rather than low-priority alarms (Breznitz, 2013; Liu et al., 2001). This is effective in chemical plants, where an alarm appears approximately every 2 min, risking that the operator would disregard a true alarm (Liu et al., 2001). The concept was also applied to car phones, which would automatically direct calls to the mailbox if the driver's mental workload was too elevated (Piechulla et al., 2003, related: Semmens et al., 2019 evaluated the best time-point to receive non-driving related information during a simulated driving task). Comparable to this are Airbus aircraft, which suppress alarms that are not critical during take-off and land-

ing, as focusing on the task at hand may be disturbed by unnecessary alarms (Airbus, 2008). A study by Welch (2011) tested a different approach to alarm fatigue in a clinical setting. Their alarms were rendered more precise by adjusting the threshold for each patient.

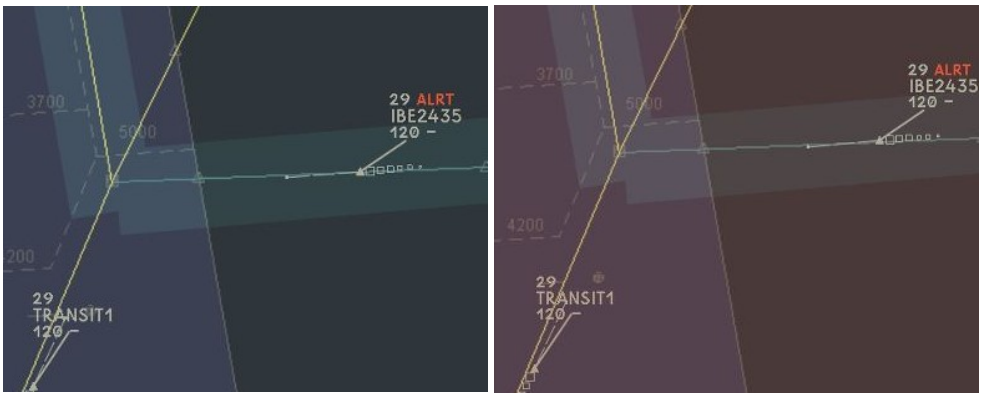

Fig. 2.9 Example of the visual alerts used as an adaptive interface in (Saint-Lot et al., 2020). The Red flash signifies the presence of an emergency aircraft to the ATC. (Picture Source: see Section V.14.1)

Like the concept of attention, the examined mitigation methods differ widely in their approach and may seem contradictory. Adaptive interactions targeting attention may best be summarized not as an effort to add or remove alerts and information but as an effort of effective attentional management. The attentional load must be carefully calibrated, respecting the operational demands, the state of a given mission, and the state of the operator (Piechulla et al., 2003).

Cognitive flexibility

Cognitive flexibility, the ability to switch between tasks and adapt to changing situations, is critical in many complex operations (Koch et al., 2018; Vandierendonck et al., 2010). Task-switching costs are well explored in the literature (see Section I.3.2.4). However, relatively little research has been conducted on adaptive systems and cognitive flexibility. Some studies still point toward some possible solutions for reducing task-switching costs. Arrington and Logan (2004) investigated whether voluntary switches could reduce task-switching costs instead of scheduled switches. Surprisingly, it was shown that task-switching costs persisted in the form of higher reaction times even if switches were voluntary. Similar effects on accuracy were seen in a follow-up study (Arrington & Yates, 2009). Making switches voluntary seems to reduce the tax on cognitive flexibility. Still, within complex operations, it is questionable if switches can be made voluntarily or have to occur at a specific time or as a reaction to critical incidents.

A promising idea is to provide operators with preparation time or cues before switching. As explored in a previous section, some studies (for example: Lorist et al., 2000; Monsell et al., 2003) showed that preparation time and cuing could reduce reaction time in trials (overall reaction time was not shorter if the added preparation time was taken into account). The effect of preparation on accuracy is not conclusively reported in the articles, as error rates were low across all conditions (possible ceiling effect), and no statistics were reported. Rubinstein et al. (2001) reported error rates in their study on cued task-switching. While not reporting any statistics regarding the significance of the data, an almost 30 % reduction in errors was seen if

trials were cued. These results make cuing task switches a promising candidate for adaptive interfaces, especially if short delays are acceptable for increased accuracy. As Petruo et al. (2018) had found, prolonged use in combination with the necessity of using working memory may be particularly detrimental to cognitive flexibility. Adaptive systems that outsource working memory might prove viable. Adaptive automation may harm cognitive flexibility under certain circumstances. Switching costs may occur when the change in Level of Automation (LOA) is so large that the tasks no longer resemble each other (Reinerman-Jones et al., 2011; Squire et al., 2006).

Situational Awareness

Situational awareness (SA), an operator's ability to integrate information to gain a complete image of a situation, is critical in complex operations (Endsley, 1996; Zhang et al., 2021). Several adaptive systems have been tested to improve operators' SA in different contexts. As with attention and cognitive flexibility, automation is much more explored than other forms of adaptation. The results are striking, as automation often negatively impacts SA (Chen & Barnes, 2014). In one study, participants either had an agent use a path determined by the operator or an automatically proposed path. Performance on the objective task measures was better with automation, but participants had a significantly decreased SA (Marquez & Cummings, 2008). In a meta-analysis of the effects of LOA on task performance, the authors offered a very vivid analogy, stating that the lumberjack effect can describe LOA: "the higher they are, the further they fall" - Onnasch et al. (2013). Their analysis of 14 studies revealed that the Level of Automation had an overall negative impact on performance.

As an alternative to automation, Scott et al. (2006) proposed an interruption assistance interface. The idea is that an operator of multiple UASs could recover SA after a distraction. Two interfaces were compared to a baseline condition with no aid. A short playback of what had recently happened was displayed, or a timeline with bookmarks of events was provided. It was shown that both types of interfaces aided in recovering SA following an interruption, with the bookmark system performing marginally better. A similar tool was developed by John et al. (2005) and Smallman and John (2003). The Change History EXplicit tool automatically detects changes in a monitoring task and then notifies and helps an operator understand these changes. The implementation increased response speed by 80 % and accuracy by 150 %. Results on interruptions have also shown an improved recovery if the interruption only partially occludes the main task (Ratwani et al., 2007).

Adaptations working on SA need to account for the time it takes a person to regain SA. Lu et al. (2017) showed that participants in a simulated driving scenario needed at least 7 seconds of video to estimate the position of other cars around their own car. While specific to cars, adaptations for regaining SA need to account for an operator's time to understand the situation. A similar situation occurs with the so-called startle effect, in which an alarm surprises a person and several seconds are required to regain SA (Diarra et al., 2023)[12].

An entirely different route was proposed by Roldán et al. (2017). To keep SA, they tested an adaptive interface and virtual reality to control multiple UASs. Adapting to the situation, the UAS with the highest risk indicator would be highlighted, and the most relevant UAS would be chosen for manual control. These conditions were presented in a conventional interface or an immersive virtual reality simulation. The results proved that virtual reality was an effective way of increasing situational awareness and reducing workload.On the other hand, the

[12] A video of a startle effect in real flight conditions https://www.youtube.com/watch?v=fQNaocQWDsY

adaptive part of the system decreased situational awareness if it was not coupled with virtual reality.

Peysakhovich et al. (2018) proposed a final adaptive approach to SA. The proposal envisions a system equipped with eye-tracking that detects fixations on objects in the cockpit and temporarily removes said objects if looked at for too long. By doing this, the pilots' attention is re-engaged, allowing for the recovery of SA.

Adaptations differ widely in their design, goals, and success. Nevertheless, the findings may perhaps best be summarized by Onnasch et al. (2014) stating: "Automation helps when all goes well, but leaving the user out of the loop can be problematic because it leads to considerable performance impairment if the automation suddenly fails." This claim generalizes to other forms of adaptive interfaces and beyond SA.

Article	Development Stage	Category of adaptation	Adaptation	Mental State	Effects	Domain
Parasuraman et al. (2009)	Implemented	Automation	Automated target recognition	SA	Improved performance and increased	UVC
Deza et al. (2017)	Implemented	Modification of interaction & task scheduling	information on sufficient image examination	AT	Increased ACC	UVC
Chew et al. (2021)	Implemented	Modification of content	selected cameras based on gaze	AT & SA	reduced perceived workload	UVC
Aricò et al. (2016)	Implemented	Modification of content	Highlight aircraft & danger	WM	attentional load reduction	Aerospace
Park et al. (2021)	Implemented	Modification of content	Operator notification about their mental state	AT	Successful recovery of operator AT	UVC
Piechulla et al. (2003)	Implemented	Modification of interaction	Suppression of phone calls	AT& WM	Monitoring traffic environment and variables of driving dynamics	Automotive
Scott et al. (2006)	Implemented	Modification of content	Summary of events after interruptions	SA	Increased SA recovery	UVC
John et al. (2005) and Smallman and John (2003)	Implemented	Modification of content	Changes made explicit	SA	Increased response speed and increased accuracy	Marine
Roldán et al. (2017)	Implemented	Modification of interaction and of content	Relevant UASs selected for manual control and high risk UASs highlighted	SA	Improved performance, and decreased SA (without virtual reality)	UVC

Tab. 2.4 Summary of adaptations for various mental states. Note: situational awareness (SA); working memory (WM); cognitive flexibility (CF); attention (AT); Uncrewed vehicle control (UVC).; Non-partial refers to studies not conducted in the context of any specific domain. This table is an updated version taken from (Hinss, Brock, & Roy, 2022a).

Article	Development Stage	Category of adaptation	Adaptation	Mental State	Effects	Domain
Saint-Lot et al. (2020)	Tested	Modification of content	Screen covered in red hue to alert	AT	Increased detection rate; Understandability increased performance	Aerospace
Taylor et al. (2013)	Tested	Adaptive Automation & Modification of content	Auditory alerts supporting visual perception cognition; Driving automation	WM	Diving automation increased WM and disengagement of driver & Auditory alerts reduced workload	UVC
Marquez and Cummings (2008)	Tested	Automation	Automation of task preparation	SA	Improved performance; decreased SA & automation bias	Aerospace
Arrington and Logan (2004)	Proposed	Task scheduling modification	Allowing for voluntary switches between tasks	CF	Switch costs reduction	Non partial
Lorist et al. (2000)	Proposed	Modification of task scheduling	Task switch cueing	CF	Reduction of switch costs	Non partial
Rubinstein et al. (2001)	Proposed	Modification of task scheduling	Task switch cueing	CF	Reduction of switch costs	Non partial
Petruo et al. (2018)	Proposed	Modification of task scheduling	Reducing working memory during task-switching	CF	Reduction of switch costs	Non partial
Peysakhovich et al. (2018)	Proposed	Interaction modification	Screen removal if looked at for too long	SA	n.a.	Aerospace

Tab. 2.5 Summary of adaptations (Continued) for various mental states. Note: situational awareness (SA); working memory (WM); cognitive flexibility (CF); attention (AT); Uncrewed vehicle control (UVC).; Non-partial refers to studies not conducted in the context of any specific domain. This table is an updated version taken from (Hinss, Brock, & Roy, 2022a).

I.2.2.5 When to adapt

Key Points

- AUIs need to be triggered by some input in combination with a rule or machine learning approach.
- Covariate metrics refer to information that is operator independent, such as weather.
- Behavioral metrics are all data that result from the operator's action.
- Physiological metrics use specific sensors to collect information on the physiological state of the operator.

The previous section covered the Adaptivity Efferential Component or the actual adaptive part of an interface. The focus of this section, the Adaptivity Differential Component and the Inferential Component determine when an interface should adapt itself (Miraz et al., 2021). We propose that metrics that allow for operator mental state estimation can be categorized into either covariate (1), behavioural (2), or physiological (3).

(1) Covariate metrics are completely independent of the operator. This can range from Time on Task (How long a participant has been performing a specific taks, TOT) to the type of mission, time of day, weather, or system state. The important aspect is that an adaptive system purely based on covariate metrics would, in the same situation, behave identically independent of any operator.

(2) Behavioral metrics then encompass all metrics based on the classical system input devices such as keyboard, computer mouse, and joystick. This includes a behavioural performance that is measurable by the system, such as reaction times or accuracy, but could also extend to conscious input, such as activating an autopilot (adaptable interfaces).

(3) The major difference between behavioural and physiological measures is the need for a specific sensor or specific configuration of a sensor to capture any physiological measure, hence creating a new system input device. Typical examples may include Electrocardiography (ECG), Electroencephalography (EEG), and eye-tracking. These sensors' signals can extract metrics in both the temporal and the spectral domains (Roy et al., 2020). Next, a decision has to be made based on the chosen metric(s). If automated, this decision can be made using a rule-based/heuristic approach, reinforcement learning, or a machine learning approach (Sikander & Anwar, 2019; Todi et al., 2021).

Covariate & Operator independent metrics

Adaptive interfaces often use covariate metrics. The time of day may trigger a smartphone to toggle to night mode (Lunn, 2022), a car may adapt its driving characteristics based on the type of road it is on (Kesting et al., 2007), or, as mentioned above, Airbus aircraft can suppress alarms in certain flight phases (Airbus, 2008). Covariate metrics are fully independent of their operators. Although this limits their capability to predict mental fatigue accurately, there are examples where this has been successfully done. Metrics such as time of day, weather, and TOT have been used to support behavioural metrics such as driving performance parameters to detect mental fatigue (Alhazmi et al., 2015). TOT was critical in a time series clustering model informed by participants' behaviour in a simulated UAS task (Roldán et al., 2017). Similarly,

in the automotive domain, mathematical models have used time of day to produce models of driver fatigue (Sikander & Anwar, 2019).

Covariate metrics being independent of their operator may be seen as a disadvantage. However, operator-independent systems also bring advantages. They do not require calibration on the operator, and the system makes far fewer assumptions about the mental state of the operator, therefore reducing the need for trust. Finally, due to their simplicity and non-obtrusiveness, they may be relatively low-cost (no external sensors needed.

Behavioral metrics

Behavioral metrics offer some unique advantages as they remain non-obtrusive and require no additional equipment while allowing for accuracy increases by considering individual behaviour. On the other hand,in contrast to covariates, they generally require a calibration phase, i.e., training a classification algorithm.

To monitor mental fatigue in drivers, steering (Alhazmi et al., 2015; Li, Li, Wang, et al., 2017; McDonald et al., 2012) and the combination of multiple covariate driving measures (Wakita et al., 2006) have proven effective (For a review on driver drowsiness, but not mental fatigue see Liu et al., 2009). Pimenta et al. (2013) tested many metrics extracted from mouse and keyboard behaviour during a task to detect mental fatigue. Six out of 15 metrics were capable of classifying fatigue. These included keydown time, time between keys, mouse acceleration, mouse velocity, time between clicks, and errors per key. As with the covariate metrics; and combining several driving metrics, also confirmed the superior performance of combining several metrics over a single one.

Physiological metrics

Research in physiological computing and passive brain-computer interfaces has identified several task-independent metrics that can be used for adaptive interfaces with high precision (Roy et al., 2020). The downsides of many systems are their intrusiveness, making them less practical, and the need to acquire training data to calibrate a system to a user and specific task. Furthermore, sensors are subject to noise, and differences in metrics (e.g. Heart Rate) may be influenced by several factors (exercise and mental effort). For an extended overview of physiological metrics that can be used for mental fatigue detection and adaptation, please refer to Section I.5.2.3.

I.2.2.6 Design Motivation & Considerations

Key Points

- The adaptation needs to be carefully calibrated to the user and situation.
- Operators need to understand and trust the system for successful use.

Understanding & Control

Beyond the question of when and how to adapt an interface, several other aspects are important for good usability. It has been shown that there is an improvement in performance and operator comfort if the operators are aware of what the adaptation is doing (Graefe et al., 2021;

Kiefer et al., 2017; Pritchett et al., 2014; Scott et al., 2006; Xing et al., 2021). For example, in the study of Hou et al. (2007) and Saint-Lot et al. (2020), understanding the adaptation resulted in significantly better performance. Some research goes further and proposes that the operator should be able to take control (deactivate the adaptation) at all times (Kiefer et al., 2017; Scott et al., 2006).

Pritchett et al. (2014) further emphasise accommodating individual character, preference, and skill differences. This point is also raised by Hou et al. (2007).

Secondary Effects

While designing adaptations, secondary effects also need to be taken into account. An adaptation may prove successful on one measure while decreasing performance on another, as in the study of Chen and Barnes (2014), where behavioural performance increased with adaptation while Situational Awareness decreased. Adaptive interfaces may also be ill-suited for some applications, especially in environments with many non-routine tasks; adaptivity may not be beneficial (Lavie & Meyer, 2010).

Trust

Trust is vital for the interaction between humans and machines. If an operator lacks trust in a system, they may decide not to engage with it and consequently not use its assistance. One catalyst for the Chernobyl disaster was that operators ignored several safety alarms that were coming from the system[13].

On the other hand, putting too much trust in a system can have similarly detrimental effects (Graefe et al., 2021). For instance, the Panamanian passenger ship Royal Majestic crew trusted their auto navigation systems too much. Complacency over 24 hours resulted in the ship running aground, inflicting costs of over 7 million dollars. Luckily, there were no fatalities (van de Merwe et al., 2016).

To avoid both over-trust and mistrust, Lee and See (2004) propose the notion of calibrated trust for automated systems. Calibrated trust occurs when the trust an operator has in a system is equal to the true trustworthiness of the system. To achieve this, three variables are identified as critical components: *Performance*, *Process* and *Purpose*. *Performance* refers to how well the automation performs and has performed. The *process* is the degree to which the operator understands how the automation operates. *Purpose* finally explains to what extent the automation performs what it was intended to do. Negligence in accommodating these dimensions during the development of an adaptive interface can result in mistrust and over-trust. This is easy to imagine in automation-based adaptations, as a low-performance system will likely result in the operator denying the automation of tasks.

To assess this, trust questionnaires have been the long-standing gold standard, however recent advances have proposed an EEG based bio-marker of trust. A 2018 study by de Visser et al. made a preliminary step towards using the EEG signal's event-related potentials (ERP) to classify trust in a system (de Visser et al., 2018). A mature classification system may provide substantial aid in designing adaptive interfaces.

The need for trust in automation is continuously explored, while questions regarding trust in adaptive interfaces remain unanswered. Trust may not be considered necessary for non-

[13] https://www.iaea.org/newscenter/focus/chernobyl/faqs

automating adaptations. For example, a car navigation system that zooms in on its map as the driver reduces his/her speed does perhaps not need any trust (Partala et al., 2006).

There may be the belief that knowledge or trust in automation extends naturally to non-automating adaptive systems. However, variables of trust still impact adaptive interfaces. In a study to investigate if predictability (akin to Lee's process dimension) or accuracy (akin to Lee's performance dimension) had a larger impact on trust in an adaptive system (Gajos et al., 2006). Participants had to perform a task on a modified computer program, where an adaptive drop-down menu with useful items was displayed. The interface presented different items depending on the condition, arranged in a 2 (high predictability, low predictability) by 2 (high accuracy; low accuracy) design. Predictability was manipulated by changing whether the interface would adapt at random times or based on the operator's behaviour. Accuracy determined whether the presented items were useful. The study showed an effect of accuracy and predictability on the subjective liking of the interface. However, the adaptation was used significantly more if it was accurate. Even this study could be argued to partially examine automation, as suggestions (the high or low accuracy conditions) presented in this study could be seen as automation.

Mental state-based adaptive interfaces have to infer the operator's current mental state. These metrics, explored extensively in section 4, determine when and how a system will adapt. The operator must trust the system's ability to predict their current mental state accurately (Fairclough, 2009). This is further exacerbated when a system decides to trigger a system adaptation. The triggering will inevitably signal the operator about the system's estimation of the current interaction. Any discrepancy between the operators' perception of the interaction and the interface will probe trust between the human and the system. If the operator does not trust the accuracy of the interface, they might decide to disengage. Research by Fairclough (2009) and Kizilcec (2016) shows that inaccurate predictions decrease trust. However, this may be mitigated with a more transparent system. The claim of increased trust through strong performance is supported by a 2016 Meta-Analysis on trust in human-robot interactions (Schaefer et al., 2016). The results of 30 studies found that after robot behaviour, the robot's capabilities were the largest predictor for trust. This likely applies equally to adaptive automation.

Several rules should be appreciated to incorporate trust during the design of an interface. Again, following the advice of Lee and See (2004), for adaptive automation, it is proposed to:

1. Design for appropriate trust, not greater trust
2. Show the past performance of the automation
3. Show the process and algorithms of the automation by revealing intermediate results in a way that is comprehensible to the operators
4. Simplify the algorithms and operation of the automation to make it more understandable
5. Show the purpose of the automation, design basis and range of applications in a way that relates to the users' goals
6. Train operators regarding its expected reliability, the mechanisms governing its behaviour and its intended use.
7. Carefully evaluate any anthropomorphize of the automation, such as using speech to create a synthetic conversational partner, to ensure appropriate trust.

Evaluation

Finally, an AUI must also be evaluated and tested. Evaluating the effectiveness of AUIs may be relatively easy in simple tasks. As the tasks become more akin to simulations of realistic, complex environments, many variables must be accounted for (Pritchett et al., 2014). The risk here is to have an improvement on some metric, be it behavioural, subjective, or physiological, while overseeing a decrement on another metric. A multimodal approach incorporating different data types and measures may be the safest way to ensure an accurate picture of an AUI performance.

Beyond performance, an AUI may also be evaluated, considering several other aspects. A system must be designed to function despite the humans not doing what they intended to (Pritchett et al., 2014). Boundary conditions also deserve attention. The inferential component of the AUI should be constructed to handle cases where the operator is close to a threshold or is perhaps oscillating between thresholds (Pritchett et al., 2014). Scalability and the possibility of increasing complexity should also be taken into consideration. Designs that do not allow this lack realism and ecological validity (Hou et al., 2007). Adaptive systems, according to this doctrine, should be built in a way that they could accommodate future additions such as multiple levels of interactions, multiple modes, incorporating individual differences, adding a hierarchy of adaptations, and allowing for multiple ways to activate and deactivate the adaptations (Höök, 2000; Lajos, 1995).

Chapter 3

Humans

I.3.1 Mental fatigue

Central Ideas

- Mental fatigue may be defined as a cognitive condition resulting from mental exertion and its associated resource depletion or loss of motivation.
- Mental fatigue should be measured using subjective, behavioural and physiological data as the absence of effects on one metric need not necessarily imply the absence of mental fatigue.
- Mental fatigue does not result in errors. The effect is indirect: It negatively affects attention, cognitive flexibility, and situational awareness, consequently resulting in errors.

One question to ask: **How can we better detect mental fatigue, and if it is so elusive, is that the best way to ensure safety?**

Mental fatigue was researched as early as 1912 (Arai, 1912; Takasuna, 2013). Tsuruko Arai Haraguchi, a Japanese doctoral student under the supervision of the famous behaviouralist Edward Thorndike (Britannica, 2023), wrote her dissertation titled "Mental Fatigue" at Teachers College of Columbia University (Jenkins, 2013). This made her the first Japanese woman to earn her PhD (Ceynar et al., 2023). Arai investigated mental fatigue in a series of experiments, during which the participant (often herself) was asked to perform a series of multiplications and remember nonsense words for 2-4 Hours. She recorded calculation times, errors, and the pulse of her participants. Her results showed that mental fatigue resulted in a performance decrement while not resulting in any physiological changes (Jenkins, 2013).

Throughout the 20th and the 21th century, mental fatigue remains a concern in many sec-

tors. In the 90ies 21% of the Federal Aviation Administration (FAA) Safety Reports mention pilot or crew fatigue (NTSB, 1999), and a 2006 study found that three out of four short-haul pilots obtained a severley fatigued rating when assessed with fatigue questionnaires (Jackson & Earl, 2006). A Centers for Disease Control and Prevention (CDC) estimation of several studies adjusted to inflation projected that fatigue had an annual cost of 815.6 billion dollars to the US workforce (Wong, 2021). This represents almost 3.5% of the US GDP (Bank, 2022). If fatigue were a government agency, it would be the 5th largest. Nested in between the smaller Department of Agriculture and the larger Social Security Administration (USAspending, 2022). If it were a country, the nation of Fatigue would have a GDP comparable to that of Switzerland or Turkey (keeping in mind that this is only the cost of fatigue occurring in the US Bank, 2022).

As mental fatigue does not result in accidents directly, it is impossible to say if a particular accident occurred due to mental fatigue. Nevertheless, countless accidents are attributed to fatigue and mental fatigue. Some of the most striking catastrophes with a strong link to mental fatigue include:

- On Corporate Airlines Flight 5966, 13 out of 15 people onboard were killed as a result of a controlled flight into terrain. The situation was attributed to pilot error, likely influenced by mental fatigue as the pilots were on their $5+^{th}$ consecutive flight of that day, already having worked for 14 hours (NTSB, 2006);
- The grounding of the tanker Exxon Valdez. The third mate of the ship was likely influenced by fatigue when he took over the navigation of the ship shortly before the grounding (NTSB, 1990);
- The Challenger Disaster, where all astronauts onboard the Challenger Space shuttle were killed during launch. The Accident report states: "The report cites operator fatigue as one of the major factors contributing to this incident. The operators had been on duty at the console for eleven hours during the third day of working 12-hour night (8:00 p.m. to 8:00 a.m.) shifts." (on science and technology House of Representatives, 1986).

In the following sections, we will explore definitions of mental fatigue (Section I.3.1.1), the theories that aim to explain mental fatigue (Section I.3.1.2), experimental designs for studying mental fatigue (Section I.3.1.3), the effects of mental fatigue (Section I.3.1.4), and how mental fatigue can be modelled using data (Section I.3.1.5).

I.3.1.1 Definitions of Mental fatigue

Key Points

- Mental fatigue is independent of physical fatigue as well as fatigue caused by the circadian rhythm.
- A prolonged mental effort causes mental fatigue.
- Mental fatigue as a construct is not directly observable. Therefore, definitions and how mental is measured has a large impact on study outcomes.

Any measurement of mental fatigue heavily depends on its definition. Unfortunately, in over 100 years of research, no consensus has emerged on defining mental fatigue. We will try to provide our own definition based on the existing ones.

To define the subconstruct mental fatigue, it is helpful first to define the superconstruct fatigue and then determine which aspects are distinct from mental fatigue. We can define mental fatigue by what it is not, before discovering what it is.

In several sources, fatigue is defined as:

> *"extreme tiredness resulting from mental or physical exertion or illness"* - Dictionary (2023)

> *"Fatigue is characterized by a general lack of alertness and degradation in mental and physical performance. Fatigue manifests in the aviation context not only when pilots fall asleep in the cockpit during flight, but perhaps more importantly, during the task-critical takeoff and landing phases of flight. Reported fatigue-related events have included procedural errors, unstable approaches, lining up with the wrong runway, landing without clearances, and poor decision-making."* - FAA (2012)

> *... a change in psychophysiological state due to sustained performance [of one or more tasks at work]* - van der Linden, Frese, and Meijman (2003)

> *"a decline in performance on a prolonged or demanding research task generally attributed to the participant becoming tired or bored with the task. The fatigue effect is an important consideration when administering a lengthy survey or test in which participants' performance may worsen simply due to the challenges of an extended task."* - APA (2018)

> *"A physiological state of reduced mental or physical performance capability resulting from sleep loss or extended wakefulness, circadian phase, or workload (mental and/or physical activity) that can impair a crew member's alertness and ability to safely operate an aircraft or perform safety-related duties."* - ICAO (2016)

Phillips et al. incorporate all of the aforementioned aspects, resulting in a complete definition of fatigue (Phillips, 2015).

> *Fatigue is a suboptimal psychophysiological condition caused by exertion. The degree and dimensional character of the condition depends on the form, dynamics and context of exertion. The context of exertion is described by the value and meaning of performance to the individual; rest and sleep history; circadian effects; psychosocial factors spanning work and home life; individual traits, diet, health, fitness and other individual states; and environmental conditions. The fatigue condition results in changes in strategies or resource use such that original mental processing or physical activity levels are maintained or reduced.* - Phillips (2015)

With these definitions, we can now pinpoint the specific aspects that do not align with any definition of mental fatigue. A first distinction needs to be made concerning exertion due to physical performance. Mental fatigue is not Physical Fatigue. It should thus not be directly influenced by physical activity, nor should it directly influence physical activity (Matthews, 2012).

Let us use an example to illustrate these two aspects to provide a clearer understanding. A weightlifter should see no difference in strength when performing an exercise, whether mentally fatigued or not. Disregarding that mental fatigue may influence concentration and precision in executing a specific exercise.

In addition, exercising should not result in decreased performance of cognitive tasks. This does not mean that any decrease in cognitive tasks following exercising refutes the independence of mental and physical fatigue. Exercise greatly affects brain activity, independent of mental fatigue (Ploughman, 2008).

Mental fatigue is not only not physical, but it should also be set apart from fatigue due to periods of extended wakefulness. This type of fatigue may be better explained by terms such as sleepiness fatigue or circadian fatigue. As with physical fatigue, sleep deprivation or the circadian rhythm are implicated in mental fatigue without them being the direct cause(Brown, 2000). Mental fatigue is, therefore, better described as acute than chronic (Borragán et al., 2019; Salomone et al., 2023). Phillips (2015) specified that the exertion must arise from mental effort and not physical or circadian factors, although these can be considered as possible mediating factors.

Definitions of mental fatigue are sometimes very similar to those of affective (e.g. tired mood) and cognitive fatigue (e.g. distractibility) (Matthews, 2012; Salomone et al., 2023). We here adopt the taxonomy of fatigue construct proposed by Matthews et al. (2012), subdividing mental fatigue into affective, cognitive and motivational fatigue (Matthews, 2012). Mental fatigue thus has a cognitive component as well as some subjective components such as a tired mood and apathy.

I.3.1.2 Theories of Mental fatigue

Key Points

- There is no overall consensus on the underlying theories of mental fatigue
- Resource depletion, motivation, large-scale brain network connectivity, as well as cost-benefit calculations seem to be major factors in explaining mental fatigue
- Network connectivity as a resource is emerging as a new angle on the issue

Resource Depletion & Motivation theories

Mental fatigue not only has many definitions, but many theories exist that aim to explain the onset and the results of mental fatigue. Early cognitive models, paralleling physical fatigue, posited that mental fatigue was a loss of energetic resources due to extensive mental work (Boksem et al., 2006; Hopstaken et al., 2016).

The concept of resource depletion is not limited to mental fatigue. As put forward by Kahneman (1973), attention would be drawn from multiple reservoirs and can become depleted over time. Additionally, successful self-regulation, such as controlling alcohol consumption, has been suggested to rely on a reservoir of self-resources (Baumeister, 2003). Resource depletion theories have been criticized for several reasons, two worth mentioning. An implicit assumption of resource depletion theories is the presence of a finite reservoir, but these theories fail to explain that these resources are renewable (Helton & Wen, 2023).

Furthermore, many studies have overturned this theory as evidence showed that external motivation could alleviate mental fatigue (Hockey, 2011; Hopstaken et al., 2015b). Similar studies have shown that mental fatigue is more associated with motivation and cost/benefit calculations than energy resources. These theories posit that as a task is continued for a prolonged period, the reward associated with said task decreases. As a reaction, mental fatigue

will be triggered to achieve two goals: i) To conserve cognitive resources that could be better allocated, and ii) to signal to the person to stop performing the task and pursue other more rewarding activities (Boksem & Tops, 2008; Qi et al., 2019).

Support for this comes, among others, from a 2015 study where participants completed seven blocks of a 2-back task, each lasting around 18 minutes (Hopstaken et al., 2015b). Breaks between blocks were limited to avoid rest. Before starting the last block, participants were informed that the last block's duration depended on their performance on that block. Participants were told the block could last 5 to 40 minutes, while the block was just as long as the others. While performance decreased during the first six blocks, a significant rebound was observed during the last block (see Figure 3.1). In this way, mental fatigue can be seen as a stopping signal that encourages people to search for more rewarding tasks.

Seeing mental fatigue as a stop signal explains the decrease in mental fatigue due to external reward observed in the study by Hopstaken et al. (2015b), but some questions remain. Why did mental fatigue increase (and performance decrease) despite the rewards not changing in the first six blocks? Attributing this to a depletion of energy resources then, however, would contradict the motivational theories that posit that the limiting of cognitive resources is a result and not the cause of mental fatigue (Hockey, 2011; Hopstaken et al., 2015b).

Motivational theories have also been criticized for their lack of evidence on a neurological level. Gergelyfi et al. (2021) showed in a functional magnetic resonance imaging (fMRI) study that there was no significant difference in reward circuit activity between a group of fatigued and non-fatigued participants. The fatigue contrast maps between a fatigued and non-fatigued group of participants did not overlap with the contrast maps of the reward circuits for the participants.

Dynamic resource control theory

Several theories have proposed novel models incorporating resource depletion and cost/benefit calculations (Christie & Schrater, 2015).

The dynamic resource control theory stands out (Christie & Schrater, 2015). The principal component of said theory is glycogen, a molecule in the human body and brain. Glycogen stores energy in the form of glucose for short to intermediate periods (Swanson et al., 1992). According to Christie and Schrater, glycogen provides neurons with the necessary energy when it is swiftly needed. As glycogen stores deplete due to sustained activity, mental fatigue increases. There is experimental evidence to support that glycogen depletion results in decreased performance on many tests and tasks (Benton & Parker, 1998; Martin & Benton, 1999), whereas increases in glycogen result in a performance boost (Benton & Owens, 1993; Manning et al., 1990). This parallels the theories of resource depletion of the mid-20th century (Boksem et al., 2006; Hopstaken et al., 2016). The decrease in performance, however, is argued to occur not only due to the unavailability of glycogen but also as a result of an effort to avoid complete depletion by the body, effectively a signal to conserve cognitive resources. Conserving cognitive resources draws some parallels to the more recent cost/benefit theories in the way that there is some form of calculation of whether a potential reward merits the expenditure of further resources.

The dynamic resource control theory is not without its limitations. The theory does not explain the absence of overlap between fatigue and reward contrast maps observed by Gergelyfi et al. (2021). The results show no overlap between the fatigue and reward contrast maps. If only by correlation, these results show that there seems to be no change in activity within the

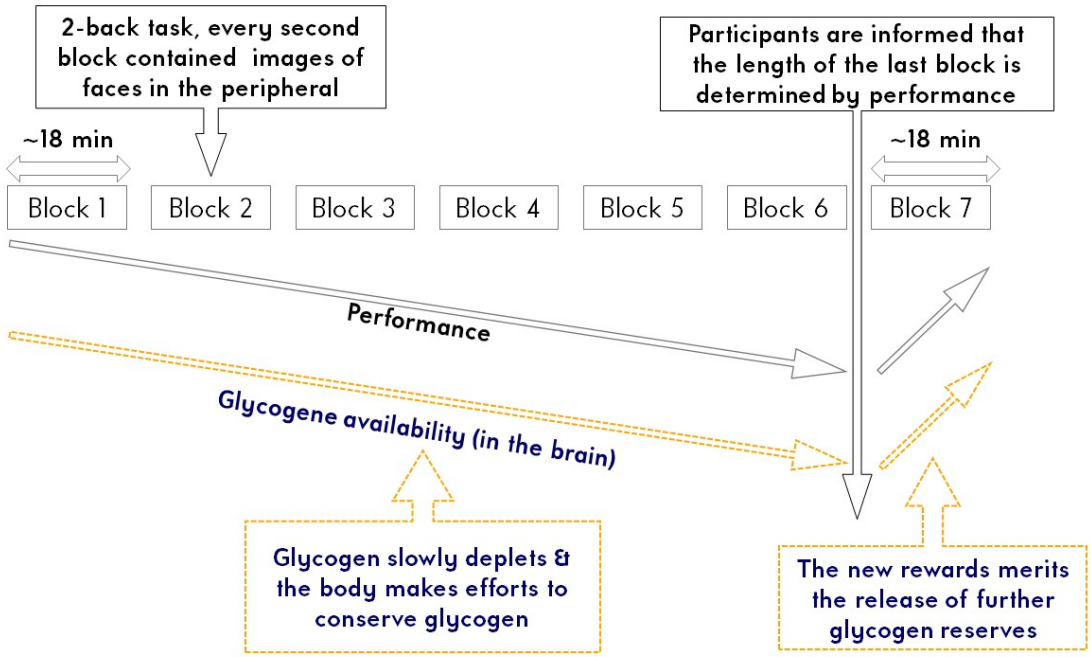

Fig. 3.1 How would the glycogen hypothesis by Christie and Schrater (2015) explain the results of the Hopstaken et al. (2015b) study? The initial decrease in performance over the first six blocks of the 2-back task is caused by the depletion of glycogen, while the body also chooses to keep some reserves for unexpected situations. As the reward for better performance increases before block seven (the negative punishment is increased), glycogen is released and allows for an increase in performance.

reward circuit due to mental fatigue. On the other hand, the same study observed a depression of activity in the task-related areas, partially supporting the glycogen hypothesis.

How would the dynamic resource control theory explain the Hopstacken study (Hopstaken et al., 2015b) with the seven blocks of N-back tasks? During the first six blocks, glycogen might deplete, resulting in the body taking action to conserve the remaining glycogen. Subsequently, performance decreases as less glycogen is made available to the neurons. Once the new reward is added in the 7th block, the body permits usage of the remaining glycogen as the benefit merits a more substantial glycogen depletion. Unfortunately, the study does not have a control group of participants doing 7 blocks of the task without a change in reward. For a visualization, see Figure 3.1.

Connectivity Theories

Mental fatigue may also be understood by investigating the large-scale integration of the entire brain. Qi et al. (2019) studied data from the human connectome to understand fatigue better. The human connectome describes how neurons and brain regions are connected (Sporns, 2012). Here, researchers investigated 29 fatigue-related studies using EEG, fMRI and functional near-infrared spectroscopy (fNIRS). The fatigued state was correlated with reduced global integration and more local segregation. These results support the global neuronal workspace theory (Dehaene & Naccache, 2001), stating that high task demands require a globally inte-

grated network. As outlined above, the connectivity of large-scale brain networks seems to be one of the determining factors of mental fatigue.

André et al. (2019) posit that mental fatigue results from the depletion of functional connectivity due to several processing units evaluating the task constraints and subject's state. The result is an effort signal that modulates the neuronal activity of the brain areas required for performing the task at hand and creates the subjective feeling of fatigue as a signal for goal-directed behaviour.

The theory proposed by Kok (2022) is somewhat similar. The authors redefine resources not as a type of "fuel" but instead as a reflection of the functional state of brain networks. This line of argument explains both resource depletion and resource renewal, the interaction with motivation, and the effect on goal-driven behaviour due to subjective feelings. While both André et al. (2019) and Kok (2022) present original angles to the issue, at the time of writing (June 2024), these theories, to the best of our knowledge, have not been tested in a dedicated experimental setting.

Hence, there is no consensus on the underlying mechanisms of mental fatigue. However, the more modern theories of connectivity and glycogen incorporate aspects of motivation and resource depletion into plausible models.

I.3.1.3 Within and between subject designs for mental fatigue assessment

In psychological or neuroscientific studies, mental fatigue is often estimated using Time-on-task (time passed doing a task; TOT). The assumption is that as participants spend more time on a task, they become increasingly fatigued. A common experimental design requires participants to perform the target task for extended periods. The analysis uses time as an independent variable with dependent variables such as physiological measures, performance, or subjective feelings. Examples for this in a within subjects design are given by Csathó et al. (2012), Dorrian et al. (2007), Guo et al. (2016), Lorist et al. (2009), Monsell et al. (2003), Plukaard et al. (2015), Vijaya K et al. (2014), and Yu, Mückschel, and Beste (2021) (see Figure 3.2 top).

The second popular approach uses a between-subjects design. Here, one group performs a mentally fatiguing but irrelevant task before performing the main task. In contrast, another group does nothing before completing only the relevant task (van der Linden, Frese, and Meijman, 2003; van der Linden, Frese, and Sonnentag, 2003, see Figure 3.2 bottom). The strengths of the within-subjects design are a smaller sample size and a less complicated setup (only one task is needed). However, learning effects may negate the effects of mental fatigue. The between-subjects design on the other hand avoids a confounding learning effect, but it is difficult to control that the 'non-fatigued' group will not also become fatigued or bored due to the non-fatiguing distractor task. Furthermore, interaction effects between the target task and either the distractor or the fatiguing task are difficult to account for.

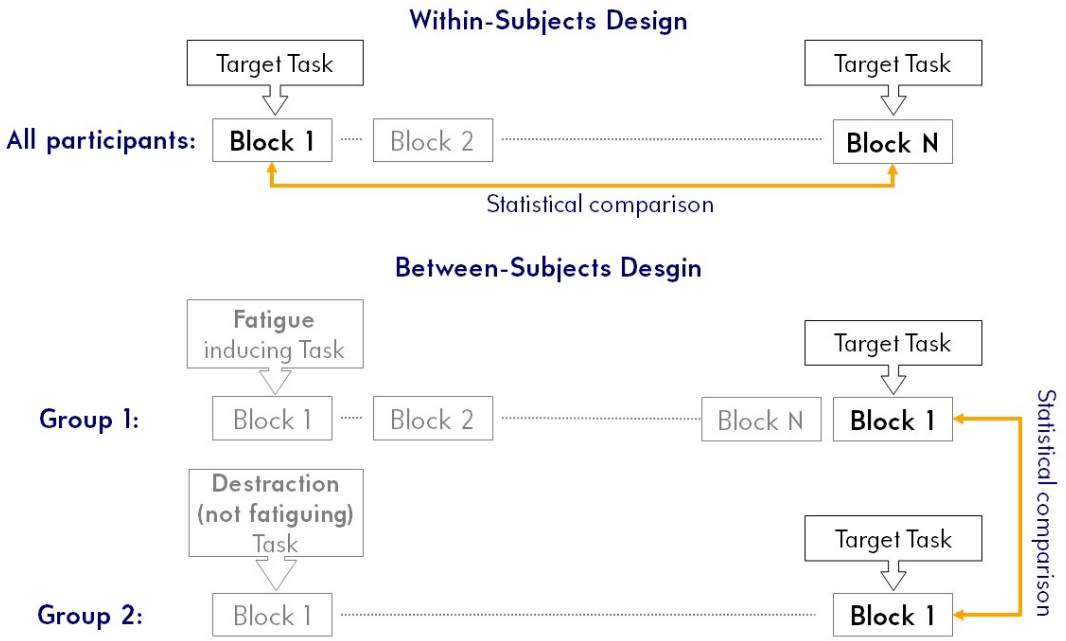

Fig. 3.2 Comparison of a prototypical within and between-subjects design for mental fatigue research. In the within-subjects design, all participants will perform the same task for an extended period of time. During this phase, there are little to no breaks. A statistical comparison is then performed between one of the first and one of the last blocks (Csathó et al., 2012; Dorrian et al., 2007; Guo et al., 2016; Lorist et al., 2009; Monsell et al., 2003; Plukaard et al., 2015; Vijaya K et al., 2014; Yu, Mückschel, & Beste, 2021). In the between-subjects design, one group will perform a fatiguing task for some time before performing the target task. A second group relaxes while the first group performs the first task. The second group then performs the same target task. Outcomes on the target task are then compared between groups (van der Linden, Frese, & Meijman, 2003; van der Linden, Frese, & Sonnentag, 2003).

I.3.1.4 Effects of mental fatigue

Key Points

- In its effects on attention, mental fatigue has a stronger impact on top-down than on bottom-up attentional processes.
- Both cognitive flexibility and situational awareness are negatively influenced by mental fatigue.

Sections I.3.1.1 & I.3.1.2 explored several mental fatigue definitions and theories. This section explores the effects of mental fatigue and its implications within human factors. The results are summarized in Table 3.1.

Cognition	Effect	Duration and Reference
Attention	Decreased top-down attention, no effect on bottom-up attention	35min Holtzer et al. (2010)
	Decreased ACC & Increased RT during visual attention	150 min Csathó et al. (2012)
	Decreased attention during driving	20 min Sun et al. (2014)
	Decreased spatial attention	63 min Guo et al. (2016)
cognitive flexibility	Increased task-switching costs (higher RT & lower ACC)	120 min Yu, Mückschel, and Beste (2021), 120 min Lorist et al. (2009), 180 min Petruo et al. (2018), 60 min Plukaard et al. (2015)
	Reduced task switching costs if switches were cued	180 min Petruo et al. (2018), 120 min Yu, Mückschel, and Beste (2021), 120 min Lorist et al. (2009)
	Increased task-switching costs if working memory was involved	120 min Yu, Mückschel, and Beste (2021)
	Decreased task-switching costs when given opportunity to prepare before switching	120 min Lorist et al. (2009)
Situational Awareness	Increased preservation	na. van der Linden, Frese, and Meijman (2003)
	Decreases planning capabilities	na. van der Linden, Frese, and Meijman (2003)
	Decreased evaluation of a simulated ATC task	120 min Vijaya K et al. (2014)
	Decreased exploitative behavior	na. van der Linden, Frese, and Sonnentag (2003)
	Increased alert fatigue	na. Blackmon and Gramopadhye (1995), Bliss et al. (1995), and Cash (2009)

Tab. 3.1 Summary of prolonged continuous use effects on cognition. Duration refers to the TOT, if applicable (within-subjects designs, else na.). Note: Reaction time (RT); Accuracy (ACC).

Attention

Attention is a broad concept that encompasses many subtypes and is critical in human-machine interaction (Roda, 2011). A simple categorization is accomplished by dividing into exogenous "Top-Down" or endogenous "Bottom-Up" processes (Theeuwes, 1991). Top-down attention is a voluntary allocation toward a specific feature or object. For example, humans can focus on just one person speaking, even in a crowded and noisy room. This is also known as the cocktail party effect (Bronkhorst, 2000). Bottom-up attention is the exact opposite; salient objects in the environment strongly guide involuntary attention. This is the capability to hear our own name

from far away despite not paying attention to that person (Katsuki & Constantinidis, 2014).

General consensus seems that mental fatigue has a much more pronounced effect on top-down attentional processes, as demonstrated by Holtzer et al. (2010). In this study, participants had to perform an Attention Networks Test. The ANT is based on a flanker task (Eriksen & Eriksen, 1974), and interpretation of the behavioural results allows for distinguishing between effects on bottom-up and top-down attention. After performing the task for a prolonged period, bottom-up attention showed no change due to TOT, whereas top-down attention performance significantly decreased. Similarly, Csathó et al. (2012) showed a decrease in performance (increase in RT and decrease in accuracy ACC) on a visual attention task, as mental fatigue increased. Interestingly, although participants showed initial performance enhancements, most likely due to learning effects, fatigue ultimately decreased performance. Further support for the effect of fatigue on attention was seen in a study by Sun et al. (2014). Over time, attention decreased, and subsequently, reaction time increased.

Mental fatigue also seems to impact spatial attention and visuo-spatial attention. In one study, participants were instructed to respond if a target letter appeared in a target position (Wickens & McCarley, 2008). Non-target letters could also appear in the target position. Both target and non-target letters could also appear in a non-target position. As fatigue increased, false negatives and false positives became more frequent. On the other hand, it was reported that significantly fewer errors occurred if letters were in non-target positions (either target or non-target letters in non-target position with response). Participants appeared to focus only on one position while committing more errors than at the beginning of the task.

In a different study by Guo et al. (2016), participants had to perform a visual task for over an hour in which they had to track a target and respond to random stimuli. The results of this dual-task setup highlighted the participant's inability to maintain fast reaction times and accuracy as mental fatigue increased. Participants paid the most attention to the targets and not the surrounding stimuli. In this setup, the optimal strategy was to only focus on target letters that required a response. However, this could also be seen as an indicator that fatigue increases the risk of attentional tunnelling. Attentional tunnelling is the mental state of an operator when they only focus on one feature or an interface, ignoring other relevant information. Depending on the definition of attentional tunnelling it may be seen as a problem of "top-down" attention where all attention is focused on one aspect while neglecting the rest of the environment (Wickens & McCarley, 2019). Alternatively, attentional tunnelling can also be defined as a problem of "bottom-up attention", where an operator cannot detect information outside of a very close set of objectives. The former definition would align with the idea of fatigue having a lesser effect on "bottom-up" attention, while the latter would contradict this notion. To the best of our knowledge, there is no research examining the link between attentional tunnelling and mental fatigue.

Cognitive flexibility, task-switching and multitasking

Several studies show that mental fatigue hinders cognitive flexibility (Lorist et al., 2005; Plukaard et al., 2015; van der Linden, Frese, & Sonnentag, 2003; Yu, Mückschel, & Beste, 2021). Cognitive flexibility is the ability of an individual to change from one task to another task rapidly. Task-switching usually incurs a cost, resulting in increased reaction time and decreased accuracy (De Jong, 1995). As complex systems are often composed of multiple sub-systems (see Lim et al., 2018 as an example of the evolution of control surfaces in cockpits), between which operators have to switch, cognitive flexibility becomes of major importance to operations and

prolonged continuous use (Squire & Parasuraman, 2010). Cognitive flexibility is often used interchangeably with task-switching. At the same time, some researchers distinguish between cognitive flexibility, seen as the ability to switch between tasks, and task-switching, the actual process of switching (Ionescu, 2012). In this manuscript, we will follow this notion.

Some recent evidence for decreases in task-switching performance due to mental fatigue comes from a study by Yu, Mückschel, and Beste (2021). In their study, 33 participants had to perform a prolonged task-switching experiment. The task consisted of trials where a rule had to be followed or a new rule had to be adapted. Behavioural results showed fatigue and task-switching effects. Over time, RT and ACC performance decreased. The surprising finding was that the difference between switching and regular trials grew over time. Mental fatigue even more profoundly affected cognitive flexibility beyond impacting general performance. Similar results were reported by Lorist et al. (2009) and Plukaard et al. (2015). Petruo et al. (2018) had used a similar paradigm. However, here, the task switching was initiated either by an explicit cue or it had to be based on working memory. Results mirrored those in the study by Yu, Mückschel, and Beste (2021). Furthermore, the study by Yu, Mückschel, and Beste (2021) showed that working memory particularly impacted the task-switching ability when participants were fatigued.

Contrary results have also been shown. Lorist et al. (2000) studied cognitive flexibility, fatigue and preparation time. The study showed that switch trials resulted in a lower performance and that fatigue impacted performance. However, an interaction between fatigue and cognitive flexibility was not found. Reaction times during switch trials were significantly lower if participants had longer inter-stimulus intervals to prepare for the switch. Even though the overall time to act (interval + reaction time) increased, the relative decrease in reaction time suggests that participants were better able to cope if given time to prepare for a switch. No effects on accuracy were seen, which may be attributed to ceiling effects, as accuracy remained very high at all times. Similar results are reported in Monsell et al. (2003).

While task-switching is strictly sequential, complex systems often require true multitasking. Task-switching and multitasking have often been investigated separately. Task-switching tasks are tested sequentially, while multitasking (almost always dual-task) involves concurrent task engagement. However, as laid out by Koch et al. (2018), there are many similarities between task switching and multitasking. Many characteristics, such as switch costs, also exist for multitasking. This has led to a proposal to see task-switching and multitasking not as different entities but as the same process with only the degree to which the task is sequential differing along a continuum (Salvucci et al., 2009). To the best of our knowledge, the link between multitasking and mental fatigue is not very well explored. Multitasking is more often investigated as a cause of mental fatigue rather than examining the effect of mental fatigue on multi-tasking performance (Kudesia et al., 2022).

Given that multitasking is argued to exist on the same continuum as task-switching, it is not unreasonable to expect that performance on multitasking may be impacted by fatigue in similar ways as task-switching is (Salvucci et al., 2009).

Situational Awareness and Other effects

Beyond attention and flexibility, it is worth looking at other aspects that may play a particular role in interacting with complex systems. The paper by van der Linden, Frese, and Meijman (2003) showed that participants' performance on a Wisconsin Card Sorting Test got significantly worse if they were subjected to a mentally fatiguing exercise beforehand. The Wisconsin Card

Sorting Test (Monchi et al., 2001) is a neurological assessment tool in which participants are presented with cards. They are instructed to sort the cards according to constantly changing unknown rules. Based on feedback, they then have to adapt their behaviour. It was shown that fatigued participants took longer to adapt to this changing of rules than non-fatigued participants, a phenomenon also known as preservation. In the same study, participants completed the Tower of Hanoi task, testing their ability to plan ahead (Rönnlund et al., 2001). Fatigued participants took significantly longer to make their first move, reflecting difficulties with initial planning. Similar effects were observed following the prolonged operation of an ATC task (Vijaya K et al., 2014). After performing about 60 minutes of simulated ATC work, participants took significantly longer for their tasks. In another study relying on a more ecologically valid design, prolonged continuous use was examined in the context of train driving (Dorrian et al., 2007). Professional train conductors had to perform two simulated 8-hour train driving shifts on different days. From no to moderate fatigue, the operators increased the number of commission errors.

For instance, they applied the brakes excessively, even more than necessary. When extremely fatigued, they no longer made errors of commission but a lot more errors of omission. A likely explanation for these errors is decreased attention. Errors of omission may also be seen as the failure to explore the environment at hand. van der Linden, Frese, and Sonnentag (2003) investigated this possibility by studying the effect of mental fatigue on systematic exploration behaviour. Participants either had to do mentally fatiguing tasks or wait before completing an experimental computerized task. The degree to which the participants used systematic exploration was observed within the complex interface. Mental fatigue proved to have an impact, reducing systematic exploration significantly.

Mental fatigue also introduces alert fatigue in humans. This phenomenon describes desensitization to safety alerts due to repeated exposure to these events (Cash, 2009). This problem has been repeatedly demonstrated in the health care sector, mining operations and construction (Blackmon & Gramopadhye, 1995; Bliss et al., 1995). Decreased adaptation to change, difficulties with planning, reduced exploration, errors of omission and alert fatigue can be seen as independent entities but may also be seen as parts of the more general concept of situational awareness (SA; Endsley, 1995). SA is a key concept that needs to be considered for any interaction with complex systems. Endsley (1995) define SA as the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future." Stanton et al. (2001) explain that SA encompasses understanding the past and current environment, reflections about potential future events and their meaning/relevance, and awareness of goals and goal-directed behaviour.

Loss of SA can have dire consequences, as perhaps best highlighted by the gruesome accident of an L-1011 Lockheed Tristar of Eastern Air Lines in Florida. The pilot, co-pilot and flight engineer of flight 401 did not realize they were losing altitude, as all three were focused on repairing a broken light bulb (O'Brien & Bull Schaefer, 2020).

SA is simultaneously a vital factor during complex operations and at perpetual risk of being undermined by mental fatigue, making it an attractive target for support through adaptive interfaces, which are explored in Section I.2.2.4.

I.3.1.5 Modelling and labeling mental fatigue

Key Points

- To build an applicable model of mental fatigue, a model needs to be consistent, of high quality and have high value to the user
- Mental fatigue can be observed on a subjective, behavioural and physiological level, each of which captures essential aspects of MF
- MF models are often built using supervised machine learning, where the labelling of the data can be of major importance
- TOT is commonly used but may have severe drawbacks with regard to its usability in a predictive model

Detecting mental fatigue in real time requires a model to predict or forecast when an individual will become fatigued. It is worth considering what constitutes a good model before exploring existing models and what can be used to create new ones. An unlikely source for an answer comes from the domain of weather forecasting. In the landmark essay *"What Is a Good Forecast? [...]"* Murphy (1993) defines three types of goodness in forecasting/modelling. (1) Consistency defines the extent to which the forecast/ model is biased. A model that purposely overestimates the likelihood of mental fatigue (or rain in weather forecasting) is inconsistent with the best possible predictions it could produce. (2) Quality refers to the degree to which the forecast/predictions correspond with the observations/ ground truth. This is often reflected in scores such as accuracy, reliability or uncertainty. Finally, (3) Value reflects the incremental benefits of forecasts. A model that predicts when a pilot will have lunch may be consistent and of high quality but will provide little benefit to the user.

With this in mind, we can now examine some of the many existing models for fatigue estimation. Mathematical models are often based on sleep schedules and circadian rhythms and focus on fatigue.

In 1995, New Zealand was the first country to allow airlines to adapt their crew scheduling based on a fatigue management risk system (FRMS, Signal et al., 2008). These scientifically driven tools allow for a better estimation of fatigue and can, therefore, create crew schedules that are better suited to prevent and mitigate fatigue. These systems incorporate policies, training, fatigue reporting, and safety management for safe operations (Bendak & Rashid, 2020; Cabon et al., 2012).

The Systems for Aircrew Fatigue Evaluation model is widely used in aviation to estimate pilots' fatigue, using crew scheduling information, location (time zones), and sleep activity (Powell et al., 2011). The Sleep, Activity, Fatigue and Task Effectiveness model is similar, while also data from an auto sleep algorithm (Hursh et al., 2004).

While these systems are gaining popularity among airlines such as easyJet and Air New Zealand (Eurocontrol, 2012) and specific military squadrons (Paul et al., 2020), this may not be the case for UAS OPS (Caid et al., 2016). The nature of the operations may not allow this, especially in conjunction with a scarce number of operators available (see Section I.2.1.5).

Systems for Aircrew Fatigue Evaluation and Sleep, Activity, Fatigue and Task Effectiveness model are used across various aviation domains, proving their effectiveness. However, these models cannot estimate the individual and temporal fluctuations in fatigue. Moreover, these models only estimate fatigue and do not focus on mental fatigue specifically. Researchers often

use participants' real-time data to estimate mental fatigue to alleviate this issue accurately. Data can be collected using many different methods and sensors. The three largest types of data are (i) subjective data, (ii) behavioural data, and (iii) physiological data. Each type comes with its own advantages and disadvantages which we will present in the following.

Subjective Data

Subjective participant data can be collected using open-questions or closed-form questionnaires. Most recognizable are the Karolinska Sleepiness Scale (KSS, Kaida et al., 2006) and the Samn-Pernelli Scale (SPS, Samn and Perelli, 1982, see Table 3.2). Both are simple scales where participants indicate their current feelings of mental fatigue. In research and building models, subjective data is often considered less reliable due to demand characteristics. Conversely, it is important to consider the face validity of questionnaires and whether they accurately reflect the intended construct (Bertrand & Mullainathan, 2001). Additional factors need to be considered when researching mental fatigue using subjective questionnaires.

As seen in Sections I.3.1.1 & I.3.1.2, mental fatigue has a significant subjective component that needs to be considered. One could even argue that one may be mentally fatigued but not feel fatigued, while it might be impossible to feel mentally fatigued without actually being fatigued. However, the onset of this feeling may only occur at a late stage of mental fatigue (Hockey, 2013). As mentioned above, subjective questionnaires have face validity, which may lead participants to assume that MF is the same as physical fatigue and apply their knowledge of physical fatigue to the questionnaire (Brown, 2000). As can be seen in Chapters 8 & 9, subjective ratings of fatigue reliably increase over time. It is difficult to disentangle the extent to which this may be due to subjective fatigue and demand characteristics. Participants have an intuitive understanding of fatigue (that you should be more fatigued at the end of an effort compared to the beginning) and may be inclined to respond not based on their actual subjective experience but on their understanding of fatigue. Furthermore, participants may be inclined to please the experimenter and thus respond according to what they think the experimenter is expecting. Furthermore the subjective experience may be biased by false memory, and therefore the time in between an experience and the subjective questionnaire relating to this experience may influence the results (DeVellis & Thorpe, 2021). For these reasons, subjective reports often have a bad reputation. But just like in consciousness research, where subjective reports were first frowned upon, but since have become pivotal, the subjective experience is an essential component of the fatigue experience and should, therefore, not be disregarded completely (Dehaene & Changeux, 2011).

Behavioral Data

As explored in Section I.3.1.4, mental fatigue can affect attention, cognitive flexibility and situational awareness. Researchers rely on behavioural metrics to measure such changes, most commonly reaction time (RT) and accuracy (ACC, see Table 3.2). Reaction time is often the more popular choice for analyzing mental fatigue. This can be attributed to reaction time being a continuous variable and not binary when looking at a single trial. Other than accuracy, reaction times are less prone to ceiling or floor effects and may be understood as a more sensitive measure (Wood & Jennings, 1976).

There are several noteworthy advantages of behavioural data over subjective data. Behavioural data is more objective and robust towards demand characteristics than subjective

data while still having face validity. Additionally, it is easier to analyze behavioural data statistically, as there is often more available information, and other than open questions, it is mostly quantitative.

Finally, behavioural data may also be more useful from an operational perspective as it provides direct information on task performance. However, unlike subjective data that will reliably show mental fatigue, many examples fail to show a decrease in behavioural performance due to fatigue. Prior studies (Hamann & Carstengerdes, 2023; Robert J. Hockey, 1997; van der Linden, 2011) & the study in Chapter 8 of this PhD thesis do not, or only partially, produce the expected behavioural decrement. Causes may be grouped into three bins: experimental design, participants, and compensatory strategies.

The experimental design will directly affect the measured behaviour. A task that is too easy or difficult may result in ceiling or floor effects. Any change due to MF may then not be observable. If the procedure requires a lot of steps to be completed before the actual task begins, participants may already be fatigued when starting the experiment. Given a specific task, choosing the task's duration, including breaks and the degree to which participants are aware of the design, will influence behaviour.

In addition, several factors originate from the participants. Learning effects (especially in a within-subjects design, see Figure 3.2) may negate any effect of fatigue if participants are not sufficiently trained on the task (a long training phase, on the other hand, increases the risk of mental fatigue at task onset). MF may also manifest across participants in several ways, therefore biasing statistics. Finally, participants are generally motivated to do well on experimental tasks. They, therefore, tend to use compensatory strategies to mitigate mental fatigue. Focusing on only the most important aspects of a task, exerting more effort, and prioritizing are possible strategies. This last aspect, in combination with individual differences, e.g. one group of participants may prioritize accuracy over reaction times and vice versa, may result in no significant net change of performance across the entire sample (Robert J. Hockey, 1997; van der Linden, 2011). Changes in behaviour may also be of very short duration. Participants may lose cognitive control for a few seconds but rebound (West & Alain, 2000).

Beyond reliability issues, there is another aspect to consider when building a model for detecting mental fatigue. In many real-world contexts, behavioural data is not available. In an experiment, we can, for example, measure the time it takes for a participant to react to a red light and then infer a degree of fatigue. This is rarely possible in a real-world application as the ground truth is often unknown. If such data cannot be obtained in real-time, a model will provide little to no value to a user (Murphy, 1993).

Physiological

Over 100 years ago, when Arai (1912) was investigating mental fatigue, one of the dependent variables was physiological data in the form of pulse measurements. Since then, many other methods of estimating mental fatigue using physiological data have evolved. This section will focus on the theoretical implications of using physiological data. For a list of all metrics that can be used to detect mental fatigue, please refer to Section I.5.2.3. Physiological data from eye-tracking, EEG, ECG, fMRI or fNIRS can be extremely valuable and is often supported by the theoretical constructs of mental fatigue (see Section I.3.1.2, see Table 3.2). Nevertheless, without understanding the underlying mechanisms, e.g., functional brain connectivity reduces when mentally fatigued (Qi et al., 2019), we must make strong assumptions to use physiological data. Many models that estimate mental fatigue rely on supervised machine learning

Data	Advantages	Drawbacks
Subjective	- Face validity - Significant component of MF - Easily obtained/analyzed	- Demand characteristics - Susceptible to a participant's naive theories - Not necessarily temporarily accurate
Behavioral	- Fewer demand characteristics - Face validity	- Susceptible to: experimental design, learning effects, individual differences, motivation, reaction time/accuracy tradeoffs & compensatory strategies
Physiological	- No demand characteristics - Sensitive to changes in MF	- Complex analysis - Dependent on labels - Requires preprocessing - Intrusiveness & data sensitivity

Tab. 3.2 Summary of the advantages and disadvantages of Subjective, Behavioral and Physiological data for mental fatigue models.

(see Section I.5.1). As this requires labelled data for the algorithm's training, researchers must decide what constitutes fatigue and what does not. Furthermore, each type of sensor has its drawbacks. FMRI provides high spatial precision but has low temporal precision, is very costly to use, and is not applicable in mobile applications (Amaro & Barker, 2006). FNIRS uses the same underlying principles of the BOLD signal and is a mobile technique but sacrifices a lot of its spatial precision (Pinti et al., 2020). EEG combines can be used in mobile environments with high temporal precision but suffers from a low SNR (see Section I.4.1.2). ECG is easy to implement but provides little information and is easily influenced by factors other than mental fatigue (see Section I.4.2). Finally, Eye-Tracking is rapidly gaining popularity and may provide valuable information, but is also easily influenced by environmental factors (see Section I.4.3).

One drawback that applies to all types of physiological data is the need to process the data. Filtering, normalizing, cutting, projecting (and all other possibilities of manipulating) data all constitute choices on the researcher's side and will forcibly alter the results of any model. Given a large array of choices, there is a likelihood of making false choices and developing a sub-optimal model. Finally, while these technologies are non-invasive (no harm or damage is inflicted on the body to record data), these sensors are perceived as much more intrusive, and their data is much more sensitive, as compared to purely behavioral measures.

I.3.2 Cognitive flexibility

Central Ideas

- Cognitive flexibility is central to complex cognition and is mostly studied using an iteration of the task-switching paradigm.
- While the concept is well studied in laboratory settings and its importance to complex operations is understood, there is very little research into cognitive flexibility in ecological settings.

One question to ask: **Does our knowledge of cognitive flexibility map onto more realistic tasks, and can it be used to improve performance?**

The following sections will provide a general overview of cognitive flexibility and task-switching protocols. The focus here is primarily on cognitive psychology rather than neuroscience. Besides exploring some of the fundamental theories of cognitive flexibility (see Section I.3.2.1), we will also take a look at the different experimental paradigms (see Section I.3.2.2) and finally, some of the implications of cognitive flexibility in complex operations (see Section I.3.2.4).

I.3.2.1 Theories of cognitive flexibility

Key Points

- Cognitive flexibility is often studied using a task-switching paradigm. Participants perform better when repeating a task rather than when they switch from one task to another.
- The cognitive flexibility interference view suggests that previous task components remain active and interfere with the current trial when a switch occurs.
- The reconfiguration view suggests that task set parameters must be retrieved from memory when a switch occurs.

Cognitive flexibility, the ability to adapt our behaviour and thoughts to the environment, is one of the most important aspects of human cognition (Vandierendonck et al., 2010). Besides working memory and inhibitory control, cognitive flexibility is considered one of the core functions of cognitive control (Diamond, 2013).

A classical paradigm to study cognitive flexibility is the task switch paradigm (Arrington et al., 2003; Cooper et al., 2015; Mayr & Kliegl, 2000; Monsell, 2003; Monsell et al., 2003; Rogers & Monsell, 1995). This paradigm requires the participant to switch between two or more relatively simple tasks continuously. Results consistently show that switching between tasks results in task switch costs (TSC) in the form of increased RT and error likelihood following a switch (Arrington et al., 2003; Cooper et al., 2015; Mayr & Kliegl, 2000; Monsell, 2003; Monsell et al., 2003; Rogers & Monsell, 1995). The existence of TSC poses the question of how these effects occur. Two main theories explain this behavioural phenomenon. The interference and the reconfiguration views (Allport et al., 1994; Logan & Gordon, 2001). The interference view was first proposed by Allport in 1994. The idea is that the components relevant to the previous

task remain activated and interfere with the current trial if a switch occurs. These components are activated in repetition trials but do not interfere as they are the same for trial n-1, and trial n (Allport et al., 1994; Monsell & Mizon, 2006). The reconfiguration view, on the other hand, proposes that task set parameters need to be activated or retrieved from long-term memory when a switch occurs (Logan, 2004).

In pursuit of understanding cognitive flexibility and validating either one of these theories, a multitude of variables, such as preparation time, previous interference residual switch costs, working memory, bivalence, similarity, and other factors, have been investigated concerning their relation to TSC (Goschke, 2000; Liefooghe et al., 2008; Logan, 2004; Mayr & Kliegl, 2000; Monsell & Mizon, 2006). For a review of the two opposing theories and a summary of the influences of factors on TSC see Koch et al., 2018; Liefooghe et al., 2008; Vandierendonck et al., 2010.

Task-switching paradigms will be explored in the next section (see Section I.3.2.2), but one fundamental aspect of these paradigms is their sequential nature. Participants perform one task and then another task. If the time between tasks is reduced to zero, then participants are multi-tasking (Koch et al., 2018). While methodologically very different, it has been argued that the underlying processes are the same (Salvucci et al., 2009). The idea is that when participants multitask, they switch between tasks sequentially. This argument is not beyond criticism, as humans can easily perform many tasks in parallel. For example, many people have no issue with driving a car, riding a bicycle, walking and conversing without stopping to drive during sentences. Schumacher et al. (2001) showed that a subset of participants could perform multiple subtasks simultaneously with no performance loss compared to only one task. Cooper (2010) explain these results and state that tasks that are perfectly independent in their processing may allow for 'perfect time sharing'. In complex operations, this may only very rarely be the case as subtasks often work towards a unified goal. It is thus reasonable to expect participants or operators to incur TSC when performing their tasks, as postulated by Salvucci et al. (2009).

I.3.2.2 Experimental paradigms of task-switching

Since its inception, the experimental paradigms for investigating cognitive flexibility have evolved considerably. The earliest paradigms, called list procedures, used a blocked design where participants would perform blocks (in some cases in the form of printed lists) of task A, task B and mixed blocks in fixed sequences (Jersild, 1927; Spector and Biederman, 1976, see Figure 3.3 A). These designs then evolved to having a single type of block, with repetition and switch trials occurring first in a fixed order (alternating run, Koch, 2003; Rogers and Monsell, 1995, see Figure 3.3 B) and then in a random order (cueing procedure, Meiran, 1996, see Figure 3.3 C). A major difference between a random and a fixed-order design is the necessity to cue the task in a random order, as participants need explicit instructions on the task to perform, especially if the stimuli are bivalent (see next paragraph). Fixed-order designs may use cueing but need not rely on it (Vandierendonck et al., 2010). Moving away even further from these designs, TSC are shown to persist even if participants can switch at voluntary moments between two subtasks (Arrington & Logan, 2004).

Other paradigms for studying cognitive flexibility exist. The Wisconsin Card Sorting Test (WCST) is a popular clinical tool for testing executive function. Participants are presented with a card that displays certain shapes. The shapes may vary in amount, colour and shape. They are then instructed to place said card onto four piles of cards. They are given feedback on whether their choice was correct. Using the feedback, the participants can then infer the rules

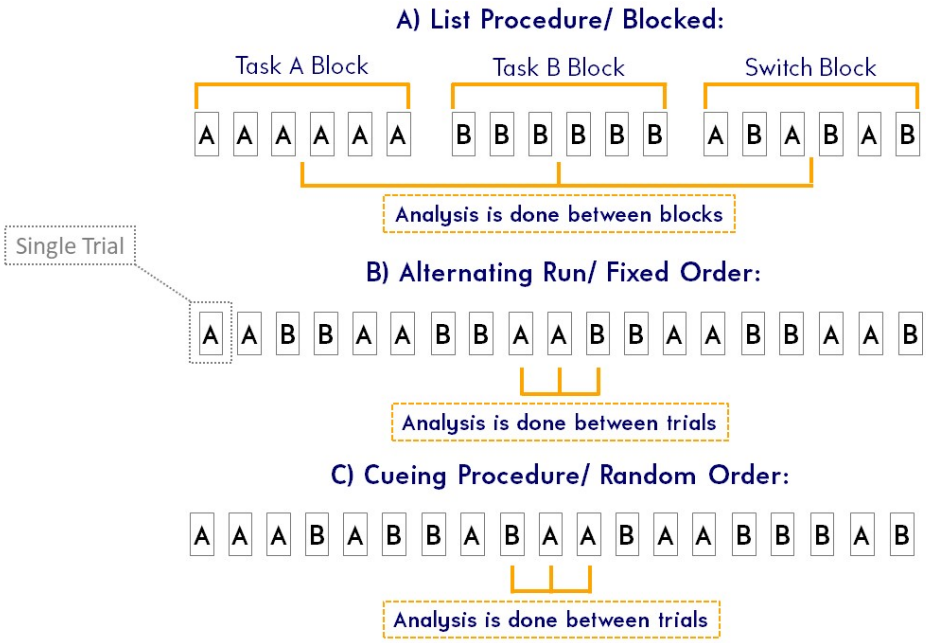

Fig. 3.3 Three typical task-switching protocols A) A blocked design: Participants perform blocks of tasks A, B and mixed blocks. Task-switching costs are estimated by comparing between blocks. B) Fixed Order design: Participants continuously alternate between tasks A and B in a fixed order. Statistical comparison is performed on a trial-by-trial basis. C) Random Order design: Participants continuously alternate between tasks A and B in a quasi-random order. Statistical comparison is performed on a trial-by-trial basis.

to follow (e.g., match colours). The rules change randomly, forcing participants to update their understanding of the task (Berg, 1948).

However, the WCST has been criticized as a tool for evaluating cognitive flexibility. Poor performance is not necessarily linked to a participant's cognitive flexibility but may also be explained by a failure to learn from feedback, forgetting goals or failure to update goals (Geurts et al., 2009).

I.3.2.3 Similarity, Bivalency, Inter-stimulus Intervals & Task-Rule Congruency

Similarity

The similarity factor is of interest for research on cognitive flexibility, especially as most fundamental research has only focused on very similar tasks (Arrington & Logan, 2004; Rogers & Monsell, 1995). Similarity is best explained from the perspective of the task set. The task set is the procedural schema in which the mental resources must be configured to perform each task (Monsell, 2003). Similarity can be seen as the distance between the task-sets of two tasks. Recent results from a study by Siqi-Liu and Egner (2023) show that learning in a task-switching context depends on the task set rather than the physical features or the stimulus presented. The study shows that participants can apply the experience from previous trials to a novel

trial, even if the presented stimulus is unknown. By investigating similarity, we can examine whether task-set information is partially transferable across tasks.

In many experimental designs, ecological validity is not always given, as task switching, in reality, may occur between vastly dissimilar tasks. Ecological validity is essential to cognitive flexibility, as research has shown that issues with cognitive flexibility have significant real-world implications, especially in the domain of Human-Machine Interaction (HMI; Barber et al., 2019; Chen, Barnes, and Harper-Sciarini, 2010; Chen and Barnes, 2014; Reinerman-Jones et al., 2011; Squire et al., 2006). If the fundamental research conducted on cognitive flexibility is to be applied to real-world contexts and problems, the question arises as to what extent these fundamental findings also apply to a realistic context. If this is not the case, then it is of interest to see the differences between task switching in a fundamental context and task switching in an ecologically valid context (Chen et al., 2022; Kleinsorge, 2004; Kleinsorge & Heuer, 1999). This question makes similarity between tasks a critical aspect of understanding cognitive flexibility, as it could be argued that classical task switch paradigms involve tasks (Vandierendonck et al., 2010) that are more similar than task switching that is required of operators in complex operations (Chen, Barnes, & Harper-Sciarini, 2010).

Some studies have investigated the effects of similarity and closely related effects on task switching (Arrington & Logan, 2004; Chen et al., 2022; Kleinsorge, 2004; Kleinsorge & Heuer, 1999; Zhu & Han, 2022). In a study by Arrington et al. (2003), participants were presented with rectangles of different shapes and colours. Participants had to switch between two sets of two similar tasks and make binary decisions. The tasks were judgments of height, width, hue, and intensity. The assumption, later validated, was that the width, height, and hue and intensity tasks were more similar. The responses within that study were univalent (each response possibility for each task was mapped onto a different key on the keyboard or a different vocal reaction). However, the stimuli were bivalent, meaning that each stimulus could be applied to all tasks (e.g., the number 4 can be used for high/low tasks and even/odd tasks).

Bivalency

A stimulus is bivalent if it can be applied to several or all tasks of a task-switching paradigm. Often, responses are bivalent as well, meaning that the participants use the same response keys across tasks. In studies where both the stimuli and the responses are bivalent (the same responses are used for all keys, and all stimuli can be applied to all tasks) an interesting effect occurs: Task-rule congruent stimuli (stimuli for which the correct response to the relevant as well as the irrelevant task are the same) are typically faster than task-rule incongruent stimuli (stimuli for which the proper response to the appropriate task is not the same as to the irrelevant task) (Goschke, 2000; Liefooghe et al., 2008; Meiran, 2008). The increased TSC during incongruent trials support the idea of interference. On the other hand, task-set inhibition, the idea that the inactive task has to be actively inhibited, supports the reconfiguration view.

This does not directly apply to the study of Arrington and Logan (2004), as here, each response was mapped to a different key, and therefore, the effect of task-rule incongruency was kept constant. In a recent study Chen et al. (2022) used a similar dimensional design of different tasks. In addition to switching between two tasks, the switch rules could also be different. The work of Kleinsorge and Heuer (1999) also uses a dimensional design, but here, instead of similarity, compatibility of the stimuli is added as a second factor. Further expanding is the study by Kleinsorge (2004). Using a factorial design, participants switched between a parity or numerical magnitude task while focusing on the numeric value or the number of

digits displayed. Therefore, the participants in this study had to switch between 4 different tasks. However, except for switches between tasks where both factors were changed, the tasks all had one factor in common (either the actual task instruction or the part of the stimulus on which to focus).

Preparation Time

Finally, the effect of preparation time or inter-stimulus intervals (ISI) has been investigated in several studies (Kiesel et al., 2010; Mayr & Kliegl, 2003; Rogers & Monsell, 1995; Rubinstein et al., 2001; Vandierendonck et al., 2010). These studies show that TSC (especially reaction time, but also accuracy) is improved if participants have more time between the instruction and the actual stimulus. It is important to note that the reduction in reaction times is always less than the difference in ISI, meaning that a longer ISI will reduce RT but overall trial time will increase. These factors are summarized in Table 3.3.

Tab. 3.3 Summary of independent variables manipulated in task-switching paradigms

Name	Description	Effect	Reference
Preparation Time/ ISI	Typically referring to the period between a trial-specific instruction and the stimulus onset	Longer ISI results in a TSC reduction	Kiesel et al. (2010), Mayr and Kliegl (2003), Rogers and Monsell (1995), Rubinstein et al. (2001), and Vandierendonck et al. (2010)
Bivalency (stimulus)	A stimuli is bivalent if the stimulus itself is independent of both tasks and, therefore, provides no information as to what task the participant is currently performing. Bivalent stimuli necessitate either a fixed schedule or an instruction before each trial.		Goschke (2000) and Liefooghe et al. (2008)
Bivalency (response)	A response bivalency occurs when the same physical response (e.g. the same button) is used across all tasks		Goschke (2000) and Liefooghe et al. (2008)
Task-Rule Congruency	Task-Rule Congruency occurs only when stimuli and response are bivalent. If a stimulus would require the same response irrespective of task, the trial is Task-Rule Congruent, else it is Task-Rule incongruent.	Task-Rule Incongruent trials increase TSC	Goschke (2000)

I.3.2.4 Cognitive flexibility in complex operations

As defined in Section I.2.1.3, complex operations or tasks often require high cognitive flexibility. This may be due to switching between tasks (Squire et al., 2006), concurrent operation of multiple vehicles (Chadwick et al., 2004), switching strategies (Squire et al., 2006; Wang & Lewis, 2007), interruptions (John et al., 2005; Sherbino & Norman, 2021) or other factors.

Research has focused on cognitive flexibility as a trait and a mediating factor for performance in complex operations rather than examining cognitive flexibility within simulated or real complex operations. For example, a 2016 study showed that cognitive flexibility is a mediating factor in the relationship between age and pilot performance (Van Benthem & Herdman, 2016). However, cognitive flexibility and task-switching remain relatively under-explored as essential components of complex operations.

One reason may be the difficulty of marrying a task-switching protocol with an ecological task. Switching tasks in controlling an Uncrewed Aerial System (UAS) or plane often requires using different devices (e.g. Yoke/ Flightstick, computers, tablets, control surfaces). Furthermore, information is often distributed, meaning a switch requires both an attentional shift and perhaps a head movement. In other cases, the modality of the information may also shift between auditory, haptic or visual. These factors are independent of cognitive flexibility but would likely bias any reaction time measurement and inflate the TSC. Furthermore, creating an appropriate 'schedule' of trials within complex operations may be difficult. Some tasks may take significantly longer and appear more frequently, and some switches between tasks may be unrealistic within the context. Pilots are less likely to change the flight path and then use the radio than vice versa (controlled air spaces require ATC clearances). Nevertheless, a limited number of studies have examined task-switching within complex operations.

A 2006 study had participants control multiple uncrewed vehicles (Squire et al., 2006). These vehicles could be controlled using different automation levels ranging from waypoint-to-waypoint navigation (low automation) to 'Superplays' (high automation), which allowed the participant to make strategic decisions involving several sub-components. The results reproduced the expected task-switching costs when participants changed between automation levels. These results are interesting for two reasons. They show that TSC exists in more complex tasks, but they also show a risk of automation. As shown here, helpful automation may have a detrimental effect on performance due to TSC.

As alluded, interruptions are common within complex operations and may impact performance due to TSC. Surprisingly, systems to support recovery from interruptions may induce more errors than no support at all (John et al., 2005). A task to monitor the airspace around a naval vessel showed that instant reply of the events (aircraft movements) resulted in more errors rather than a simple resumption of the tasks (John et al., 2005). This also means that using an adaptive interface or automation may result in additional difficulties in cognitive flexibility.

It should also be noted that difficulties with cognitive flexibility are not beyond conscious perception. A 2017 study used semi-structured interviews to collect data from 73 UAS Operators. Over 30 % of both pilots and sensor operators reported often having trouble shifting during their work (Scheiman et al., 2017).

Chapter 4

Sensors

Central Ideas

- EEG
  - records the electrical activity of the brain
  - very valuable data of high dimensionality
  - metrics are often either in the frequency or time domain
  - low signal-to-noise ratio
- ECG
  - records electrical cardiac activity
  - easy use for both researchers & participants
  - metrics in both time and frequency domain
- Eye-Tracking
  - modern eye trackers use infrared cameras to record eye movements
  - able to extract metrics from eye movements, pupillometry and gaze
  - reliability is very dependent on the equipment

One question to ask: **How can we realistically use these technologies to learn how operators do their tasks in real-time?**

I.4.1 EEG

Key Points

- Electrical activity in the brain results at least in part from a neuron's action potential in which a signal is sent to connected neurons.
- The resulting signal may be recorded by electrodes acting as a voltmeter.
- The signal needs preprocessing due to the low signal-to-noise ratio and can then be analyzed in many different ways, among which are frequency and time domain analyses.

Electroencephalography measures the brain's electrical activity due to the collective increased activation of millions of neurons[1]. The following subsections will provide the reader with an overview of the structure of the brain (see Section I.4.1.1), the EEG signal (I.4.1.2), how EEG data can be preprocessed (see Section I.4.1.3) and finally how the data may be analyzed (see Section I.4.1.4).

I.4.1.1 Mechanisms

The brain and the spinal cord comprise the central nervous system. The brain may be subdivided into five sections due to 5 different folds in the neural tube that shape the brain during embryonic development. The *brainstem* is the evolutionarily oldest part of the brain, positioned rostral to the spinal cord and comprises the *Mylencephalon*, *Mesencephalon* and parts of the *Metencephalon* (Joynt, 2024; Pinel, 2008). It is most often associated with basic functions such as respiration and regulation of heart rhythms (see Figure 4.1).

Dorsal and partially wrapping around the posterior section of the brainstem, we find the *cerebellum* (Latin for little brain). Beyond its involvement in fine motor tuning, this part of the *Metencephalon* is also involved in conditioning and conditioned learning (Fine et al., 2002). The *Diencephalon* includes the *Thalamus*, *Hypothalamus*, and *Epithalamus*. These regions are involved in sensory integration and neurotransmitter regulation.

Finally, the *Telencephalon* sits dorsally on the cerebellum and encompasses the *Diencephalon*. It is the evolutionarily youngest part of our brain and is involved in higher-order cognition, including sensory processing, problem-solving, emotion, and decision-making (Arnould-Taylor, 1998). More medial sections of the cerebrum include structures such as the *limbic system* and the *basal ganglia*.

The outside layer of the *cerebrum* is called the *cerebral cortex*. This is the walnut-shaped aggregation of billions of neurons and their connecting axons. 90 % of this structure comprises the 2-4mm thick, 6-layered *neocortex*. In layers III, V and VI, we find *large cortical pyramidal neurons* with their *apical dendrites* arranged perpendicularly to the surface of the *neocortex*. These pyramidal cells' activity is the EEG signal's main generator/source (Pinel, 2008).

Pyramidal cells and action potentials

To understand how pyramidal cells create current, it is worth looking at the anatomy and functioning of such a pyramidal cell. Many pyramidal neurons are comprised of three main

[1] In fact neurons are active at all times, firing one action potential per second on average (depending on many factors). Increased or elevated activity thus refers to an increased frequency of action potentials.

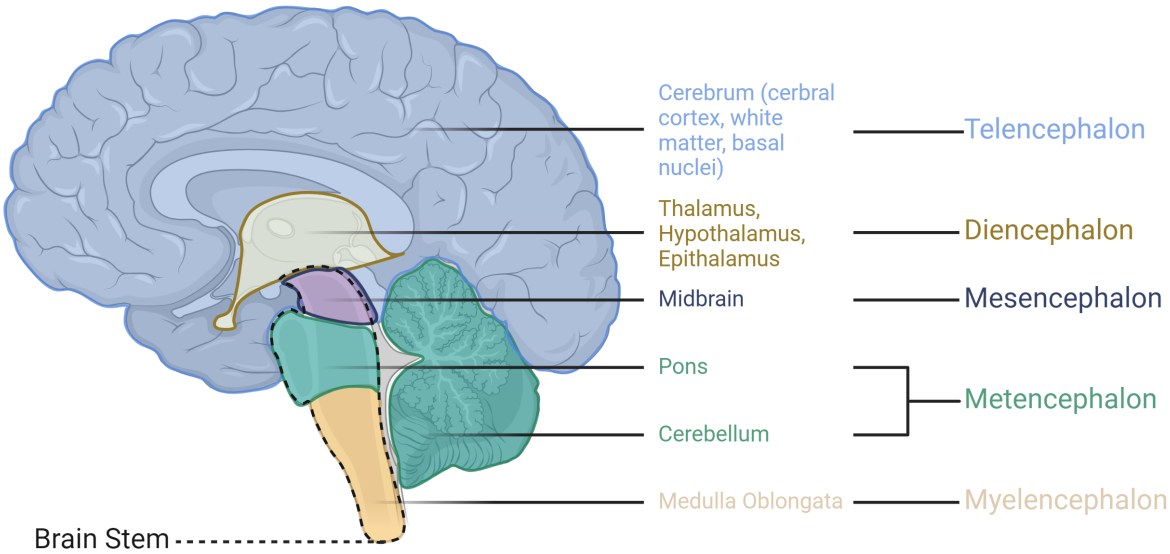

Fig. 4.1 Basic human brain anatomy divided by the five folds of the neural tube. Figure created using Biorender.

sections: *Dentrites*, which allow connections from other neurons to this neuron's body, the *Soma* or cell body, and finally the *Axons* with their *Synaptic Terminals*, which transfer information to the *Dentrites* of other neurons (Hodgkin & Huxley, 1952). When resting, a neuron has a slight negative charge ($\tilde{}65mV$) on its inside compared to the surrounding cerebrospinal fluid. Along the cell membrane, many different channels allow particular chemicals to enter or exit the cell. As these particles are often either negatively or positively charged, their movement can change the relative membrane potential.

Due to several factors (sensory stimulation, a signal from another neuron), some *sodium channels* in the neuron membrane may open. These allow the influx of positively charged *Na+* molecules. The inside of the neuron is now becoming less negatively charged. If the membrane potential change does not exceed -55 mV, the neuron will return to its negative membrane potential of -65 mV. If it exceeds -55 mV, an action potential will occur. The cell depolarizes as this change in membrane triggers more *sodium channels* and the slower *K+ channels* to open. The *K+ channels*, once open, result in *K+* leaving the neuron, deplolarizing, the cell.

The influx of *Na+* creates a significant positive membrane potential that is then reduced by repolarization as the slower *K+ channels* allow *K+* to escape from the neuron. This spike in potential first occurs locally in one area of the neuron before travelling down the cell's axon by a process called saltatory conduction.

Many axons in the brain are myelinated, meaning that fatty tissue covers the outside of the axon. *Myelination* can drastically increase the speed by which an action potential can travel. Once the action potential reaches the synaptic terminal, it will cause voltage-activated CA^2+ *channels* to open, resulting in an influx of CA^2+. This component signals *vesicles* (containers) of neurotransmitters located in the neuron to move to the membrane and release said neurotransmitters outside of the neuron into the *synaptic cleft*.

From the synaptic cleft, neurotransmitters may dock onto another neuron's dendrites. Neurotransmitters docking onto a neuron can have many effects. Excitatory neurotransmitters such

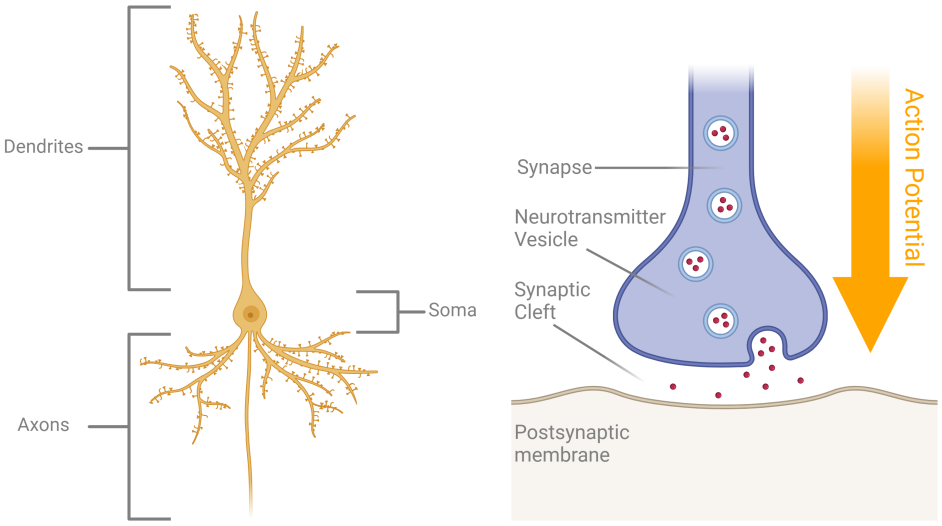

Fig. 4.2 Right: Schematic of a pyramidal neuron of the neocortex. Left: Synaptic cleft. Vesicles of neurotransmitters are released and can be taken up by the postsynaptic terminal. Figure created using Biorender. (Picture Source: see Section V.14.1)

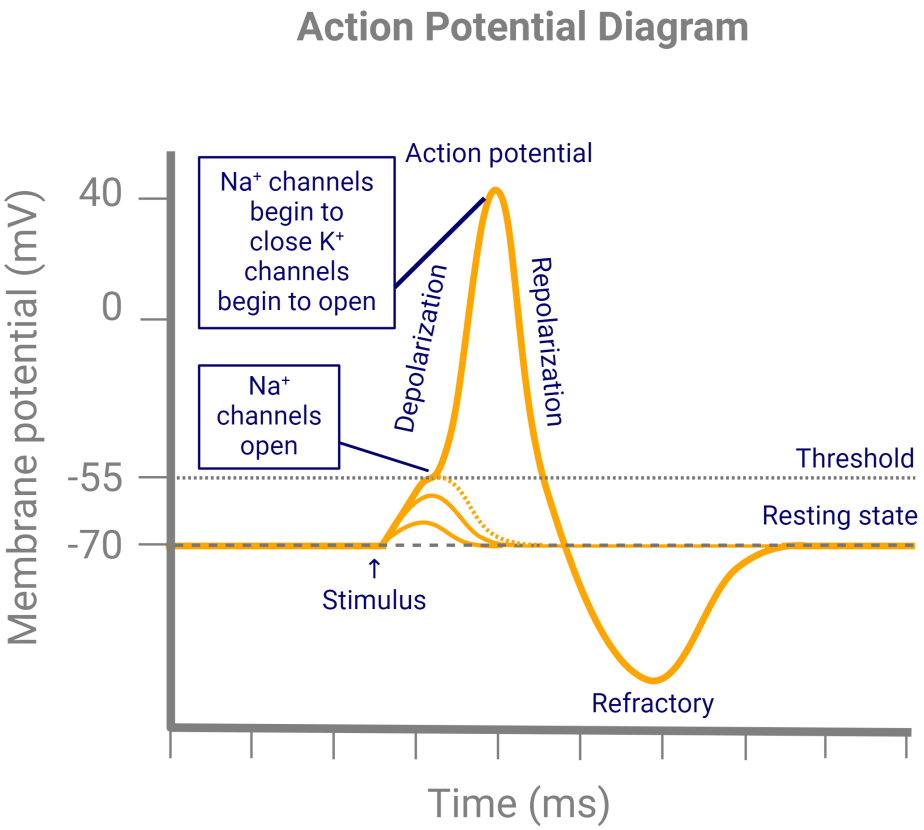

Fig. 4.3 Schematic of the process of an action potential. Figure created using Biorender.

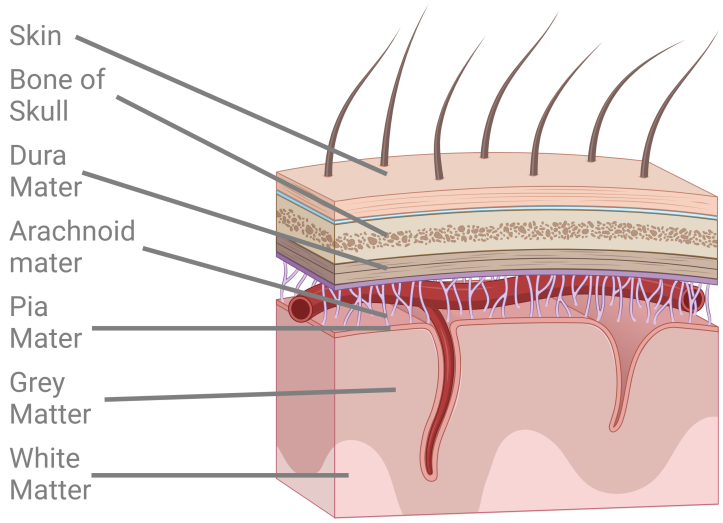

Fig. 4.4 Anatomy of the meninges. The electrical current has to traverse all of these different matters before being recorded by an electrode. The result is noise and mixing of the signal. Figure created using Biorender. (Picture Source: see Section V.14.1)

as *Glutamate* will result in more positive membrane potential in the postsynaptic neuron, increasing the likelihood of that neuron also discharging an action potential. This is referred to as an excitatory postsynaptic potential. Inhibitory neurotransmitters such as *GABA*, on the other hand, result in a more negatively charged membrane potential, reducing the likelihood of an action potential. This is called an inhibitory postsynaptic potential (Cuevas, 2019).

As stated above, EEG measures electrical currents. In pyramidal neurons, these currents result from action potentials and mostly the post-synaptic potentials. If multiplied by millions or billions, these changes in membrane, action, and postsynaptic potentials will result in a strong current that an EEG can measure. The signal propagates through the cortex, cerebrospinal fluid, as well as *meninges* – several layers of tissue protecting the brain and the skull. Especially the skull with its low conductivity, will severely distort and attenuate any signal (Chang-Hwan, 2018). It is important to note that most of the EEG signal originates from cortical neurons close to the cortex, as the electrical current follows an inverse square law. For every doubling in distance, the current reduces by a factor of four.

How does an EEG measure current?

An EEG can be understood as a collection of several voltmeters. Voltage is the electrical pressure or (electric) potential difference between two loads. In a simple electrical circuit with a battery and a lightbulb, a voltmeter attached before and after the lightbulb would read the voltage drop due to the lightbulb (Storr, 2022). Attaching the same voltmeter to the scalp would read the electrical potential difference between the two placements if it were very sensitive. If one area were positively charged at that moment while the other had a negative charge, we would read a significant voltage difference. If both were equally charged, the voltage would be zero (even if both are positively or negatively charged). So, an EEG will always measure the

difference between two different sources. This may be the neighbouring electrode (*Sequential Montage*), a dedicated reference electrode for all channels (*the ground electrode*), or surrounding electrodes (*laplacian method*). The signal measured by EEG electrodes is too weak, so amplifiers amplify the voltage between 1000-100.000 times (Hari & Puce, 2023).

While the basic physics of EEG remains the same, several types of electrodes exist. Passive electrodes collect the signal and pass it to the amplifier. These electrodes are generally cheaper but require a cleaner signal, as the unamplified signal is susceptible to noise while travelling through the cable to the amplifier. This led to the development of active electrodes containing a pre-amplifier circuit within the electrode, thus making the signal more robust while travelling to the amplifier.

EEG among other neuroimaging techniques

Several different techniques exist for measuring brain activity, either directly or indirectly. Functional near-Infrared Spectroscopy (fNIRS) and functional magnetic resonance imaging (fMRI) both measure brain blood oxygenation, and fMRI, in particular, offers a very high spatial resolution. ECG and eye-tracking measure peripheral activity, while EEG and MEG measure the electrical activity as a result of neuronal activity (see Figure 4.5.

EEG is one of the most widely applied noninvasive neuroimaging techniques in research and clinical applications (see Figure 4.5). In contrast fMRI or MEG (see Figure 4.5) often require entire rooms to house the large apparatuses that are needed for data collection. EEG data recording only requires an amplifier (the size of a toaster, or significantly smaller), the EEG electrodes and a computer. Not only does this make EEG significantly cheaper than other methods, but it also makes them more accessible, requiring less training and expertise, as well as allowing for dynamic and highly mobile experiments (Straetmans et al., 2022). EEG data collection is often more comfortable for participants as it does not require being in claustrophobia-inducing environments or injecting radioactive materials (Positron Emission Tomography). The acquisition is also silent, further enhancing participant comfort and allowing researchers to study audition. Finally, EEG has one of the highest temporal resolutions of all neuroimaging techniques (Chang-Hwan, 2018).

These advantages come at the cost of a meagre spatial resolution. Signals from many brain regions mix as they travel from the origin through the scalp to the electrode. Furthermore, an electrode can only represent the mean activity of an area. In addition, the EEG's activity is overwhelmingly representative of the pyramidal cells in the neocortex. The electrical current that is measured follows an inverse square law. Current decreases quadratically as distance increases. Deeper brain regions, such as the thalamus, must produce currents several magnitudes larger than those of the neocortex to be recorded by EEG electrodes. A significant disadvantage for participants and experimenters is the long preparation time it takes to place electrodes and ensure a clean signal (Pinel, 2008). This is particularly true for wet electrodes, where a conductive gel needs to be applied between the scalp and the electrode. Dry Electrodes without the gel requirement exist; however, their signal-to-noise ratio (SNR) is even lower than that of normal electrodes, which already suffer from often noisy signals.

I.4.1.2 The EEG signal

The components of the recorded EEG data that originate from cerebral activity typically lie in the frequency band of 0.01 - 100 Hz, although high-frequency gamma components exceed this threshold. These are better recorded with intracranial recordings (Gardner et al., 2007).

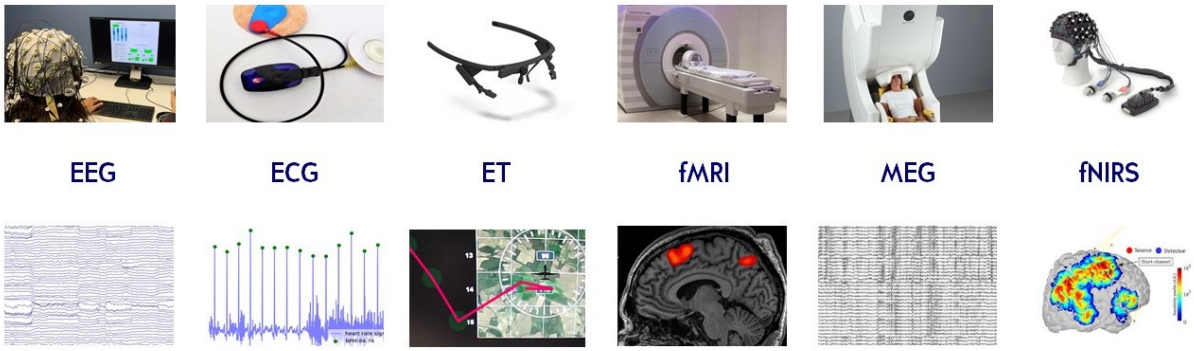

Fig. 4.5 Examples of different neuroimaging methods, as well as ECG and eye-tracking, with examples of their respective signals. (Picture Source: see Section V.14.1)

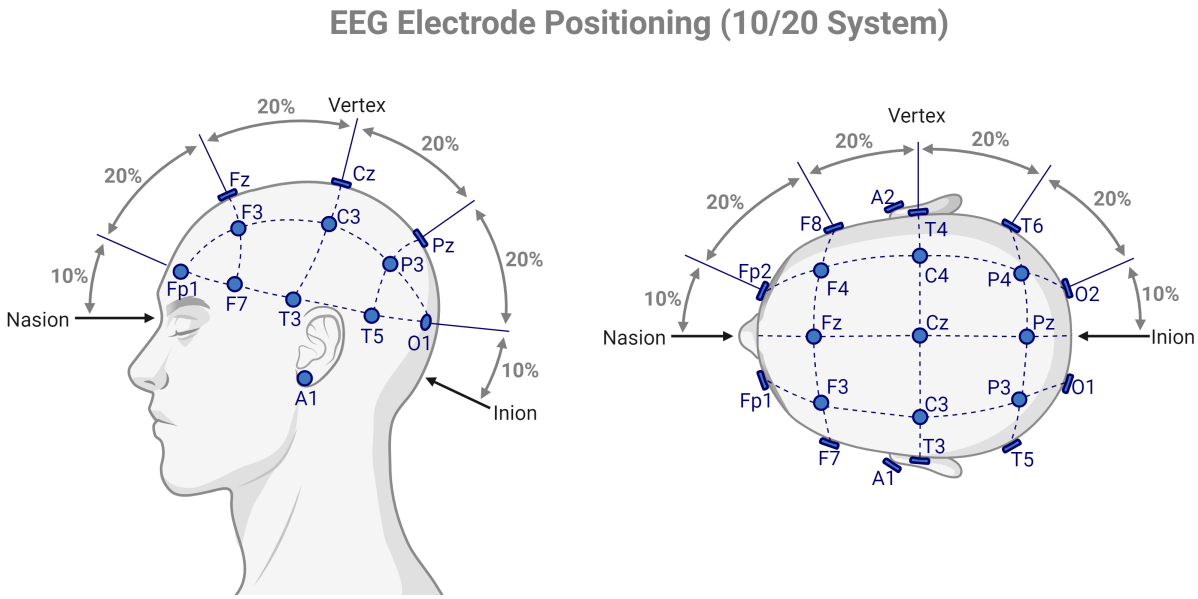

Fig. 4.6 Schematic of the 10-20 electrode positioning system. For a 64 electrode setup, distances of 10% are used (also called the 10-10 system). Figure created using Biorender. (Picture Source: see Section V.14.1)

This is because the power distribution of the EEG frequencies follows a $1/f^1$ distribution (Goncharova et al., 2003). EEG is typically recorded from arrays of 32 to 128 electrodes, with some high-density recordings using up to and exceeding 1000 electrodes (Krishnan et al., 2018). Electrode placement is commonly achieved using the international 10-20 or 10-10 system. Here, the distance between specific skull landmarks, such as the nasion, inion and ears, is divided into 10 or 20 intervals at which electrodes are placed (Jasper, 1958).

The frequency components of the EEG signal are often dividend to 5 frequency bands delta (0;1–4 Hz), theta (4–7.5 Hz), alpha (7.5–13 Hz), beta (14–30 Hz) and gamma (>30, but typically <100 Hz), see Section 4.7

If recorded with high-pass filters above 0.3 Hz, activity in the delta band rarely occurs in

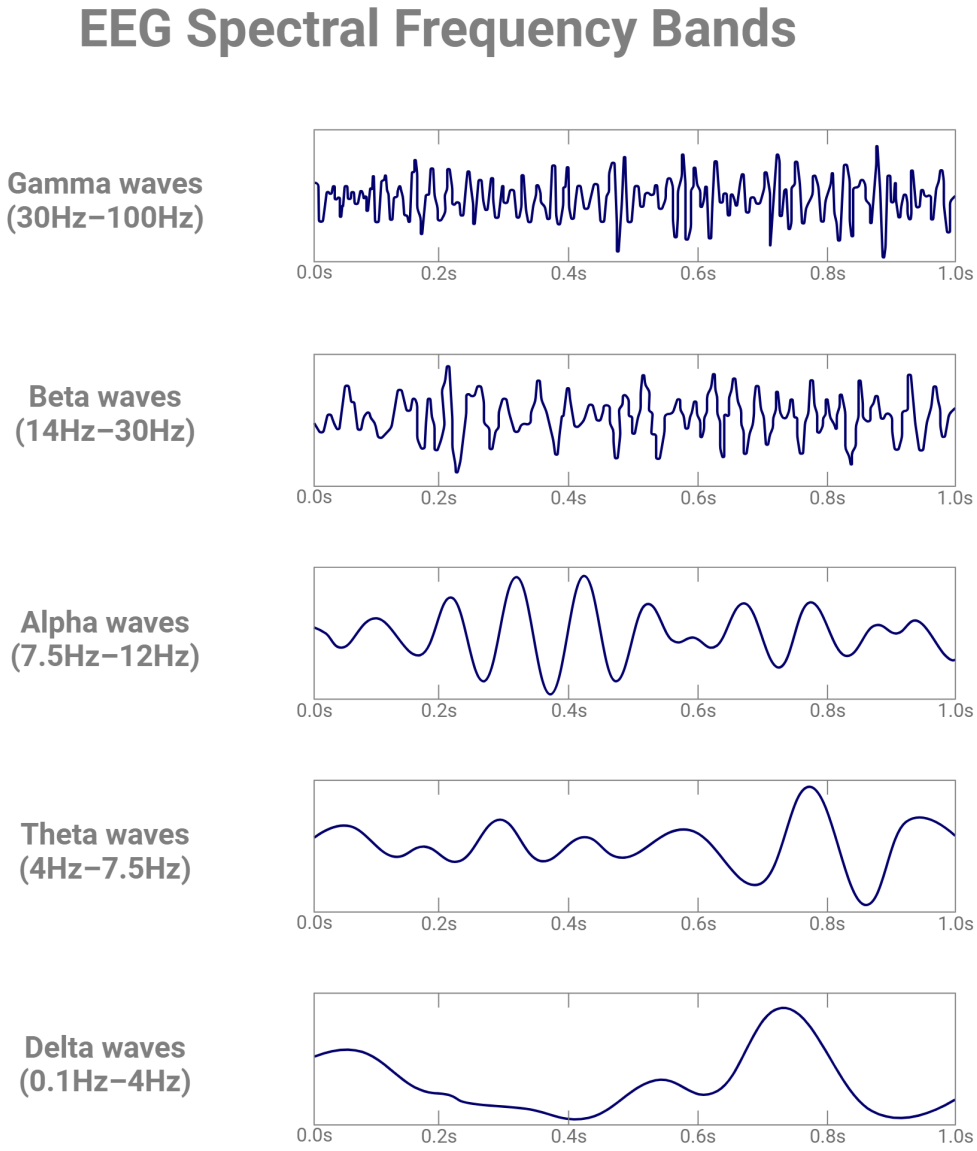

Fig. 4.7 Visualization of the EEG frequency bands. Figure created using Biorender. (Picture Source: see Section V.14.1)

healthy adults outside of deep sleep (Hari & Puce, 2023). It has also been related to synchronising brain activity with autonomic functions, reward-oriented behaviour, attention and motivation (Knyazev, 2012). It has been proposed that delta waves interact with all these functions in an inhibitory manner (Harmony, 2013). The origin of delta waves lies mostly in the medial frontal cortical regions, the insula, nucleus accumbens and tegmental brainstem area (Knyazev, 2012).

Theta oscillations are understood to occur in the hippocampus and are commonly associated with spatial memory processes, navigation and mechanisms of cognitive control (Beppi et al., 2021; Cavanagh & Frank, 2014). Oscillations at the higher end of the spectrum, often recorded in more posterior electrodes, correlate with movement speed and have been associated with spatial navigation. Slower oscillations, located more anteriorly, have instead been associated with non-spatial cognitive processes (Goyal et al., 2020). Theta has also been correlated with higher demand for information processing, memory load and manipulation (Meyer et al., 2019; Sauseng et al., 2010).

The alpha rhythm was the first to be discovered by Hans Berger in 1929 and is generally inversely related to increased cortical activation (Beppi et al., 2021; Berger, 1929). Occipital electrodes often increase alpha activity as visual attention is reduced (Lopes da Silva, 2013). The alpha rhythm is also inversely related to increased demands of alertness and attention (Bollimunta et al., 2011). Both increases and decreases in alpha power have been associated with mental fatigue (Sauseng et al., 2010). The changes in alpha frequency may be modulated by task-complexity (Kamzanova et al., 2011; Wascher et al., 2014). The oldest theory regarding the alpha rhythm proposed that alpha activity is related to cortical idling (Adrian & Matthews, 1934). This view has been challenged repeatedly, as alpha rhythms have been associated with information coding and attention modulation (Lopes da Silva, 2013). For a review, see Lopes da Silva (2013).

The beta frequency band is observed in motor, somatosensory, visual and olfactory areas and is most associated with sensory-motor functions (Beppi et al., 2021). The frequency band activity originates from the basal ganglia and reaches the cortex via the thalamus (Holgado et al., 2010). It has also been associated with vigilance and attention. One experiment showed that participants reacted faster when the spontaneous beta activity occurred (Dustman et al., 1962). This resulted in the proposition that beta brain waves can be used as an index of alertness. Alternatively, the beta rhythm has been proposed to maintain the status quo and to support control and cognition (Engel & Fries, 2010). For a review see Cannon et al. (2014).

As stated earlier, the gamma band is often divided into the low components from 30-100 Hz and the high components (beyond 100 Hz). The gamma band has been associated with attention and memory (Tallon-Baudry et al., 1999), and may play a facilitatory or inhibitory role in perception and cognition (Fries et al., 2007). Research on gamma frequencies is made more difficult as the signal is easily contaminated by the electrical activity elicited by microsaccades of the eyes (Yuval-Greenberg et al., 2008).

Sources of Noise

EEG is a susceptible measurement, making it susceptible to many sources of noise originating from the participant or the environment. We will not explore all possible artefacts but highlight some of the most common or representative artefacts. For a more comprehensive list, please refer to Table 4.1.

The eyes are a significant source of noise in the EEG signal. Electrooculography (EOG) artefacts are electrical potentials that propagate across the skull and contaminate the EEG signal. Their origin is disputed but may be attributed to three mechanisms. (1) When moved,tThe positively charged cornea and the negatively charged retina create a dipole that changes the electrical field. This is referred to as cornea-retinal dipole movement. (2) Alternatively, it has been proposed that retinal dipole movement causes the noise in an EEG. The dipole here originates from differences in charge across the retina. (3) The third possible explanation is that eyelid movements are the source of EOG artefacts (see Section I.4.3.1, Croft and Barry, 2000). Eye movement artefacts contaminate mostly frontal electrodes due to their proximity to the source.

Muscle artefacts are similar sources of noise. Head, face, body movements, and swallowing occur during most EEG recordings with wakeful people, with a signal in the 15-30 Hz range sometimes going as low as 2 Hz (Goncharova et al., 2003; Sörnmo & Laguna, 2005; Urigüen & Garcia-Zapirain, 2015). The artefact can often be detected across the entire scalp and may be more difficult to characterize as they occur less frequently than other artefacts (Goncharova et al., 2003).

A particular case of muscle artefacts occurs from cardiac activity. Due to the regular heart rhythm, ECG artefacts (0.3-6 Hz range, Tamburro et al., 2021) have a very regular pattern that can be mistaken for EEG activity (Urigüen & Garcia-Zapirain, 2015).

Artifact type	Source	Description
Eye blink	Ocular	Internal/Physiological
Eye movement	Ocular	Internal/Physiological
REM Sleep	Ocular	Internal/Physiological
Scalp contractions	Muscle	Internal/Physiological
Glossokinetic artifact	Muscle	Internal/Physiological
Chewing	Muscle	Internal/Physiological
Talking	Muscle	Internal/Physiological
ECG	Cardiac	Internal/Physiological
Swallowing	Muscle	Internal/Physiological
Respiration	Respiratory	Internal/Physiological
Galvanic Skin Response	Skin	Internal/Physiological
Sweating	Skin	Internal/Physiological
Electrode movement	Instrumental	External/Extra-physiological
Electrode Impedance Imbalance	Instrumental	External/Extra-physiological
Cable movement	Instrumental	External/Extra-physiological
Electromagnetic coupling	Electromagnetic	External/Extra-physiological
Powerline Electrical	Electrical	External/Extra-physiological
Head movement	Movement	External/Extra-physiological
Body movement	Movement	External/Extra-physiological
Limbs movement	Movement	External/Extra-physiological

Tab. 4.1 Artifact types and sources, from Kaya (2019) based on Islam et al. (2016) and Sazgar and Young (2019).

I.4.1.3 Preprocessing

EEG data preprocessing is a complex undertaking that depends on many factors, such as equipment, signal quality and planned analysis. The following section provides a very simple overview of some common steps taken in EEG preprocessing. In addition, best practices should be followed for EEG research (and in any other domain, Niso et al., 2022).

Rereferencing

As EEG measures the relative amplitude of an electrode compared to the reference electrode, any noise on that electrode induces noise in the measured electrode. By subtracting the mean of all electrodes from each electrode, the noise of the reference electrode can be reduced. This causes two issues: noise injection and rank deficiency. Noise from any electrode is now injected into each electrode (de Cheveigné & Arzounian, 2018). Alternative statistical estimation methods have been proposed to mitigate this reference effect (Lepage et al., 2014). Average referencing also removes one component of the overall EEG signal by subtracting the average. This means the data becomes rank-deficient, posing issues when working with covariance matrices. To avoid this, adding an artificial channel of zeros to the data before referencing and deleting that channel after the referencing resolves this issue (Kim et al., 2023).

Filters

Filters are an essential component in the preprocessing of EEG data. Notch filters remove line noise from electrical outlets (Leske & Dalal, 2019), high-pass filters remove slow drifts (Kappenman & Luck, 2010), low-pass filters can remove muscle artefacts (Oken, 1986) and band-pass filters can be used to extract the power of a frequency band (de Cheveigné & Nelken, 2019). For an overview of how filters work and how they should be employed, de Cheveigné and Nelken (2019) provide an overview that is nothing short of brilliant. Filters produce output by calculating the weighted sum of several data points. Depending on how each data point is weighted, the resulting signal is filtered in any of the aforementioned ways. A moving average, where each point is calculated to be the average of itself and a specific subset of points around it (e.g. 10 points before and after, or 5 points before), is a form of a low-pass filter, where high-frequency components are removed (see Figure 4.8). This means each point in the filtered signal depends on several points in the unfiltered signal. It also means that each input data point influences several filtered data points. Filters are either causal, which means that the output only depends on previous data points, or they are acausal, in which case they depend on previous and future data points. Choosing a filter carries the risk of both losing information and introducing signal-mimicking noise. When analyzing Event-Related Potentials (ERP) (see Section I.4.1.4) or other temporal features, there is the risk of blurring temporal or causal relations. For frequency analyses, the risk of spurious oscillations is of serious concern. An impulse or a step in the signal may be converted into noise with an oscillatory feature by the filter (de Cheveigné & Nelken, 2019).

Cleaning Algorithms

Many different cleaning or denoising algorithms exist, promising to improve the SNR. The following paragraphs aim not to provide an extensive overview of all existing algorithms but rather some of the most common approaches and the approaches used in this PhD work. For a

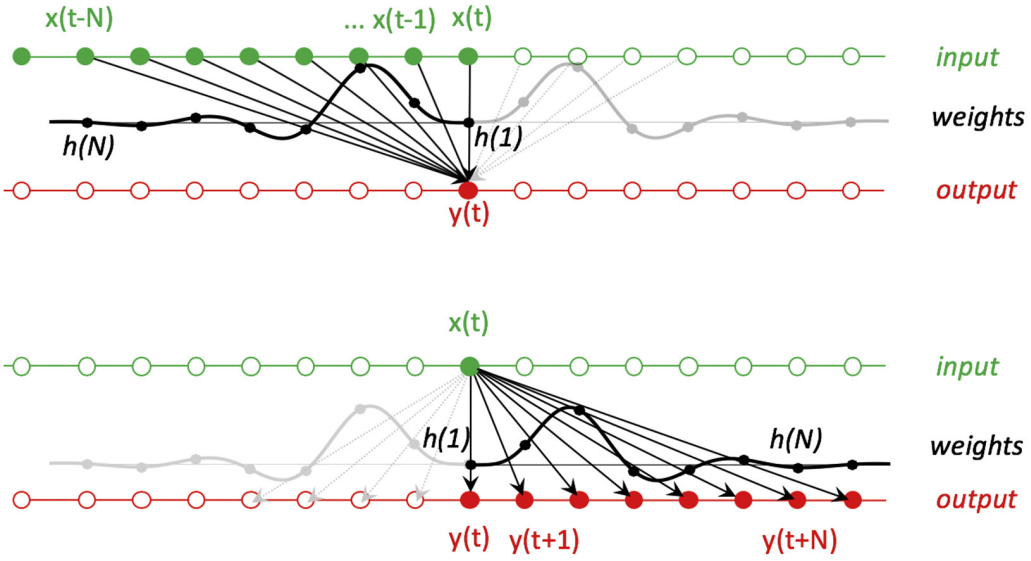

Fig. 4.8 Figure taken from de Cheveigné and Nelken (2019): Top: each sample of the output y is the sum of samples of the input x weighted by the impulse response h. For a causal filter, only past or present samples of the input contribute (black). Future samples can contribute to an acausal filter (gray). Bottom: another way of describing this process is that each sample of the input x affects multiple samples of the output y, with a weight determined by the impulse response h.

more extensive overview of denoising algorithms, please refer to Kaya (2019), Minguillon et al. (2017), and Urigüen and Garcia-Zapirain (2015).

Linear Regression algorithms Linear Regression algorithms assume that the recorded EEG signal is a sum of the clean EEG signal and one or more artefact signals. Using Electrooculography (EOG), Electromyography (EMG), ECG reference channels or artefact templates, the noise can be inferred and then subtracted from the overall signal. The technique has been widely used to remove either EOG or ECG artefacts (Croft & Barry, 2000; Waser & Garn, 2013). Linear Regression Algorithms have mostly been replaced by more efficient algorithms, such as PCA or ICA, that do not require dedicated reference channels (Minguillon et al., 2017; Urigüen & Garcia-Zapirain, 2015).

Blind Source Separation In Blind Source Separation (BSS), a matrix of signal sources is estimated from the overall EEG signal. Sources of signal not originating from brain activity can then be removed from the signal. PCA and ICA are the most common algorithms, with ICA being considered more reliable. A likely explanation comes from the assumptions that both PCA and ICA make. An ICA assumes that the data is a mixture of independent sources and aims at finding statistically independent components. As noise is assumed to be uncorrelated from brain activity, this assumption is likely met by EEG data (Minguillon et al., 2017). Artifact Subspace Reconstruction (ASR) is a popular algorithm that automates this process and can be used for online and offline data (Chang et al., 2018).

Source Decompostion Source Decomposition aims to decompose the signal of each EEG channel into its waveforms. This is done independently on each channel and assumes that a single function can represent every source of noise or signal (Minguillon et al., 2017; Urigüen & Garcia-Zapirain, 2015).

Data Clamping Outliers present a large risk for machine learning algorithms, as they disproportionally skew the data. Recently, a simple method of handling these extreme outliers has outperformed more complex data transformation algorithms. To clamp data points, the EEG data is standardized, and all data points exceeding a certain standard deviation threshold (e.g. 20 STDs (Défossez et al., 2023), or 5 std (Benchetrit et al., 2023)) are set equal to the value of the threshold. To compare this approach with a more sophisticated algorithm, see Défossez et al. (2023).

I.4.1.4 Analysis

There are countless ways to analyze and use EEG signals to understand the brain and human cognitive functions. As with preprocessing, the following paragraphs are not exhaustive but highlight some of the most common ways to analyze EEG data.

Time Domain - ERP

The analysis of event-related Potentials (ERP) presents a powerful neuroscience tool. The basic assumption is that a specific stimulus will evoke a characteristic brain response. The event is repeated several times in many trials to find the signal in the noise. The signal of each electrode can then be averaged across trials by time-locking the event to the same time point for each trial. Since we assume that the noise is random but not the signal, adding more trials will average out the noise (see Figure 4.9). The remaining waves are called potentials and are assumed to originate from brain activity (Luck, 2005). The origins of these potentials are debated. The traditional view sees ERP potentials as the activity being superposed to the independent EEG activity (Tallon-Baudry et al., 1996). Alternatively, ERPs may also result from phase resetting, where spontaneous neural activity aligns its phase to a specific event (Congedo & Lopes da Silva, 2018; Jansen et al., 2003).

Several events can be in an experimental paradigm to create ERPs. One common paradigm is an oddball paradigm (Squires et al., 1975). In this paradigm, two auditory or two visual stimuli are presented randomly. One stimulus occurs infrequently and may require the participant to respond. Using this paradigm, an event-related potential can be reliably observed over the parietal electrodes. 300 ms after the stimulus presentation, a distinctive positive component, the P300, can be observed (Dehais et al., 2019). In the oddball paradigm, the size of the positive deflection is larger for the infrequent stimulus compared to the frequent one. ERP analysis allows for investigating many cognitive processes by comparing latencies, amplitudes and occurrences of different potentials as a response to different events.

Time-frequency analysis is a specific type of ERP analysis that allows the investigation of frequency components related to an ERP. For a summary, see the work of Chang-Hwan (2018).

Frequency Domain

In frequency analysis, the strengths of the different powerbands (see Figure 4.7) are compared across location, time, condition or other factors. Frequency analysis is used to assess mental

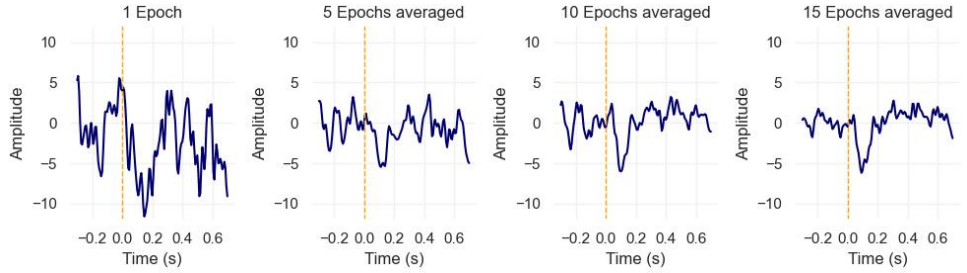

Fig. 4.9 ERP example: Signal in response to an auditory stimulus (yellow line) of 5 electrodes averaged over an increasing number of epochs.

workload (Chikhi et al., 2022), fatigue (Tran et al., 2020), emotions (Rahman et al., 2021), sleep (Zhao et al., 2021) and many other aspects of cognition. Beyond simply comparing individual powerbands across conditions, the use of ratios such as alpha over theta or others has become a popular means of combining the information of several powerbands into one (Raufi & Longo, 2022).

EEG data must first be converted from the temporal to the frequency domain to calculate the power of a frequency band. We can use the Fourier transform to calculate a signal's power spectrum. However, the original Fourier transform assumes an infinite and continuous signal. When digitalized, EEG measurements are neither discrete nor infinite. The Discrete Fourier Transform (DFT) does not make these assumptions and can thus be used for EEG signals. Here, it is important to observe the Shannon theorem, which states that the highest observable frequency is equal to the Nyquist Frequency, which in turn is equal to the sampling frequency divided by two (de Cheveigné & Nelken, 2019). In practice, the sampling frequency should be at least three times as high as the acquisition frequency (Chang-Hwan, 2018). The Fast Fourier Transform is mathematically identical to the DFT but avoids some redundant calculations and is thus faster.

Two issues need to be considered during spectral analysis. **Aliasing** occurs when a signal has a high-frequency component that masks as a different frequency due to the sampling rate (see Figure 4.10).

Leakage occurs when the window of a recorded signal perfectly matches the periodic cycle of the signal. The result is that a specific frequency component may leak into neighbouring components. Leakage can be reduced by multiplying the signal with a *Hanning Window* (see Figure 4.11).

A final factor that will influence the results of any spectral analysis is the window length, as this determines the step size between frequencies. The step size is the acquisition frequency divided by the window size. For example, a two-second epoch with a 500 Hz Acquisition Frequency results in $\frac{acquisition_freq}{epoch_length} * acquisition_freq = freq_step$, $\frac{500Hz}{2sec} * 500pt = 0.5Hz$ Steps. To avoid awkward frequency steps, epochs that are multiples of second lengths are recommended.

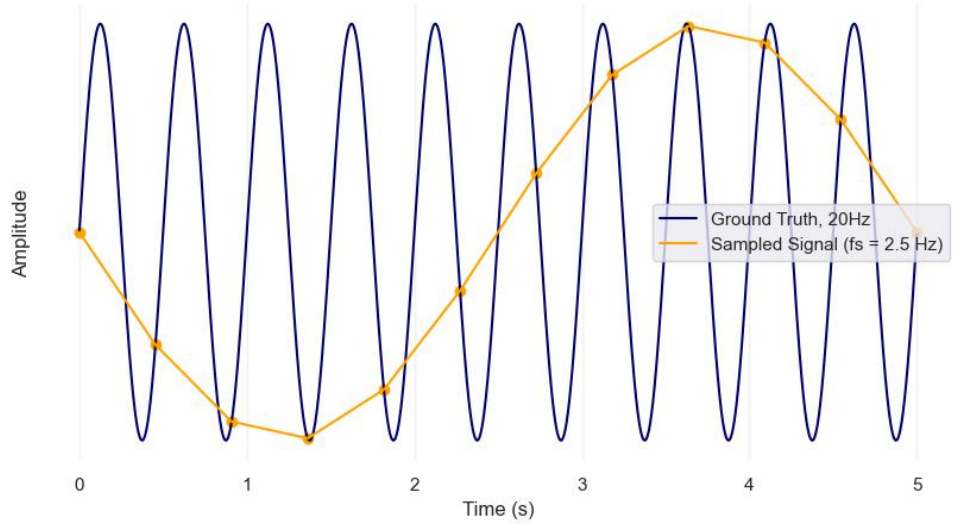

Fig. 4.10 Example of aliasing: The true frequency of the signal is made as the acquisition frequency is lower than the true signal.

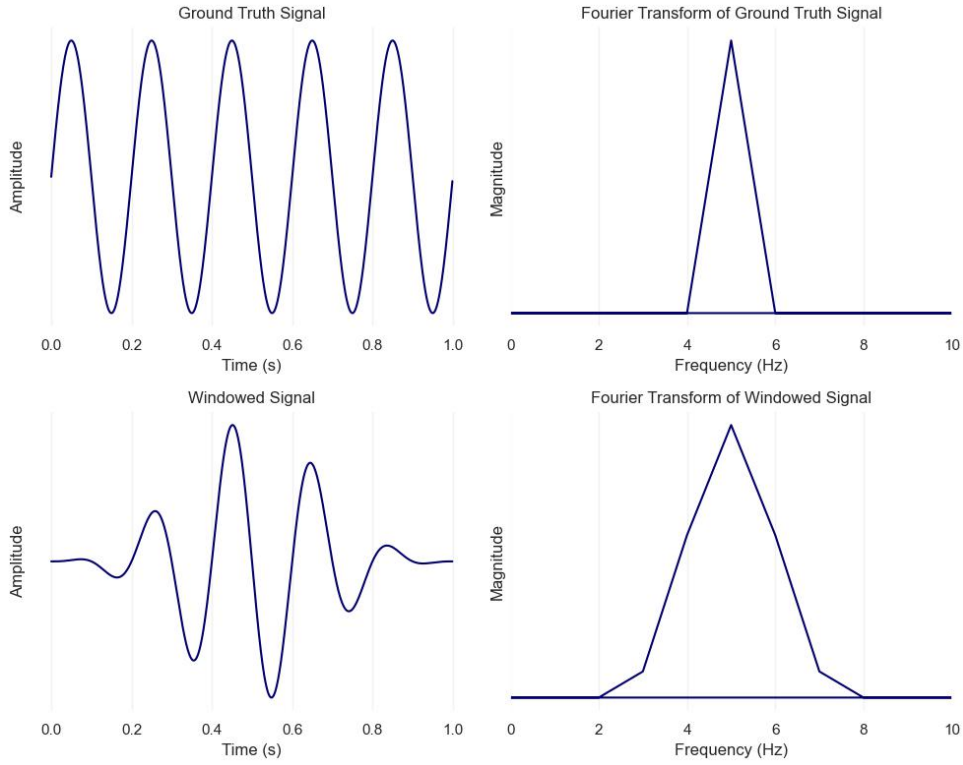

Fig. 4.11 Example of signal leakage: With an infinite signal (top left), the Fourier transform would only show the actual frequency (top right). However, as the signal is not infinite, we must apply a window function to the data (bottom left). The result is a slight misrepresentation of the data, as the frequency has leaked widened, bottom right).

I.4.2 ECG

Key Points

- The heart rhythmically produces electrical currents that contract the different muscles of the heart to pump blood through the body.
- Heart-Rate and Heart-Rate Variability metrics may inform on several mental and physiological states.
- Frequency components of the ECG signal are often associated with sympathetic and parasympathetic involvement.

With constant pumping, the heart produces comparatively strong electrical currents propagatings through the human body. The following subsections will detail how muscle activity can be used to infer something about mental processes. We will first detail the anatomy and mechanisms of the heart (Section I.4.2.1) and how the data can then be processed and analyzed (see Section I.4.2.2).

I.4.2.1 Mechanisms

The Heart

The heart rhythmically pumps oxygen-enriched blood from the lungs into general circulation and unoxygenated blood to the lungs to enrich it.

In most cases, the pumping process is initiated by so-called *pacemaker cells*, producing currents that spread through the heart, resulting in muscle contracting. These cells are located in the *sinoatrial node*. The current produced by the *pacemaker cells* propagates through the *right atrium (upper chamber)* and into the *left atrium*. As this happens, the muscles contract and pump blood. Next, the current reaches the *atrioventricular junction*, a relay connecting to the *ventricles (lower chambers)* via two bundles (one for the left and one for the right). The current propagation then results in the contraction of the *ventricles* (Goldberger & Gold-berger, 2013).

ECG electrodes can only record relatively large currents, like those produced by the heart muscle pumping, rather than those originating from the *sinus node* or *the atrioventricular junction*.

An ECG can pick up cardiac activity by measuring voltage differences. The *myocardial muscle cells* of the heart carry a negative charge inside them, with a positive charge just outside of the cell. when stimulated, these muscles depolarize and become positively charged on the inside and negatively charged on the outside. When the heart muscle contracts, the *myocardial cells* contract successively. An electric current is created at the junction between the contracted, depolarized cells that carry a negative charge on the outside and the resting cells with their positive outside charge. This is the electrical current that an ECG can measure (for a detailed explanation, please refer to Section I.4.1.1). An entire heartbeat follows the *PQRST* sequence. The contraction of the *atria* is represented as a P wave, and the ventriclal stimulation is captured by the *QRS* complex. The *S* segment and *T* wave then represent the return of the *ventricular muscle* to its resting state (Goldberger & Gold-berger, 2013).

As electrodes will only record electrical activity parallel to the line drawn between both electrodes, the positioning of electrodes influences what part of the cardiac activity is recorded

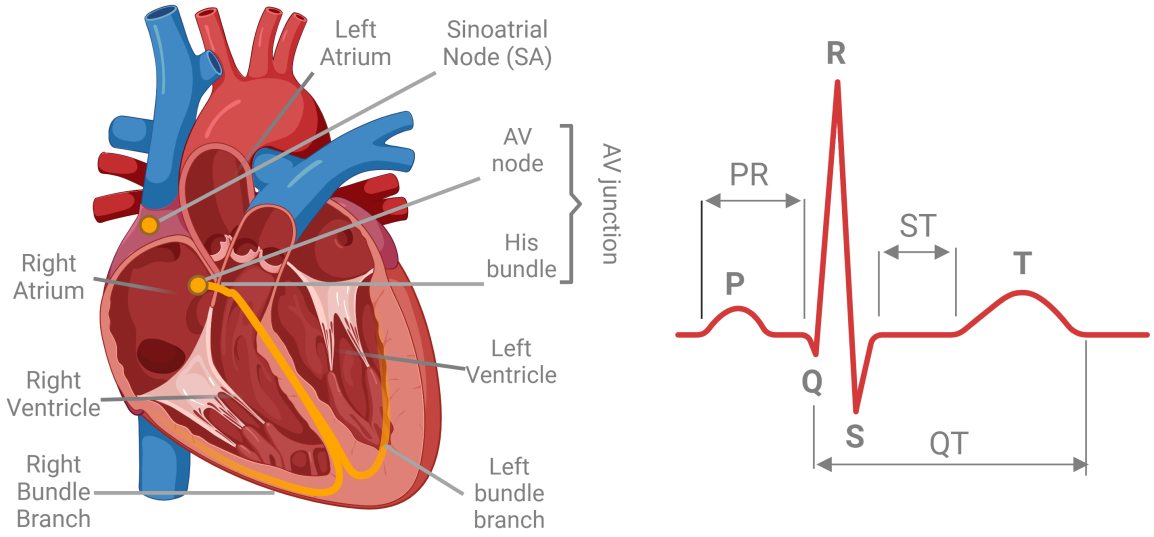

Fig. 4.12 The basic anatomy of the heart and the PQRST complex. Figure created using Biorender. (Picture Source: see Section V.14.1)

by the ECG. In non-clinical applications concerned with heart rate and heart rate variability metrics, the placement is less important, as heartbeats will be captured almost entirely irrespective of the electrode placement (see Figure 4.13).

I.4.2.2 Analysis & Processing

Preprocessing

Unlike EEG, the SNR ratio in an ECG signal is considerably stronger, reducing the need to preprocess the data before analysis. Still, extra physiological sources of noise may disturb the signal (see Section 4.1, Kaplan Berkaya et al., 2018). A common first step is to filter the data using a bandpass filter between 0.1-100 Hz (Goldberger & Gold-berger, 2013). The data may then be downsampled to reduce memory requirements or maintain consistency across databases.

Analysis

Heart-Rate and Heart-Rate Variability features The signal is transformed into RR intervals to compute heart rate (HR) and heart rate variability (HRV) features (see Table 4.2). These intervals describing the time differences between successive QRS complexes can be found using a

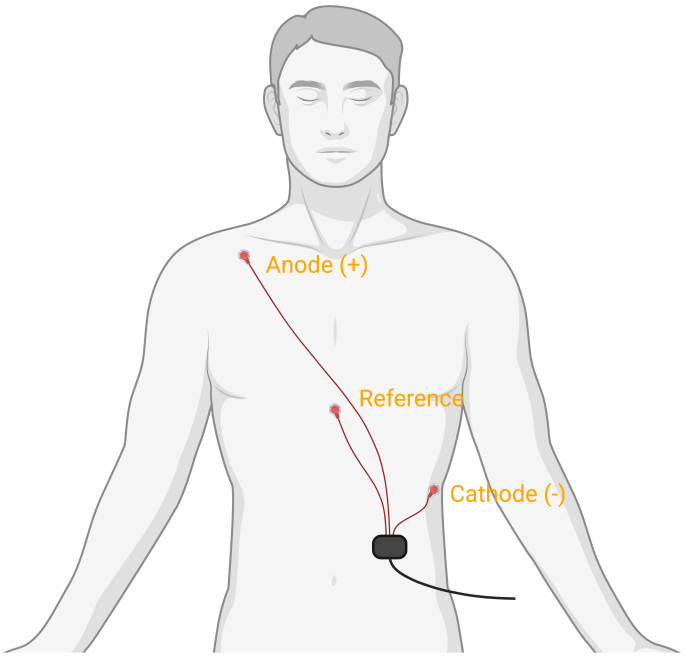

Fig. 4.13 A 3-electrode setup for a simple non-clinical ECG measurement. Many ECG systems do not even require a ground electrode and simply measure the voltage change between anode and cathode. Figure created using Biorender. (Picture Source: see Section V.14.1)

Domain	Metric	Explanation
Time Domain	Heart Rate (HR)	Beats per minute
Time Domain Variability	SDNN/SDRR	Standard deviations of the RR/NN interval
	RMSSD	Root mean square of successive RR interval differences
	pNN50	Percentage of adjacent NN intervals with more than 50ms difference
Frequency Domain	VLF	Very Low-Frequency Power (0-0.004 Hz)
	LF	Low-Frequency Power (0.04-0.15 Hz)
	HF	High-Frequency Power (0.15-0.4 Hz)
	LF/HF	Low to High Frequency Ratio

Tab. 4.2 Summary of usual ECG metrics.

peak-detection algorithm. Next, several metrics can be calculated. The heart rate describes the rate of heartbeats in a specific time interval, usually in beats per minute (BPM). Several metrics can represent HRV. The SDNN and SDRR are the simple standard deviations of RR and NN intervals. RR intervals are calculated from the QRS complex peak to the next QRS complex peak. NN intervals are the distance between two successive normal heartbeats. In normal ECG, both metrics measure almost the same thing (Malik, 1996). Only if abnormal beats are removed do the metrics differ. In most cases, 5 minutes of ECG must be recorded for a reliable SDNN estimate. There are two more common metrics for HRV estimation. HRV may be calculated as the Rrot mean square of successive RR interval differences (RMSSD). For the RMSSD, a 5-minute interval is considered the gold standard, but ultra-short periods of 10 seconds of the RMSSD have been used for mental state estimation (Salahuddin et al., 2007). Finally, the pNN50 calculates the percentage of adjacent NN intervals that differ from each other by more than 50 ms. Although the RMSSD and the pNN50 are strongly correlated, the RMSSD is considered the more robust measure. For a full overview of many HRV metrics, please refer to the work of Shaffer and Ginsberg (2017).

Frequency metrics Like in EEG, the ECG signal can also be analyzed by estimating the frequency components (see Table 4.2). Here, the ECG signal is transformed into the frequency domain before extracting band powers of several frequency bands and other metrics. The signal is cut into three frequency bands: The very-low-frequency band (VLF, 0-0.004 Hz), the low-frequency band (LF, 0.04-0.15 Hz) and the high-frequency band (HF, 0.15-0.4 Hz). From these, the absolute and relative powers and the LF/HF power ratio can be computed.

The VLF band is recorded over a 5-minute minimum period, and low power within this band is associated with arrhythmic death (Bigger et al., 1992) and Post-Traumatic Stress Disorder (Shah et al., 2013). The LF band can be recorded over 2 minutes, and power in this band is generated by activity within both the parasympathetic nervous system (PNS) and sympathetic nervous system (SNS). The HF band is also called the respiratory band and is conventionally recorded over a minimum 1-minute period. Power in this band reflects parasympathetic activity and HR variations related to the respiratory cycle (HR increases during inspiration and slows down when expiring). Low power in the HF bands is associated with stress, panic, anxiety or worry.

The LF/HF ratio estimates the ratio between SS and PNS activity (Shaffer et al., 2014). The idea is that LF power is generated by the SNS and HF by the PNS. A low value reflects parasympathetic dominance or tend-and-befriend behaviours. When the LF/HF ratio is high, sympathetic activity and fight-or-flight behaviours are observed. This notion has been criticized for several reasons. The LF power is influenced by both PNS and SNS activity. Furthermore, respiration may be a large confounder of the LF/HF ratio (Billman, 2013).

I.4.3 Eye Tracking

Key Points

- Eyes have several distinctive movement types that can be measured using ET:
  - Saccades are rapid movements towards objects of interest.
  - Fixations occur when focusing on an object without little movement.
  - Smooth Pursuit is the following of a moving object.
  - Pupil dilation is the result of environmental and physiological factors.

The eye is a fascinating organ. Here, we will explore the anatomy of the eye (Section I.4.3.1) and how it inevitably results in the need for eye movements (Section I.4.3.1). Next, some methods for measuring eye movements (Section I.4.3.2) and the different types of eye movements (Section I.4.3.3) and metrics will be detailed.

I.4.3.1 Mechanisms

The Eye

The eyes consist of several gross anatomical components. The *Sclera* is the white part of the eye, with the colourful *iris* and the *pupil* in the middle. The *pupil* is the eye's opening that allows light to enter the *eyeball* and stimulate the *retina* at the posterior end of the *eyeball*. Both the *iris* and *pupil* are covered by a transparent surface that fades into the *sclera* called the *cornea* (Fig. 4.14). Three pairs of muscles generate eye movements, and information from the retina is carried to the base of the brain via the *optic nerve* (Pouget, 2022a). The *pupil* can contract and dilate using two opposing muscles: the *radial dilator muscle* for opening the pupil and the *constrictor muscle* for closing (Mathôt, 2022).

When looking into the eye, we can see the *retina*. The largest visual feature of the *retina* is the *optic disk* or *blind spot*, where the nerve fibres of the eye exit the *retina*. In the middle of the *retina*, we find a big spot around 4 mm in diameter, called the *fovea*. The *fovea* is responsible for detailed vision due to a difference in *photoreceptor* distribution. In most of the *retina*, so-called *rods* are predominant. These *photoreceptors* do not perceive colour but are extremely light-sensitive. The *fovea* only contains slower, more precise cones, allowing colour vision. *Cones* are also easily fatigued when exposed to the same stimulation (the same light pigments stimulating the same cells) (Bear et al., 2007). The *fovea* covers less than 1% of the surface of the *retina*, while 50% of the *visual cortex* is dedicated to the processing of information from the *fovea*.

Why we move our eyes

Understanding the anatomy of the eye already allows some conclusions as to why eye movements are important. For detailed vision, we must direct our eye to allow the *fovea* to capture the light emitted by the object of interest. Eye movements then allow us to keep said objects stationary should there be motion in the visual scene or focus on different objects entirely. To have our eyes converge on close or far objects, we also need them to move inward or out-

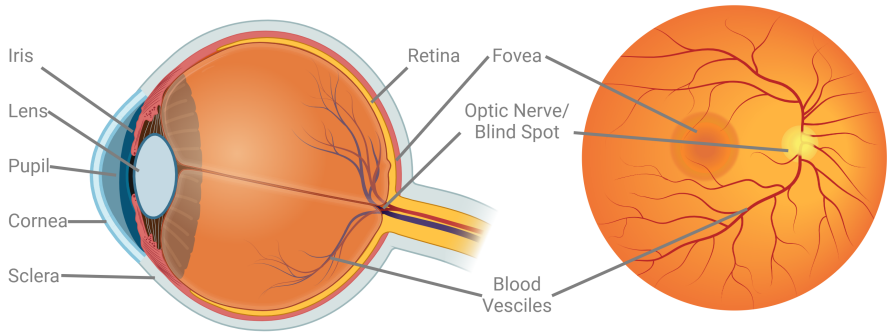

Fig. 4.14 The basic anatomy of the eye and the retina. Figure created using Biorender. (Picture Source: see Appendix V.14.1)

ward (towards or away from the nose). Finally, the eyes can rotate along the visual plane and compensate for some head movements (Land, 1999).

I.4.3.2 How to measure movements

The interest in eye movements has resulted in continuously developing techniques to measure eye movements.

Scleral Search Coils

are specialized contact lenses with a wire coil on the outside of the lens (see Figure 4.15). By placing the participant with the lens into a magnetic field, any eye movement will create a current by induction. This current can then be measured and used to monitor eye movements. This more than 60-year-old technique still offers very high temporal and spatial resolutions. However the technique is very discomforting to the participant (Hutton, 2022; Pouget, 2022a).

Electrooculography (EOG)

EOG uses the differences in currents between the retina and the more anterior eye sections to record some relative degree of movement(see Figure 4.15). EOG measurements only allow for relative measurements of eye movements (Pouget, 2022a), but can nevertheless provide useful basic metrics and are useful for reducing noise in an EEG recording.



Fig. 4.15 Examples of some methods for measuring Eye movements: Far Left: Scleral Search Coil: a Lens with electrical wiring that allows for highly precise, if uncomfortable, measurements. Left: EOG electrodes that can measure horizontal and vertical movement. Right: Pupil Labs Eye-tracking glasses for mobile 3D gaze estimation. Far Right: GazePoint GP3 Eyetracking bar for gaze estimation on computer screens. Figure created using Biorender. (Picture Source: see Appendix V.14.1)

Video-based Systems

These constitute the most popular form of eye-tracking (ET). By using an infrared camera directed at the eye, modern eye-trackers can obtain a relatively good spatial and temporal resolution (Pouget, 2022a). The ET shining infrared light allows the camera to capture two landmarks in the eye (see Figure 4.15): The corneal reflection point and the pupil. The internal algorithms can use these two points to determine gaze (Hutton, 2022). Both computer-mounted and head-mounted systems for ET exist. Generally, the head-mounted systems are less precise (Hutton, 2022) (see Figure 4.15).

Gaze Estimation Most models either use 2D or 3D estimation. The biggest difference between eye trackers with 3D and 2D estimation is that some of the 3D eye trackers require multiple cameras. This produces higher accuracies compared to 2D models (Kar & Corcoran, 2017), but is computationally more expensive. All video-based eye trackers require some form of calibration; a common way to do this is with a 5-point calibration, where participants are instructed to look at 5 points distributed across a computer screen. Many companies that produce eye trackers do not publish their algorithms that estimate gaze from the infrared video feed.

I.4.3.3 Types of eye movements & other metrics

Saccades

Saccades are rapid eye movements that follow a ballistic speed trajectory with a very high velocity of up to 100 deg/sec. These movements are very short and usually take anywhere between 30-120 ms. Once a saccade is initiated, it cannot be modified (in the case of a moving target, for example). As these movements are extremely fast, vision is suppressed during the movement. Saccades aim to orient the fovea towards important objects in the visual field. They interact with fixations (see Section I.4.3.3) to allow for exploration of the visual field (Pouget, 2022b).

The number of saccades, their velocity, amplitude and peak amplitude are some common metrics used to research cognitive states such as mental fatigue (Di Stasi et al., 2013; Skaramagkas et al., 2023).

Fixations & fixation related movements

Fixations are almost motionless periods between 100-1000 ms (mostly between 200-600 ms) that occur at an approximate frequency of 3 Hz (Pouget, 2022a). Fixations allow the collection and detection of detailed information. Common fixation metrics include the number of fixations, duration, fixation density and repeat fixations (Skaramagkas et al., 2023). During fixations, small movements are necessary to avoid fading due to the cones fatiguing (see Section I.4.3.1, Martinez-Conde, 2022). These movements are micro-saccades, tremors and drifts. Micro-saccades are similar to normal saccades, just on a smaller scale. Research has even shown that both movements are triggered by the same neuronal circuits, supporting the notion that a continuum between saccades and microsaccades exists (Martinez-Conde et al., 2013; Otero-Millan et al., 2008). Unlike microsaccades, tremors are involuntary shaking movements in the 50-100 Hz range (Ko et al., 2016; Pouget, 2022a). Occular drift is a random walk of the eye, first assumed to only occur at very low speeds of up to 10 degrees per second, with more recent evidence proposing average speeds of around 50 deg/second (Ko et al., 2016; Rucci & Poletti, 2015).

Smooth Pursuit

During smooth pursuit, the eyes track a moving object (Pouget, 2022a). These movements require a target and can have up to 100 deg/sec speeds. Despite the name, said movements are not smooth and include many corrective saccades to avoid fading (Lencer & Trillenberg, 2022).

Pupillometry

The dilation and contraction of the pupil occur not only in reaction to light, but pupil size has been associated with the luminosity of the attended stimulus (rather than the overall luminosity of the visual field), fatigue (Dawson et al., 2014; Hopstaken et al., 2016; Wang & Xu, 2016; Wu et al., 2015) and many other aspects of cognition. These cognitive effects on pupil size are partially explainable by the PNS and SNS governing the degree of dilation (Beatty et al., 2000). In a working memory task, pupils dilate more, with increasing memory load (Husta et al., 2019). Unfortunately, pupil size may be confounded by several aspects other than cognitive state and luminosity. Age, humidity and temperature can significantly affect pupil size (Skaramagkas et al., 2023). Pupil size can still be useful as a metric for SNS and PNS engagement and emotional, cognitive or sexual arousal (McIntire et al., 2014).

Eye blinks

Eye blinks are semi-involuntary actions where the eyelids are rapidly closed and reopened. The average duration is between 0.1 and 0.4 sec. Blink rate is influenced by factors such as humidity, temperature, brightness, physical activity (Ousler et al., 2008), but also emotional and cognitive processes (Skaramagkas et al., 2023). From eye blinks, several metrics such as blink rate, blink velocity, and blink duration may be extracted (Borghini et al., 2014; Di Stasi et al., 2013; Joseph & Murugesh, 2020; Maffei & Angrilli, 2019; Roy et al., 2016).

Other Eye Movements & metrics

Two other types of movements are noteworthy besides the eye movements discussed earlier: Torsional Eye Movements, which are movements of the eye around the viewing axis to compensate for head movements, as well as Vergence Movements which allow for the focus to shift from near to far objects (Pouget, 2022a).

Chapter 5

Algorithms

I.5.1 Machine Learning

Central Ideas

- Machine learning covers all types of algorithms that allow computers to learn and generalize from data
- Approaches are often categorized into supervised, unsupervised and reinforcement learning

One question to ask: What type of algorithm is best suited for physiological data and prediction of human performance?

Machine learning is the field of study interested in equipping computers with the ability to learn automatically (Sarker, 2021). A more formal definition:

> *A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E.* - Mitchell (1997)

All types of Machine Learning algorithms share a common goal: to generalize beyond the initially shown data, also known as the *training set* (Domingos, 2012). Machine learning may be subdivided into four domains based on the data these algorithms are fed and how they treat it (see Figure 5.1).

Supervised machine learning can be said to learn a *function* that maps some *input* to *output*. To do so, the algorithm is shown some pairs of matched *input* and *output*, commonly referred to as *labels* (Han et al., 2012). The goal of these algorithms is to predict said labels.

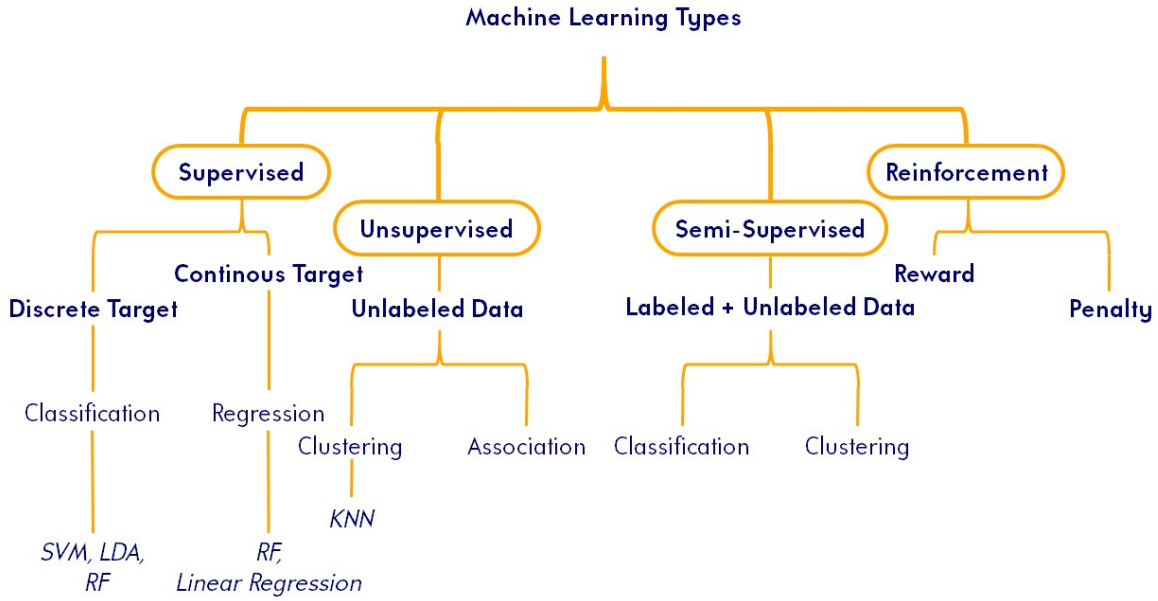

Fig. 5.1 Taxonomy of different types of machine learning. Based on Sarker (2021).

If the *output data* is continuous, the algorithm is called a *regression algorithm*. Regression algorithms are, for example, used to predict stock prices (Cakra & Distiawan Trisedya, 2015). *Classification algorithms* are utilized for discrete output data. They can be used to determine if cell clusters on X-rays are cancerous or not (Rezaei, 2021) or in Brain-Computer Interfaces (BCI) to determine e.g. if a person is paying attention or not, based on EEG data (Belo et al., 2021). Classification algorithms themselves again come in several different flavours: In *binary classification*, there are only two labels; in *multi-class classification*, there are more than two; *multi-label classication* is used when each input data entry is associated with several labels (Sarker, 2021).

While supervised algorithms need data with labels, *unsupervised algorithms* aim to show trends and structures among data without any labels. Unsupervised Machine learning is, for example, useful in political and social research to identify groups of voters (Waggoner, 2020). This is done using clustering, where data points are grouped based on commonality. The most popular algorithm for this is the K-Means clustering algorithm. This algorithm groups data points closest to a shared mean (Cam & Neyman, 1967). Unsupervised machine learning is not limited to clustering. Association algorithms aim to uncover relationships between variables. This is, for example, useful in detecting patterns of consumer purchasing behaviour. While clustering concerns finding similar data points, association algorithms are more interested in the relationships between dataset variables. Association Rule mining, for example, determines how often two items co-occur in a dataset, how likely this happens, and the degree of dependency between the variables (Sarker, 2021).

A third group of machine learning algorithms are so-called *semi-supervised algorithms* (Sarker, 2021). This marriage of supervised and unsupervised algorithms is particularly useful when very little labelled data is available, but unlabelled data is abundant. Text classification is a good example for this (Duarte & Berton, 2023).

Input 1	Input 2	Input n	Labels
13,5	'high'	...	1
87.6	'high'	...	0
nan	'low'	...	1
-26	'high'	...	1
0.25	'low'	...	0

Tab. 5.1 Example of how data could be structured, several input columns with different data types and missing values. As well as one label column with all the output.

Finally, *Reinforcement Learning algorithms* optimize a certain behaviour. Algorithms are rewarded or punished based on a certain observed outcome (Sarker, 2021). These algorithms are roughly divided into model-based algorithms, which aim to minimize the error between an algorithm's prediction of the environment and reality. Model-free algorithms, on the other hand, aim to maximize the reward obtained. These algorithms may be used to find the optimal strategy in a boardgame (Xenou et al., 2019).

The upcoming sections will detail all the necessary steps for creating a supervised machine learning pipeline (see Section I.5.1.1), as well as some of the common issues in machine learning (see Section I.5.1.2).

I.5.1.1 Creating a supervised machine learning pipeline

The section highlights the main aspects and steps for creating and using a supervised classification pipeline to analyze new and unseen data.

Data & labels

As mentioned in the previous section, a supervised machine-learning algorithm needs labels and training data. Data entries are commonly represented as a vector of features, and the label, in the case of classification, is assigned a class (see table 5.1 Domingos, 2012).

Preprocessing

Data is rarely perfect and requires some form of cleaning and preprocessing before it can be used in a machine-learning algorithm. In physiological data, this can mean using filters and other preprocessing algorithms (see Sections I.4.1.3 & I.4.2.2). Other aspects include formatting, data interpolation and outlier rejection as detailed below:

- It is important to handle the format of data. As an example, postal codes can be transformed into *strings* or categories (e.g. as a predictor of socioeconomic status) and not treated as actual numbers (Boardman & Robert, 2000).

- Missing data is another aspect that needs to be addressed during machine learning (Sarker, 2021). Entries may be deleted, interpolated or replaced if some data is missing.

- Finally, outliers pose a large risk to machine learning algorithms, as these are often quite sensitive to skewed data (Sarker, 2021). Outliers can be rejected, or the data can be transformed so that outliers are less extreme (see data clamping section I.4.1.3).

Feature Selection

From the clean data, the appropriate features can then be selected. An algorithm needs to be fed interesting features to predict some labels efficiently. Prior knowledge and data-driven approaches can be used to do this. Analysis of Variance (ANOVA) is a powerful statistical method that can give general insight into the relationship between features and labels (Sarker, 2021). Alternatively, correlations can also show relationships between features and labels. Feature Selection can reduce the dimensionality of the data (the number of features that the algorithm has to learn), making the algorithm faster and more precise.

Feature Extraction

Feature extraction further serves the goal of dimensionality reduction. Here, we want to create new features from the ones we have (Sarker, 2021). Feature extraction can be as easy as taking the average of several measurements or merging multiple strongly correlated features. Another popular approach is to use Principal Component Analysis (PCA). This dimensionality reduction method transforms correlated variables into their uncorrelated components (Sarker, 2021).

Classification

The following section will highlight some popular classifiers for physiological computing, including the Linear Discriminant Analysis, the Support Vector Machine and the Random Forest classifier. Deep neural networks and other methods, such as logistic regression, are popular but are beyond the scope of this manuscript and not used in the contributions. For an overview, see Sarker (2021) and Singh, Hussain, et al. (2021).

Linear Discriminant Analysis (LDA) is a simple linear classifier that assumes the classes have a Gaussian density and equal covariances between classes. The algorithm is similar to an ANOVA or a linear regression (Sarker, 2021). The probability of any point belonging to any class can be calculated by calculating the mean and distribution of each class. The decision boundary runs along the points where the probability of belonging to each class is equal (see Figure 5.2). LDAs are quick and perform well if the assumption of equal covariances is not violated.

SVM Support Vector Machines (SVM) are some of the most popular algorithms in machine learning. One of their strengths is their capacity to deal well with high dimensional data (Maglogiannis, 2007). The algorithm works by constructing a hyperplane between the data classes that maximizes the distance between the hyperplane and both classes (see Figure 5.3). The data points of each class closest to the hyperplane are called the support vectors, giving the classifier its name. Therefore, the algorithm performs poorly if the target classes overlap (Sarker, 2021). The SVM classifier is linear, but the data can be projected into a higher dimensional space where a linear hyperplane can be constructed; this allows for using the classifier in both a linear and non-linear fashion.

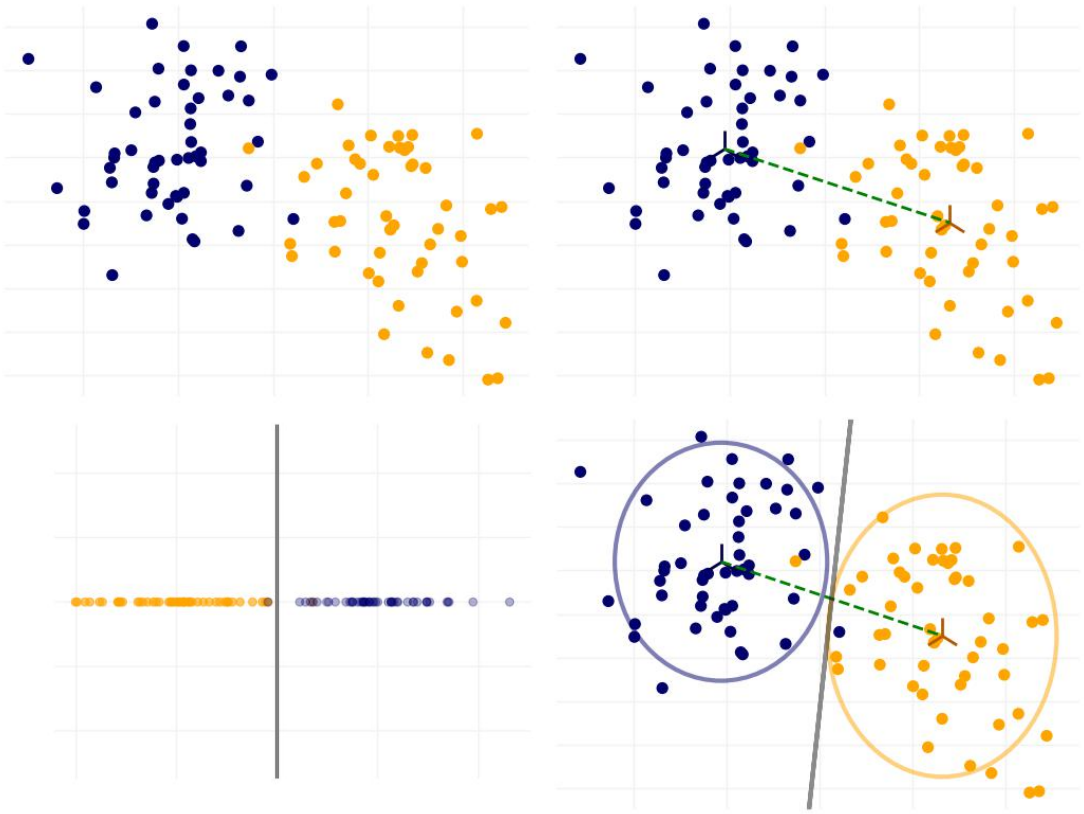

Fig. 5.2 Visual intuition of an LDA: (1) Datapoints of two classes are plotted into a 2-dimensional coordinate system. Each dimension represents a feature; (2) The mean (tripoint), standard deviations (ovals) and the projection line between the two means (green dotted line) are calculated; (3) The data is projected along the projection line, which now runs along the X-axis. The position of each point is mapped onto this line and represents an estimated probability that the data belongs to either class 1 or class 2. At equal probabilities, a decision boundary (grey line) is calculated; (4) The data can now be projected into its original form. Figures created using Python, with the Seaborne, Matplotlib and Sklearn Toolboxes.

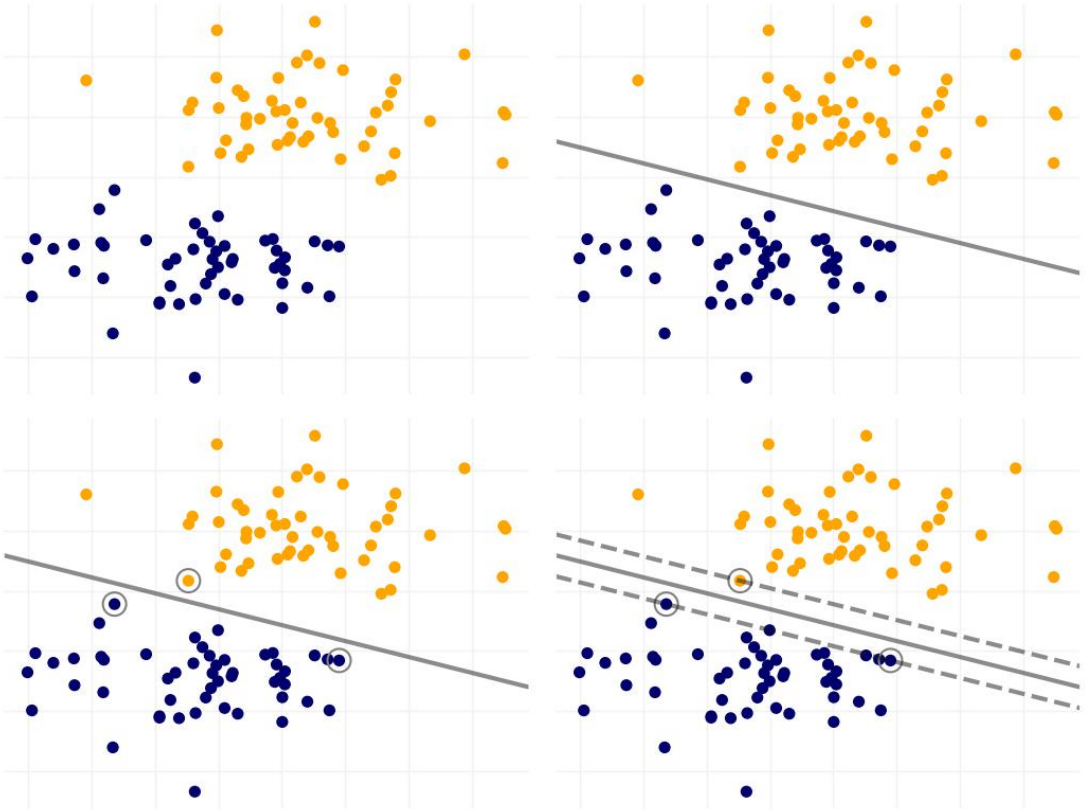

Fig. 5.3 Visual intuition of a linear SVM: (1) Data is plotted with each class being represented by a different colour; (2,3) A separating hyperplane is found (grey line) that maximizes the margin of the hyperplane to the closest data points (support vectors, grey circles). (4) We can plot the parallel lines running through the support vectors from these points. The support vectors are the only data points influencing the SVM. Figures created using Python, with the Seaborne, Matplotlib and Sklearn Toolboxes.

Decision Trees & Random Forests Decision trees construct a set of rules to determine the class of some data. The result is a flowchart (or tree) with each branch representing a data-based decision (see Figure 5.4). Decision trees tend to overfit data (see Section I.5.1.2). To avoid this, the random forest classifier merges many trees, each trained on a random subset of the data and takes the average vote of each tree as a result. The popularity of this algorithm is due to its ability to handle large datasets and outperform simple decision trees (Srivastava et al., 2023). One major disadvantage of random forest classifiers is that their models are difficult to interpret, and it can be difficult to understand why a certain model makes certain decisions. To mitigate this, Born-Again trees, based on the outcome of a random forest classifier, have been proposed (Vidal & Schiffer, 2020). These algorithms use the information of a trained Random Forest to increase explainability and to improve understandability.

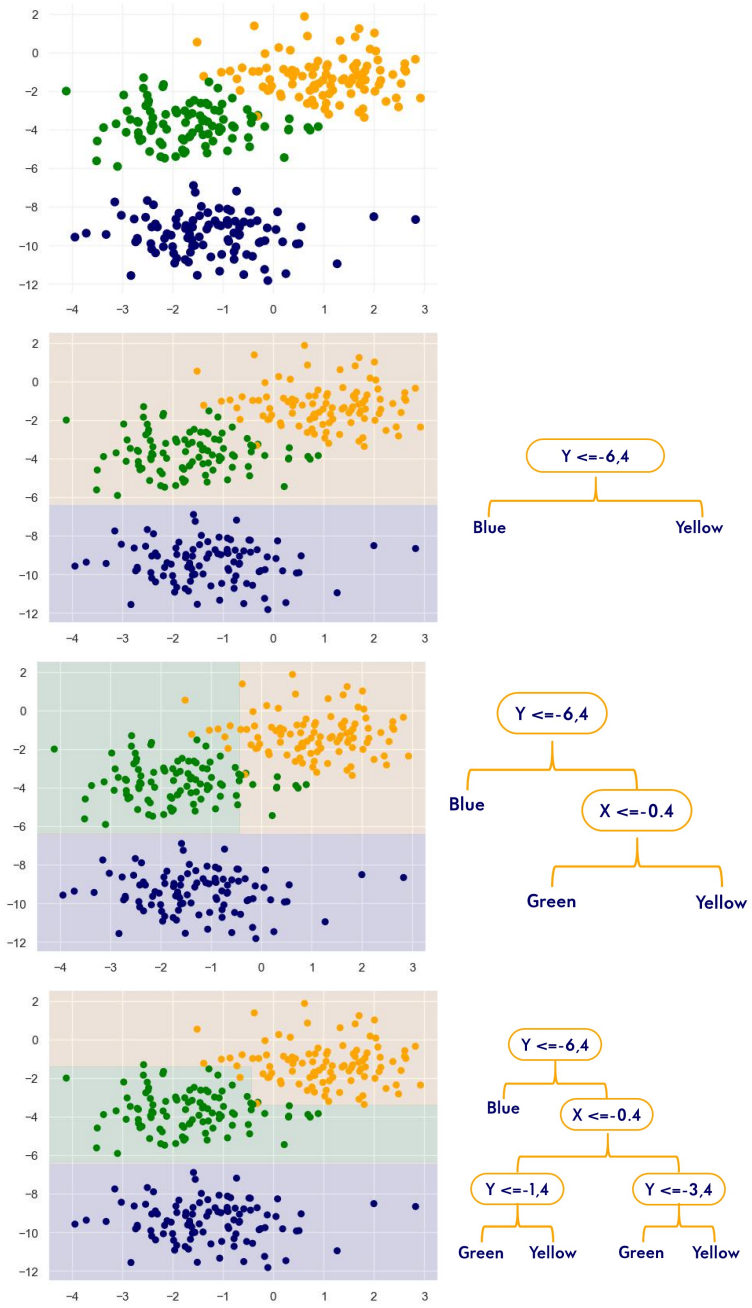

Fig. 5.4 Visual Intuition of a Decision Tree classifier. (1) First row: A random dataset with three clusters is created (2) We fit a decision tree with only one decision branch. The classifier is capable of distinguishing the blue data from the other but cannot separate the classes due to the limited decision nods; (3) Adding a branch allows for a crude separation. (4) The third branch is necessary to get more accuracy. Figures created using Python, with the Seaborne, Matplotlib and Sklearn Toolboxes.

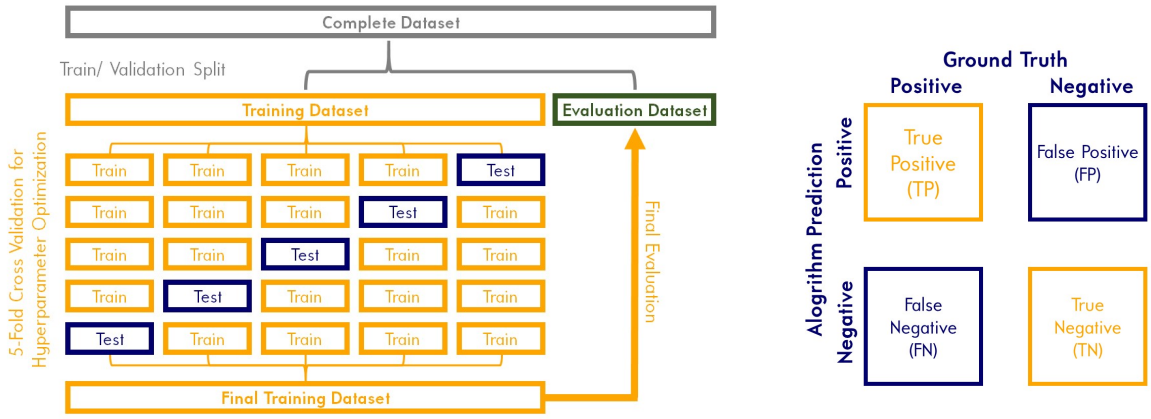

Fig. 5.5 Right: Example how data may be split for testing, training and cross-validation; Left: Representation of the different outcomes of predictions.

Tuning

Many classifiers have one or more hyperparmeters[1]. The parameters for the machine learning algorithm must be set manually, as they are not learned from the training data. Examples of hyperparameters are the number of trees in a random forest, the number of branches in a tree algorithm, which kernel to use in an SVM or the degree of regularization. Setting or tuning these hyperparameters can be done manually, but in many cases, a search algorithm may be more adept at finding the optimal combination of parameters (Wu et al., 2019).

Grid-Search is an algorithm that tests all possible combinations of the given hyperparameters by training and evaluating a model for each combination. This can be very time-consuming in large models, or with many parameters. As all possible combinations are tested, it offers the best results (Wu et al., 2019).

Bayesian Optimization algorithms iterate over many combinations of hyperparameters, choosing the combinations based on previous results. As the algorithm does not evaluate all possibilities, it is faster and better suited for models with many hyperparameters (Wu et al., 2019).

Evaluation

In classification, each prediction is one of four cases (see Figure 5.5). To illustrate, we can use the example of an algorithm stating whether cancerous cells are shown on an X-ray picture. If cancerous cells are on the image and the algorithm detects them, this is a true positive (TP). If the algorithm does not detect the presence of cancerous cells and there are none, the case is called a true negative (TN). Should the algorithm not detect the presence of cancerous cells, it's called a false negative (FN). Finally, detecting cancerous cells without any cancerous cells is called a false positive (FP) (Fananas & Soria, 2024)[2].

[1] Hyperparameters are not limited to classifiers but can also refer to the frequencies used in filtering or other steps within a machine learning pipeline.

[2] Part of the here presented literature and knowledge were gathered by Miguel Ángel Fañanás & Xavier Soria; two ISAE-SUPAERO students who, under the supervision of Raphaëlle N. Roy and Marcel F. Hinss conducted their

The most recognized and used evaluation metric in machine learning is accuracy (see Equation 5.1), the ratio of correct predictions to overall predictions. If the number of classes is unbalanced, this metric no longer accurately represents the performance of a specific algorithm. In this case, it may provide very encouraging results, despite the algorithm not outperforming the chance level.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{5.1}$$

Precision is the ratio of true positive predictions to all positive predictions (see Equation 5.2). A high precision score means that the model had very few false positives. In cancer detection based on X-rays, this would mean that when the algorithm stated that an X-ray showed a cancerous growth, the probability of it being cancerous was very high (Srivastava et al., 2023; Tohka & van Gils, 2021).

$$Precision = \frac{TP}{TP + FP} \tag{5.2}$$

Recall or sensitivity, on the other hand, measures the rate of true positive predictions to the overall number of positive predictions (see Equation 5.3). A high recall indicates very few false negatives. This would mean that the algorithm never 'overlooked' cancerous X-rays for cancer detection.

$$Recall = \frac{TP}{TP + FN} \tag{5.3}$$

The F1 score is an alternative to accuracy for evaluating the performance of a model (see equation 5.4). This score incorporates both precision and recall. The advantage of the F1 score is its robustness to unbalanced datasets.

$$F1 = 2 \times \frac{TP}{2 \times TP + FP + FN} = \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}} \tag{5.4}$$

When evaluating the performance of an algorithm, there is an inclination to state that any accuracy (if appropriate) above the chance level (in a balanced binary classification = 50%) shows that the model can classify between the two classes. However, this does not take statistical modelling into account. A confidence interval should be established around the chance level to be statistically significant. Only accuracies above this level should be considered to outperform a random model significantly. To do so, the appropriate adjusted chance levels, based on trial numbers, should be reported (Combrisson & Jerbi, 2015; Mueller-Putz et al., 2008).

I.5.1.2 Issues in Machine Learning

Overfitting

The goal of machine learning algorithms is to generalize beyond the data that was seen before. When, for example, a classification algorithm performs very well on the training set but fails to perform above the chance level on unseen data, the model is overfit. This can be understood as the algorithm detecting patterns when there are none (Domingos, 2012). Certain algorithms, such as decision trees, are more prone to overfitting (see Section I.5.1.1). Overfitting can be left undetected if the entire training dataset is used to train and evaluate the algorithm. A remedy for this is K-Fold cross-validation (see Figure 5.5, Wu et al., 2019). This procedure

12-month research project investigating the use of Eye-Tracking and Cardiac measures for fatigue detection and performance estimation.

subdivides the training dataset into K equally large subsets. A model is then trained on all but one training subsets and evaluated on the remaining subset. The performance is recorded, and the procedure is repeated K times with each of the K subsets used as a testing set once. The average performance of all folds then shows how well the model can generalise to new data (James et al., 2017).

Many algorithms also allow for regularization. This technique prevents algorithms from overfitting by introducing penalties if the classifier fits the parameters too close to the training data. It thus encourages simpler models (Ying, 2019).

Leakage

Somewhat related to the issue of overfitting is the danger of leakage. Suppose the data on which the algorithm is trained and evaluated are not correctly separated. In that case, the performance of the final evaluation may be inflated, as the algorithm has already seen the testing data. Again, cross-validation can protect from this (Domingos, 2012; James et al., 2017).

Curse of Dimensionality

The dimensionality of data refers to the number of features an algorithm is given to train on. As dimensionality increases, more training data is required as the training space (the number of possible combinations of features) increases exponentially. Some data with 2 features with ten different values each result in a training space of 100 different combinations. For each added feature (assuming they only have ten different possible values), the training space will increase by tenfold (Domingos, 2012). It is because of this that feature selection and extraction are so important. When the training data is limited, reducing the number of features can make algorithms much more effective (Sarker, 2021).

I.5.1.3 Conclusion

This Section explored some of the fundamentals of machine learning, focusing on supervised machine learning and classification. These techniques will be used to classify different mental states using EEG, eye-tracking and ECG features in the Contributions part and more specifically in Chapters 8, 9 & 10.

I.5.2 Physiological Computing

Central Ideas

- Physiological Computing and passive Brain-Computer Interfaces are valuable tools for mental state detection in Human-Machine Interaction.
- Performance estimation presents a powerful alternative to mental state estimation.
- Transfer learning is a promising solution for issues of moving brain-computer interface technology out of the lab.

One question to ask: **Can we use Physiological Computing or BCIs to estimate operators' performance in a useful and realistic way?**

In the movie 2001, A Space Odyssey, the computer and artificial Intelligence HAL 9000 is, among other things, capable of inferring the emotions of the astronauts aboard the spaceship (Kubrick, 1968). While not real, the ideas in Stanley Kubrick's movie somewhat anticipated the development of systems collecting covert data to infer something about the observed human. Today, the development of Physiological Computing Systems (PCS) is moving towards making this prediction a reality.

PCS collect covert physiological data to inform a computer/system or machine on the operator's state (Fairclough, 2009, 2017). The following sections will provide an overview of the differences and commonalities between PCS and Brain-Computer Interfaces (Section I.5.2.1), the issues of Mental State estimation & Performance estimation (Section I.5.2.2), metrics for mental fatigue and performance estimation (Section I.5.2.3), and finally the concept of transfer learning as a potential solution to tackle variability issues (Section I.5.2.4).

As proposed by Fairclough (2009), a physiological computing system consists of 3 main components:

- One or more sensors that capture physiological activity, such as EEG, ECG, Eye-Trackers, fNIRS, etc. (see Chapter 4).
- A system for analysis, this may be based on absolute (e.g. heart-rate above 120) or relative thresholds (e.g. heart rate increased by 20%) as well as categorization (e.g. electroencephalography (EEG) signal classified as reflecting a high mental workload using machine learning).
- A response by the adaptive system, for example, by adapting or simplifying the tasks or informing the operator on their mental state.

Physiological computing systems with their adaptive component are closely related to adaptive user interfaces (AUI). The three components proposed by Miraz et al. (2021) are very similar to those presented by Fairclough (2009). Small differences are that the referential component of an AUI may use non-physiological data, which is not the case in PCS. Moreover, a PCS does not necessarily need to adapt an interface as it may also provide feedback to the operator.

I.5.2.1 Physiological Computing Systems & Brain-Computer Interfaces

Key Points

- Physiological Computing Systems and passive Brain-Computer Interfaces aim to improve human-machine interaction by allowing the machine or the operator insight into the mental state of the operator.
- While BCIs will always measure cerebral activity, physiological computing may also be done with peripheral sensors such as electrocardiogram (ECG) and eye-tracking (ET).

In 1929, Hans Berger recorded the first EEG signal in humans (Berger, 1929). Seventy years later, at the end of the 20th century this technology has become a staple in research and medical diagnosis. It was about to be used in a radically new way.

Researchers demonstrated for one of the first times that the EEG signal could allow people who had lost all voluntary motor control to communicate with the world (Birbaumer et al., 1999). By voluntarily changing their brain activity, two patients could produce binary signals that could be converted to letters, words and sentences (Kübler, 2020). This, among other developments, led to Brain-Computer Interfaces, an entire discipline devoted to using the brain's activity to control, adapt or communicate through computers.

While physiological computing systems may also use data collected by sensors, not measuring the activity of the central nervous system, BCIs exclusively rely on brain activity (Lotte & Roy, 2019).

Today, most BCIs are classified as active, reactive or passive, as detailed below.

Active BCIs

In active BCIs, the participants voluntarily and overly perform mental activities such as imagining body movements or doing mental calculations (Khan et al., 2014). A BCI can translate the different types of activity or their absence and presence into instructions or answers (see Figure 5.6). This type of protocol has been revolutionary for people suffering from locked-in syndrome or other pathologies that result in the loss of all motor control. Active BCIs have allowed these people to regain the ability to interact with the world (Kübler, 2020).

Reactive BCIs

Reactive BCIs exploit the cerebral responses to visual or auditory stimuli. Here, participants do not need to do anything but pay attention to certain stimuli.

The iconic P300 speller probes for attention by having participants focus on a specific letter on a virtual keyboard. In either a random or predetermined pattern, single letters will flash up. The flash on an attended letter will elicit an event-related potential (ERP). This P300 component (a positive deflection, mostly found in peripheral electrodes and occurring around 300 ms after a stimulus) can then be detected in the EEG signal and linked to the specific letter (see Figure 5.7, Clerc et al., 2016).

Steady State visually evoked potentials (SS-VEP) and Code Visually Evoked Potential (Code-VEP) are two other reactive BCI protocols. Flickering an object in a particular frequency will increase the frequency's power in the occipital lobe (Ladouce & Dehais, 2024; Lim et al., 2015). By flickering different objects at different frequencies, algorithms can determine where the participant's visual attention is directed. Code VEP constitute a more recent development where the steady state flickers have been changed for codes or patterns of flashes that allow for a more reliable and faster transfer of information (Cabrera Castillos et al., 2023).

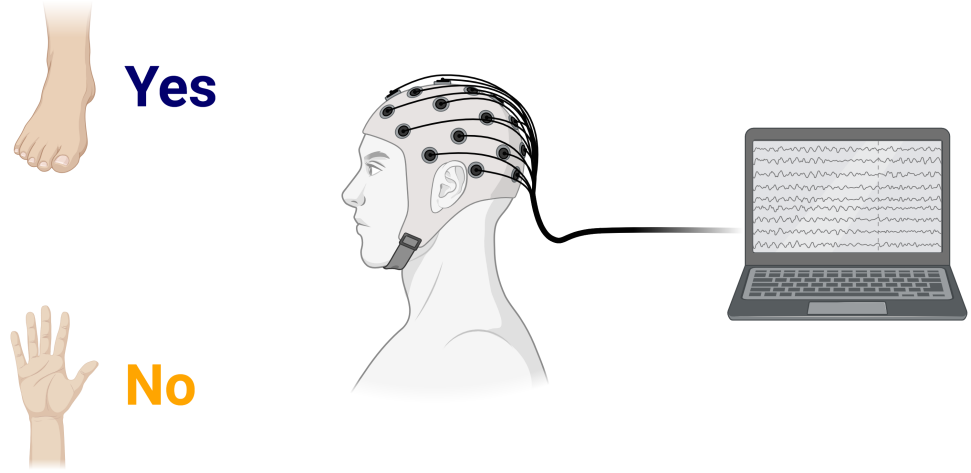

Fig. 5.6 Motor Imagery protocol: To Respond to a question, the participant has to either image foot or hand movements. Figure created using Biorender. (Picture Source: see Section V.14.1)

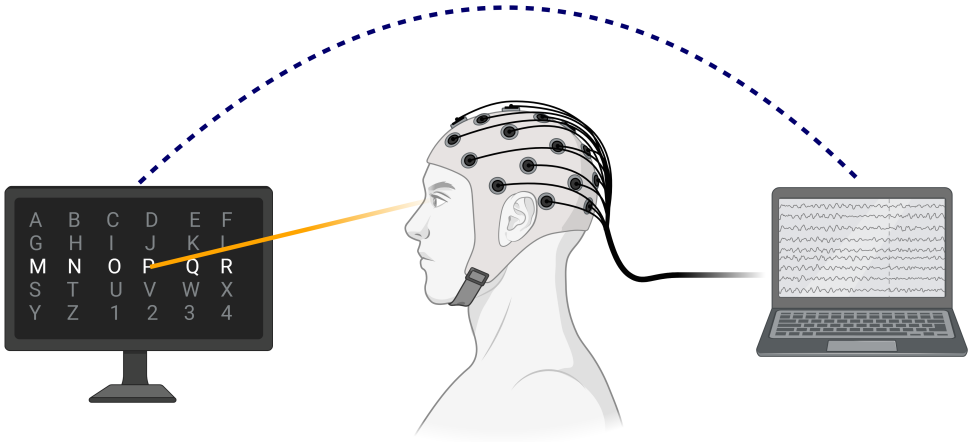

Fig. 5.7 P300 speller: The participant focuses on the letter they want to type. Randomly rows and columns of letters flash (indicated in white). The algorithm detects the P300 component when the attended letter flashes. Figure created using Biorender. (Picture Source: see Section V.14.1)

Passive BCIs

Passive BCIs (pBCI) aim to monitor a user's mental state and allow for feedback or adaptation to accommodate the current mental state of the user (Clerc et al., 2016; Lotte & Roy, 2019). Research has proven the feasibility of such systems, detecting fluctuations in mental workload (Aricò et al., 2017; Brouwer et al., 2012; Hogervorst et al., 2014), attention and fatigue (Sun et al., 2014). Krol, Laurens R. and Zander, Thorsten O. (2017) propose that passive BCIs should be further subdivided into four types of systems: mental state assessment, open-loop adaptation, closed-loop adaptation, and automated adaptation, as detailed below.

Mental State Assessment encompasses all systems for measuring a mental state. For example, this is useful when evaluating user interface designs (Frey et al., 2014).

Open-Loop Adaptation goes further and can give users feedback on their current mental state while interacting with an interface or doing a task (Zander et al., 2017). This could indicate workload or be a warning when certain thresholds are exceeded (see Figure 5.8).

Closed-loop adaptation constitutes the next step, where the task or interface dynamically adapts to the current mental state. This could mean adaptive automation as in (Aricò et al., 2016), adjusting the difficulty of a game to ensure optimal workload (Ewing et al., 2016). The defining difference between open and closed-loop systems is that closed-loop systems influence the next task iteration based on the current estimated mental state.

Automated Adaptation goes beyond closed-loop adaptation. These systems are capable of learning the behaviour of the user and can then simplify the interactivity. In a study by (Zander et al., 2014), error and satisfaction-related cerebral activity were used to teach a system a user's preferences in two-dimensional cursor control. Over time, the system could predict where the user wanted to move the cursor.

While this taxonomy shows the breadth of different types of pBCIs that exist, the differences between categories need to be carefully considered. For example, a system that adapts an interface automatically, that on the other hand, can also be controlled manually (as is proven to be advantageous for adaptive interfaces, see Section I.2.2.6) may be considered both open and closed loop.

Hybrid BCIs

In addition to active, reactive and pBCIs, it is important to mention hybrid BCIs. The term hybrid BCI is used for two types of systems (Lim et al., 2015). For simplicity, we will refer to them as Sensor-Hybrid-BCI and Activity-Hybrid-BCI.

Sensor-Hybrid-BCIs augment the information they obtain from the user by adding additional sensors (collecting either peripheral or cerebral data). This can be the use of Electrooculography (EOG, see Section I.4.3.2) electrodes to improve the EEG signal by removing artifacts or other sensors such as ECG and Eye-Tracking to provide better data (Lim et al., 2015).

On the other hand, **Activity-Hybrid-BCIs** use one single sensor to detect two types of cerebral activity simultaneously. For example, the data collected during an SS-VEP protocol may be used to measure workload simultaneously (Dehais et al., 2022).

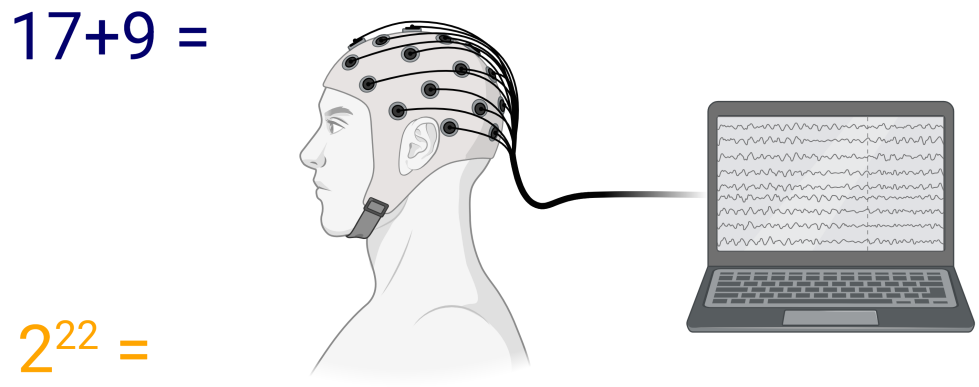

Fig. 5.8 Passive BCI protocol: Mental workload is estimated when the participant performs either easy or difficult mental calculations. $2^{22} = 4194304$ Figure created using Biorender. (Picture Source: see Section V.14.1)

Definitions of adaptive interfaces, pBCIs and physiological computing overlap at certain points but not others. A pBCI can be an adaptive interface, but it is not if it is only used to collect data. Adding an eye-tracker may turn an EEG-based system into a Sensor-Hybrid-BCI and a Physiological computing system. On the other hand, removing the EEG would result in the system no longer being a BCI. Furthermore, as discussed in Section I.2.2.2, definitions of AUIs vary. If an interface which provides feedback is considered an adaptive interface, in that case, the aforementioned distinctions of passive BCIs no longer hold (Zander et al., 2014). Open and closed loop systems, in this case, would be considered the same category (see Section I.5.2.1).

This state-of-the-art has revealed the plethora of definitions and taxonomies applicable to pBCIs and PCS and. In this work we will consider that BCIs and PCS are systems equivalent to the afferential and inferrential components of an AUI, with the adaptation itself being the efferential component (see Section I.2.2.2, Miraz et al., 2021)[3]. The absence of sensors directly measuring cerebral activity (e.g. EEG and fNIRS) renders a system a PCS, while a mixture of peripheral sensors and cerebral sensors means that a system constitues a Sensor-Hybrid-BCI[4].

Physiological computing systems face several challenges. PCS often assume a one-to-one relationship between the recorded data and the inferred underlying mental state or process.

[3]Often pBCI include an adaptive component, and therefore may be considered both a pBCI and an AUI. For consistency, we will refer to such systems as AUIs in this manuscript.

[4]Publications sometimes use the term Sensor-Hybrid-BCI (or Hybrid-BCI) for systems that collect cerebral data only, but with multiple sensors (e.g. EEG and fNIRS, Dehais et al. (2018)). For the sake of this taxonomy, we will group this type of BCI with the regular BCIs and limit Sensor-Hybrid-BCIs to systems collecting cerebral and peripheral data simultaneously

However, this relation may be one-to-many, many-to-one or many-to-many (Cacioppo & Tassinary, 1990).

There are also ethical concerns. If an interface adapts to the user's mental state, coworkers or other persons could perceive that change and infer the mental state of the operator (Fairclough, 2009). Another issue for PCS that needs to be considered is data handling following an incident. In commercial flights, all relevant data is recorded in the so-called flight data recorder, commonly known as the black box[5] (despite the box not being black). If incorporated in a future aircraft, the pilot's data recorded by a PCS could be used in an accident investigation. This may discourage pilots from using such a system for fear of being held accountable for their physiology.

I.5.2.2 Performance estimation vs Mental fatigue estimation

Both passive BCIs and Physiological Computing Systems often aim to detect certain mental states such as high mental workload or mental fatigue (see Section I.3.1, Aricò et al., 2017; Fairclough, 2009). The underlying argument for detecting said mental states is their correlation with erroneous or sub-optimal behaviour by the operator. Thus, by detecting, e.g. a high mental workload, the system may adapt itself to reduce workload and increase safety (Zander et al., 2017). This approach has certain drawbacks. Mental states as constructs are not observable and vary across definitions (Brouwer et al., 2015; Phillips, 2015). Furthermore, mental states depend on the current context and tasks. Differing task instructions can result in differing brain activity (Heitz, 2014). A high mental workload during an N-Back task may not be comparable to a high workload during a Stroop task (Kirchner, 1958; Stroop, 1935). Finally, mental states are not always strong predictors of performance. In the case of mental fatigue, evidence suggests that participants, using, e.g. compensatory strategies, can uphold performance despite fatigue (Hamann & Carstengerdes, 2023; Robert J. Hockey, 1997; van der Linden, 2011). So, all these aspects must be accounted for when constructing a mental state-based pBCI with a complex system (Krol, Laurens R. & Zander, Thorsten O., 2017).

For mental fatigue, time-on-task (TOT) has likely become the most popular way of attributing labels to training data (see Section I.3.1.3 & Figure 3.2). Following this approach, an hour-long continuous experiment may be segmented into four blocks of 15 minutes. The first 15 minutes may be labelled "non-fatigued", and either the last or second block could be labelled "fatigued". Studies using this or a similar approach have shown impressive classification accuracies of over 90% (see Figure 5.9, Huang et al., 2018; Shahbakhti et al., 2022; Shen et al., 2008).

Yet, we must ask if this approach has practical value for future real-world applications. Again, the underlying argument for using the TOT approach is that over time, mental fatigue increases, resulting in a decrease in performance. Therefore, classifying TOT can be used in a system to increase safety. However, we suggest that the argument presented above is flawed. There are two possible cases: 1) If the first statement were not true and mental fatigue would not increase over time, then using TOT labelling would have no practical value. 2) If mental fatigue does increase with TOT, then there is no need for a sophisticated physiological computing system, as TOT itself could be used directly to estimate the degree of fatigue.

Alternatively, a pBCI or PCS could forego the mental state aspect and try to predict a participant's behaviour directly. Performance estimation has been proven to work in several contexts and does not suffer from any of the aforementioned issues (Alhazmi et al., 2015; Blaha et al.,

[5] https://skybrary.aero/articles/flight-data-recorder-fdr

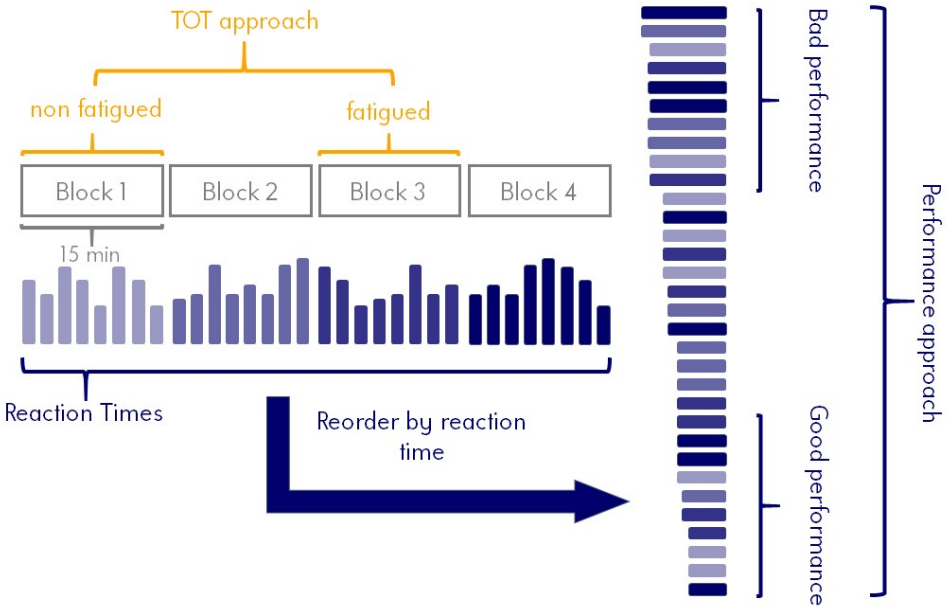

Fig. 5.9 Visual comparison of TOT-based labelling and performance labelling

2016; Ko et al., 2015; Müller et al., 2008). It works by assigning labels to the physiological data, using the recorded behavioural data of participants, such as reaction time, accuracies and misses (see Figure 5.9). Performance prediction allows direct observation of the variable we want to optimize with a pBCI, but also faces challenges. To predict performance, we first need to define good and bad performance.

Many tasks, such as the Stroop or N-back tasks, involve some measure of correctness and reaction time (Kirchner, 1958; Stroop, 1935). Reaction time or accuracy may be considered a valid performance measure in these cases, but only the combination into a global score will provide a complete picture of performance. Combining scores, on the other hand, raises questions about how to weigh each metric. Here, it needs to be considered that these measures are not orthogonal (Heitz, 2014). As mentioned above, the issue gets more complex as we move away from very controlled tasks and move towards more ecologically valid measures that may include several different reaction times and accuracies.

A related challenge is then how these cases should be labelled. Imagine, for example, an experiment where participants continuously perform a task for one hour. A global score of performance is assigned to each minute of the task. The value is continuous from 0 (bad) to 1 (good). How can the data now be divided into a 2-class problem? The ten worst minutes of performance versus the remaining 50 minutes, the ten best minutes versus the ten worst minutes, the good half versus the bad half, or values exceeding a threshold (e.g. >0.8) or subceeding another (e.g.<0.2) are all present plausible approaches. The metric calculation and label assignment issue is further complicated when algorithms are tested to classify the data. Does a chance level classification accuracy mean the algorithm doesn't work or the label assignment is sub-optimal?

In many cases, predicting any change in performance may be helpful, whether it is the likelihood of committing an error, missing a trial or the speed at which a participant responds.

Still, trying out unlimited combinations of labels may also create a global performance score not because of its usefulness but because of its ability to be classified.

One popular approach to performance estimation is the Error Related Negativity Event-Related Potential (ERP, see Section I.4.1.4 for an explanation of ERPs). This signal is a large negative deflection following an error of commission (Luck & Kappenman, 2011; Nieuwenhuis et al., 2001). While the ERN has proven useful in understanding errors, human factors and Brain-Computer Interfaces (Chavarriaga & Millan, 2010; Nieuwenhuis et al., 2001; Scheffers & Coles, 2000), for realistic implementations, some issues remain. The ERN occurs after an error is committed, so it cannot prevent operators from making an initial error. Nevertheless, where participants are more likely to commit errors following an error, rebound effects still provide some potential value to the ERN (Danielmeier & Ullsperger, 2011). The larger issue is that the ERN occurs following an error of commission. It will thus not be useful for errors of omission, such as missing instructions. Finally, the ERN has mostly been studied using well-defined tasks such as the Flanker task (Eriksen & Eriksen, 1974). To our knowledge, the degree to which ERNs are evoked in more ecological settings is unknown.

I.5.2.3 Metrics for mental fatigue detection

Key Points

- Physiological metrics for mental fatigue detection depend on the task, available sensors and labelling approach.
- EEG-based mental fatigue detection is often based on frequency metrics or Riemannian covariance metrics.
- Both eye-tracking and ECG give valuable mental fatigue metrics, as do sensors that measure posture and other metrics.

The difference between performance estimation and mental fatigue estimation is largely an issue of labelling. If we assume that errors occur due to mental fatigue in a given experiment, the metrics that can be used for fatigue estimation should also predict errors. The following section will explore some metrics used in various applications and domains.

In Section I.2.2.5, we proposed subdividing fatigue detection metrics into covariate/operator independent, behavioural and physiological (see Table 5.2). The following section will focus only on physiological metrics.

EEG

Frequency EEG-based detection of mental fatigue has, in large part, been based on using spectral features and the ratios between features (Borghini et al., 2014; Bose et al., 2020; Cao et al., 2016; Dimitrakopoulos et al., 2017; Fan et al., 2015; Huang et al., 2016; Jahanpour et al., 2020; Ko et al., 2016; Ko et al., 2015; Lee et al., 2014; Lorist et al., 2009; Roy et al., 2013; Roy et al., 2014a; Shahbakhti et al., 2022; Sheng et al., 2018; Ting et al., 2010; Tran et al., 2020; Trejo et al., 2015; Vijaya K et al., 2014; Wang et al., 2018; Wascher et al., 2014; Yu, Antonik, et al., 2021). The general trend within these frequency bands is to observe an increase in power of the low-frequency bands (Delta, Theta and Alpha) while seeing a decrease in amplitude in higher frequencies such as Gamma and Beta (Wascher et al., 2014). However, the opposite effects have

Type	Sensor/Task	Measure/Metric	Reference
Covariate	Time	TOT&TOD	Alhazmi et al. (2015), Rodríguez-Fernández et al. (2017), and Sikander and Anwar (2019)
Subjective	Scales	KSS & SPS & VAS	Kaida et al. (2006) and Samn and Perelli (1982)
Behavioral	Driving	Steering corrections	Alhazmi et al. (2015), Li, Li, Li, et al. (2017), and McDonald et al. (2012)
	Computer	Mouse and keyboard	Pimenta et al. (2013)
Physiological	EEG	Spectral Power	Borghini et al. (2014), Bose et al. (2020), Cao et al. (2016), Dimitrakopoulos et al. (2017), Fan et al. (2015), Huang et al. (2016), Jahanpour et al. (2020), Ko et al. (2016), Ko et al. (2015), Lee et al. (2014), Lorist et al. (2009), Roy et al. (2013), Roy et al. (2014a), Shahbakhti et al. (2022), Sheng et al. (2018), Ting et al. (2010), Tran et al. (2020), Trejo et al. (2015), Vijaya K et al. (2014), Wang et al. (2018), Wascher et al. (2014), and Yu, Antonik, et al. (2021)
		Frequency Ratios	Bose et al. (2020), Fan et al. (2015), Monseigne et al. (2019), Sheng et al. (2018), Wang et al. (2018), and Wascher et al. (2014)
		ERP	Hopstaken et al. (2016), Möckel et al. (2015), and Wang et al. (2016)
		Covariance	Hajinoroozi et al. (2017), Roy et al. (2014a), and Wriessnegger et al. (2021)
		Entropy	Fan et al. (2015) and Sheng et al. (2018)
		Connectivity	Sun et al. (2014) and Wang et al. (2018)
	Occular activity	Blinks	Borghini et al. (2014), Charbonnier et al. (2016), He (2013), Ko et al. (2020), Roy et al. (2014b), Shahbakhti et al. (2022), Wang and Xu (2016), and Wu et al. (2015)
		Saccades	Di Stasi et al. (2012, 2015, 2016) and Morris and Miller (1996)
		Fixation and Gaze	Ji et al. (2004), Leske and Dalal (2019), Roy et al. (2016), and Wang et al. (2016)
		Pupillometry	Hopstaken et al. (2016), Wang et al. (2016), and Wu et al. (2015)
	Cardiac activity	Heart rate	Ahn et al. (2016) and Lal and Craig (2001)
		Heart rate variability	Borghini et al. (2014), Huang et al. (2018), Li (2015), Melo et al. (2017), Ting et al. (2010), Wang et al. (2018), Wu et al. (2015), and Zhao et al. (2012)
	Other	fNIRS	Ahn et al. (2016) and Dehais et al. (2018)
		Body posture	Ansari et al. (2021), Furogori et al. (2005), and He (2013)
		Respiration	Lee et al. (2014) and Li (2015)

Tab. 5.2 Examples of metrics modulated by mental fatigue that can be used for adaptive system design. Note: To the best of our knowledge, no adaptive interfaces use subjective fatigue questionnaires to activate or deactivate an adaptive interface. Nevertheless, such a system is easily envisionable.

also been observed (Lal & Craig, 2001; Lal & Craig, 2002). The use of ratios such as $\frac{\alpha+\theta}{\beta}$ (Bose et al., 2020; Fan et al., 2015; Monseigne et al., 2019; Sheng et al., 2018; Wang et al., 2018), $\frac{\alpha+\theta}{\alpha+\beta}$ (Fan et al., 2015; Sheng et al., 2018), $\frac{\alpha}{\beta}$ (Fan et al., 2015; Sheng et al., 2018) and $\frac{\theta}{\beta}$ (Bose et al., 2020; Fan et al., 2015; Sheng et al., 2018) have all proven to detect mental fatigue. The major advantage of frequency-based MF detection is that it can be done continuously as it is event-independent.

ERP Some studies have used the characteristics of ERP components to detect mental fatigue (Hopstaken et al., 2016; Möckel et al., 2015; Wang et al., 2016). Hopstaken et al. (2015b) detected mental fatigue through a decrease in the P3 amplitude. However, this decrease may be caused by decreased attention due to increased mental fatigue, not directly by mental fatigue. As discussed in Section I.5.2.2, the ERN has also been proven useful for detecting errors (Chavarriaga & Millan, 2010).

While ERPs produce distinctive features in the EEG signal, some external event needs to occur for them to be triggered. This may occur naturally in complex operations, but the system relies on these events to infer the operator's state. If the interface elicits these events, they may be perceived as annoying. A promising opportunity for ERP-based mental fatigue and performance estimation would be using Activity-Hybrid-BCIs (see Section I.5.2.1), coupled with a P300 speller or a VEP (Dehais et al., 2022).

Covariance metrics - Riemannian approaches The covariance matrix of an EEG signal has developed into one of the most popular metrics in BCI research (Barachant et al., 2012; Bougrain et al., 2021; Congedo et al., 2017; Roy et al., 2022; Roy et al., 2014a). It has also proven to be capable of detecting changes in mental fatigue (Hajinoroozi et al., 2017; Wriessnegger et al., 2021). The success of covariance-based methods is in no small parts due to the discovery that projecting covariance matrices onto the Riemannian manifold allows for very accurate classification (Barachant et al., 2010). For a complete introduction, please see Congedo et al. (2017).

The intuition of this method is as follows: The covariance matrix of an EEG epoch contains information on the variability of each electrode and the correlation between two electrodes. High correlations are assumed to occur due to functional connectivity. Therefore, it also represents some measure of connectivity (Cansiz, 2023). Covariance matrices fulfill certain characteristics referred to as Semi-Positive Definite. This endows the matrix with certain properties, allowing it to be projected onto a curved subspace called the Riemannian Manifold. Here, every covariance matrix is represented by a single point on the manifold. The subspace is curved, so measuring the distance between two points on the manifold (two different covariance matrices) is more appropriate by measuring the distance along the manifold and not the straight line that goes through the manifold. This is the Riemannian distance or geodesic distance. This distance can be used to classify between classes directly, or the data may be projected into the tangent space (Congedo et al., 2017). The Tangent space is the immediate environment around a point on a manifold that appears flat, much like the earth appears flat when only looking at one's surroundings. In the tangent space, one can use any classifier like an LDA or an SVM (see Section I.5.1.1). The impact of this method on BCIs is hard to understate. For a review of the method and its use, please see (Congedo et al., 2017; Yger et al., 2017).

Other Beyond frequency, time, and covariance metrics, some studies have explored using other metrics to detect mental fatigue in the EEG signal. Both connectivity measures (Sun et

al., 2014; Wang et al., 2018) and entropy have shown promising results (Fan et al., 2015; Sheng et al., 2018).

Eye-Tracking

Eye tracking for mental fatigue detection is effective as eye movements seem susceptible to change under conditions of mental fatigue. Several studies have shown a decrease in saccade velocity, and in particular in peak-saccade velocity when mentally fatigued (Di Stasi et al., 2012, 2015, 2016; Morris & Miller, 1996).

While blinks have been used extensively, blink rate, blink duration, and Percentage Closed (PERCLOS) have all proven also to be sensitive to changes in mental fatigue (Borghini et al., 2014; Charbonnier et al., 2016; He, 2013; Ko et al., 2020; Roy et al., 2014b; Shahbakhti et al., 2022; Wang & Xu, 2016; Wu et al., 2015).

Gaze and Fixations have also characterised mental fatigue (Ji et al., 2004; Leske & Dalal, 2019; Roy et al., 2016; Wang et al., 2016).

Finally, decreasing pupil size due to prolonged task performance has also been used to detect mental fatigue (Hopstaken et al., 2016; Wang et al., 2016; Wu et al., 2015).

It should be noted that similar metrics are also used to measure mental workload, thus attention needs to be paid to the experimental protocol to ensure that mental fatigue rather than changes in mental workload are observed (for an overview of eye-tracking metrics in mental workload, estimation see Marquart et al., 2015 & Tao et al., 2019).

ECG

Heart-rate variability metrics, in particular, seem to provide the most robust ECG-based estimator of mental fatigue (Borghini et al., 2014; Huang et al., 2018; Li, 2015; Melo et al., 2017; Singh et al., 2022; Ting et al., 2010; Wang et al., 2018; Wu et al., 2015; Zhao et al., 2012). Certain studies have detected mental fatigue by monitoring heart rate changes (Ahn et al., 2016; Lal & Craig, 2001). In addition, Ting et al. (2010) also used frequency metrics (Low Frequency: LF & High Frequency: HF) to estimate optimal performance during a two-hour complex task experiment.

A recent review of Heart-rate variability (HRV) metrics as mental fatigue markers supports this, finding that the low-frequency component of the HRV was the most conclusive marker for mental fatigue, followed by the HRV metric RMSSD. Both LF and RMSSD are found to increase with mental fatigue (Csathó et al., 2024).

Other

Physiological metrics are not limited to EEG, ECG and Eye-Tracking. In actual flight conditions, Dehais et al. (2018) classified cognitive fatigue based on functional near-infrared spectroscopy (fNIRS; blood oxygenation measurement) and EEG connectivity measures. Ahn et al. (2016) showed that mental fatigue increased frontal oxygenation in a non-sleep deprived group. In a sleep deprived group TOT resulted in a decrease of frontal oxygenation. As with the other measures, combining several metrics also seems to improve estimation (Ahn et al., 2016). Furthermore, the researchers have also found that respiration rate (Lee et al., 2014; Li, 2015) and posture can determine TOT and mental fatigue during driving simulations (Ansari et al., 2021; Furogori et al., 2005; He, 2013).

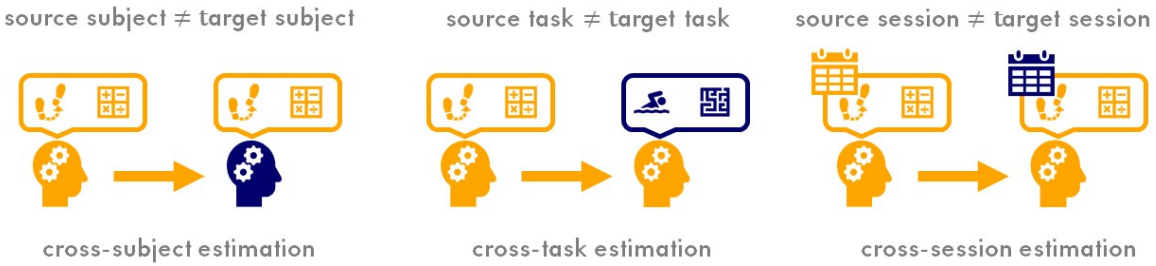

Fig. 5.10 Visual representation of cross-subject, cross-task and cross-session adaptation. Based on (Wan et al., 2021). Blue highlights the change in domain.

I.5.2.4 The variability issue and transfer learning

Most machine learning methods assume that the training and testing data are distributed equally. For example, we may train an algorithm to detect mental workload using the heart rate and HRV. If we train the algorithm on a normal person with a resting heart rate of 60 BPM, the algorithm might learn that an HR over 65 means the participant has a high workload. If we apply this algorithm to an athlete with a resting HR of 40, then we would most likely not find that the algorithm performs well at predicting the mental workload of the athlete.

Particularly, the EEG signal, as it is non-stationary, non-linear and non-Gaussian, suffers from this issue (Subha et al., 2010). This forces us to apply transfer learning if we want to apply knowledge from one person to another or from one person on one day to the next (see Figure 5.10). Additionally, electrodes may not be placed the same way across participants and days, and cleaner or dirtier hair may further influence the EEG signal. Studies have also shown that factors such as lifestyle, gender and age are associated with BCI performance (Ahn & Jun, 2015). Many BCI experiments train their algorithm immediately before testing it to avoid transfer learning. Mental fatigue has a large temporal component, so training a model immediately before testing it would inevitably mean that the participant is fatigued at the end of the training session. Even if we want to estimate performance, we need a sufficient training period because early errors committed during the experiment may be due to the novelty of the task and not the participant making errors.

Transfer learning is a domain within machine learning that aims to tackle this problem. This can be quite easy for univariate signals such as ECG, but multivariate, non-stationary signals such as EEG present a bigger challenge (Wan et al., 2021). The issue of transfer learning was one of the major motivations for creating the COG-BCI database, to allow for cross participant and cross-session mental state estimation using several different cognitive tasks (see Appendix V.14.7, Hinss, Jahanpour, Somon, et al., 2022a). The central idea of transfer learning is to apply algorithms that can transfer information across domains. Domains can be seen as spaces where a measurement of data occurred. Each session of each participant doing a specific task is a unique domain. Transfer learning is applying algorithms that allow the use of information obtained in one domain in another. The domain on which the algorithm is trained is called the source domain, and if testing is done on another domain, this is called the target domain (Liu et al., 2020; Wan et al., 2021).

Many approaches for transfer learning in EEG exist. Domain adaptation aims to change the data distribution in the target domain to more closely fit the source domain (Santana et al., 2019).

Another approach is using improved common spatial pattern algorithms (CSP). A CSP

builds spatial filters from frequency band power features (Blankertz et al., 2007). These filters maximize the variance in one class and minimize them in the other class. The variance can then be learnt as a feature for a machine learning algorithm. Improvements in this approach, such as regularization I.5.1.2 (Cheng et al., 2017) or using Bayesian statistics that assume that spatial patterns across subjects share the same subspace have allowed for CSP to be applied as a transfer learning technique.

Many approaches take advantage of subspace learning, which follows that the high-dimensional data of both the source and target domain can be represented in a low-dimensional subspace. The Riemannian manifold is an example of a subspace (see Section I.5.2.3).

Finally, special deep neural networks trained on one domain and feeding some target domain data can also allow transfer learning. Many transfer learning algorithms require feeding some target domain data to an algorithm. In the case of mental fatigue, this may be difficult again as these algorithms often require data from both labels (e.g., fatigued and non fatigued). While it is easy to assume that a participant is not fatigued at the beginning of the experiment, fatigued data would only show up after the point when we would like to use the transfer learning. A test of three different transfer learning algorithms and their performance on two types of labels can be found in Appendix V.14.6.

This state-of-the-art work has unveiled several shortcomings in the literature. Little to no solutions exist for mitigating or managing mental fatigue in UAS operations. The importance of cognitive flexibility and its interaction with mental fatigue is especially understudied. While there is an understanding that passive Brain-Computer Interfaces and Physiological Computing systems with an adaptive interface component are likely to support operators, these systems have not yet been implemented. Lastly, there is a need to investigate the usefulness of transfer learning for cross-subject mental fatigue estimation, which will be addressed in the various chapters of the Contributions part.

Part II
Contributions

Chapter 6

Double Task Switching: Cognitive flexibility and mental fatigue

Central Ideas

Similarity between tasks is an understudied factor in research on cognitive flexibility. This behavioural experiment had 31 participants perform a task switch paradigm in which participants were required to switch between 4 tasks of varying similarity. The experiment was constructed in a way that simultaneously allows for investigating the impact of mental fatigue and task-rule congruency on the participants. The results indicate that similarity between tasks substantially impacts performance with different effects on reaction time and accuracy. While learning effects may have negated the impact of mental fatigue across the 5 experimental blocks, a significant decrease in performance was observed within blocks. Furthermore, the exploratory analysis proposes a novel interaction between task-rule incongruent trials and the task of the previous trial. These results support the notion that neither the interference view of cognitive flexibility nor the reconfiguration view are fully adequate at explaining task switch costs if similarity is added as a factor. The presented study presents strong evidence that fundamental findings in the domain of cognitive flexibility may not map linearly to more ecological settings where tasks are often more dissimilar.
Recommended Reading:

- Cognitive flexibility (Section I.3.2)
- Effects of mental fatigue on cognitive flexibility (Section I.3.1.4)

One Question to ask: **Is switching between more dissimilar tasks more or less difficult, and what does this tell us about more ecological experimental protocols.**

II.6.1 Motivation and Research Goals/ Questions

As we have seen before, Uncrewed Aerial System (UAS) Operators have to routinely switch tasks (see Sections I.2.1.3 & I.2.1.5), which is rendered more difficult when operators are mentally fatigued (see Section I.3.1.4). Unfortunately, cognitive flexibility and task switching are poorly understood in ecological contexts. Fundamental protocols for measuring cognitive flexibility and task switching usually present the stimuli in the same field of vision to reduce eye movements that could confound any behavioural metrics. If tasks are in different areas, the reaction time during a switch trial may be partially attributable to the eye movement and not the task switching costs. Many studies use bivalent responses, where the same keys are used to respond to all tasks, and if that is not the case, then the tasks are constructed in a way that reduces motion by the participant. Finally, the task sets are often very similar, meaning the tasks do not differ much from each other (see Section I.3.2.2 3.3).

None of the above is the case in UAS Operations (UAS OPS, see Sections I.2.1 & I.2.1.5). Switching tasks requires looking at different screens, changing modalities (e.g. flightstick to trackball) and performing very different tasks.

While it is safe to assume that changing modalities or looking at a screen when switching tasks would only increase task-switching costs (Posner, 1980,TSC), the relation between TSC and task similarity is less understood (see Section I.3.2.3).

The here presented study was designed to probe this by constructing a protocol that tests the effect of task similarity in a fundamental protocol.

To do so, the "Double task-switching protocol" uses two pairs of two tasks (4 tasks in total) to test similarity. To account for effects due to eye movements, the stimuli were presented in the same visual field. To account for modality changes, the stimuli were bivalent, meaning that the responses by the participants were on the same keys irrespective of task.

In a realistic environment, operators can often perform any task at any time, even if instructed to do only one task. We wanted to account for this by investigating conflicts elicited by other tasks. The stimuli here had different degrees of task-rule congruency conflict to account for this. Stimuli were sometimes unambiguous with regard to what task the participant had to perform. Sometimes, the stimuli were associated with several tasks, but the correct response was the same irrespective of the task. In other cases, the stimuli were associated with several tasks, but the information was conflicting.

NOTE: For the remainder of this contribution, we will no longer use the terms bivalency and univalency for the sake of simplicity. In the experimental design, seven different levels of task-rule congruency were implemented.

- *Trials in which **only** the relevant information for the given task is presented (univalent) are **task-rule neutral trials**.*
- *Trials in which additional information is presented, but the information is in **line with the correct response** (bivalent) are **task-rule congruent trials**.*
- *Trials in which additional information does **not align with the correct response** (bivalent) are **task-rule incongruent trials**.*

This study asks:" Task-switching in realistic environments often involves switching between dissimilar tasks. Is this more or less difficult than switching between similar tasks?"

Analysis Group	Hypothesis	Expected effect
Task-Switch Costs	Following a switch of the task, RT on said trial is higher than the average RT on trials of the same task where previously no switch occurred	The effect of trial-type (switch vs repeat) is significant in the Mixed Multilevel Regression (higher RT for switches)
	Following a switch of the task, ACC on said trial is lower than the average ACC on trials of the same task where previously no switch occurred.	The effect of trial-type (switch vs repeat) is significant in the Mixed Multilevel Logistic Regression (lower ACC for switches)
Similarity effects	For all switch trials, RT will be longer if the switch occurs between dissimilar tasks	The interaction effect between trial-type (switch vs repeat) and similarity is significant in the Mixed Multilevel Regression (higher RT for dissimilar switches)
	For all switch trials, ACC will be lower if the switch occurs between dissimilar tasks	The interaction effect between trial-type (switch vs repeat) and similarity is significant in the Mixed Multilevel Logistic Regression (lower ACC for dissimilar switches)
Task-Rule Congruency effects	A task-rule incongruent stimulus will result in a higher RT regardless of whether the trial is a switch trial or a repetition trial.	The effect of task-rule congruency (task-rule congruent vs task-rule incongruent) is significant in the Mixed Multilevel Regression (higher RT for incongruent trials)
	A task-rule incongruent stimulus will result in a lower ACC regardless of whether the trial is a switch trial or a repetition trial.	The effect of trial-type (task-rule congruent vs task-rule incongruent) is significant in the Mixed Multilevel Logistic Regression (lower ACC for task-rule incongruent trials)
Mental fatigue	As TOT increases, the RTs will increase regardless of whether the trials are switch trials or repetition trials.	The covariate of TOT is significant in the Mixed Multilevel Regression (higher RT over time)
	As TOT increases, error likelihood will increase regardless of whether the trials are switch trials or repetition trials.	The covariate of TOT is significant in the Mixed Multilevel Logistic Regression (lower ACC over time)
	Following a switch of the task, RT on said trial is higher than the average RT on trials of the same task where previously no switch occurred	The interaction effect between trial-type (switch vs repeat) and TOT is significant in the Mixed Multilevel Regression (higher RT for switches as TOT increases)
	Following a switch of the task, ACC on said trial is lower than the average ACC on trials of the same task where previously no switch occurred.	The interaction effect between trial-type (switch vs repeat) and TOT is significant in the Mixed Multilevel Logistic Regression (lower ACC for switches as TOT increases)

Tab. 6.1 Summary of the expected results with corresponding hypotheses

II.6.2 Materials and methods

II.6.2.1 Ethics Statement

Ethical consent was obtained from the Comité d'Ethique de la Recherche—CER at the Université de Toulouse (ID 2022–521) on the 2nd of June 2022. Data collection occurred from January until April of 2023.

II.6.2.2 Design

The DTS protocol consists of four subtasks: two based on the numeric value of the displayed stimuli and two based on their visual appearance. The numeric subtasks consist of the LOW/HIGH and EVEN/ODD tasks (see Figure 6.3). For the LOW/HIGH task, participants must indicate, using a keystroke, whether a single number is higher or lower than 5. The number 5 is never present in the task, and all LOW/HIGH task stimuli are on the left. In the EVEN/ODD task, participants must decide if a stimulus (again, a single number, but located on the right side) is even or odd. Again, participants are instructed to respond with a keystroke. For the visual tasks, participants have to either perform the COLD/HOT task, where they have to decide

whether the stimulus is presented in either a hot (red/orange hue) or a cold colour (blue/green hue), or they have to perform the vertical/horizontal task. Here, a pattern's orientation must be determined (see Figure 6.2 for an example array of stimuli for each condition). Participants must perform consecutive trials, where the task and stimuli change quasi-independently from trial to trial. For each trial, participants are presented with 3 quasi-independent pieces of information. They are presented with an instruction for the current trial in the form of the task name (e.g. vertical/horizontal) 300 ms before the actual stimulus. Two stimuli then follow the instructions as signs or numbers that may vary in value, pattern and colour. Based on the stimuli and the instruction, participants are instructed to respond as fast as possible, using either the S or the L key on a keyboard (keys were selected as they are located at a comfortable distance on the keyboard to avoid straining the hands of the participants).

There are 168 different conditions for each trial (4 Tasks * 3 switch conditions * 7 Task-Rule Congruency Conditions * 2 Response types). However, as differences in the performance of individual tasks and response types are not hypothesized to impact participants' performance, 21 main conditions (3 switch conditions * 7 Task-Rule Congruency Conditions) remain.

Task switch conditions

Each trial (n) is classified into one of the three Task switch Conditions depending on the task in trial n-1. If the task remains the same, the trial is a repeat trial. Suppose the task switches but remains within the same domain (Numeric: LOW/HIGH to EVEN/ODD or Visual: vertical/horizontal to COLD/HOT and vice versa). In that case, it is considered an internal switch trial. Finally, switches from one domain into the other (LOW/HIGH or EVEN/ODD to either VERTICAL/ HORIZONTAL or COLD/HOT and vice versa) are considered external switch trials (see Figures 6.1 & 6.2).

	Response 1	Response 2	Neutral
Task 1 Low / High	1 2 3 4	6 7 8 9	$ % & €
Task 2 Even / Odd	2 4 6 8	1 3 7 9	$ % & €
Task 3 Vertical / Horizontal	4 4 4 4	4 4 4 4	4 4 4 4
Task 4 Cold / Warm	4 4 4 4	4 4 4 4	4 4 4 4

Fig. 6.1 Examples of different stimuli for all subtasks of the task. *Note: Task 3 Horizontal and Vertical may be difficult to see in this picture. Please refer to Figure 6.2.*



Fig. 6.2 Larger example of stimuli with cue relevant for Task 3. Left: vertical orientation. Right: horizontal orientation.

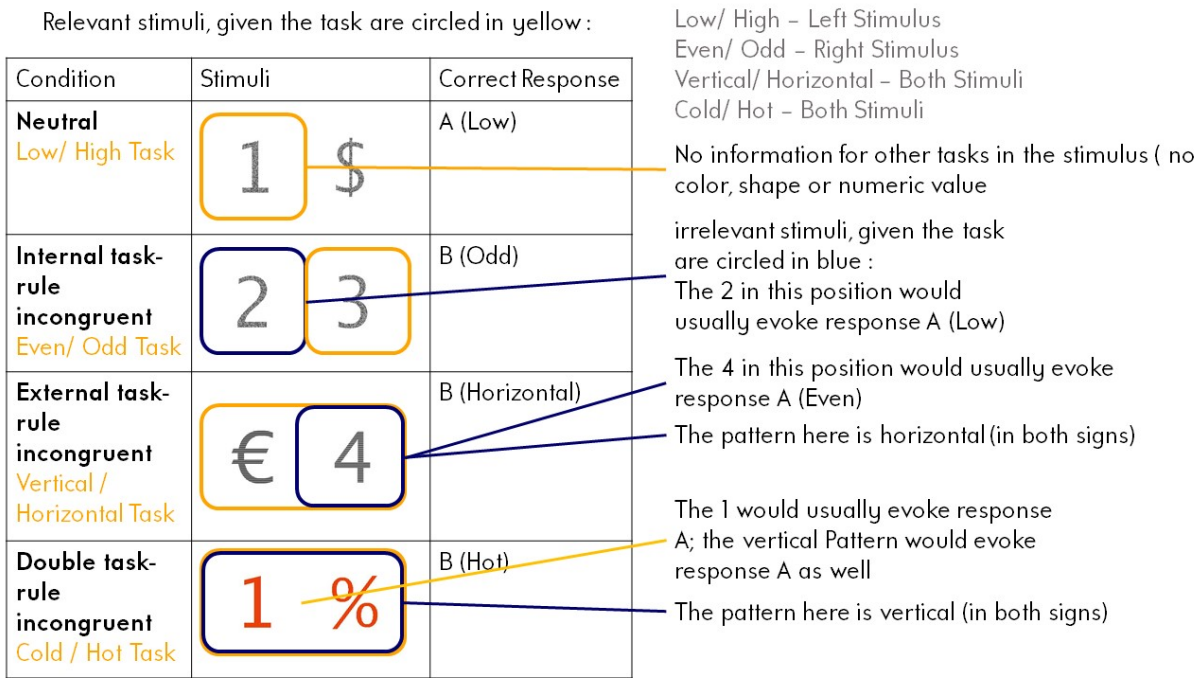

Fig. 6.3 Examples of the different task-rule congruency conditions.

Task-Rule Congruency Conditions

Depending on the presented stimuli, a trial may be more or less task-rule congruent. Task-rule congruency occurs when the task-relevant information is supplemented by task-irrelevant information that is biased towards the correct or the incorrect answer. 7 different conditions exist. In the task-rule neutral condition, the stimuli only present one piece of information, the relevant one. In Figure 6.3, the stimuli in the task-rule neutral condition provide no information pertinent to each of the 3 non-relevant tasks.

The *internal task-rule incongruent condition* adds information indicating the opposite response

(E.g. clicking L instead of S). This information is relevant for the other task of the same domain (the similar task). The information is therefore *internal* (same domain) and *task-rule incongruent* (bias towards the wrong response). In the example, the participant has to perform the even/odd task, which is always based on the right number. Therefore, 3 is the relevant number; the correct answer would be a click on the letter 'L'. The other stimulus (the number 2) is relevant only to the high/ low task and would click on the letter' S' if the participant were performing said task.

The *external task-rule incongruent condition* is similar; the only difference is that the non-relevant task is not from the same domain but is one (and always only one) of the tasks from the other domain.

The *double task-rule incongruent condition* combines the two previous conditions, where the other internal task and one (and always only one) of the external tasks bias the participant towards an incorrect response.

3 more conditions were added to the experiment. These mirror the task-rule incongruent conditions. However, the task-irrelevant information is congruent with the task at hand (Logan & Gordon, 2001). In the *external task-rule congruent condition*, this would, for example, mean that a participant would be instructed to do the low/high task and see a '3' as a left stimulus in a blue hue. The value 3 would mean that the number is low (corresponding to the letter 'S' on the keyboard), and the colour is cold (also corresponding to the letter 'S' on the keyboard).

Added stimuli will always be either task-rule congruent or task-rule incongruent.

During task-rule congruent trials, it is impossible to determine if a participant responded correctly due to the relevant task or the irrelevant information. Nevertheless, the inclusion of these conditions was necessary to be able to study the effect of congruency.

The locations of the relevant stimuli for the numeric tasks were chosen to be low/high on the left and even/odd on the right for the following reasons. Should participants perceive two single digits not as, e.g. seven and one but as seventy-one, this would not change the responses. As the even/odd relevant number is located (right or second), it determines whether a number is odd or even, independent from the number to the left of it (that is to say, both 1 and 71 are odd numbers). Similarly, the size of the first number is somewhat independent of the second number (7 is larger than 5, but 71 is also larger than any number in the fifties).

Trial Setup

Each trial begins with a fixation cross in the middle of the screen for 350-850 ms (600 ms with a 250 ms jitter). This is followed by trial instructions presented for 400 ms. These instructions give participants the current task in the form of the task's name (e.g. LOW/HIGH). Following a 300ms delay, participants are presented with the stimulus (350 ms), where they can respond with a keystroke. If participants take longer than 350 ms to respond, the stimulus disappears. Once a response is recorded, a new trial begins (see Figure 6.4). The task was computed using PsychToolBox-3 (http://psychtoolbox.org/) in Matlab and is presented on a desktop computer. Participants are not randomized (within-subjects design). However, the trials that participants will complete are quasi-random. That is to say, trials will appear in a random order, which ensures a balanced design and equal probabilities of each trial occurring.

II.6.2.3 Sample size

The approximate range of required participants was determined using GPower (Faul et al., 2009). The reported statistics by Arrington and Logan (2004) and Cooper et al. (2015) ranged

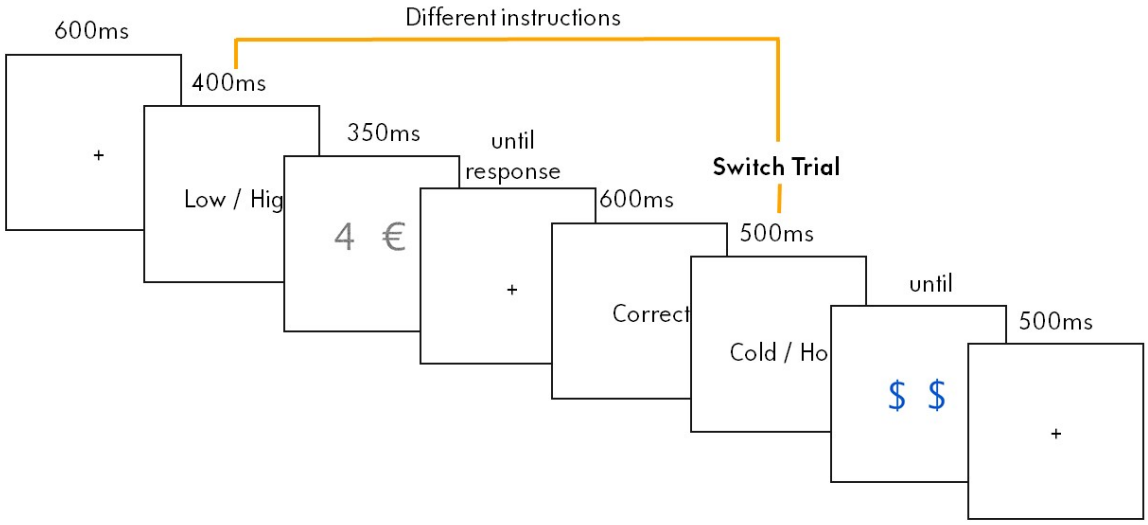

Fig. 6.4 Structure of a trial.

from 0.84 to 1.81 (f-squared). For a one-tailed test with an alpha of 0.05, the sample size required ranges from 15 to 25 participants. These numbers are only approximations, as the effects we intend to observe are not precisely the same as the ones reported in the aforementioned studies. From 31 Participants that were recruited, 2 participants had to be discarded for non-completion of the study. Hence, the final sample comprised 29 participants (17 females, mean age 23.9 years). The scores on the Edinburgh Handedness Questionnaire (EDI) range from 4 (right-handed) to 20 (left-handed). 25 participants had EDI scores indicating right-handedness (5-10), with three participants scoring as ambidextral (11-14) and one participant as left-handed (15-20). A visual representation of the demographics can be found in Figure 6.5.

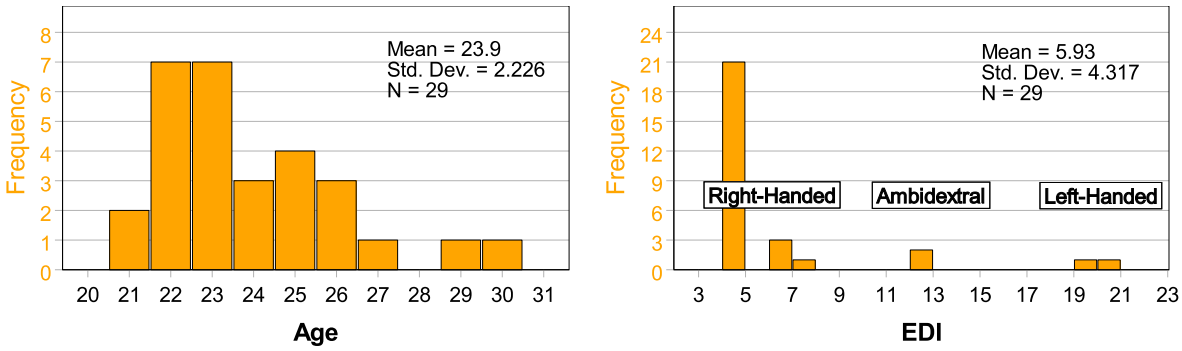

Fig. 6.5 Summary of demographics: Left: Distribution of age; Right: Distribution of handedness as assessed by the EDI (Veale, 2014).

II.6.2.4 Recruitment inclusion criteria and compensation

Participants were recruited by e-mail and via paper announcements posted in the buildings of ISAE-SUPAERO. Participants had to be between 18 and 30 years old, have obtained a level of study baccalaureate minimum, have normal or corrected-to-normal vision and hearing, be affiliated with social security, and have signed an informed consent form. Participants were prohibited from participating if they met any of the following criteria: Protected persons, presence of known neuropsychological disorders, significant visual or hearing impairment, current treatment with psychotropic medication or substances, positive g-test or being a nursing woman. In addition, due to the nature of the experiment, colourblind people were excluded from participation. As recommended to minimize the risk of an invasion of privacy, all criteria (inclusion and/or exclusion) were listed together in the inclusion criteria sheet, and potential participants were asked whether they met all the criteria rather than asking them whether they met each criterion one by one. Due to the nature of the experiment, participants were also tested for colourblindness before signing the inclusion/exclusion criteria. To do so, the Ishihara test for colourblindness was employed (Hardy et al., 1945). Participants were compensated in gift vouchers (Illicado, illicado.com). Upon completion, participants received vouchers worth 20€ (8€ for the first training session and 12€ for the more extended testing session). Participants were informed that dropping out after the first session would not impact their right to receive the 8€ voucher.

II.6.2.5 Procedure

After the participant's arrival for the first session, the participant was asked to complete the Ishihara colourblindness test before filling in the informed consent and information sheets. The participants were told they were permitted to ask questions at any time and were also asked in what language they would like to perform the task (French or English). They then filled in the Demographics questionnaire, the Edinburgh Handedness Questionnaire (EDI), St Mary's Hospital Sleep Questionnaire (QSN) and a KSS and the Samn-Perelli Fatigue (SPS) 7-point Likert scale (Samn & Perelli, 1982). Next, the participants were familiarized with the DTS protocol and performed a training session until they acquired adequate skills. For the training session, participants performed at least 5 blocks (96 trials each, 4 minutes with feedback, 3,5 minutes without feedback) of the DTS protocol. During the first and second blocks, participants received feedback following each trial to increase their learning rates. Subsequent blocks excluded the feedback to mimic the experimental conditions completely. Additional training blocks were added for some participants, ensuring adequate task performance. Additional training was added if the last completed block of trials showed significant improvement in RT over the preceding block. Extra blocks were only added with the explicit consent of the participant. After the training session, the participants filled in a KSS as well as an SPS scale (see Supplementary materials V.14.2.4 V.14.2.5).

The second session was scheduled no longer than 14 days after the first session. A second session was required to observe the effects of mental fatigue. This is necessary because if the entire experiment were performed in one session instead of two, we would expect participants to be slightly tired at the onset of the main task following the long training period. Would we reduce the length of the training period to reduce this effect, due to the complexity of the task, this might result in strong learning effects during the main experimental task.

The procedure of the second session was mostly similar to that of the first session. After filling in an QSN, participants performed a much shorter second training session intended to

help the participants remember the task. Participants completed half a training block of trials (48 trials) with feedback and another half without feedback. Next, participants filled in a KSS and an SPS scale and performed the experimental DTS protocol as detailed above. Participants completed 5 Blocks of 336 trials each. In between blocks, participants had a short break of 28 seconds, as proposed by Lim et al. (2016), to avoid recovery from the TOT effects. Upon completing the task, participants finished the experiment by responding to a KSS and an SPS scale, before being thanked for their participation and paid (See Figure 6.6).

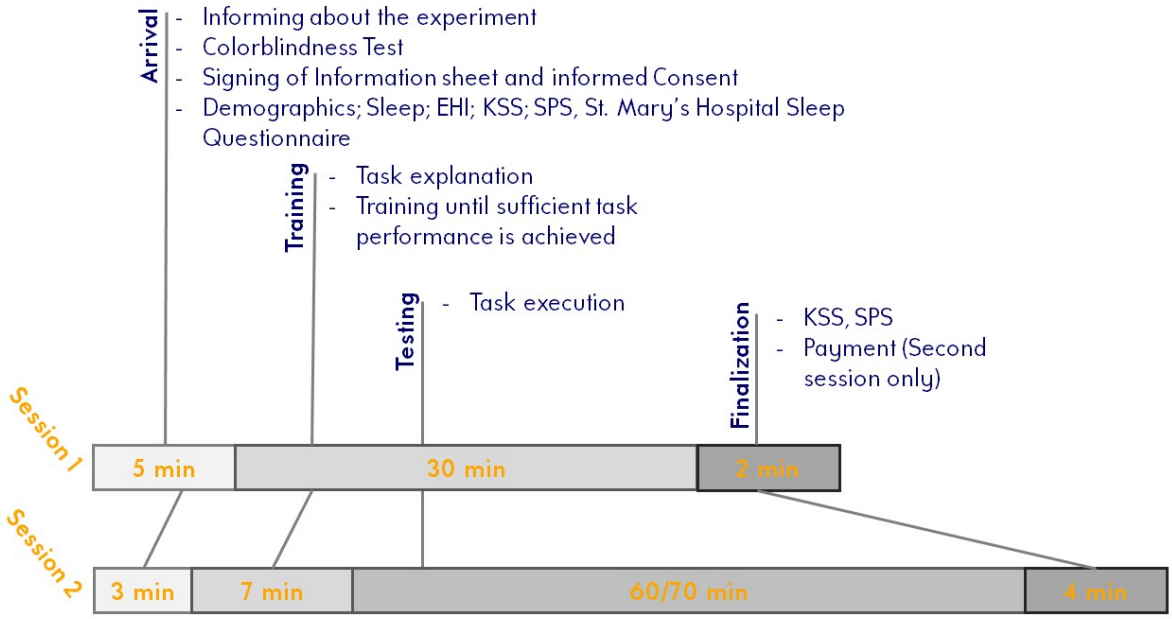

Fig. 6.6 Schema of the experimental procedure.

II.6.2.6 Variables

Independent Variables

The following stimuli were used in this study:

- Numeric Value: The value of the presented stimuli was manipulated. Participants saw combinations of the numbers 1,2,3,4,6,7,8 and 9 and symbols %, &, € and $.
- Color: The stimuli colour was manipulated to either be hot (red/orange hues), cold (blue/ green hues) or neutral (grey hues).
- Pattern: The pattern of the stimuli was manipulated and could either show a horizontally oriented pattern, a vertically oriented pattern, or a random pattern.
- Task: The task participants had to complete given the stimuli could be adapted, as seen in Figures 1 and 2. The four tasks were low/high, even/odd, cold/hot, and vertical/horizontal.
- Task-rule congruency: Depending on the stimuli, the experiment manipulated the degree of task-rule congruent/incongruent information presented. See Figure 6.3.

The following covariate was considered: TOT, *i.e.* how long the participant has been performing the task, was incorporated into the analysis as an indicator of mental fatigue in the form of the factor blocks.

Dependent Variables

The following measures were recorded:

- ACC: The number of correct answers during the task.
- RT: The RT on each trial. RT here was measured as the time from stimulus presentation to response (keystroke).
- Subjective Questionnaire Data: The data from several subjective questionnaires. All questionnaires were presented to the participants in a paperless digitalized form.

II.6.2.7 Questionnaires

Participants filled in several questionnaires, including a demographics questionnaire, the Edinburgh handedness questionnaire (EDI, Veale, 2014), the Ishihara Colorblindness Test (Hardy et al., 1945), the Karolinska Sleepiness Scale (KSS, Kaida et al., 2006), the Samn-Perelli Sclae (SPS, Samn and Perelli, 1982) and the St Mary Hospital Sleep questionnaire (QSN, Ellis et al., 1981). Details on all questionnaires can be found in Appendix V.14.2.

II.6.2.8 Analysis Plan

The general inference criterion is a p-value of $p < .05$. In multiple comparisons, we adjusted that criterion according to the adjusted Bonferroni method. Assumptions for each statistical test were checked and accounted for if not satisfied. Exploratory Analysis along with their results and a discussion of the findings are detailed in Appendix V.14.4.

Outliers

Two steps were employed to detect and reject outliers. First, trials with RTs above 5 seconds were rejected and counted as misses. In addition, an outlier detection was performed based on the interquartile range criterion. Trials with RT outside the 1.5x criterion were excluded. This was done for trials grouped by condition.

Primary Analysis

The analysis was performed using a mixed (multi-level) regression, as this allows for accounting for the effects of participants and tasks. A first mixed multilevel regression was performed on the dependent variable of RT (in ms). The independent variables were trialtype (3 Levels), task (4 Levels), block (5 Levels) and rask-rule congruency (7 Levels). The dependent variable was RT (in ms).

In a second Mixed multilevel logistic regression, ACC was analyzed. The independent variables were Trialtype (3 Levels), Task (4 Levels), Block (5 Levels) and Task-Rule Congruency (7 Levels). The dependent variable was ACC (Binary). To account for participants and task effects, both were dummy-coded and included as IV effects in the Mixed Regression for both

analyses. The analysis was first performed with a random intercept + Autoregressive Moving Average (ARMA)/ Autoregressive of Order 1 (AR1) to determine the most fitting covariance structure. Random slopes were added if the fixed model was reduced or followed a test for slope variance of the IV effects.

Expected Results: The expected results based on the hypotheses are detailed in Table 6.1.

II.6.3 Results

II.6.3.1 Outliers

From the initial 48720 trials, 52 were removed as the RT of the participants was above 5 seconds. This pertains to 0.11 % of the data. Then, using the 1.5 interquartile range criterion, 7.22 % further per cent of the data were removed. These data were distributed across all types of trials, therefore not biasing the analysis in any direction.

II.6.3.2 Reaction Time

The mixed regression, with task-rule congruency, block, task, and switch, showed several novel results. The covariance structure chosen for the analysis was AR1, as it presented a significantly better model than the more simple ID ($X^2(1) > 25, p > .001$). All main and two-way interaction effects were significant (switch: $F(44531, 2) = 1432.98, p < .001$, task: $F(43428, 3) = 1634.11, p < .001$, task-rule congruency: $F(43871, 6) = 21.81, p < .001$, block: $F(12425, 4) = 174.41, p < .001$, switch*task: $F(44455, 6) = 109.04, p < .001$, switch*congruency: $F(43461, 12) = 5.54, p < .001$, switch*block: $F(44419, 8) = 5.2, p < .001$, task*task-rule congruency: $F(43589, 12) = 9.42, p < .001$, task*block: $F(44335, 12) = 8.67, p < .001$, task-rule congruency*block: $F(43731, 24) = 4.18, p < .001$).

Switches

As expected, repetition trials had the fastest RTs with 624 ms (std err: 12) (repeat vs internal switch: $t(43996), p < .001$, repeat vs external switch: $t(45030), p < .001$). Internal switch trials were also significantly faster (723 ms, std err: 12) than external switch trials (740 ms, std err: 12) (internal vs external switch: $t(45118), p < .001$). RTs across the types of switches are represented by the bars in Graph 6.7.

Tasks

In terms of RT, all tasks differed from each other significantly. Participants were the fastest in the visual tasks (cold/hot 596 ms, vertical/horizontal 675 ms) as compared to the numeric tasks (low/high 728 ms, even/odd 784 ms) (low/high vs even/odd $t(45026), p < .001$, low/high vs vertical/horizontal $t(42438), p < .001$, low/high vs cold/hot $t(42356), p < .001$, even/odd vs vertical/horizontal $t(42660), p < .001$, even/odd vs cold/hot $t(42553), p < .001$, Vertical/ Horizontal $t(44843), p < .001$). The bars in Graph 6.8 represent RTs across tasks.

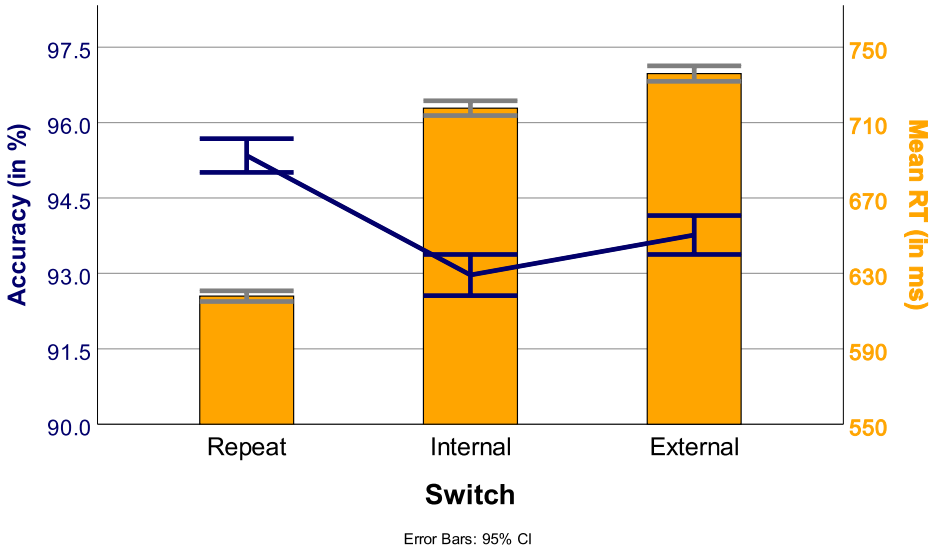

Fig. 6.7 Effect of the type of switch trial on performance. Bars: RT (in ms); Lines: ACC (in %), with 95% confidence intervals.

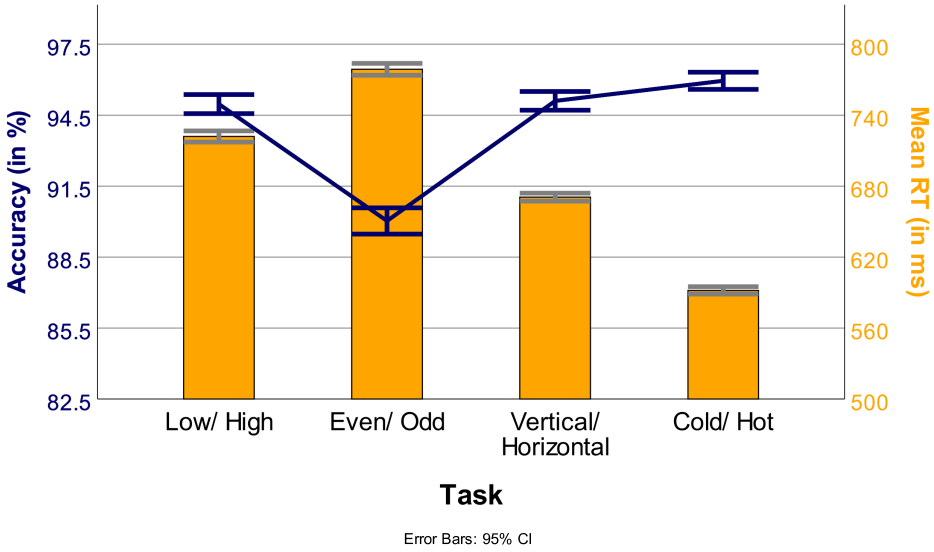

Fig. 6.8 RTs (in ms, Bars) and ACC (in %, Line) across all tasks, with 95% confidence interval.

Block

Across blocks, RT decreased from 739 ms (Std. Err=13) to 653 ms (Std. Err=12). The pairwise comparisons between all blocks were significant at a p-value of $p > .01$. RTs across blocks are represented by the bars in Graph 6.9.

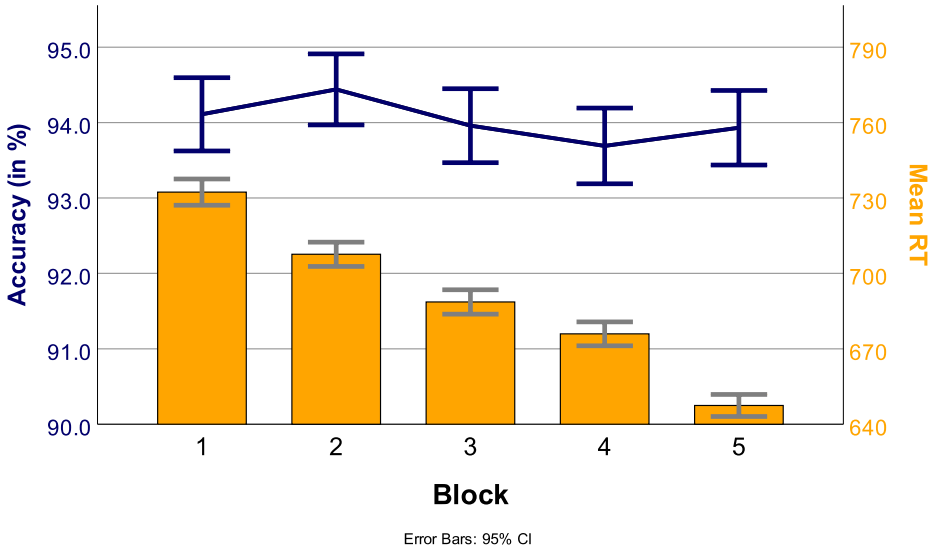

Fig. 6.9 RTs (in ms, Bars) and ACC (in%, Line) across blocks, with 95% confidence interval.

Task-rule congruency

RT was also influenced by task-rule congruency. Participants reacted fastest to neutral stimuli (679 ms Std. Err=13). The longest RTs occurred when participants had stimuli that were either double task-rule congruent or both internally and externally task-rule incongruent (708 ms (both), Std. Err=1). Most pairwise comparisons were significant at $p > .05$ except for neutral vs externally task-rule congruent ($t(43690), p = 1$), internally task-rule incongruent vs double task-rule incongruent ($t(44110), p = 1$), internally task-rule incongruent vs double task-rule congruent ($t(44144), p > .05$, externally task-rule incongruent vs internally task-rule congruent ($t(43986), p = 1$), externally task-rule incongruent vs double task-rule congruent ($t(44199), p > 0.1$), double task-rule congruent vs double task-rule incongruent ($t(44186), p = 1$). The bars in Graph 6.10 represent RT as influenced by task-rule congruency.

II.6.3.3 Accuracy

The mixed regression revealed several significant effects, with task-rule congruency, block, task, and switch on the correct or incorrect response dependent variable. The covariance structure chosen for the analysis was ID, as it presented a significantly better model than AR1 ($X^2(1) > 25, p > .001$). Except for block, all main effects were significant (switch $F(45123, 2) = 40.07, p < .001$, Task $F(45123, 3) = 154.48, p < .001$, task-rule congruency $F(45123, 6) = 4.9, p < .001$, block $F(45123, 4) = 1.299 p > .3$).

Switches

As expected, repetition trials were the most accurate (95,1 %, Std. Err.=0.3%) (repeat vs internal switch $t(45123), p < .001$, repeat vs external switch $t(45123), p < .001$). Other than with RTs, internal switch trials were less accurate (92.8% ms, Std Err.=0.3%) than external switch trials

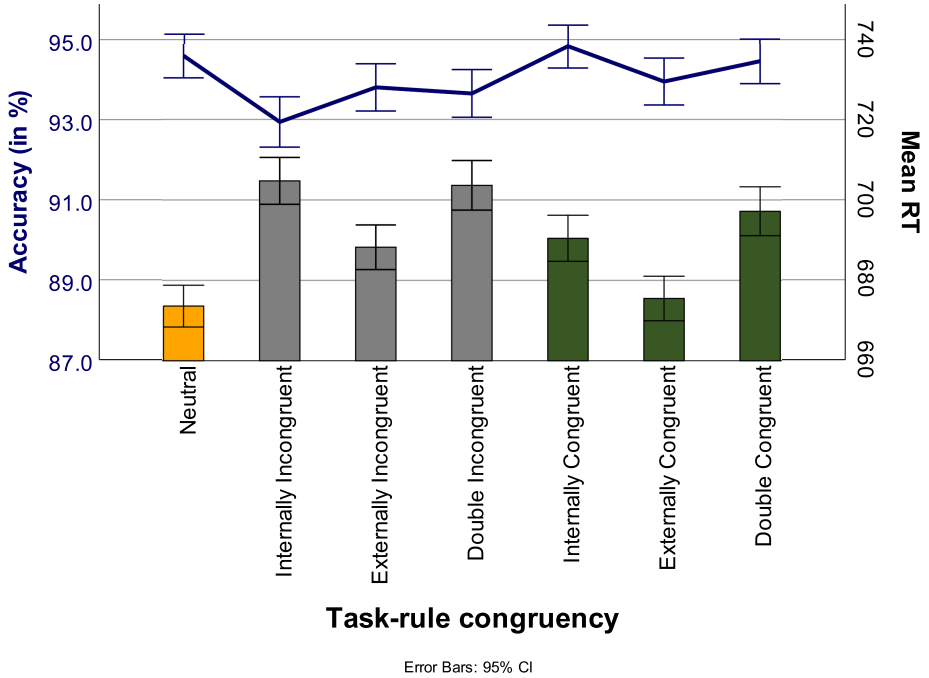

Fig. 6.10 RTs (in ms, Bars) and ACC (in %, Line) across different types of task-rule congruency, with 95% confidence interval.

(93.5% ms, Std Err.=0.3%) (internal vs external switch $t(45123), p < .05$). Accuracies across the types of switches are represented by the line in Graph 6.7.

Tasks

In terms of ACC, most tasks differed from each other significantly. Participants were most accurate in the visual tasks (cold/hot 95.8%, vertical/horizontal 94.9%) compared to the numeric tasks (low/high 94.7%, even/odd 89.9%). The only tasks that did not differ from each other significantly were low/high and Vertical/ Horizontal (low/high vs vertical/horizontal $t(42557), p = 1.0$). Accuracies across tasks are represented by the line in Graph 6.8.

Block

The main effect of block was not significant ($F(14052, 4) = 12.37 p > .2$). Accuracies across blocks are represented by the line in Graph 6.9.

Task-Rule Congruency

ACC was also influenced by congruency. Responses were most accurate when presented with internally task-rule congruent stimuli (94.6 % Std. Err=1.1%). Post-hoc tests showed that only internally task-rule incongruent trials resulted in a significant decrease in performance compared to neutral trials (92.8 % Std. Err=1.1%; $t(45123), p < 0.001$). Accuracies dependent on task-rule congruency are represented by the bars in Graph 6.10.

II.6.4 Discussion

Cognitive flexibility is essential in complex operations (see Section I.3.2.4). While much fundamental research has explored cognitive flexibility (see Section I.3.2.1), studies remain focused on simple settings rather than ecological tasks. One significant difference between fundamental task switching protocols and ecological applications is that fundamental tasks are generally similar. In contrast, real-world applications may require operators or persons to switch between very dissimilar tasks. The goal of this experiment was to investigate to what extent similarity between tasks influences Task-Switch costs. In addition, a particular focus was laid on aspects of mental fatigue.

The protocol required participants to switch between two pairs of two tasks. Each pair's two tasks were more similar to those of the other pair. RTs, as well as accuracies, were recorded.

The KSS and SPS questionnaires showed increased mental fatigue on a subjective scale following the testing session but not following the training session.

Contrary to our expectations, the participants' performance did not decrease significantly across blocks. RTs improved significantly. This is most likely attributable to a flaw in the protocol by not adequately accounting for learning effects. While any observable trend in the decrease of ACC between blocks was non-significant, this was not the case for the within-block (timepoint) analysis. Through a single block, participants tended to make increasingly more errors. This supports the notion that mental fatigue was induced during the task. The most likely reason this did not extend between blocks is the recovery period between blocks.

The main interest of this study was to investigate the effect of similarity on cognitive flexibility. The results indicate that Task-Switch costs (both RT and ACC) were significantly impacted by similarity. It would have been expected that trials with the most errors would also have the largest RTs. This was not the case. Participants struggled not to make errors during internal switches (between similar tasks) but remained capable of responding rapidly. External switches (between tasks) required the longest RTs but were more accurate. A possible explanation may be that participants used a compensation strategy during the external tasks to reduce errors. However, the task instructions did not explicitly state that ACC was more important than RT. Furthermore, if this were the case, one would expect internal switch trials to be as accurate as external switch trials were. Finally, participants were unaware that the design considered internal or external switches to be two different classes, further undermining the idea of a compensatory strategy.

An alternative explanation is linked to the interference and the reconfiguration view (see Section I.3.2.1). Following the reconfiguration view, if the same task is repeated, the activated elements and representations remain active, no reconfiguration occurs, and responses are faster. In the case of similar switches, some elements/ representations may be capable of remaining activated, resulting in faster RTs. For example, participants need not pay attention to the numeric values in the visual tasks but in both numeric tasks. So when reconfiguring across dissimilar tasks, more elements must be retrieved/ deactivated.

On the other hand, the interference view would assume that dissimilar trials are easier since there is more interference between similar tasks. So, these results support neither the interference nor the reconfiguration view. The reconfiguration view explains why similar tasks would result in faster responses but would also expect lower accuracies when participants switch to dissimilar tasks. The interference view would expect higher RTs and more errors in similar tasks, as a previous trial may interfere with the current one. The suggestion has been made previously that the components to be retrieved depend on the pair of tasks and impact

TSC (see Section I.3.2.1, Monsell et al., 2000; Rubinstein et al., 2001). Suppose, however, that both theories are partially true. How could a combined interference and reconfiguration view explain the difference in switch costs between trial types?

In a similar switch, some components of the task set remain the same. So, reconfiguration does not change these components, which may positively impact the RT. However, the instructions from the previous trial may still interfere with the current trial, slightly bias the decision-making process, and result in a higher likelihood of error. Monsell et al. (2000) made similar arguments in combining the interference and reconfiguration views. Their argument is based on residual switch costs, an increased RT when switching even if participants are given abundant time to prepare for a switch. According to Monsell et al. (2000), these costs may be explained by interference.

Besides similarity and mental fatigue, conflicting information presented during trials was also investigated during this study. As expected, the fastest RTs were recorded when only relevant information (neutral trials) was given. Neutral trials were not significantly different regarding ACC from the most accurate trials (internally task-rule congruent). While both task-rule congruent and task-rule incongruent trials increased the RTs, the effect was much more pronounced when trials were task-rule incongruent. A possible explanation for the increase in RTs in congruent trials may be that participants struggled to identify relevant stimuli when presented with multiple bits of information. The differences between internal, external, and double trials were similar for task-rule congruent and task-rule incongruent conditions. In both cases, external information interfered the least (fastest RTs) with the participant's decisions. Concerning ACC, especially internal task-rule incongruent information seemed to increase errors. Differences between tasks were also observed. low/high & odd/ even numeric tasks were significantly more difficult for participants than the visual tasks (vertical/horizontal & cold/hot). Regarding both ACC and RTs, even/odd was the most difficult, while cold/hot was the most straightforward task. The balanced design of the experiment, as well as the inclusion of this factor in all analyses, mitigates any potential bias on other analyses.

Several limitations have to be taken into account when interpreting the results of this study. The effects of mental fatigue did not accumulate across blocks as expected. In addition to learning effects likely to have had the most substantial impact on performance, compensatory strategies may as well have influenced the performance. While the link between mental fatigue and performance is well-established, it is not uncommon to detect the expected behavioural change over time (see Section I.3.1.5 & Chapter 8, Hamann and Carstengerdes, 2023). Maintaining or improving on performance may also be a factor of motivation that can overpower mental fatigue. It may also be an increase in effort as a response to mental fatigue that negates the decrement in performance.

The significant differences in performance between tasks are no limitation to the results presented here as the design was balanced, and the factor was considered during the statistical analysis.

Lastly, it should be noted that a minor bug in the code was detected following the data collection. In double task-rule congruent or double task-rule incongruent trials, the same external task was presented in each of these trials from block two on. However, a post-hoc comparison of these trials in the first and the remaining blocks showed no effect of this configuration on the data. This was tested with a non-parametric test by comparing RTs and accuracies of relevant trials to search for any difference in performance.

It would be interesting to consider replicating the design in the future using non-bivalent response keys. This would be easily achievable by assigning different keys to each task. Doing

so, would allow for an investigation into the source of errors. Do participants respond using the keys of the previous task, or do the errors occur while using the current task-assigned keys?

II.6.5 Conclusion

This study on how task similarity affects cognitive flexibility has yielded several noteworthy findings. As similarity decreases between tasks, participants get slower in switching; similar functions, on the other hand, produce more errors. One interpretation of these findings is that the interference view of cognitive flexibility is better suited to explain changes in error-making. In contrast, the reconfiguration view applies better to measures of RT. These findings also show that results from fundamental research are not readily applicable to more ecological settings where similarity between tasks is more scarce than in traditional task switch paradigms. To the best of our knowledge, the effect of task-rule congruency convergence has not yet been reported in the literature. If reproducibleF, this effect would have implications for our understanding of task switch and necessitate adapting future paradigms to account for it. With only two tasks in our study, this effect is not observable. In conclusion, this chapter promotes a shift toward adopting fundamental paradigms of task switches in ecological settings while proposing novel aspects that shape the understanding of the mechanisms of cognitive flexibility.

This contribution was submitted and accepted as a preregistered study with PLOS ONE. The final article has been accepted and is currently in production (see Appendix 13, Hinss, Brock, and Roy, 2023a; Hinss, Brock, and Roy, 2024)

Chapter 7

Double Task Switching II: Countermeasures for performance improvement

Central Ideas

The conception of this study was to provide a preliminary investigation into visual alerts as a countermeasure for task-switching costs. As implemented here, visual alerts are a flash draping the entire experimental screen in a colourful hue for a short period of 35 ms. Their goal is to inform participants that the upcoming trial in a task-switching experiment (based on the DTS protocol) is a switch trial rather than a repetition trial. The within-subjects design used three conditions (control, single alert & double alert) to investigate different types of visual alerts. Results show no significant increase in performance in conditions due to visual alerts. The absence of results is likely attributable to a combination of a study design with many variables and relatively few participants. A future study with a simplified design and larger sample may allow for a more thorough investigation into the fundamental aspects of visual alerts.
Recommended Reading:

- DTS (Chapter 6)
- Cognitive flexibility (Section I.3.2)

One Question to ask: **Does the mere presence of Visual Alerts facilitate cognitive flexibility?**

II.7.1 Motivation and Research Goals/ Questions

The DTS experiment (see Chapter 6) provided several interesting results. Switching between similar tasks results in lower task-switching costs than switching between more dissimilar

ones. In complex situations such as UAS operations, the operators are frequently required to switch between very dissimilar tasks (see Section I.2.1). Finding a way to reduce task-switching costs is a path towards safer operations. Longer preparation times and inter-stimulus intervals have been shown to reduce TSC (see Section I.3.2.1). This study aims to test whether this effect can be exploited to reduce TSC in the context of the DTS protocol. Here, we used a simplified protocol version and added two conditions: single and double alert. The single alert is a simple, colourful flash across the entire screen on which the task is presented. This flash only occurs if the upcoming experimental trial is a switch trial. In the double alert, the flash colour depends on the switch type. There is one colour for internal switches and another for external switches.

II.7.2 Methods

II.7.2.1 Ethics Statement

Ethical consent was obtained from the Comité d'Ethique de la Recherche—CER at the Université de Toulouse (ID 2022–521) on the 2nd of June 2022. Data collection occurred from the 4th of September until the 11th of November of 2023.

II.7.2.2 Design

The DTS protocol consists of 4 subtasks: 2 based on the numeric value of the displayed stimuli and two based on the visual appearance of the stimuli.

Participants perform the LOW/HIGH or EVEN/ODD tasks as the numeric subtasks. For the LOW/HIGH task, participants must judge whether a single number is higher or lower than five and indicate their response with a keystroke. The number 5 is never present in the task, and all LOW/HIGH task stimuli are always presented as the left stimulus. In the EVEN/ODD task, participants must decide if a stimulus (again, a single number, but presented as the right stimulus) is even or odd. Again, participants are instructed to respond with a keystroke.

Participants perform the COLD/HOT task or the VERTICAL/HORIZONTAL task for the visual tasks. In the COLD/HOT task, participants must identify whether the stimulus is a hot colour (red/orange) or a cold colour (blue/green). In the VERTICAL/HORIZONTAL task, a pattern's orientation must be determined (see Figure from previous contribution 6.3 for an example array of stimuli for each condition).

During the task, participants must perform consecutive trials, where the task and stimuli will change quasi-independently from trial to trial. For each trial, participants are presented with three quasi-independent pieces of information. They are presented with instructions for the current trial in the form of the task name (e.g., VERTICAL/HORIZONTAL, presented before the actual stimulus) and two stimuli in the form of signs or numbers. Based on the stimuli and the instruction, participants are instructed to respond as fast as possible, using either the S or the L key on a keyboard (keys were selected as they are located at a comfortable distance on the keyboard to avoid straining the hands of the participants).

Task Switch Conditions

Each trial (n) is classified into one of the three Task Switch Conditions depending on the task in trial n-1. If the task remains the same, the trial is a repeat trial. Suppose the task switches but

remains within the same domain (Numeric: LOW/HIGH to EVEN/ODD or Visual: VERTICAL/HORIZONTAL to COLD/HOT and vice versa). In that case, it is considered an internal switch trial. Finally, switches from one domain into the other (numerical to visual and vice versa) are considered external switch trials (see Figures from previous contribution 6.1 & 6.2).

Alert Conditions

The protocol for this experiment is divided into six blocks belonging to one of three conditions.

The "No Alert" condition presents the task instructions independent of the task-switch condition. The instructions are presented 300 ms before the stimuli are presented for all trials.

In the "Single Alert" condition, a short 35 ms coloured flash across the computer screen appears 700 ms before the instructions are shown. There are no visual alerts for repeat trials, and the flash colour is the same for both internal and external switches. In this experiment, three different colours were used as flashes, chosen so as not to overlap with the hot/cold colours of the stimuli.

- Brown (R: 135, G: 99, B: 71)
- Green (R: 117, G: 236, B: 149)
- Purple (R: 242, G: 142, B: 242)

For the "Double Alert" condition, participants receive a visual alert similar to the single alert condition. However, internal and external switches have different colours. There are no visual alerts for repeat trials. Randomly, participants were assigned to one of three color schemes:

Single Alert: Brown Double Alert 1: Green, Double Alert 2: Purple

Single Alert: Purple Double Alert 1: Brown Double Alert 2: Green

Single Alert: Green Double Alert 1: Purple Double Alert 2: Brown

Unfortunately, a bug resulted in four participants having a colour scheme in which they were presented with a colour scheme other than the one they were assigned to. We regrouped these participants in the analysis. The colour scheme and this group did not affect the statistics, as the effects were insignificant.

Trial Setup

Each trial began with a fixation cross in the middle of the screen for 350-850 ms (600 ms with a 250 ms jitter). In the task switching trials during either the single or double alert, the corresponding colorful flash was then presented for 35ms. In the control condition, this was omitted. A delay of 650 milliseconds was followed by trial instruction (400 ms). Finally, participants were presented with the stimulus (350 ms), where they could respond with a keystroke. If participants took longer than 350 ms to respond, the stimulus disappeared. Once a response was recorded, a new trial began (see Figure 7.1). The task was computed using PsychToolBox-3 (http://psychtoolbox.org/) in Matlab[1] and was presented on a desktop computer . Participants were not randomized (within-subjects design). However, the trials that participants

[1] https://fr.mathworks.com/products/matlab.html

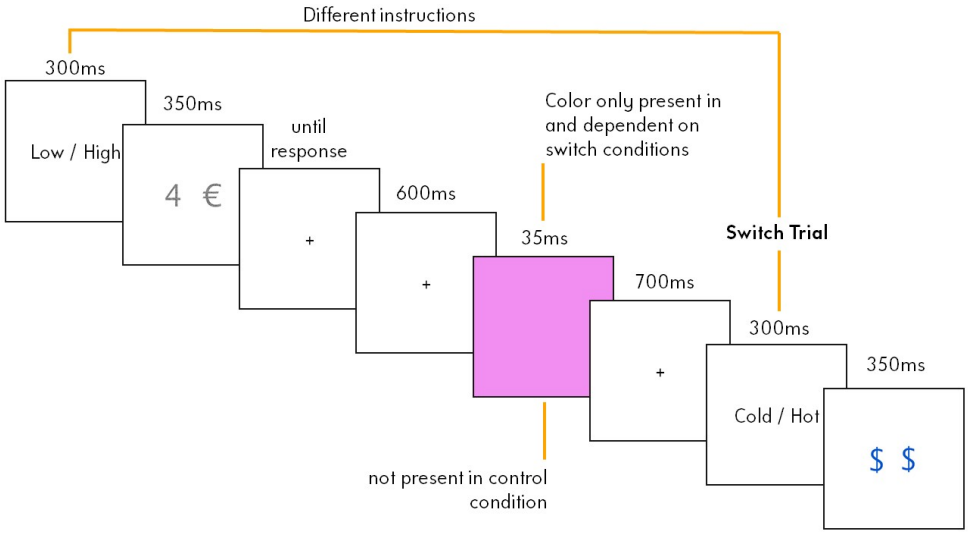

Fig. 7.1 Trial setup for the DTS-II experiment. Color of the visual alert depends on the condition as well as the assigned color scheme.

completed were quasi-random. That is to say, trials appeared in a random order, which ensures a balanced design and equal probabilities of each trial occurring.

II.7.2.3 Participants

For this study, 25 participants (9 female participants, mean age 23.96 years std = 2.76) were recruited. Ten participants experimented in English, while the others chose French. The scores on the Edinburgh Handedness Questionnaire (EDI) range from 4 (left-handed) to 20 (right-handed). 3 participants had EDI scores indicating left-handedness (5-10), with one participant scoring as ambidextral (11-14) and 21 participants as right-handed (15-20).

II.7.2.4 Recruitment inclusion criteria and compensation

Participants were recruited by e-mail and via paper announcements in the institute buildings at ISAE-SUPAERO. Participants had to be between 18 and 30 years old, have obtained a level of study baccalaureate minimum, have normal or corrected-to-normal vision and hearing, be affiliated with social security, and have signed an informed consent form. Participants were prohibited from participating if they met any of the following criteria: presence of known neuropsychological disorders, significant visual or hearing impairment, current treatment with psychotropic medication or substances, positive g-test or being a nursing woman. In addition, due to the nature of the experiment, colourblind people were excluded from participation. As recommended to minimize the risk of an invasion of privacy, all criteria (inclusion and/or exclusion) were listed together in the inclusion criteria sheet, and potential participants were asked whether they met all the criteria rather than asking them whether they met each criterion one by one. Due to the nature of the experiment, participants were also tested for colourblindness before signing the inclusion/exclusion criteria. To do so, the Ishihara test for colourblindness was employed (Hardy et al., 1945). Participants were compensated in gift



Fig. 7.2 Procedure for the DTS-II study. The NASA-TLX and RSME were asked following each block.

vouchers (Illicado, http://illicado.com/). Upon completion, participants received a voucher worth 15 €.

II.7.2.5 Procedure

After the participants had arrived, they were asked to complete the Ishihara colourblindness test before filling in the informed consent and information sheets. The participants were informed that they were permitted to ask questions at any time, and were also asked in what language they would like to perform the task (French or English). They then filled in the Demographics questionnaire, the EDI, a KSS, and the Samn-Perelli Fatigue (SPS) 7-point Likert scale (see Appendix V.14.2). Next, the participants were familiarized with the DTS protocol in the form of a PowerPoint presentation. They then performed a familiarization and a training protocol. Participants performed one block of 48 trials in the three conditions in the familiarisation protocol. Following a response, they were given visual feedback ('Correct' vs 'Incorrect') on their performance. Next, they performed the training. The blocks were identical to the experimental ones, except for removing the visual feedback. Participants performed 3 training blocks. Upon completion of the training, the participants were asked if they wanted to add additional training before starting the experiment. None of the participants requested additional training. Before starting the experiment, participants filled in a KSS and SPS scale (see Appendix V.14.2).

Each block started with instructions on which type of condition the upcoming block would be in. Each block consisted of 96 trials and took 5-6 minutes to complete. After completion, participants filled in a RSME scale and a NASA-TLX (see Appendix V.14.2.8) scale before either starting with the next block or taking a break before the next block.

The participants completed the subjective questionnaire battery after completing the 6th and last block. Participants were then thanked for their participation and compensated with the vouchers. The procedure is detailed in Figure 7.2.

II.7.2.6 Variables

Independent Variables

The following stimuli were used in this study:

- Numeric Value: The value of the presented stimuli was manipulated. Participants saw combinations of the numbers 1,2,3,4,6,7,8 and 9 and symbols %, &, € and $.

- Color: The stimuli colour was manipulated to either be hot (red/orange hues) , cold (blue/ green hues) or neutral (grey hues) .
- Pattern: The pattern of the stimuli was manipulated and could either show a horizontally oriented pattern, a vertically oriented pattern, or a random pattern.
- Task: The task participants had to complete given the stimuli could be adapted, as seen in Figures 1 and 2. The four tasks were low/high, even/odd, cold/hot, and vertical/horizontal.

Dependent Variables

The following measures were recorded:

- ACC: The number of correct answers during the task.
- RT: The time from stimulus presentation to response (keystroke).
- Subjective Questionnaire Data: The data from all questionnaires (see Section II.7.2.7). All questionnaires, except the three open questions, were presented to the participants in a paperless digitalized form.

II.7.2.7 Questionnaires

Participants filled in several questionnaires, including a demographics questionnaire, the Edinburgh handedness questionnaire (EDI), the Ishihara Colorblindness Test, the Karolinska Sleepiness Scale (KSS), the Samn-Perelli Scale (SPS) and the St Mary Hospital Sleep questionnaire (QSN). Details on all questionnaires can be found in Appendix V.14.2. Following each block, participants also filled out a NASA TLX questionnaire. At the end of the study, several further subjective questionnaires were used. Participants were presented with VAS and multiple-choice questions for each condition (No Alert, Single Alert, Double Alert). Questions included:

- VAS
  - How difficult was the task in the `Condition`? (0 = Very Easy, 100 = Extremely Difficult)[2]
  - How difficult was it to use and comprehend the Visual alerts in the `Condition`? (0 = Very Easy, 100 = Extremely Difficult)
  - How useful were the Visual alerts in the `Condition`? (0 = Not Useful at all, 100 = Very Useful)
- Multiple Choice
  - Would you prefer if the Visual Alerts in the `Condition` were presented in a different modality?
    - Auditive | Haptive | Neither
  - Would you prefer if the Visual Alerts in the `Condition` were presented in a different color?

[2]This was the only question asked in all three conditions, while the following questions were only asked regarding the single and the double alert.

* Green | Red | Blue | Black | White | Other | No

Finally, participants were asked to respond to 3 written questionnaires.

- Please note down 3 things you liked about the visual alert:
- Please note down 3 things you disliked about the visual alerts:
- If you have any other remarks please feel free to note these here:

II.7.2.8 Analysis Plan

The general inference criterion is a p-value of $p < .05$. In multiple comparisons, we adjusted that criterion according to the adjusted Bonferroni method. Assumptions for each statistical test were checked and accounted for if not satisfied.

Outliers

Two steps were employed to detect and reject outliers. First, trials with RTs above Five seconds were rejected and counted as misses. In addition, an outlier detection was performed based on the interquartile range criterion. This was done for trials grouped by condition.

The analysis was performed using a Mixed (Multi-Level) Regression, as this allows for accounting for the effects of participants and tasks. A first mixed multilevel regression was performed on the dependent variable of RT (in ms). The independent variables were Trialtype (3 Levels), Condition (3 levels) and Task (4 Levels). The dependent variable was RT (in ms). TOT was included as a covariate.

In a second Mixed multilevel logistic regression, ACC was analyzed. The independent variables were Trialtype (3 Levels), Condition (3 levels) and Task (4 Levels). The dependent variable was ACC (Binary). TOT (time since task onset and/or trial number) was included as a covariate. In order to account for participants and task effects, both were dummy-coded and included as IV effects in the Mixed Regression for both analyses. In order to best determine the most fitting covariance structure, the analysis was first performed with a random intercept + ARMA (or AR1). Random slopes were added if the fixed model was reduced or followed a test for slope variance of the IV effects.

We also used ANOVAS to analyze the subjective results to compare the conditions.

II.7.3 Results

II.7.3.1 Reaction Time

The mixed regression, with colour scheme, task, and switch and condition, showed several exciting results. The covariance structure chosen for the analysis was ARMA11, as it presented a significantly better model than the more simple AR1 ($X^2(1) > 25, p > .001$). All main effects except for color scheme were significant (Switch: $F(13240, 2) = 487, 92p < .001$, Task: $F(12881, 3) = 932.48, p < .001$, Condition: $F(324, 2) = 19.39, p > .001$, colour scheme: $F(25, 3) = 5.91, p = .512$). Interactions between Switch and Task as well as Order and Switch were also significant (Switch*Task: $F(13233, 6) = 19.86, p < .001$,Order*Switch: $F(13236, 6) = 5.91, p < .001$).

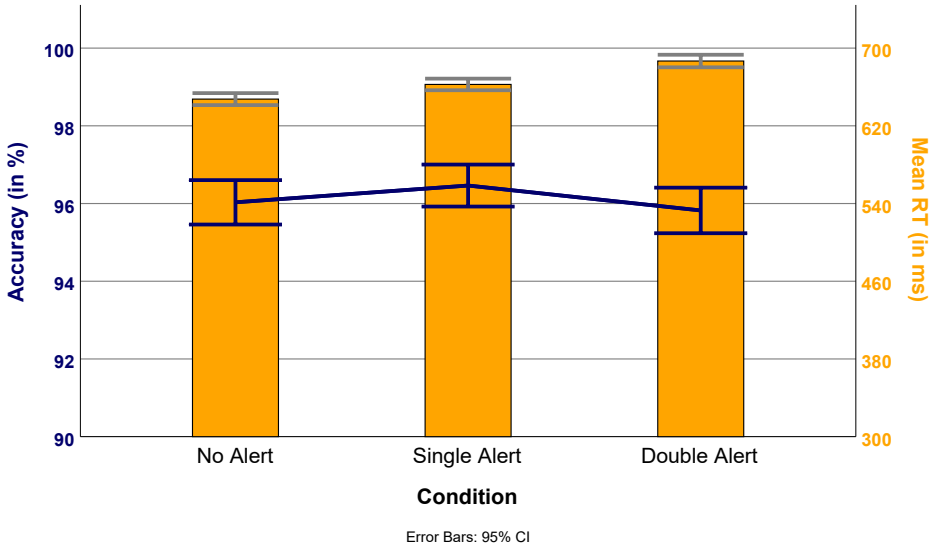

Fig. 7.3 Effect of the condition trial on performance. Bars: RT (in ms); Lines: ACC (in %), with 95% confidence intervals.

Switches

As expected, repetition trials had the fastest RTs with 612 ms (std err: 24) (Repeat vs Internal switch: $t(13110), p < .001$, Repeat vs External switch: $t(13288), p < .001$). Internal switch trials were not significantly faster (702 ms, std err: 24) than External switch trials (709 ms, std err: 24) (Internal vs External switch: $t(13365), p = .103$). RTs across the types of switches are represented by the bars in Graph 7.4.

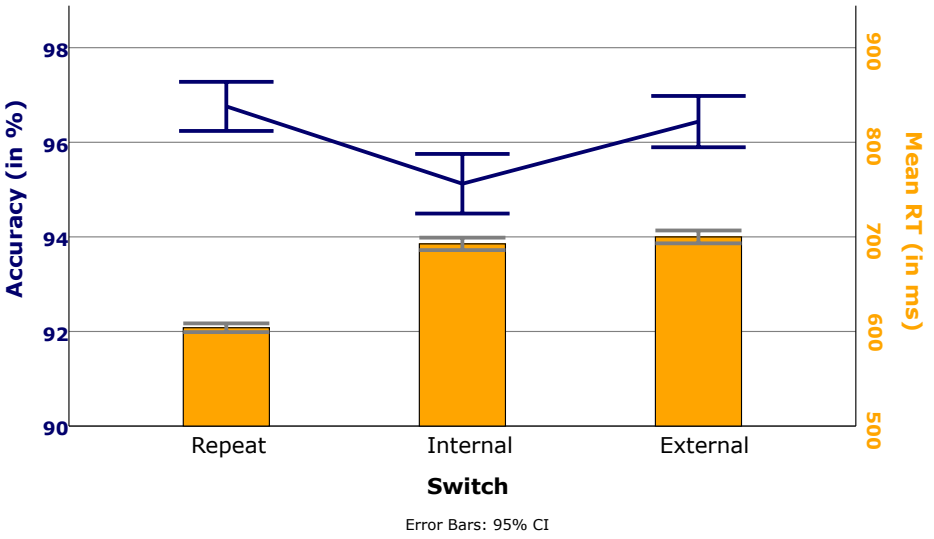

Fig. 7.4 Effect of the type of switch trial on performance. Bars: RT (in ms); Lines: ACC (in %), with 95% confidence intervals.

Condition

Other than expected, the no alert condition was with 656 ms (std err: 23) significantly faster than the Double Alert condition (No Alert vs Double Alert: $t(248), p < .001$). The Single Alert (671 ms, std err: 23) condition was not significantly different from the no alert condition (No Alert vs Single Alert: $t(425), p = .035$). However, it was significantly faster than the Double Alert condition (Double Alert vs Single Alert: $t(345), p < .001$). RTs across conditions are represented by the bars in Graph 7.3.

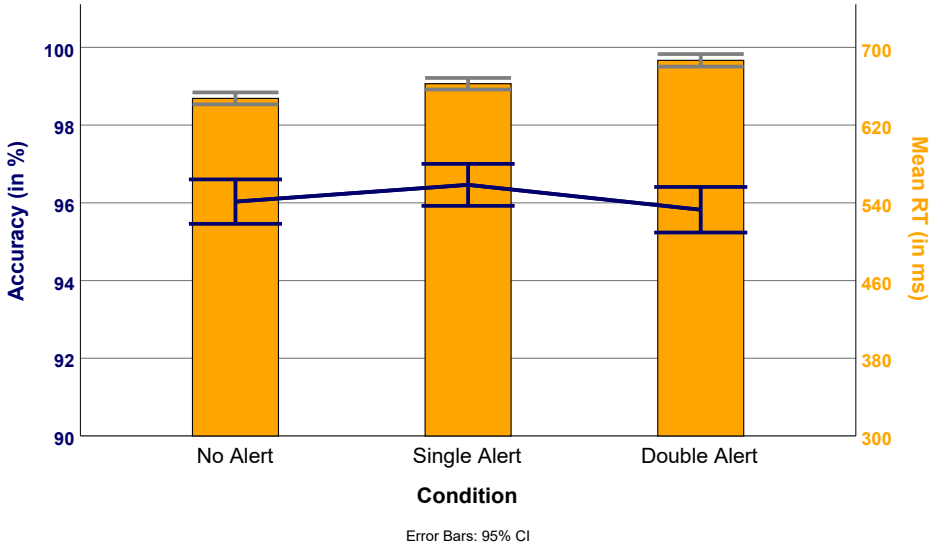

Fig. 7.5 Effect of condition on performance. Bars: RT (in ms); Lines: ACC (in %), with 95% confidence intervals.

Tasks

In terms of RT, all tasks differed significantly from each other. Participants were the fastest in the hot/cold task (555 ms, std err 24), followed by the low/high task (689 ms, std err: 24). Participants were slower on the vertical/horizontal task (715 ms, std err: 24) and slowest in the even/odd task (739 ms, std err: 24). These differences were significant (low/high vs even/odd $t(13318), p < .001$, low/high vs vertical/horizontal $t(12362), p < .001$, low/high vs cold/hot $t(12236), p < .001$, even/odd vs vertical/horizontal $t(13025), p < .001$, even/odd vs cold/hot $t(12733), p < .001$, Vertical/ Horizontal vs cold/hot $t(13260), p < .001$). The bars in Graph 7.6 represent RTs across tasks.

II.7.3.2 ACC

The mixed regression revealed significant task and switch effects on the correct or incorrect response dependent variable. The covariance structure chosen for the analysis was AR1, as it presented a significantly better model than ARMA11 ($X^2(1) > 25, p > .001$), and as the Hessian Matrix of the ARMA11 model could not converge (Switch $F(13408, 2) = 7.80, p < .001$, Task $F(13111, 3) = 53.91, p < .001$, Condition $F(4788, 6) = 0.86, p = .423$, colour scheme $F(25, 4) = 0.292$ $p = .831$).

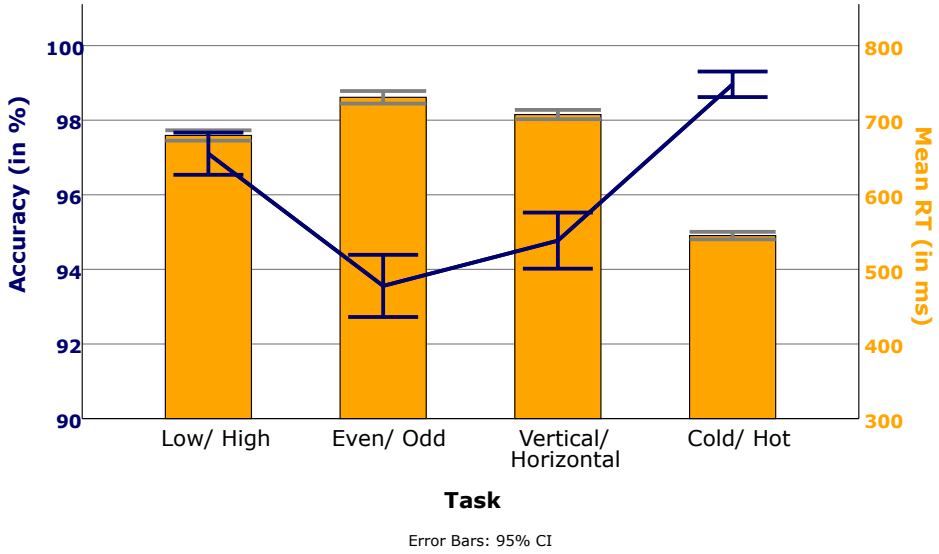

Fig. 7.6 RTs (in ms, Bars) and ACC (in %, Line) across all tasks, with 95% confidence interval.

Switches

As expected, repetition trials were the most accurate (96.8 %, std err: 0.7%) (Repeat vs Internal switch $t(13405), p < .001$), but they did not differ significantly from external switch trials (Repeat vs external switch $t(13432), p = .744$). Other than with RTs, Internal switch trials were less accurate (95.1% , std err: 0.7%) than external switch trials (96.3% ms, std err: 0.7%) (internal vs external switch $t(13335), p = 0.021$). Accuracies across the types of switches are represented by the line in Graph 7.4.

Tasks

In terms of ACC, most tasks differed significantly from each other. Participants were most accurate in the cold/hot task (98.9%, std err 0.7%), followed by the High/Low Task (971%, std err 0.7%), the Vertical/ Horizontal task (94.7%, std err 0.7%) and the even/odd task (93.5%, std err 0.7%). The only tasks that did not differ from each other significantly were even/odd and Vertical/ Horizontal task (even/odd vs vertical/horizontal $t(12780), p = .050$). Accuracies across tasks are represented by the line in Graph 7.6.

II.7.3.3 Subjective Analysis

The RSME and NASA-TLX (see Appendix V.14.2) scores asked following each block did not differ significantly between conditions or blocks. The comparison between conditions at the end also showed no significant difference. KSS scores increased significantly from pre-task to post-task ($F(24, 2) = 6.44, p < .001$). Training did not significantly alter fatigue when comparing KSS scores before and after training ($t(-.0364), p = 1$). SPS scores showed a similar trend ($F(24, 2) = 8.28, p < .001$). Training did not significantly alter fatigue when comparing SPS scores before and after training ($t(-1.09), p = 0.84$, see Figure 7.7). Participants did not indicate a significant preference to change the modality of the alarms.

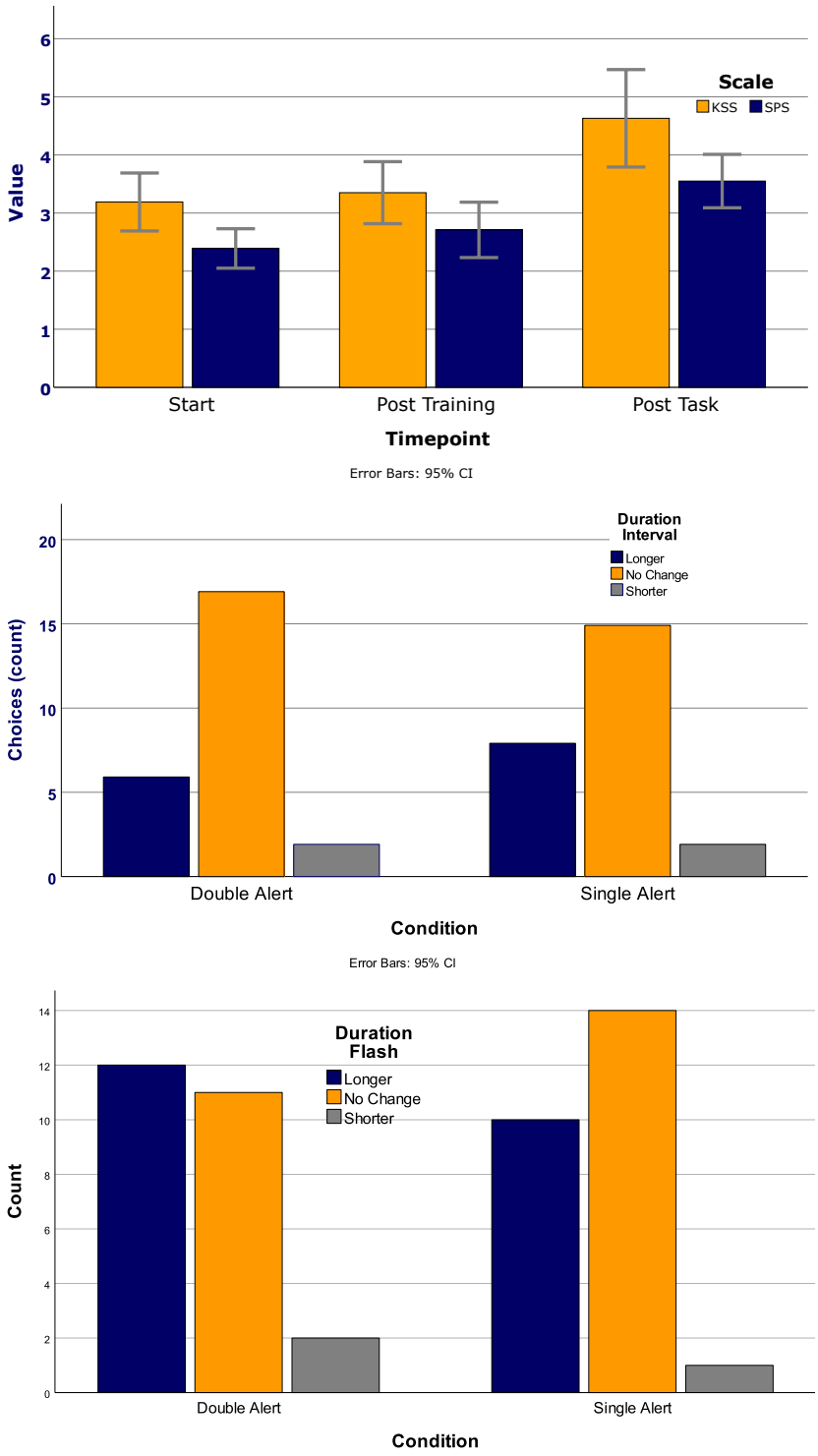

Fig. 7.7 Top Left) Subjective Fatigue over time as measured by the SPS an KSS scales. Top Right) Subjective preferences for the interval duration. Bottom Left) Subjective preferences for the flash duration.

The questionnaires at the end also gave further insight into how participants perceived the visual alerts. The double alert seems to have been too difficult to comprehend, while participants stated that the single alert allowed them to remain more vigilant. Participants would have preferred a longer period between the flash and the stimuli, however the difference between the frequency of people preferring a longer interval as compared to a shorter one or no change was not significant ($\chi^2(2) = 0.411$, $p = 0.814$). Participants also noted that the visual alerts were, at times, visually fatiguing.

II.7.4 Discussion

Contrary to our hypothesis, the visual flashes did not improve performance. Several explanations possibly explain this. First and foremost, it could be possible that the visual flashes in the way they were implemented do not facilitate cognitive flexibility. This could be because the period between the flash and the stimulus was not long enough. The fact that two of the tasks were also colour-dependent may have perturbed participants despite the chosen colours not overlapping with those used for visual alerts. As participants were not instructed to use the visual alerts but were made aware of them occurring, they may have also learnt to ignore them or even found them distracting. Alternating between them in between blocks may have also been challenging for participants.

Furthermore, the study was likely underpowered. When comparing the behavioural results to those of the DTS study (see Chapter 6), we observe larger standard errors and significant effects in that study are not reproduced. The DTS-II had three different visual alert conditions.

Yet the absence of proof in the current study does not mean visual alerts may not be helpful in interface adaptation. Interfaces also require a sizeable attentional shift when operating on several screens, in which the visual alerts are likely helpful. Furthermore, as a large body of literature shows, awareness of adaptations and trust in them significantly predicts performance (see Section I.2.2.6). In this study participants were not informed that the visual alerts may be useful. Another possible reason for the visual alerts not improving results is that participants were vigilant enough not to need the support of the visual alerts, as explained in Section I.2.2.3. A future study with a simplified design and a larger sample size may validate that visual flashes aid with cognitive flexibility.

Chapter 8

Remot3e: Research into mental fatigue classification and labelling

Central Ideas

A central component of search and rescue missions is the heavily operator-dependent visual search of survivors. These operators suffer from mental fatigue (MF), and detecting this is a critical step in implementing future systems. To the best of our knowledge, MF has seldom been evaluated using a realistic visual search task. In addition, there exists an accuracy discrepancy between studies that use the popular time-on-task (TOT) and performance metrics for labels. Yet, to our knowledge, they have never been directly compared. This study was designed to address both issues: the use of a realistic task to elicit MF during a monotonous visual search task and the choice of labelling type used for intra-participant fatigue estimation. Over 4 blocks of 15 minutes, participants had to identify targets while their cardiac, cerebral (EEG), and eye-movement activities were recorded. The results of mental state estimation show that an algorithms' capability to detect MF depends strongly on how MF is defined. Using TOT, high classification accuracies are obtained (i.e 99.3%). On the other hand, if MF is estimated with behavioural performance, a metric with a greater operational value, accuracies return to chance level (i.e. 52.2%).
Recommended Reading:

- Mental fatigue (Section I.3.1)
- Sensors (Chapter 4)
- Algorithms (Chapter 5)

One question to ask: **What are the differences between performance or TOT-based mental fatigue estimation?**

II.8.1 Motivation and Research Goals/ Questions

As explored in the state-of-the-art, Uncrewed Aerial Systems (UAS) are a diverse asset that can be used in diverse missions such as agriculture and powerline inspections in the civilian sector and surveillance, search and rescue missions, port protection, intelligence gathering, and strike missions in military applications (see Sections I.2.1.3 & I.2.1.2)[1].

Visual search using onboard cameras, infrared optics, or radars is a central aspect of these missions, as operators must continuously observe and identify targets, search for anomalies, and detect changes.

Among other factors, mental fatigue and its impact on attention are some of the many human factor issues in UAS operations (see Sections I.2.1.4 & I.3.1.4). Therefore it is interesting to investigate how mental fatigue can affect visual search performance during UAS operations.

Fan et al. (2015) had participants perform a visual search experiment for 60 minutes. However, the realism of this experiment was limited. Participants were shown abstract shapes rather than photographs. Still, the results showed increased reaction time with increasing time-on-task (TOT). Unfortunately, to the best of our knowledge, the link between mental fatigue and visual search remains underexplored apart from this work.

Beyond this experiment serving as a test for the implementation of a more realistic search task and its interaction with mental fatigue, we further aim to explore the issues of Time-On-Task (TOT) and performance-based labelling for Physiological computing systems (PCS) and passive Brain-Computer Interfaces (see Secions I.5.2.2, I.5.2.1 & I.5.2.1).

Several hypotheses were formulated, as listed in table 8.1. The hypotheses are divided into data validation and mental state estimation. The goal of the task is to elicit mental fatigue; we thus expect some established effects of mental fatigue to occur in the participants (see Section I.3.1.4). The mental state estimation hypotheses refer to our goal to compare different classifiers and types of labels. The following sections will detail the experimental procedure of a more realistic visual search task, how EEG, ECG and eye-tracking data were recorded simultaneously and how these were then used for classification with various pipelines applied using the two types of labelling.

II.8.2 Material and methods

II.8.2.1 Participants and ethical approval

Fifteen participants participated in this experiment (3 females, 2 left-handed, 26 +/- 5 years). Participants had to be between 18 and 60 years old, have obtained a level of study baccalaureate minimum, have normal or corrected-to-normal vision and hearing, be affiliated with social security, and have signed an informed consent form. Participants were prohibited from participating if they met any of the following criteria: Protected persons, presence of known neuropsychological disorders, significant visual or hearing impairment, current treatment with psychotropic medication or substances, positive g-test or being a nursing woman. As recommended to minimize the risk of an invasion of privacy, all criteria (inclusion and/or exclusion) were listed together in the inclusion criteria sheet, and potential participants were asked whether they met all the criteria rather than asking them whether they met each criterion one by one. None of the participants reported a history of prior neurological disorders. All subjects

[1] The work on this contribution was done in collaboration with Emilie Jahanpour, working as a Post Doc at ISAE-SUPAERO. She programmed the search task and performed the data acquisition.

Goal	Data	Hypothesis
Data validation	Subjective	1. MF reports ↑ with TOT
	Behavioral	2. TOT ↑, performance ↓
	Physiological	3. TOT ↑,HR ↓ 4. TOT ↑, HRV ↑ 5. TOT ↑, EEG alpha band power ↑ 6. TOT ↑, EEG theta band power ↑
Mental state estimation	Physiological computing	7. TOT-based classification > performance-based classification 8. Feature fusion classification accuracy ↑ 9. Inclusion of motor artifacts in EEG classification accuracy ↑

Tab. 8.1 Summary of hypotheses

gave their written informed consent in accordance with the Declaration of Helsinki. This study was carried out under the recommendations of the Local Research Ethics Committee (redacted for double-blind review).

II.8.2.2 Procedure

First, participants were asked to read the information sheet and sign the consent form. Then, the ECG and EEG electrodes were placed. The participants were asked to sit comfortably in their chair and face the screen. The eye tracker was placed according to the participants' position to frame their gaze and be at the right distance. All streams were synchronized via the Lab Streaming Layer (LSL, Kothe et al., 2014) and recorded with LabRecorder. The task, instructions, and all the questionnaires – in the participants' native language– were launched automatically on Matlab 2021a[2]. The experiment began with a demographic questionnaire. To estimate their sleepiness and feeling of fatigue, the *Karolinska Sleepiness Scale* (KSS, Åkerstedt and Gillberg, 1990) and *Samn-Perelli Fatigue* (SPS, Samn and Perelli, 1982) questionnaire were used. All questionnaires are detailed in Appendix V.14.2.

A 1-minute training session with instructions was presented to the participants to familiarize them with the task. Before beginning the blocks, participants were asked to indicate their level of fatigue using two visual analogue scales (VAS, see Appendix V.14.2.6).

Participants then performed four 15-minute blocks of the experimental task. Before each block, the eye-tracker was re-calibrated if necessary. After each block, participants received feedback on their performance, and filled in the two VAS scales. This was followed by a 1-minute rest, after which the participants started the next block. Upon completing the last block, participants again filled in the KSS and SPS scale. Next, the sensors were removed from the participants. Finally, they were paid and thanked for their participation.

[2]https://fr.mathworks.com/products/matlab.html

II.8.2.3 Materials

EEG

EEG data were recorded using a 64-active AG-AgCl electrode system with an ActiCHamp amplifier (Brain Products, Gmbh). Electrodes were placed using the international 10-20 system (Jasper, 1958). Data were recorded at 500 Hz and impedances were kept below 50 kΩ.

ECG

A 3-electrode auxiliary channel for the ActiCHamp recorded the participants' electrocardiogram. The data was recorded at 500 Hz.

Eye-tracker

Eye movements were recorded using a Gazepoint GP3 stationary eye tracker at 60 Hz[3]. The eye-tracker was calibrated using 5 points at the beginning of the task, and in between blocks if necessary.

Target detection task

The Target Detection task (Figure 8.1) was presented in four 15-minute blocks separated by a one-minute break. In the task, 16 (4x4 grid) images extracted from the NORB dataset (LeCun et al., 2004) were presented at all times. These images were recently used for studying pilot-UAV interaction and mental state estimation (Singh, Chanel, & Roy, 2021). Participants were instructed to click on pictures containing humans (i.e. survivors in a search and rescue task) as fast as possible. A single image was updated every 250-1250 ms. The target-to-non-target ratio of images was 1/5. The first 16 images and the last five images of each block contained no targets, as these would have been almost impossible to correctly identify and click on at the beginning because the image would have changed rapidly and at the end because the task stops in a few seconds. Reaction times, as well as accuracy, were recorded. The task was programmed using MATLAB 2021a, and synchronized using LSL (Kothe et al., 2014).

II.8.2.4 Database quality assessment

This section details the methods used to assess whether the data reflects the expected behaviour of the participants with regard to mental fatigue.

Subjective Data Analysis

The SPS and KSS scores from before and after the task were compared using a two-sided paired samples t-test. The VAS-S and VAS-F, acquired before the first block and then following the completion of each block were compared using multiple one-way ANOVAs.

[3]https://www.gazept.com/product/gazepoint-gp3-eye-tracker/



Fig. 8.1 Screenshot of the target detection task created using images extracted from the NORB database (LeCun et al., 2004).

Behavioral analysis

To assess behavioral performance mean reaction times of each correct click for each block were divided into 3 timepoints (of 5 min each) and compared using both block and timepoint as a factor. The same was done for accuracy. Repeated measures ANOVAs were used.

Preprocessing & Feature Extraction

ECG Peak detection from the continuous data of each epoch was performed by the Heplab slow detect function (algorithm detailed in (de Carvalho et al., 2002)). Three metrics were then calculated: the heart rate (HR) in beats per minute (BPM), the root mean square of successive differences between normal heartbeats (RMSSD), as a measure of HRV, and the standard deviation of normal sinus beats (SDNN), another metric of HRV (see Figure 8.2).

EEG The EEG data were preprocessed using a 50 Hz notch filter, thanks to the CleanLine[4] toolbox with default parameters (see Figure 8.2). The average power of the alpha (8-12 Hz),

[4]https://www.nitrc.org/projects/cleanline

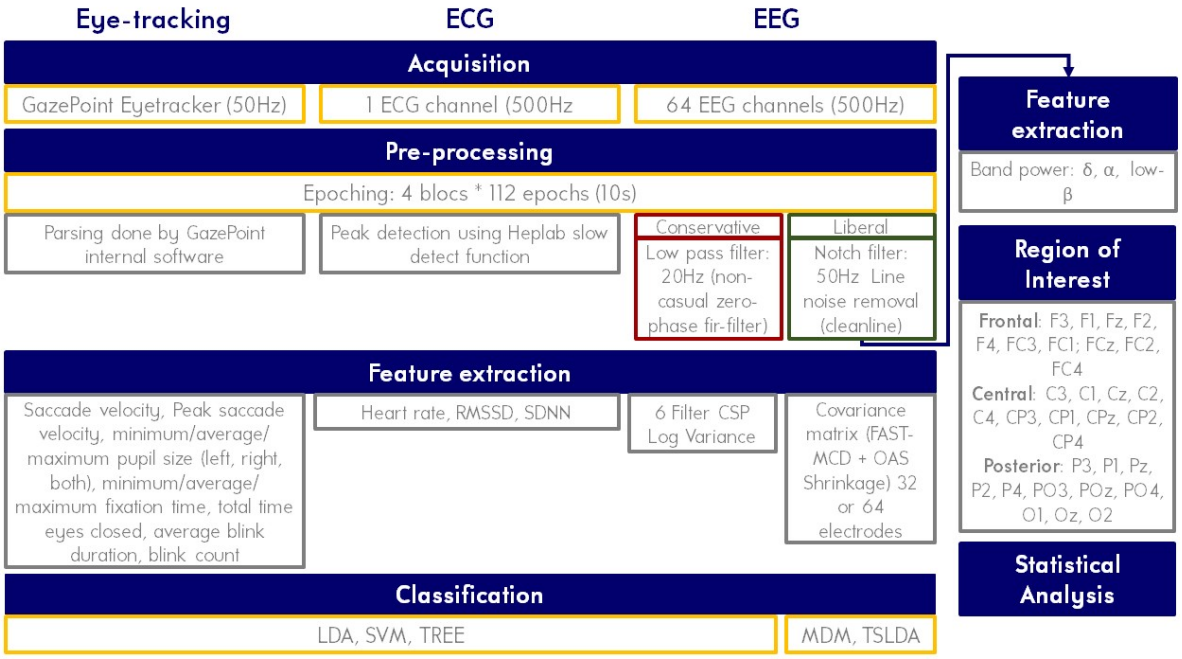

Fig. 8.2 Schematic view of the data processing pipelines evaluated in this study. The conservative pipeline discardes frequencies that may be contaminated by muscle artefacts (above 20Hz). The liberal pipeline includes these frequencies.

theta (4-8 Hz), and low beta (13-20 Hz) were extracted for analysis from the following electrode clusters or Regions of Interest (ROI):

- Frontal: F3, F1, Fz, F2, F4, FC3, FC1, FCz, FC2, FC4
- Central: C3, C1, Cz, C2, C4, CP3, CP1, CPz, CP2, CP4
- Parietal: P3, P1, Pz, P2, P4, PO3, POz, PO4, O1, Oz, O2

Statistical Analysis

To objectively assess the database's quality and analyze the ECG and EEG data at the group level, repeated measures ANOVAs were performed with block and timepoint as factors and the region of interest (ROI). Timepoints pertain to 5-minute segments of each block to obtain a higher temporal precision as well as to allow the investigation of the presence of any effects that occur within rather than between blocks. Furthermore, a supplementary analysis compared the best and worst trials for each EEG frequency band.

II.8.2.5 Mental fatigue estimation

Overview

The following subsections will detail the preprocessing and feature extraction steps of the EEG, ECG, and Eye-Tracking data, before highlighting the different classification labels and classi-

fiers that we evaluated in this study (tables 8.2 and 8.3 & figure 8.2). All data manipulation and classification steps were performed using MATLAB 2021a with the EEGlab[5]; covariancetoolbox[6] and Heplab[7] extensions. Preprocessing and feature extraction were performed on an epoch-to-epoch basis for all sensors in order to avoid spillage of information from training into testing sets (or vice versa). Epochs were 8 seconds long. Two factors informed this decision. Robust estimation of heart rate variability (HRV) metrics require around 10 seconds to be estimated reliably (Salahuddin et al., 2007). Furthermore, the CleanLine Notch filter proposes a 4-second sliding window, requiring the epoch length to be divisible by four (Delorme & Makeig, 2004). 8-second epochs produce 25% more trials, which is advantageous for classification. Five Participants were excluded from the mental fatigue estimation analysis due to too many missing data points.

Preprocessing & Feature Extraction

ECG The same preprocessing steps and features as for the database quality assessment were used (see Fig. 8.2 for a summary of the machine learning pipelines).

EEG Two different preprocessing pipelines were employed: i) A conservative pipeline aimed towards only keeping cerebral activity from the EEG data, and ii) a second, more liberal pipeline that did not account for motor activity artefacts. The goal of using two preprocessing pipelines was to examine to what extent motor artefacts impact EEG data classification (see Hypothesis. 9 in Table 8.1). In both cases, the preprocessing was kept minimal to keep as much signal as possible and to avoid introducing great computational and temporal demands for the preprocessing and classification steps. The conservative pipeline only used a low pass filter set to 20 Hz (non-causal zero-phase fir-filter with a passband edge of 20 Hz and a cutoff frequency of 22.5 Hz). In the liberal pipeline, only a 50 Hz notch filter was employed to clean the line noise. The CleanLine toolbox with default parameters was used.

In both pipelines, the same metrics were extracted. Covariance matrices were estimated using the FAST-MCD (Minimum Covariance Determinant) method by Rousseeuw and Driessen (1999). OAS shrinkage was applied to the covariance matrix (Chen, Wiesel, et al., 2010). This combination allows for a robust estimation of the covariance matrix, especially if the number of observations is small as compared to the number of electrodes. These calculations were performed using either the 64-electrode grid or a 32-electrode subset following the standard 20-10 convention (Jasper, 1958). The use of 32 and 64 electrodes tested the value of added electrodes. Using 32 electrodes reduces the number of variables in a covariance matrix from 2080 to 400, again allowing for a more robust estimation of the covariance matrix, but there is a loss in spatial information as the electrode grid is less dense. Alternatively, a common spatial pattern algorithm was applied to the continuous EEG data of each epoch (Casso et al., 2021). The six most discriminating CSP filters were kept, and their log variances were used as features for classification. The spatial filters were estimated based on the training set and then applied to the classification data.

Eye-Tracking Data from the eye tracker required little to no preprocessing. Minimum, maximum, and average fixation times were calculated for each epoch. Fixations were detected by

[5]https://sccn.ucsd.edu/eeglab/index.php
[6]https://github.com/alexandrebarachant/covariancetoolbox
[7]https://pandelisperakakis.info/heplab/

the GazePoint eye-tracker and defined as the absence of eye movement. However, there is no explicit information about GazePoint's thresholding. Minimum, maximum, and average pupil size for each eye and the average of both eyes were used as metrics. The total time of eyes closed, average blink duration, and overall blink count were calculated. For the classification, either all metrics or a subset were used. The subset (from here on referred to as the "reduced" eye-tracker) omitted six pupillometry metrics (minimum, maximum, and average for the left and right eye separately), only keeping the average of both eyes, hence reducing the dimensionality by six dimensions. As the Gazepoint eye-tracker does not supply any saccade metrics, we created a simple method for saccade detection detailed in Appendix V.14.3.

Classification labels

One primary goal of this paper is to examine the differences between assigning classes to data (labelling) based on the TOT or the subject's performance. To compare both approaches, the data were labelled in 9 different ways. All possible binary combinations of the four experimental blocks were used to create 6 out of the nine labels (e.g. 1vs3 corresponds to block 1 vs block 3). The performance-based labels used reaction time to define performance. Trials of each participant were first ordered by reaction time and then labelled as "strong performance" if the average reaction time during the trial was in the lowest 20, 30, or 50 % of all trials. If a trial was in the highest 20, 30, or 50 %, the data was labelled as "poor performance". For the 20 & 30 % labels, the data that was neither classified as strong nor poor was disregarded.

This information is summarized in table 8.2. Adjusted chance levels were calculated according to the number of trials in each class following the work of Mueller-Putz et al. (2008):

Label	Class	Description	Trials per class	Adjusted chance Levels (%)		
				.05	.01	.001
1vs3	TOT	Block 1 vs 3	112	61.36	68.18	72.73
1vs4	TOT	Block 1 vs 4	112	61.36	68.18	72.73
2vs4	TOT	Block 2 vs 4	112	61.36	68.18	72.73
1vs2	TOT	Block 1 vs 2	112	61.36	68.18	72.73
2vs3	TOT	Block 2 vs 3	112	61.36	68.18	72.73
3vs4	TOT	Block 3 vs 4	112	61.36	68.18	72.73
20%	Perf	Best Worst 20%	90	63.88	69.44	75
30%	Perf	Best Worst 30%	140	60.71	66.07	69.64
50%	Perf	Best Worst 50%	224	58.43	61.79	66.29

Tab. 8.2 Summary of all labels with the adjusted chance levels for several p-values and number of trials per class per participant, calculated according to Mueller-Putz et al. (2008).

Classification

Five different classifiers were used for evaluation in an intra-subject manner: 3 classifiers – Linear Discriminant Analysis (LDA), Support Vector Machine (SVM), and Decision Tree (TREE) – were used on all data (EEG, ECG, and Eye-Tracking), while the Minimum Distance to Mean

(MDM)[8] and the Tangent space LDA (TsLDA) were only used for the covariance matrices calculated from the EEG Signal. All of these algorithms have shown proficiency in detecting mental states using physiological data (Hu & Min, 2018; Roy et al., 2013; Zhao et al., 2011).

The data were split into a training and validation set using an 80/20 ratio. The hyperparameters of the classifiers were then optimized using Bayesian optimization. The optimized classifiers were tested using a 5-fold cross-validation of the training set. Once the best hyperparameters were obtained, the classifiers were retrained using the entire training dataset. For the LDA, both the delta and gamma parameters were optimized. Standardization, polynomial order, box constraint, kernel scale, and kernel function were optimized for the SVM. Finally, the optimizable hyperparameters of the Tree classifier were the maximum number of splits, the minimum leaf size, the number of variables to sample, and the split criterion. The TsLDA and MDM were not optimized. The implementation of the TsLDA used here did not allow for the optimization of hyperparameters and the MDM does not have any hyperparameters. As this is a binary classification problem with balanced classes, the overall accuracy was chosen as an evaluation metric for the performance of each combination of classifiers and metrics.

Analysis of mental fatigue estimation performance

As highlighted in table 8.3, three separate analyses of mental fatigue estimation were performed:

- A general analysis of all pipelines and labels: Here a 2-way ANOVA (Pipeline *labels) was employed. The goal was to determine the overall best pipelines.
- A comparison of all pipelines using the LDA, SVM, and TREE classifiers: Factors included in the 3-way ANOVA were classifier, label, and metric.
- Finally, a similar analysis was performed with the classifiers using Riemannian methods: Using a 4-way ANOVA the effects of classifier, preprocessing, electrode count, and label were examined.

These 3 separate analyses were necessary as not all combinations of all levels of each factor were possible (e.g. TsLDA with ECG data).

II.8.3 Results

II.8.3.1 Subjective measures

As illustrated by Figure 8.3 (left), both the SPS and KSS scores indicated significant increases in subjective measures of fatigue (KSS: $t(14) = -4.599, p < .001$, SPS: $t(14) = -4.294, p < .001$). Cognitive fatigue (VAS-F) and self-rated sleepiness (VAS-S) significantly increased following the completion of the task ($F(1.67, 15.07) = 16.56 p < .001, \eta_p^2 = .65$). Results are summarized in Figure 8.3 Bottom (see Hypothesis. 1 in Table 8.4).

[8] The MDM classifier is a simple classifer that determines the class of the data, by determining which class mean is closest to the datapoint.

Metrics	Dim	Classifier					Description
		LDA	SVM	Tree	MDM	TsLDA	
EEG(CSP) (Conservative)	6	✓	✓	✓			CSP based EEG metrics, conservative preprocessing
EEG(CSP) (Liberal)	6	✓	✓	✓			CSP based EEG metrics, liberal preprocessing
EEG(Covariance) (Conservative) 64 electrodes	64*64				✓	✓	Covariance based EEG metrics, conservative preprocessing
EEG(Covariance) (Liberal) 64 electrodes	64*64				✓	✓	Covariance based EEG metrics, liberal preprocessing
EEG(Covariance) (Conservative) 32 electrodes	32*32				✓	✓	Covariance based EEG metrics, conservative preprocessing
EEG(Covariance) (Liberal) 32 electrodes	32*32				✓	✓	Covariance based EEG metrics, liberal preprocessing
EyeTracker (full)	17	✓	✓	✓			All Eye-Tracking metrics
EyeTracker (red)	11	✓	✓	✓			Reduced Ey-Tracking metrics
ECG	3	✓	✓	✓			All ECG metrics
EEG(CSP) + EyeTracker (full)	23	✓	✓	✓			CSP based EEG metrics and all Eye-Tracking
EEG(CSP) + EyeTracker (red)	17	✓	✓	✓			CSP based EEG metrics and reduced Eye-Tracking metrics
EEG(CSP) + EyeTracker (red)	17	✓	✓	✓			CSP based EEG metrics and reduced Eye-Tracking metrics
EyeTracker (full) + ECG	20	✓	✓	✓			All Eye-Tracking and ECG metrics
EyeTracker (red) +ECG	14	✓	✓	✓			Reduced Eye-Tracking metrics and all ECG metrics
EEG(CSP) + EyeTracker (full) + ECG	26	✓	✓	✓			CSP based EEG metrics, all Eye-Tracking and all ECG metrics
EEG(CSP) + EyeTracker (red) + ECG	20	✓	✓	✓			CSP based EEG metrics, reduced Eye-Tracking metrics and all ECG metrics

Tab. 8.3 All combinations of classifiers and metrics tested with feature vector dimensions. Note: The Riemannian methods (MDM and TsLDA) used a covariance matrix of all electrodes (32 or 64).

II.8.3.2 Behavioral measures

As illustrated by Figure 8.4, there was a significant behavioral effect of the blocks on the percentage of correct answers ($F(1.85, 25.88) = 9.597.37, p < .001, \eta_p^2 = .41$). Participants had lower scores during block 1 compared to the other blocks (B1-B2 $MD = -5.57, p < .001$; B1-B3 $MD = -5.54, p < .001$; B1-B4 $MD = -5.84, p < .001$). Furthermore, there was a significant effect of timepoint on the percentage of correct answers ($F(2, 28) = 8.86, p < .01, \eta_p^2 = .39$). The score was higher for the first timepoint of each block (T1-T2: $MD = 3.85, p < .01$; time1-time3: $MD = 4.50, p < .01$). Lastly, the mean reaction time showed a significant effect of timepoint ($F(2, 28) = 7.37, p < .01, \eta_p^2 = .34$) with significant differences between timepoints T1-T2 and T1-T3 (T1-T2 $MD = -0.22, p < .01$, T1-T3 $MD = -0.24, p < .01$, see Hypothesis 2 in Table 8.4).

II.8.3.3 Physiological measures

ECG

Figures 8.5 & 8.6 illustrate the significant decrease in HR across blocks ($F(30, 3) = 5.401, p < .005, \eta_p^2 = .351$), however within blocks the HR increased

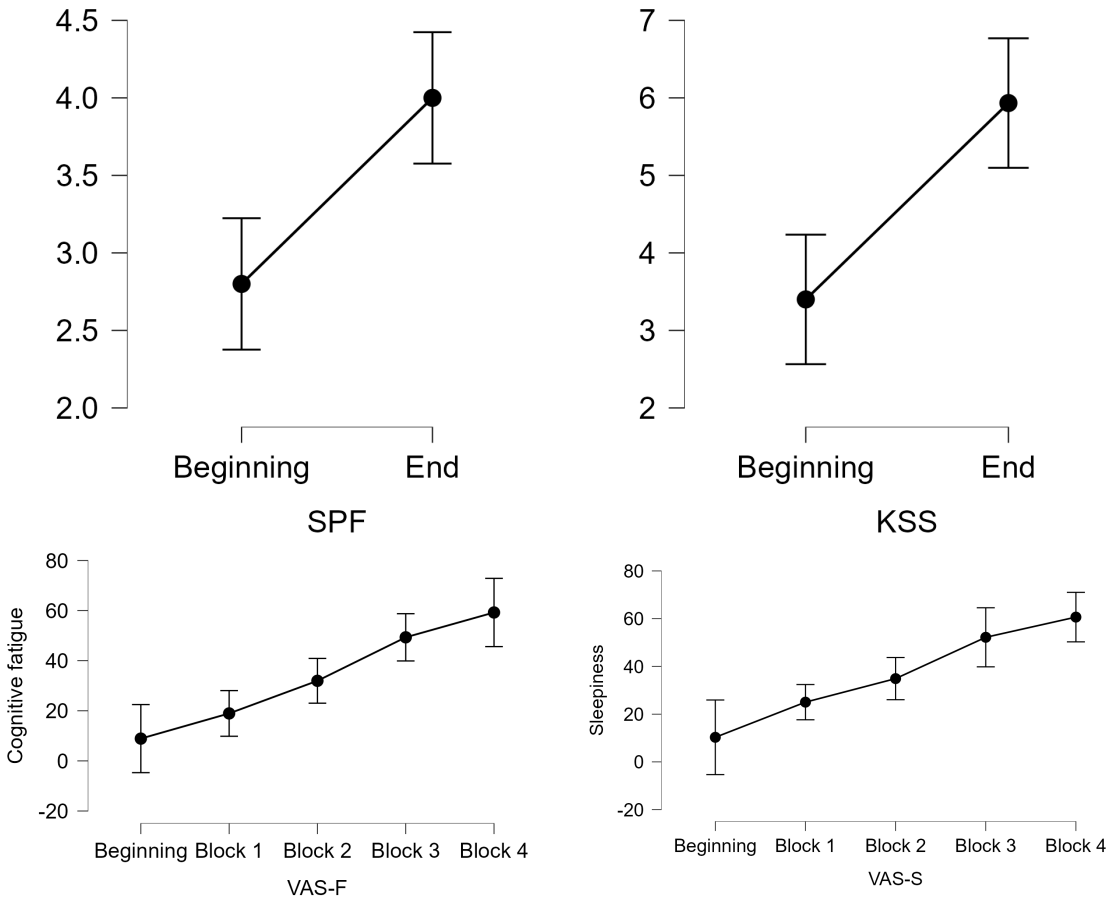

Fig. 8.3 Summary of subjective results: top left: Samn-Perelli Scale of Fatigue (Samn & Perelli, 1982); top right: Karolinska Sleepiness Scale (Åkerstedt & Gillberg, 1990); bottom left: Self-rated cognitive fatigue (VAS-F); bottom right: Self-rated sleepiness (VAS-S).

($F(30,2) = 4.68, p < .05, \eta_p^2 = .319$). HRV (RMSSD and SDNN) increased across blocks without any significant change within blocks (RMSSD: $F(30,3) = 3.336, p < .05, \eta_p^2 = .250$; SDNN: $F(30,3) = 8.061, p < .001, \eta_p^2 = .446$, see Hypothesis 3&4 in Table 8.4).

EEG

The EEG signal (Figures 8.7 & 8.6 bottom) showed no significant difference across blocks or timepoints on EEG power. The main effect of ROI was significant for all three frequency bands. The frontal ROI displayed the lowest power, followed by the central and parietal ROIs (alpha: $F(36,3) = 18.93, p < .001, \eta_p^2 = .612$, theta $F(36,3) = 45.185, p < .001, \eta_p^2 = .659$, low beta $F(36,3) = 26.599, p < .001, \eta_p^2 = .689$). A comparison between epochs of best and worst performance showed no significant impact of performance on the EEG signal (alpha: $F(36,1) = 0.763, p > .3, \eta_p^2 = .060$, theta $F(36,1) = 0.324, p > .5, \eta_p^2 = .026$, low beta $F(36,1) = 1.961, p > .2, \eta_p^2 = .138$, see Hypotheses. 5 & 6 Table 8.4).

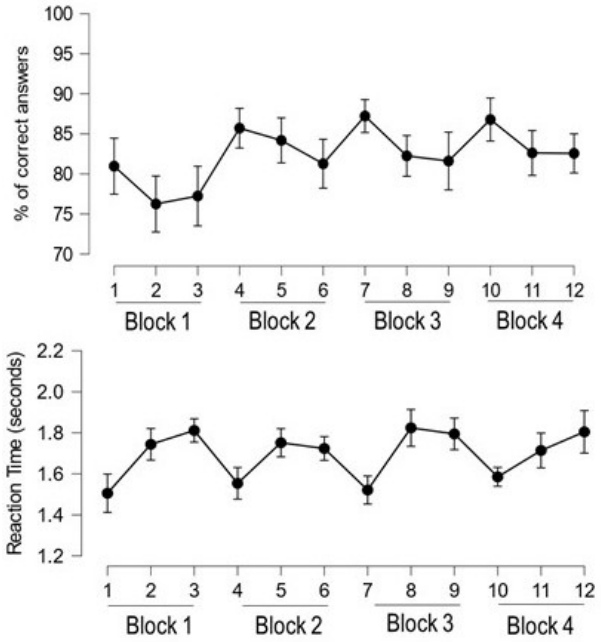

Fig. 8.4 Top: Performance accuracy across timepoints & blocks; Bottom: Average reaction time of the first correct answer across timepoints & blocks.

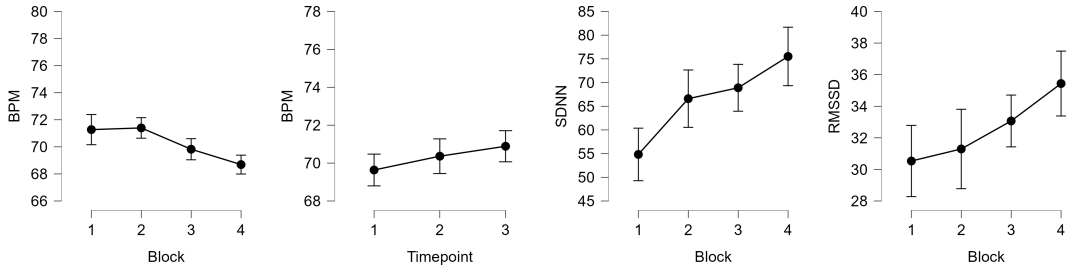

Fig. 8.5 Overview of cardiac results: far left: Heart Rate (HR) across blocks; center left: Heart Rate (HR) across timepoints; center right : Heart Rate Variability (HRV, SDNN) across blocks; far right: Heart Rate Variability (HRV, RMSSD) across blocks.

II.8.3.4 Mental fatigue estimation

Overall results

There were significant main effects of both the pipeline ($F(43,9) = 21.36, p < .001, \eta_p^2 = .205$) and label ($F(8,9) = 640.85, p < .001, \eta_p^2 = .59$) factors (Figures 8.8 & 8.9 Left). The interaction was also significant ($F(34,9) = 1.9, p < .001, \eta_p^2 = .155$). Regarding the different labels, the results clearly show that performance-based labelling results in chance-level accuracies, while TOT labelling results in elevated accuracies. Furthermore, the effect of TOT on classification accuracy is supported by the observation that a longer interval between TOT results in higher accuracies (label 1vs2 vs label 1vs3: $t(14) = -8.91, p < .001$). Block 1 vs 4 had the highest average accuracy of 84.49% while Block 1 vs 2 had an average accuracy of 74.59 (strongest performing label of two consecutive blocks, $t(14) = 12.7, p < .001$)). As for the pipelines, several observa-

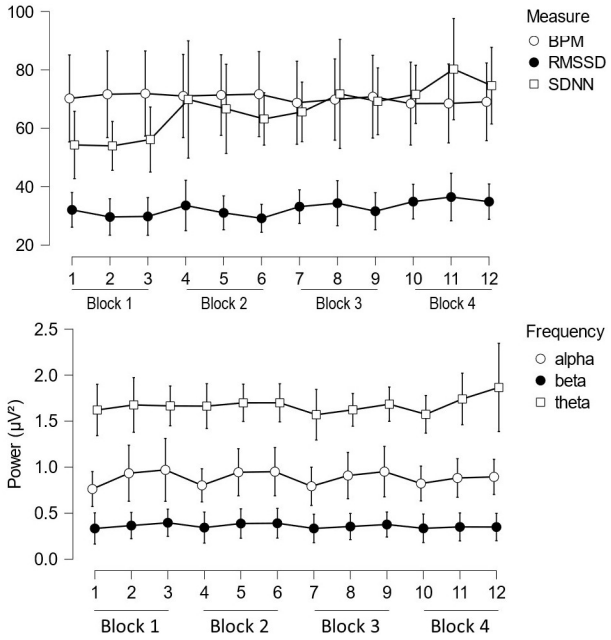

Fig. 8.6 Top: Summary of ECG results, showing heart rate (HR in BPM) & HRV (SDNN, RMSSD); Bottom: Summary of the spectral power EEG power from all frequency bands across blocks/timepoints.

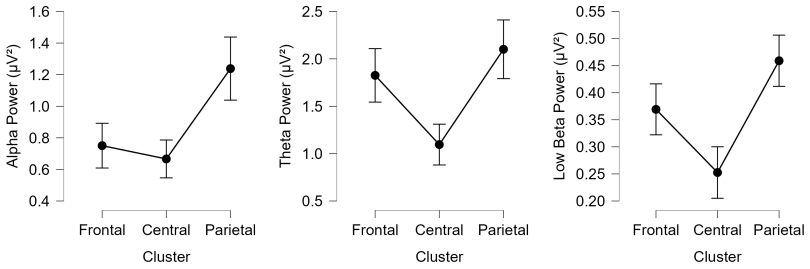

Fig. 8.7 Summary of the EEG frequency powers across clusters. Left: Average alpha power; Center: Average theta power; Right: Average low beta power.

tions may be made. EEG-based classification had the strongest performance with up to 99.32% accuracy (LDA, SVM EEG with liberal preprocessing). The TsLDA classifier performed similarly to the CSP-based EEG pipelines (e.g.: TsLDA Liberal 64 electrodes vs SVM EEG Liberal $t(14) = 0.79, p = 1.00$). The weakest pipelines relied solely on ECG, followed by the pipelines using EEG and ECG together. While the eye-tracker on its own achieved accuracies of up to 86.36%, the addition of either EEG or ECG showed a tendency of slightly increased accuracy (e.g.: LDA EyeTracker (full) + ECG vs LDA EyeTracker (full) $t(14) = 1.12, p < .001$ see Hypothesies 7 & 8; see Table 8.4).

Non-Riemannian classifiers

As illustrated by Fig. 8.9 Center, there was a significant main effect of classifier ($F(2,9) = 4.09, p < .017, \eta_p^2 = .003$). The SVM was the strongest one ($t(9) = -2.86, p < .05$). This trend is robust across all different metrics ($F(11,9) = 40.90, p < .001, \eta_p^2 = .0.134$) and labels

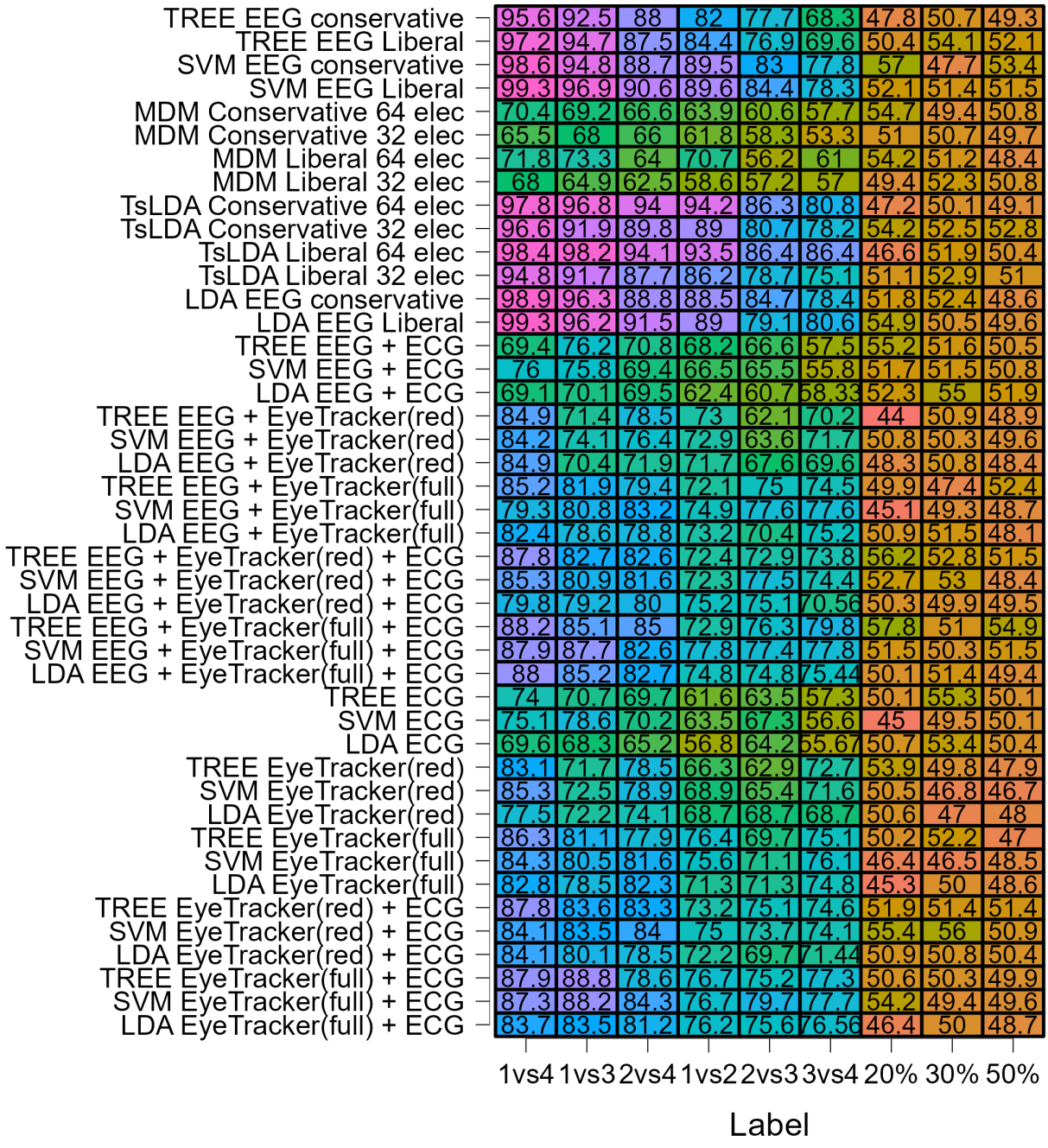

Fig. 8.8 Heatmap of mental fatigue estimation results. All pipelines are displayed on the Y axis with labels on the X axis (the first 6 labels for TOT, the last 3 for performance estimation). Values represent mean accuracies in percent.

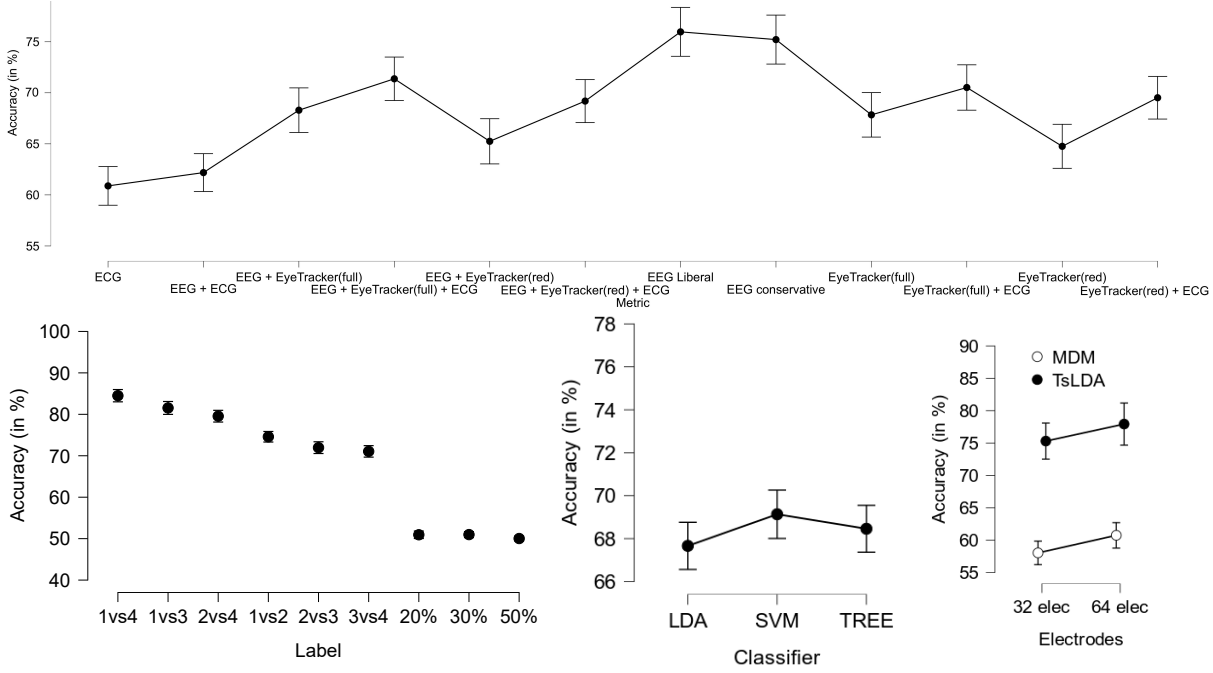

Fig. 8.9 Top: Impact of the choice of metrics for non-riemannian classifiers. Left: Impact of label type ; Center: Impact of classifier (non-Riemannian) on accuracy. Right: Impact of electrode setup –32 or 64– and Riemannian classifier.

($F(8,9) = 499.08, p < .001, \eta_p^2 = .578$). The results regarding the differences between labels are preserved in the non-Riemannian classifiers (see Figure 8.9 Top).

Riemannian classifiers

MDM and TSLDA differed significantly from each other ($F(1,9) = 567.36, p < .001, \eta_p^2 = .467$). The MDM had an inferior performance. Using 64 as compared to 32 electrodes improved performance ($F(1,9) = 13.61, p < .001, \eta_p^2 = .021$; Figure 8.9 Right). Preprocessing using the conservative or liberal method had no significant impact on the classification accuracy ($F(1,9) = 0.0014, p > .5, \eta_p^2 = .000$ see Hypothesis 9 Table 8.4). The differences between labels were preserved.

II.8.4 Discussion

In this experiment, pBCIs and physiological computing systems were tested on participants performing a 1-hour visual search task, emulating some of the tasks that UAS operators are confronted with. This experiment aimed to investigate to what extent EEG, ECG, and eye-tracking can be used to estimate mental fatigue and, more specifically, to evaluate the impact of basing the classification labels on TOT or operator performance. As expected, the protocol design elicited mental fatigue in the participants, as revealed by the KSS, SPS, and VAS scales. Indeed, the questionnaires indicated a linear increase in subjective fatigue and sleepiness over time (see Hypothesis 1 Table 8.4). However, contrary to our expectations, the behavioural analysis revealed a relatively stable performance across blocks 2-4, following an increase in perfor-

Goal	Data	Hypothesis	Validated
Data validation	Subjective	1. MF reports ↑ with TOT	✓
	Behavioral	2. TOT ↑, performance ↓	X
	Physiological	3. TOT ↑,HR ↓	✓
		4. TOT ↑, HRV ↑	✓
		5. TOT ↑, EEG alpha power ↑	X
		6. TOT ↑, EEG theta power ↑	X
Mental state estimation	Physiological computing	7. TOT-based classification > performance-based classification	✓
		8. Feature fusion ↑ classification accuracy	✓
		9 . Inclusion of motor artifacts in EEG ↑ classification accuracy	X

Tab. 8.4 Summary of hypotheses along with an additional column that summarizes whether each hypothesis was confirmed (✓) or rejected (X)

.

mance in block one (see Hypothesis 2 Table 8.4). The latter is most likely attributable to learning effects. The stable performance across blocks 2-4 can be attributed to the experimental design, which allowed for ceiling effects. The performance decrement observed by the reduction of accuracy within time points shows an effect of fatigue on the task. Breaks between tasks may have resulted in the participant being able to recover from said effects. Based on the literature and our hypotheses, analysis of the ECG metrics showed decreases in heart rate and increases in HRV. Furthermore, the increase in Heart rate within blocks may be explained by a decrease in vigilance, necessitating more mental workload within a task (see Hypotheses 3 & 4 Table 8.4). Yet the EEG results showed no significant difference across blocks. The lack of significant EEG results may partly be explained by the ceiling effects of the participant's performance (see Hypotheses 5 & 6 Table 8.4). As shown by the subjective measures, the task nevertheless elicited mental fatigue as measured by four scales (KSS, SPS, and VAS questionnaires, see Appendix V.14.2.4, V.14.2.5 & V.14.2.6). As discussed in Section I.3.1.5, the absence of behavioural indicators of mental fatigue does not necessarily indicate an absence of mental fatigue. In addition, the small sample size is perhaps why the within-block variations of frequency power bands were not significant (see Fig. 8.6 Bottom). Yet, for intra-subject estimation, the number of trials and epochs was sufficient to achieve interesting results.

Multiple pipelines using EEG, ECG, and eye-tracking and, most notably, two labelling data systems were employed to estimate each participant's current state of mental fatigue. Time-on-task-based MF estimation showed strong results, especially using EEG data. Both the Riemannian Tangent space LDA and the regular LDA with common spatial filtering achieved accuracies of over 90%. Using only ECG data (heart rate and heart rate variability metrics) had the weakest performance with accuracies ranging from 55% - 78% depending on the classifier and which blocks were used for labelling (see Hypothesis 7 Table 8.4). Moreover, the results presented here show that using an eye-tracker combined with an ECG can result in a minimal viable system that achieves an accuracy of up to 88% (see Hypothesis 8 Table 8.4). The aforementioned results used TOT as the method of labelling. Shifting attention towards the performance-based labels, a very different picture emerges. Using TOT, many pipelines far

exceed the minimum chance level threshold, while performance-based labelling results in only a few pipelines narrowly surpassing the $p < 0.05$ thresholds of 53-56%. What are some potential reasons why the same pipelines produce so vastly different results? The ceiling effects of performance may indicate a lack of sufficient difference between good and bad phases during the task. This could mean that the classifiers detect no difference between the data as there is none. While there is most likely some truth, this does not explain why the classifiers work when based on TOT labels. This counterargument is further supported by the results of the physiological data indicating no significant differences between blocks.

Assuming, due to the stable behavioural performance of participants, no real difference across time or phases of performance occurred within the participant's brain. This raises the question of the origin of the discriminable features used by the TOT classifiers. Instead of changes in the brain due to mental fatigue, these could be slow drifts of the signal due to many factors (Kaya, 2019; Urigüen & Garcia-Zapirain, 2015). The inherent non-stationarity of the EEG signal supports this argument.

II.8.4.1 Limitations

It should be noted that the results of this study are limited due to several factors. A relatively low sample size will likely have impacted the statistical analysis. The task difficulty should be adapted in future studies to avoid ceiling effects of performance. These effects may have also biased the machine learning pipelines, especially performance-based ones. This should be expanded upon in future works by testing different tasks, allowing for better performance discrimination. Eight-second epochs represent a very short period for the reliable estimation of HRV metrics (Salahuddin et al., 2007; Shaffer & Ginsberg, 2017). This choice allowed for a better comparison between sensors. The results show that the classification algorithms could extract valuable information from these metrics.

Preliminary results of this contribution were presented at the 2023 BCI Meeting. The complete results are currently under review for the BCI Meetings special edition of the Journal of Neural Engineering (see Appendix 13, Hinss, Jahanpour, Brock, and Roy, 2024; Hinss, Jahanpour, Brock, and Roy, 2023).

Chapter 9

UASOS: Moving towards ecological validity

Central Ideas

Passive brain-computer interfaces (pBCIs) developed within the neuroergonomic field frequently aim to improve safety by augmenting human-machine interaction. To accomplish said goal, many pBCIs classify mental states such as mental workload or mental fatigue. An alternative is to forego mental states and aim to predict performance. Despite its drawbacks, we argue that performance estimation is a more goal-oriented approach than mental state estimation. In a realistic experiment, 25 participants had to control an uncrewed aerial system for two hours, continuously switching between target search and navigation. EEG classification accuracies based on mental states and performance were compared. With a Tangent Space Logistic Regression, we could predict an increased likelihood of lapses in the form of missing instructions with an above-chance level accuracy of 62.09 %.
Recommended Reading:

- The Double Task Switching Protocol (Chapter 6)
- Mental fatigue (Section I.3.1)
- Cognitive flexibility (Section I.3.2)
- Sensors (Chapter 4)
- Machine Learning (Section I.5.1)
- Physiological Computing (Section I.5.2)

One Question to ask: **Can we recreate a task-switching protocol in a more ecological setting and can we predict performance using EEG?**

II.9.1 Motivation and Research Goals/ Questions

UAS operations oftentimes revolve around two aspects: navigation and visual search (see Sections I.2.1.3 & I.2.1.5)[1]. We have also seen that these operations are impacted by mental fatigue and its impact on cognitive fflexibility (see Sections I.2.1.4 & I.3.1.4). The Double-Task Switching protocol showed us that switching between more dissimilar tasks is more difficult than switching between similar tasks (see Chapter 6).

The results of Remot3e highlighted the difficulties in performance estimation using physiological sensors when compared to the TOT approach (see Chapter 8).

This contribution aimed to create an experimental paradigm/ task that accurately captures the fundamental aspects of UAS operations with regard to mental fatigue, cognitive flexibility, navigation and visual search. To do so, it would have sufficed to have a setup where participants alternate between a navigation task and a search task. The underlying assumption is that task-switching costs are not only attributable to participants switching between screens and modalities of response but also partially due to the dissimilar (or external) task switch. Instead, we recreated the main aspects of the double-task switching protocol by having two main tasks with two subtasks or modes each. This creates a much more realistic and challenging task for the participant. It also allows us to validate the results of the DTS study further (although we acknowledge that switching modalities of response and information presented on different screens are confounding variables).

Beyond recreating UAS operations (UAS OPS), we also aimed to investigate further the differences between TOT and performance-based data labelling for pBCIs. As previously explained in Section I.5.2.2 mental state estimation may have significant drawbacks as compared to performance estimation.

This experiment tested whether an EEG-based pBCI can predict meaningful performance metrics from participants performing a UAS piloting task. We aim to illustrate how label assignment impacts classification accuracies, particularly in the absence of an absolute performance definition, for such a realistic task.

II.9.2 Task Development

II.9.2.1 Overview

UASOS[2] (Uncrewed Aerial Systems Operator Simulator) was built to assess the effects of mental fatigue on operator performance during long-endurance UAS missions. This simulator is not the first for UAS missions, and software such as AirSim provides a highly realistic and complex simulation (Shah et al., 2017). The advantage of UASOS is the ability to control all variables in the simulation precisely, allowing for reproducible research designs. To allow for the study of cognitive flexibility and task-switching, UASOS consists of two concurrent tasks in which participants must fly a UAS and search for survivors. These tasks are modelled after realistic military operations and environments.

In the search task (SRC), participants are instructed to search for humans or vehicles on a 3x3 grid of images taken from the NORB database (LeCun et al., 2004). Target pictures are

[1] The development of the UASOS task was done by Nhat Tien Phan and Vincenzo Maria Vitale who completed a 3-month and 6-month Internship at ENAC and iSAE-SUPAERO respectively. Marcel F. Hinss, Raphaëlle N. Roy and Anke M. Brock supervised the students

[2] https://github.com/Marcels-2-Neurons/uasos-project

selected with a corresponding click on a numpad. Optical filters can distort the pictures using a custom *imgtreat.py* module powered by OpenCV modules (Bradski, 2000). The two modes of the SRC task require participants to either search for humans present on the images or to search for ground vehicles.

In the Navigation task (NAV), participants must follow instructions by either using a joystick and flying towards a specific heading or using a trackball and selecting the appropriate waypoint. In the heading mode of the NAV task, participants receive a heading instruction (e.g. 350) in each trial. Using a joystick, they then turn the UAS in said direction. Participants are instructed to turn the UAS in a direction that minimizes the angle of the turn (e.g. when heading north and an eastern heading is requested, a turn towards the left is faster than turning west). For the WPY mode, they receive a waypoint consisting of a letter and a number (e.g. F13). They must choose the corresponding waypoint using a trackball on a grid overlaying the navigational display. To allow experimenters to investigate cognitive flexibility, UASOS is implemented to have participants perform one task at a time, with frequent instructions to switch between tasks. For all tasks, reaction times and the correctness of responses are recorded. The instructions on the flight director widget tell the participant which mode to perform at the onset of each trial. This design is based on the Double Task Switching Protocol (see chapter 6, Hinss, Brock, and Roy, 2023a).

II.9.2.2 Characteristics

UASOS was built with Python 3.8.10 using PsychoPy[3] (Peirce et al., 2019),

A novel aspect of the UASOS is the Driver/Driven Gear architecture, allowing for a synchronized exchange of status and data packets between both the Search (*SRCWin*) and the Navigation Task Windows (*NAVWin*). Two separate instances communicate through a Remote Object *PyroServer* in the middle exchanging 2 types of packets (see supplementary Appendix V.14.5):

- The *Status()* object transfers the actual experimental status, integrating the graphical mode to draw, timestamp & pause the actual state of each graphical instance.

- The *Packet()* object transfers each iteration's instructions to record the measured behavioural metrics.

Synchronization between both tasks is achieved by setting up a Remote Object Local Server using Pyro4[4] which sockets the commands and timings from the SRC Window to the NAV Window. To record physiological data simultaneously the simulator sends Lab Streaming Layer (LSL, Kothe et al., 2014) markers for any relevant events (see Figure 9.1). In addition, all behavioural data is recorded. Records include reaction times for both tasks, accuracies, answers to questionnaires and specific information on errors.

[3] https://www.psychopy.org/download.html
[4] https://pyro4.readthedocs.io/en/stable/

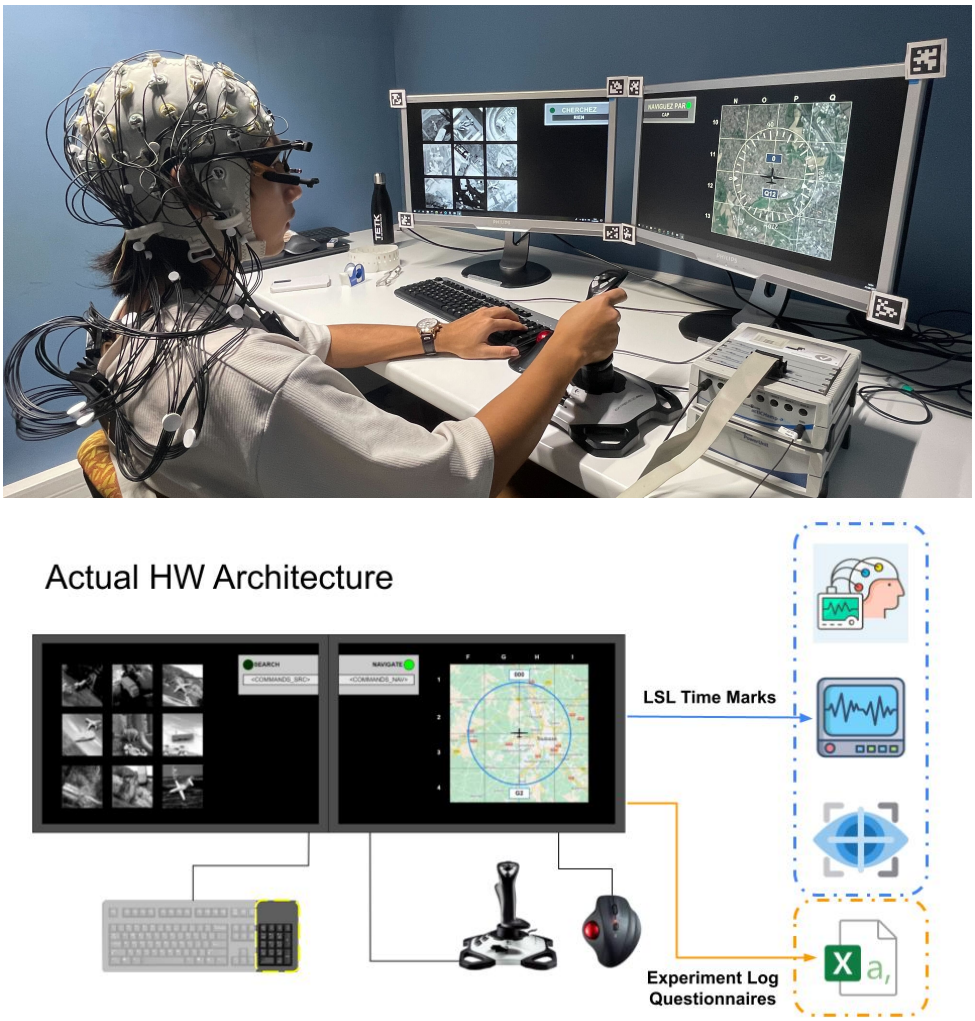

Fig. 9.1 Experimental setup: Participant is equipped with an EEG Cap, ECG sensor (not visible) & eye-tracking glasses. The SRC task is displayed on the left side of the screen, while the NAV task is on the right side. Joystick, numpad & trackball are also visible on the table. The Flight Director (grey widget at the top of both screens) instructs the participant to use the NAV task in heading mode.

II.9.3 Methods

II.9.3.1 Participants

25 Participants (7 female, mean age 23.54 years (std 2.7), 11 English speaking & 14 French speaking) were recruited and completed the experiment. From the subsequent analysis, one participant had to be dropped due to inconsistencies and missing data in the recordings.

II.9.3.2 Procedure

Participants who agreed to participate signed the informed consent forms and were equipped with the EEG, ECG and Eye-Tracking sensors. Next, participants completed a battery of questionnaires. They then completed a training phase of 16 minutes before a five-minute resting state was recorded (30-second intervals of eyes open and closed). They then started the 120-minute main phase of the experiment. After completing the main phase, participants filled in another battery of RSME, KSS, SPS, and VAS questionnaires. The ethical committee of Toulouse (Comité éthique de l'Université de Toulouse) approved the experimental protocol (Dossier 2022-501).

II.9.3.3 Materials

Task

The UASOS task (Fig. 9.1) combines some of the fundamental aspects of UAS operations with a task-switching protocol to allow the investigation of cognitive flexibility (Hinss, Brock, & Roy, 2023a). The task requires participants to alternate between tasks on a trail-based system. On average, every 7 seconds (with a +-1000ms jitter, ~1020 trials during the main condition), written instructions appear on a widget in the middle of the participant's visual field to indicate the current task. To ensure adequate performance, small pretests were conducted to calibrate these parameters. Participants work on two main tasks, with two modes each. The Navigation task (NAV) requires the participant to navigate the UAV either using headings (heading mode, HDG) or waypoints (Waypoint mode, WPY). The design was balanced with an equivalent number of trials in all tasks and modes.

Questionnaires

Participants answered 5 questionnaires at varying moments. At the beginning of the experiment, participants completed a demographics questionnaire, an Edinburgh handedness questionnaire (EDI), a Karolinska Sleepiness Scale (KSS), and a Samn-Perelli Fatigue Scale (SPS). Participants also filled in the Rating Scale Mental Effort (RSME) (Ghanbary et al., 2016) as well as two Visual Analogue Scales (VAS) scales: cognitive fatigue (VAS-F) and drowsiness (VAS-D). The VAS scales were also filled in at 19-minute intervals during the main experimental phase. At the end of the experiment participants again filled in the KSS, SPS, VAS and RSME scales. Details on all questionnaires are in Appendix V.14.2.

EEG

Using an active AG-AgCl electrode system with an ActiCHamp amplifier (Brain Products, Gmbh), EEG data was recorded from 64 electrodes. The international 10-20 system was used

for electrode placement Jasper (1958). Data were recorded at 500 Hz, and impedances were kept below 50 kΩ. Data was streamed and synchronized using the LabStreamingLayer (LSL, Kothe, 2015).

As part of the data validation, we performed a frequency analysis of the EEG data independently from the mental state prediction. For this, a zero-padded channel was added to the EEG data before an average referencing, with a subsequent removal of the zero-padded channel. Extreme values were clamped following the method proposed by Défossez et al. (2023). The data was then cut into 5-second non-overlapping epochs. The power of each frequency band was calculated by band-pass filtering the signal and calculating the root mean square for each electrode. Using the parameters suggested by Wascher et al. (2014), the power of the theta (4-8Hz) alpha (8-12Hz) bands were extracted. For the statistical analysis the data was then averaged into 10-minute epochs over three clusters of electrodes (i) Frontal: F3, F1, Fz, F2, F4, FC3, FC1, FCz, FC2, FC4; (ii) Central: C3, C1, Cz, C2, C4, CP3, CP1, CPz, CP2, CP4; and (iii) Parieto-Occipital: P3, P1, Pz, P2, P4, PO3, POz, PO4, O1, Oz, O2.

II.9.3.4 Performance and Mental State classification

EEG preprocessing:

The EEG data for each participant was cut into 5-second non-overlapping epochs, to allow for robust covariance matrix estimation, and to be independent of the task trials. The epoched data was then referenced and filtered between 2-36 Hz using the *mne.filter()* function. Data points of each channel that exceeded 20 std of the robustly scaled data were clamped to the value equal to +/- 20 std; for a detailed explanation of this method, see Défossez et al. (2023). Finally, the data was resampled to 125 Hz[5].

Label creation

The performance metrics used misses (Miss), reaction times (RT) and accuracies (ACC) of all subtasks. Across subtasks, all values were first normalized to give equal importance to each subtask. Next, averages of misses, RT and accuracies were calculated. For each value, the best and worst 33% were used to assign labels—the global performance (OVR) score combined misses, RT, and accuracy. Three different mental state-based labels were created. The time-on-task used the first and last 33% of each recording, respectively. The VAS scores were used for the other two approaches. Using the drowsiness and cognitive fatigue scales, the blocks corresponding to the most extreme values of each scale were used to label the data. Adjusted chance levels were calculated for each label-type based on Mueller-Putz et al. (2008).

Classification

The data from each participant was divided into an 80/20 split for training/testing datasets. Next, the covariances were computed using OAS or LWF, and the data was projected to the target space. We compared performances of logistic regression (Log Reg), Support Vector Machine (SVM) and Random Forest (RF) classifiers. Hyper-parameters for each classifier were optimized using 5-fold cross-validation using Bayesian search.

[5]ECG-based classification was not performed at this point, to focus only on cerebral data

II.9.3.5 Statistical Analysis

The general inference criterion is a p-value of $p < .05$. In multiple comparisons, we adjusted that criterion according to the adjusted Bonferroni method. Assumptions for each statistical test were checked and accounted for if not satisfied. Outlier detection was performed based on the interquartile range criterion. This was done for trials grouped by condition.

Subjective

To analyse the subjective results, we compared the SPS and KSS scores from the beginning to the end. For this, we used a paired samples T-Test. We also performed a one-way repeated measures ANOVA for both VAS scales. For one participant, the questionnaires at the end of the experiment were not recorded.

Behavioral

The behavioural analysis was divided into three sections: (i) Search Task, (ii) Navigation Task - Heading Mode, and (iii) Navigation Task - Waypoint mode. Overall, analysis was not possible due to the differences between tasks.

For the search task, reaction times, F1 (see Section I.5.1.1) score and misses were used as dependent variables in repeated measures within-subjects ANOVA with Task (searching humans or searching vehicles) as an independent variable and switch (repeat, internal or external) as well as time on task (19-minute blocks) for repeated measures.Accuracies and F1 scores were calculated using only trials that were not misses.

To analyse the heading task reaction times, turning direction misses and deviation were the dependent variables of repeated measures within-subjects ANOVA with switch (repeat, internal or external) and TOT (19-minute blocks) for repeated measures.

The Waypoint mode was evaluated using reaction time, correct choices, and misses as dependent variables. Again, switch (repeat, internal or external) and TOT were the independent variables for repeated measures.

EEG Frequency

The extracted powerbands were compared across blocks using repeated measures ANOVA for each cluster.

ECG metrics

The metrics were compared across 10-minute blocks using a repeated measures ANOVA for each metric.

Classification

To analyze the classification results, a 2-way ANOVA with factors classifier and label-type was performed on the dependent variable of accuracy.

II.9.4 Results

II.9.4.1 Subjective Data

Scores for the KSS, RSME and SPS measures all showed significant increases in values comparing the beginning and the end of the experiment KSS: $t(23) = -6.912$, $p < .001, d = -1.411$; SPS: $Z = -4.000$, $p < .001, r = -0.933$; RSME: $t(23) = -6.380$, $p < .001, d = -1.302$). The assumption of normality was violated for the SPS test (Shapiro-Wilk $W = 0.875$, $p = 0.007$); therefore, the Wilcoxon result is reported. VAS scores on both cognitive fatigue and drowsiness showed linear increases over time (Cognitive: $F(2.374, 49.849) = 25.979$, $p < .001$, Greenhouse-Geisser corrected,
$\eta^2 = 0.553$ and $\eta_p^2 = 0.553$; Drowsiness:$F(4.294, 90.184) = 12.159$, $p < .001$, Greenhouse-Geisser corrected, $\eta^2 = 0.367$ and $\eta_p^2 = 0.367$, see Figure 9.2 a-d).

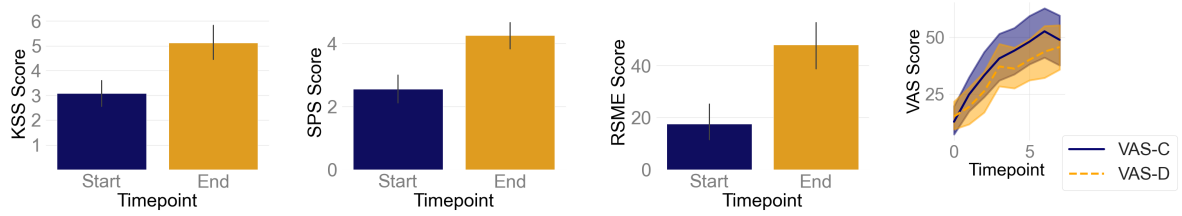

Fig. 9.2 Subjective results: a) KSS Scores comparing the beginning and end of the experiment. b) SPS Scores comparing the beginning and end of the experiment. c) RSME Scores comparing the beginning and end of the experiment. d) VAS scores throughout the experiment for both the Cognitive Fatigue and the Drowsiness scale. Timepoint 0 is before the start and before the training, and then each subsequent point occurred every 19 minutes into the experiment. The last point occurred after the completion of the experimental phase. Shading indicates a 95% confidence interval.

II.9.4.2 Behavioral Data

SRC Task

The analysis of the search task showed significant effects of trial type and task on all dependent measures of reaction time F1 score and misses, except for the main effect of switches on misses. Participants took the longest when switching tasks from the Navigation task. Between internal task switches (searching humans or vehicles) and repetition trials, there was no difference (RT: $F(2, 48) = 50.143$, $p < .001$, Greenhouse-Geisser corrected, $\eta^2 = 0.676$, Repetition - Internal: $t(24) = 1.673$, $p = 0.322$, Repetition - External: $t(24) = 6.54$, $p < .001$, External - Internal: $t(24) = 10.228$, $p < .001$). See Figure 9.3 a. The F1 score differed significantly between all groups(F1: $F(2, 48) = 454.572$, $p < .001$, Greenhouse-Geisser corrected, $\eta^2 = 0.950$, Repetition - Internal: $t(24) = 29.6$, $p < .001$, Repetition - External: $t(24) = 18.57$, $p < .001$, External - Internal: $t(24) = 4.5$, $p < .001$). The F1 score was highest for repetition trials, followed by external and internal switches. See Figure 9.3 b. Misses showed no significant differences between groups (Misses: $F(2, 48) = 3.820$, $p = 0.054$, Greenhouse-Geisser corrected, $\eta^2 = 0.137$). See Figure 9.3 c.

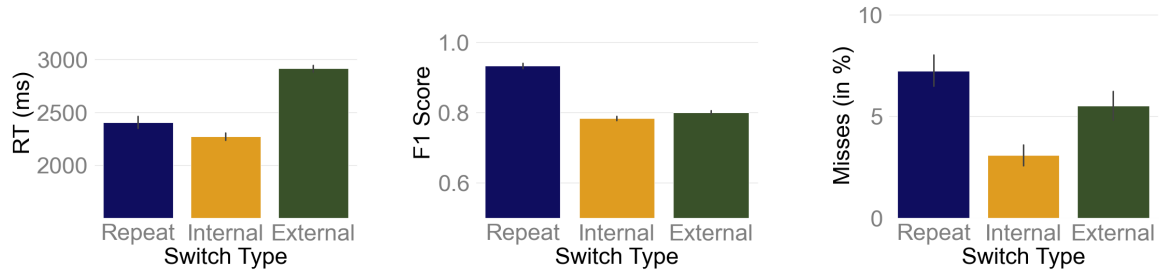

Fig. 9.3 Behavioral results of the SRC task: a) Reaction time by switch type. b) F1 scores by switch type. c) Misses by switch type.

Searching humans resulted in significantly larger reaction times and more misses (RT: $F(1,24) = 214,872$, $p < .001$, Greenhouse-Geisser corrected, $\eta^2 = 0.900$Misses: $F(1,24) = 34.308$, $p < .001$, Greenhouse-Geisser corrected, $\eta^2 = .588$). Surprisingly, the F1 score was slightly higher for searching humans (F1: $F(1,24) = 108.011$, $p < .001$, Greenhouse-Geisser corrected, $\eta^2 = .818$) . Time did not significantly affect performance on any metric (RT: $F(4.117,120) = .381$, $p = .827$, Greenhouse-Geisser corrected, $\eta^2 = .0160$; F1: $F(4.056,120) = 2.268$, $p < 0.066$, Greenhouse-Geisser corrected, $\eta^2 = 0.086$ Miss: $F(2.024,) = .751$, $p = .479$, Greenhouse-Geisser corrected, $\eta^2 = .030$.), see Figure 9.4 a-d.

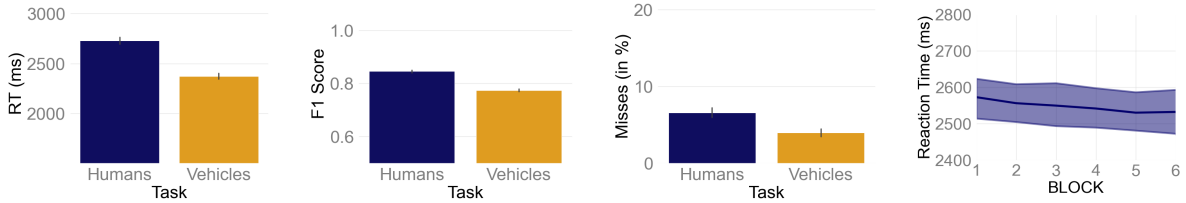

Fig. 9.4 Behavioral Results of the SRC task: a) Reaction time by Mode. b) F1 scores by Mode. c) Misses by mode. d) Reaction Times over blocks

NAV task HDG mode

Reaction times in the HDG mode of the SRC task were significantly influenced by task-switching (RT: $F(1.560,60) = 5.936$, $p < .005$, Greenhouse-Geisser corrected, $\eta^2 = .198$). External switches took significantly longer than repetition trials (Repeat - External: $t(24) = 3.092$, $p = .015$). Reaction times did not differ significantly between internal and repetitions or external switches (Repeat - Internal: $t(24) = 2.487$, $p = .061$, Internal - External: $t(24) = .181$, $p = 1.000$; see Figure 9.5 a). Neither deviation, vorrect turn, nor misses were significantly affected by switch-type (deviation: $F(1.237,60) = 3.1$, $p = 0.08$, Greenhouse-Geisser corrected, $\eta^2 = 0.114$, correct turn: $F(1.848,60) = .157$, $p = .699$, Greenhouse-Geisser corrected, $\eta^2 = .014$, Miss: $F(1.117,60) = 2.109$, $p = .159$, Greenhouse-Geisser corrected, $\eta^2 = .081$; see Figure 9.5 b-d).

RT, correct turn and misses were all influenced by time (RT: $F(4.1,60) = 41.711$, $p < .005$, Greenhouse-Geisser corrected, $\eta^2 = 0.635$, correct turn: $F(4.207,60) = 3356$, $p = .011$, Greenhouse-Geisser corrected, $\eta^2 = .123$, miss: $F(3.327,60) = 4.185$, $p = .007$, Greenhouse-Geisser corrected, $\eta^2 = .148$). Contrast analysis revealed significant cubic effects for RT, correct turn and

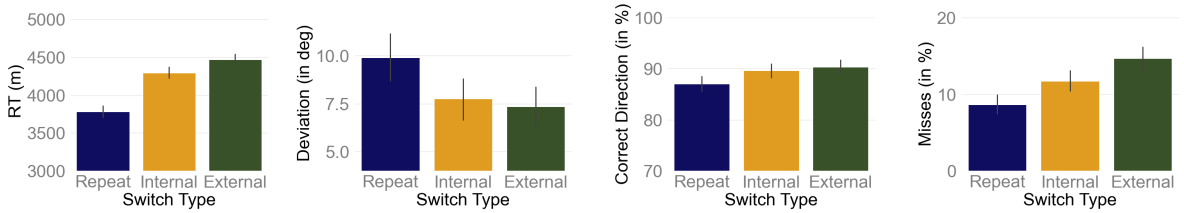

Fig. 9.5 Behavioral Results of the NAV task HDG mode: a) Reaction time by switch type. b) Course deviation (in deg) by Switch. c) Correct turn direction (in %) by switch type. d) Misses (in %) by switch type.

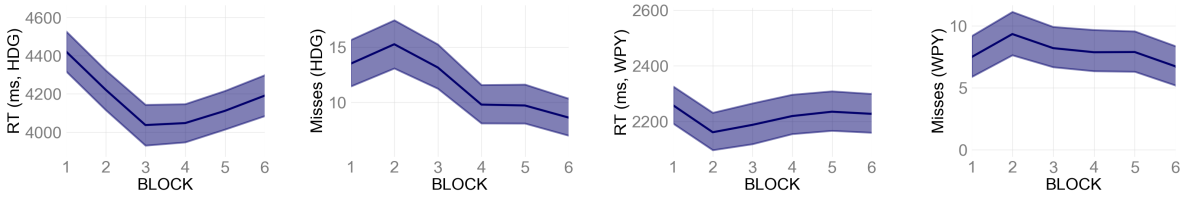

Fig. 9.6 Behavioral Results of the NAV task: a) HDG: Reaction times over time b) HDG: Misses over time. c) WPY: Reaction times over time. d) WPY: Misses over time.

misses (RT: $F(5,60) = 99.569$, $p < .005$, Greenhouse-Geisser corrected, $\eta^2 = 0.806$, correct turn: $F(5,60) = 7.627$, $p = .011$, Greenhouse-Geisser corrected, $\eta^2 = .231$, miss: $F(5,60) = 5.253$, $p = .031$, Greenhouse-Geisser corrected, $\eta^2 = .180$, see Figure 9.6 a & b).

NAV task WPY mode

Both Reaction times and misses were influenced by switch type in the Waypoint mode. Accuracy was not influenced by switches (RT: $F(1.770,60) = 59.877$, $p < .001$, Greenhouse-Geisser corrected, $\eta^2 = 0.714$, Accuracy: $F(1.744,60) = 2.025$, $p = .150$, Greenhouse-Geisser corrected, $\eta^2 = .078$, Miss: $F(1.421,60) = 6.515$, $p = .008$, Greenhouse-Geisser corrected, $\eta^2 = .214$). See Figure 9.7 a-c. Reaction times differed across all conditions. The slowest trials were external switch trials followed by internal switches and repetitions (External - Internal: $t(24) = 8.291$, $p < .005$, Internal - Repeat: $t(24) = 3.966$, $p = .002$, External - Repeat: $t(24) = 9.552$, $p < .005$). See Figure 9.7 a. Misses were significantly higher during external switches as compared to repetition trials (External - Repeat: $t(24) = 3.217$, $p < .005$; see Figure 9.7 c). Reaction times were also influenced by time (RT: $F(3.078,60) = 2.707$, $p = .05$, Greenhouse-Geisser corrected, $\eta^2 = 0.101$). Contrast tests revealed a significant cubic effect (RT: $F(1,60) = 10.395$, $p = .004$, Greenhouse-Geisser corrected, $\eta^2 = 0.302$; see Figure 9.6 c & d).

II.9.4.3 EEG

The alpha frequency band showed significant effects throughout the experiment. In all three clusters, an increase in alpha power was observed (Frontal : ($F(1,4.6) = 3.184$, $p = .012$, Greenhouse-Geisser corrected, $\eta^2 = 0.117$; Central: ($F(1,6.051) = 3.995$, $p < .001$, Greenhouse-

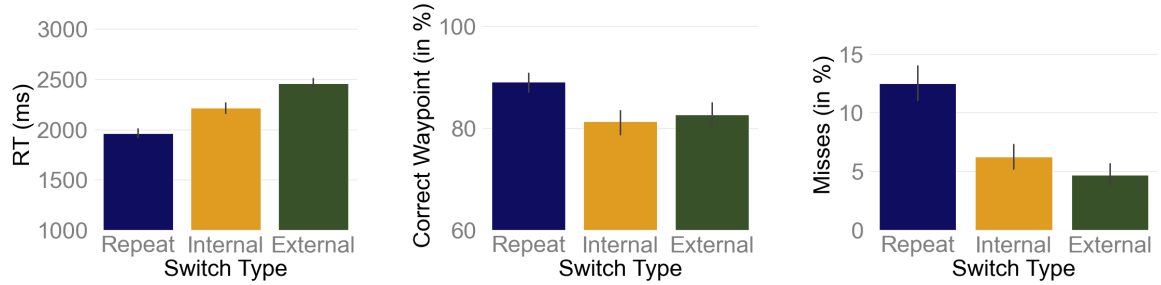

Fig. 9.7 Behavioral Results of the NAV task WPY mode: a) Reaction time (in ms) by switch type. b) Correct Waypoint choice (in %) by switch type. c) Misses (in %) by switch type

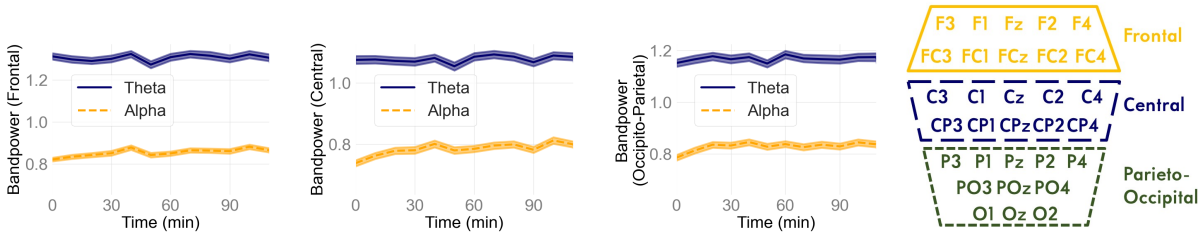

Fig. 9.8 Frequency power in the theta and alpha power bands over time for the a) Frontal Cluster; b) Central Cluster; c) Parieto-Occipital Cluster. d) Topology of the 3 clusters.

Geisser corrected, $\eta^2 = 0.143$; Parieto-Occipital: ($F(1, 4.899) = 2.508$, $p = .035$, Greenhouse-Geisser corrected, $\eta^2 = 0.095$). Theta power did not show any significant change over time (see Figure 9.8 a-d).

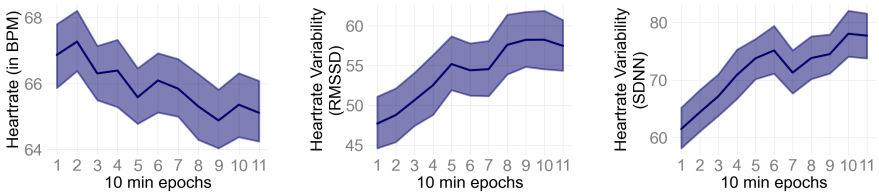

Fig. 9.9 Cardiac metrics over time in 10-minute epochs: a) Heart Rate in BPM b) HRV in RMSSD c) HRV in SDNN

II.9.4.4 ECG

ECG metrics showed significant effects over time. The heart-rate decreased significantly, while the variability increased (HR : $(F(1,5.57) = 2.93,\ p = .012$, Greenhouse-Geisser corrected, $\eta^2 = 0.118$; RMSSD: $(F(1,5.91) = 8.379,\ p < .001$, Greenhouse-Geisser corrected, $\eta^2 = 0.276$; SDNN:$(F(1,5.32) = 9.259,\ p < .001$, Greenhouse-Geisser corrected, $\eta^2 = 0.296$, see Figure 9.9 a-c).

II.9.4.5 Classification

The analysis showed a significant effect of label type ($F(6,24) = 342.12,\ p < .001,\ \eta^2 = 0.81$). There was no significant effect of classifier ($F(7,24) = 2.398,\ p > .09,\ \eta^2 = 0.01$). The highest classification accuracy was obtained using the TOT labels, with an average accuracy of 94.86% across all classifiers. However, both VAS scales, the performance labelling based on misses and accuracy, also performed above their respective chance levels. This resulted in a 62.09 % accuracy for detecting misses using the tangent space logistic regression (see Figure 9.10 a).

II.9.5 Discussion

The results of this study globally recreate the behavioural effects observed in the DTS protocol (see Chapter 6). Repetition trials are faster as compared to switch trials. With regard to accuracy, some metrics show better performance in switch trials (misses in the SRC and WPY task). Focusing on other metrics the results show that during switch trials, participants made more errors when choosing a waypoint, when turning the UAV, or when searching for visual targets. Here, switching between more dissimilar tasks is regarded as even more difficult. Other than in the DTS (see Chapter 6) experiment, external task switches require participants to shift their gazes from one screen to another. Participants also had to shift their hands between the joystick and the trackball when switching modes in the navigation task. Contrary to expectations, reaction times in the SRC task were lower for internal switches than in repetition trials. One possible explanation is the experimental protocol, which does not update all images during each trial. This means that if a participant has detected a human on one image, and that image is not updated during the next trial, they do not have to reselect it if the task did not change. Participants may scan images several times while searching, e.g. a vehicle, but they might still recognize humans and mentally note that image. When switching, they do not have to search; they only have to verify that that image has not been updated. This would also explain the high number of misses during repetition trials. If participants rely on previously searched images, they might miss newly added images during repetition trials. An alternative explanation is that

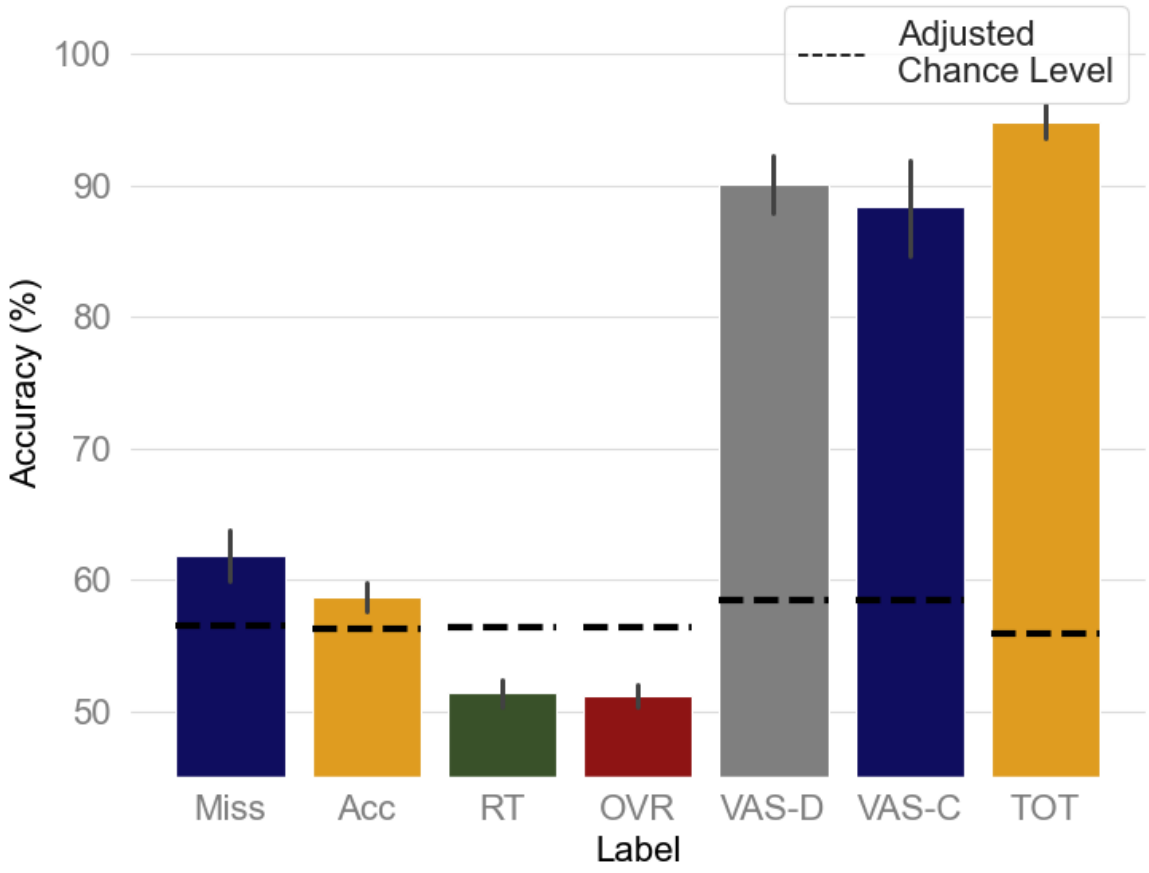

Fig. 9.10 Classification results accuracy depending on label type. Horizontal dotted lines indicate the adjusted chance level according to Müller et al. (2008).

the switching between searching humans and vehicles does not tax cognitive flexibility. In that case, an internal switch trial would be easy for participants to perform as they need to select only one image during the trial so that it does not count as a miss. One target is likely easily visible, which may explain the quick reaction. In repetition trials, on the other hand, participants may have already selected the 'easy' targets and now have to search for more distorted targets, requiring effort and, if not finding a target, resulting in more misses.

The higher reaction times of searching humans are explicable due to vehicles' long and straight contours being easily detectable. The results of the Navigation task perfectly mirror those of the DTS protocol.

Both ECG and EEG metrics were shown to be sensitive to TOT, which, together with the subjective results, indicate that participants were actually fatigued by the protocol.

To use pBCIs in complex environments, the output of a pBCI needs to have some predictive value. The results presented here compare mental-state and performance estimations using EEG data. The subjective and EEG analyses both point toward increased mental fatigue over time. Yet, while some behavioural metrics, such as reaction times in the navigation task, seem to show a similar trend, variability in overall performance is not best explained by TOT.

While mental state estimation using Time-On-Task or subjective metrics as ground truth

performs considerably above chance level, our algorithms could also predict misses with an above chance level likelihood. The algorithm's success with TOT metrics may be attributable to the observed alpha power increase often associated with mental fatigue (Boksem et al., 2005; Wascher et al., 2014).

It may also be due to slow drifts and the non-stationarity of the EEG signal (Urigüen & Garcia-Zapirain, 2015). Subjective fatigue scores increased over time, creating similarities between the TOT and VAS labels. The absence of stronger effects in the spectral analysis may have been attenuated due to the complexity of the task (Kamzanova et al., 2011). The reaction time-based performance estimation had the lowest accuracy. One possible explanation is that longer reaction times may reflect several processes that are then mixed up. Slow reaction times may be due to fatigue (Csathó et al., 2012) or a speed-accuracy tradeoff (Wood & Jennings, 1976). The moderate success of the performance estimation based on misses suggests that spectral EEG features, especially theta power, are sensitive to lapses (Peiris et al., 2006). Future work could evaluate incorporating Bayesian updating, which may further improve the performance estimation.

II.9.6 Conclusion

This study highlights the challenges and possibilities of EEG-based performance estimation. The differences between definitions of performance highlight the importance of label assignment. In our opinion, performance scores should (i) be defined as a priori, (ii) be explainable, and (iii) provide real-world value. This experiment presents an open soucre protocol and environment that pushes the ecological validity of task-switching protocols further towards realistic, complex operations.

The protocol and open access code for the UASOS task were presented at the Human-Robot Interaction Conference 2024. The complete results will be presented at the 2024 Graz BCI Meeting (see Appendix 13, Hinss, Vitale, Brock, and Roy, 2024; Hinss, Vitale, Phan, et al., 2024).

Chapter 10

UASOS-II: Real-Time implementation of solutions

Synopsis

This study follows the previous experiments to study whether cognitive flexibility during long-endurance UAS missions can be improved. To do so, the previous UASOS protocol was adapted, and a group of participants performed the 2-hour main experimental phase with a visual alert adaptation. The visual alerts informed participants of upcoming task-switching events. Results showed a weaker subjective increase in mental fatigue compared to a control group. Furthermore, behavioural results showed improved reaction time and accuracy in some performance metrics. Future work may improve the efficacy of this adaptive interface by improving certain of its aspects. Recommended Reading:

- The UASOS study (Chapter 9)
- The DTS-II study (Chapter 7)

One question to ask: **How do visual alerts impact performance during a realistic UAS task?**

II.10.1 Motivation and Research Goals/ Questions

The DTS studies (see Chapters 6 & 7) as well as the UASOS study (see Chapter 9) proved the negative impact of task-switching on performance in terms of both accuracy and reaction times. The DTS-II study tested if visual alerts could reduce task-switching costs. The absence of effects in this study is likely attributable to an underpowered design but also because participants received no instructions on using the visual alerts.

For these reasons, we decided to test the visual alerts within the more realistic UASOS environment. Based on participants' subjective feedback, we only focused on the single alert

condition. To reduce breaks during the two-hour main experimental phase, we also reduced the frequency of the visual alerts from 19-minute to 30-minute intervals. Participants were assigned to either a control condition in which the visual alerts were explained in the training phase but never activated in the main experimental phase. Participants in the visual alerts condition received the same training, but the visual alerts occurred following a predetermined schedule during the two-hour main experimental phase. The visual alerts were not always active to mimic an implementation as an adaptive interface. In the training, the participants were instructed on how to use the visual alerts.

II.10.2 Methods

II.10.2.1 Participants

30 participants were recruited for the experiment, 15 in each condition. Due to technical issues with the EEG measurement, 2 participants (one in each condition) had to be excluded from the EEG and ECG analysis. The mean age of the participants was 25.6 years (SD 7.61), with 1/3 being female participants and 16.7 % of the participants choosing to complete the experiment in English (the remaining participants completed the experiment in French).

II.10.2.2 Procedure Changes

The procedure of this experiment was only adapted in minor ways from the UASOS experiment. The following two subsections will highlight the overall changes and how the visual alerts condition was implemented. The ethical committee of Toulouse (Comité éthique de l'Université de Toulouse) approved the experimental protocol (Dossier 2022-501).

Overall changes

- **Eye Tracking** was removed from the protocol due to hardware issues.
- **Visual Analogue Scales** were presented every 30 minutes instead of every 19 minutes to reduce their impact on mental fatigue, as participants tended to use them as small breaks.
- A **visual alerts widget** was added to the flight director widget. This dynamically changes colour from grey to orange during the task if the visual alerts are activated. The widget aimed to inform participants whether the visual alerts were activated (see Figure 10.1).
- In the **overall training phase**, all participants received 5 minutes of training without the visual alerts and 5 minutes with the visual alerts. Participants were blind to their condition.

Visual Alerts Condition

For participants in the visual alerts condition, a quasirandom schedule determined the activation of the visual alerts (active around 40% of the task). When the visual alerts were activated, this triggered two changes. First, the visual alerts widget was turned on informing the participant of their activation. The activation of the visual alerts themselves depended on the upcoming trial. During repetition trials, when the previous task was the same as the upcoming

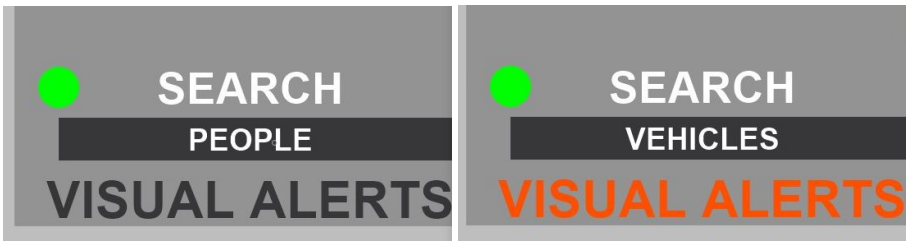

Fig. 10.1 Flight director widget for the Search task. Left: visual alerts are deactivated; Right: visual alerts are activated, and the participant is informed of this via the amber hue of the Visual Alerts widget.

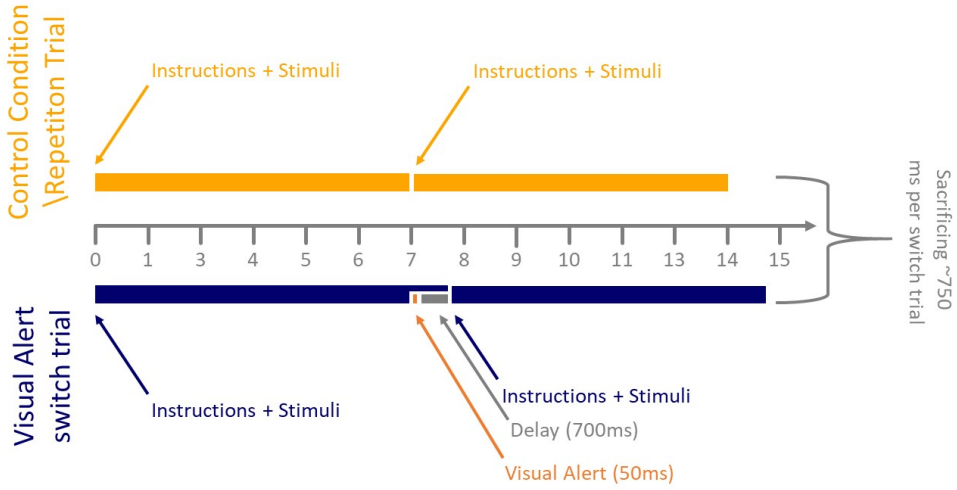

Fig. 10.2 Comparison of two trials. Top: Visual Alerts are deactivated, or it is a repetition trial. Participants receive instructions and have seven seconds to react (+-1000 ms jitter). Bottom: Participants receive instructions, and at the end of the trial, the new instructions are delayed by 750 ms. A 50 ms flash appears, followed by a 700 ms delay.

one, nothing happened. 700 ms before switch trials (when changing from one task to another), a transparent amber ■ (R: 254, G:80, B:0; Opacity: 50%) flash would simultaneously appear on both computer screens (for 50 ms)

Schedule The visual alerts were activated depending on a preprogrammed schedule in the visual alert condition. To be realistic to what a BCI might predict, the schedules were calculated based on the performance of an individual participant of the control condition.

To do so, a global performance score based on misses and accuracies was calculated for each trial. For this, the miss and accuracy metrics of each of the 4 tasks were combined. The scores for misses and accuracies were individually normalised to avoid one task having a larger influence than the others. As each trial was only one task and either had an accuracy or was a miss, the corresponding value was inserted as the global performance score. The scores were then averaged over 1-minute epochs. This was done as the visual alerts turning on and off more frequently risked perturbing the participants. As explained in Section I.5.2.2 (see Figure

5.9) the 40 best and 40 worst % were then attributed labels. This was calculated on all but the first 10 minutes of the main experimental phase, as learning effects risked strongly influencing the performance calculation. In the first ten minutes, as well as epochs that were not assigned a label (the middle 20%) the visual alerts were deactivated to reduce visual strain). With the removal of the 20% and the ten first minutes, the visual alerts were activated a total of 40 % of the time per participant.

II.10.2.3 Analysis

The general inference criterion is a p-value of $p < .05$. In multiple comparisons, we adjusted that criterion according to the adjusted Bonferroni method. Assumptions for each statistical test were checked and accounted for if not satisfied. Outlier detection was performed based on the interquartile range criterion. This was done for trials grouped by condition.

Subjective

To analyse the subjective results, we compared the SPS; RSME and KSS scores from the beginning to the end. We used a 2 way Split Plot ANOVA (Timepoint, Condition) for this. We also performed a two-way Split Plot ANOVA for both VAS scales.

Behavioral

The behavioural analysis was divided into four sections: (i) Search Task, (ii) Navigation Task - Heading Mode, and (iii) Navigation Task - Waypoint mode and (iv) Overall analysis. Two different analysis strategies were pursued. To test for within-subject effects of the visual alerts, periods of visual alerts being activated vs periods of visual alerts deactivated were compared within the visual alerts condition. For this we used a repeated measures ANOVA and only included participants of the visual alerts condition. We also compared the effects of the baseline and the visual alert conditions to examine global effects. This was done with a split-plot ANOVA.

For the search task, reaction times, F1 score and misses were used as dependent variables in repeated measures within-subjects ANOVA with Task (searching humans or searching vehicles) as an independent variable and switch (repeat, internal or external) as well as time on task (30-minute blocks) for repeated measures.

To analyse the heading task reaction times, turning direction misses and deviation were the dependent variables of repeated measures within-subjects ANOVA with switch (repeat, internal or external) and TOT (30-minute blocks) for repeated measures.

The Waypoint mode was evaluated using reaction time, correct choices, and misses as dependent variables. Again, switch (repeat, internal or external), and TOT were the independent variables for repeated measures.

EEG Frequency

The extracted powerbands were compared across blocks using repeated measures ANOVA for each cluster.

ECG metrics

The metrics were compared across 10-minute blocks using repeated measures ANOVA for each metric.

Exploratory Time of Day Analysis

Participants arrived at the lab either in the morning at around 09:00 (start of the main experimental phase at around 10:00) or in the afternoon at around 13:00 (start of the main experimental phase at around 14:00). In a supplementary analysis, this factor was included as a covariate. It is not in the main analysis as the sample was imbalanced (20 participants in the afternoon & 10 participants in the morning).

II.10.3 Results

Subjective

The subjective ratings showed significantly increased fatigue (SPS, KSS, VAS-C and VAS-D) over time in both conditions (SPS: $F(29,1) = 23.12$, $p < .001$, $\eta^2 = 0.45$, KSS: $F(29,1) = 11.99$, $p = .002.$, $\eta^2 = 0.3$, VAS-C: $F(29,5) = 20.516$, $p < .001$, Greenhouse-Geisser corrected, $\eta^2 = 0.423$, VAS-D: $F(29,5) = 14.286$, $p < .001$, Greenhouse-Geisser corrected, $\eta^2 = 0.338$). There was no significant difference between the groups. The interaction effect (Condition*Timepoint) on the SPS measure showed a significantly larger increase in fatigue for participants in the control group without visual alerts (Condition*Timepoint: $F(29,1) = 5.78$, $p = .023$, $\eta^2 = 0.17$, see Figure 10.3). Subjective mental workload (RSME) was significantly higher following the task as compared to before the task (RSME: $F(29,1) = 64.07$, $p < .001$, $\eta^2 = 0.70$, see Figure 10.3).

II.10.3.1 Behavioral

Search Task

Between Condition Analysis

Reaction Time Reaction times depended on the factors Task, Switch as well as the interaction between Switch*Task and Block*Switch (Task: $F(1,29) = 263.38$, $p < .001$,$\eta^2 = 0.904$, Switch: $F(2,29) = 91.21$, $p < .001$, Greenhouse-Geisser corrected, $\eta^2 = 0.77$, Switch*Task: $F(5,29) = 10.84$, $p < .001$, Greenhouse-Geisser corrected, $\eta^2 = 0.28$, see Figure 10.4 left). Searching humans was significantly slower than searching vehicles. Repeat and internal switch trials were equally fast and significantly faster than external trials (Repeat*Internal:$t(30) = 0.54$, $p = 1$; Repeat*External: $t(30) = -11.42$, $p < .001$; Internal*External: $t(30) = -11.96$, $p < .001$).

F1 The F1 accuracy score was significantly influenced by the factors Task, Switch as well as the interaction between Switch and Condition (Task: $F(1,29) = 79$, $p < .001$,$\eta^2 = 0.75$, Switch: $F(2,29) = 49.86$, $p < .001$, Greenhouse-Geisser corrected, $\eta^2 = 0.95$, Switch*Condition: $F(3,29) = 4.87$, $p = .019$, Greenhouse-Geisser corrected, $\eta^2 = 0.95$, see Figure 10.4 center). Searching humans was significantly more difficult. Repeat trials were more accurate than both internal and external switch trials, which did not differ from each other (Repeat*Internal:

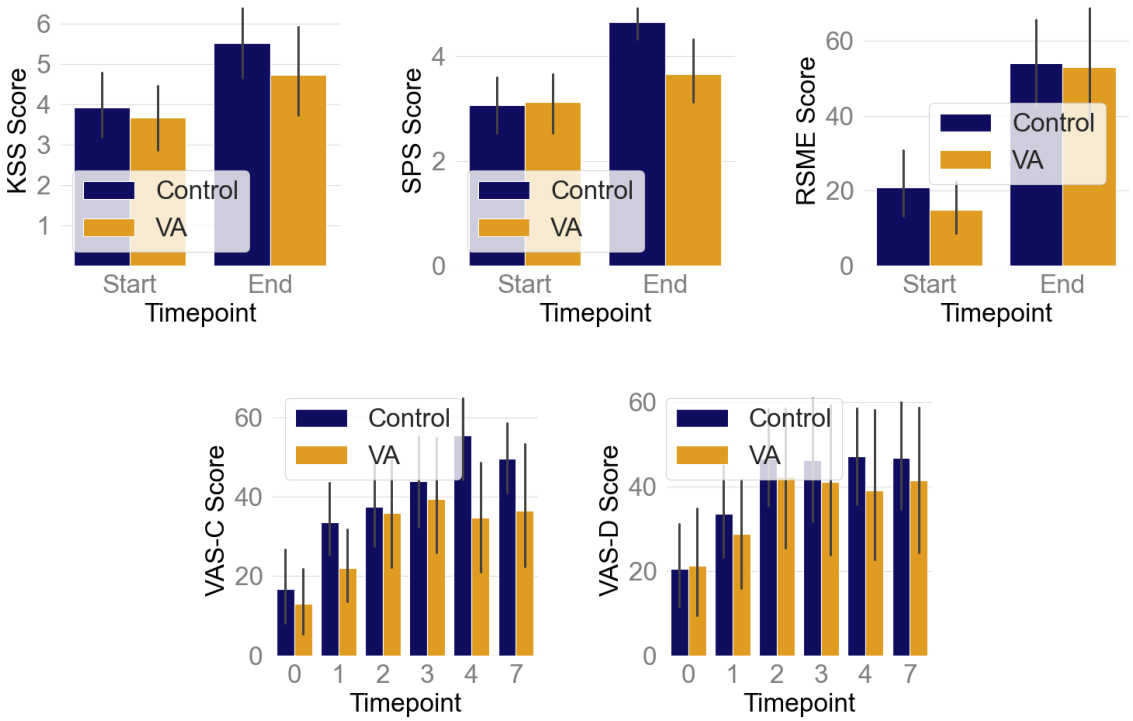

Fig. 10.3 Results of the Subjective questionnaires (in order from right to left, top to bottom): KSS, SPS, RSME VAS-Cognitive, VAS-Drowsiness.

$t(30) = 27.69, p < .001$; Repeat*External: $t(30) = 26.72, p < .001$; Internal*External: $t(30) = -0.96, p = 1$). T-Test of trials separated by switch type, showed no significant effect of condition (Repeat: $t(30) = 0.37, p = 1.$; Internal: $t(30) = -2.06, p = .67$; External: $t(30) = -2.55, p = .21$).

Misses The frequency of misses in the Search task was significantly influenced by the interaction between Task*Condition, Block*Task and Block*Switch (Task*Condition: $F(3, 29) = 9.73$, $p = .014$, $\eta^2 = 0.549$, Block*Task: $F(7, 29) = 17.21$, $p < .001$, Greenhouse-Geisser corrected, $\eta^2 = 0.68$, Switch*Block: $F(11, 29) = 4.09$, $p = .02$, Greenhouse-Geisser corrected, $\eta^2 = 0.34$, see Figure 10.4 right). T-tests of the Task*Condition were not significant.

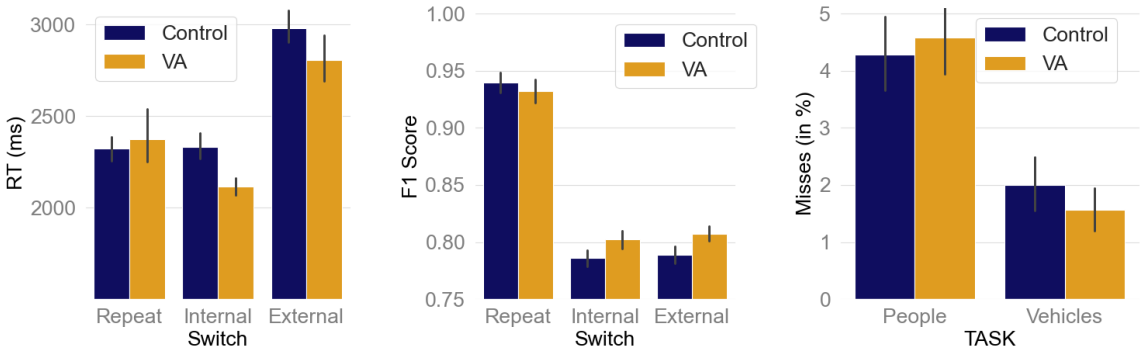

Fig. 10.4 Results of the between subjects analysis of the search task

Within Condition Analysis

Reaction Time Reaction times depended on the factors Visual Alert, Task, Switch as well as the interaction between Switch*Task (Visual Alert: $F(1,14) = 37.75$, $p < .001$,$\eta^2 = 0.77$, Task: $F(1,14) = 107.44$, $p < .001$,$\eta^2 = 0.91$, Switch: $F(2,14) = 23.15$, $p < .001$, Greenhouse-Geisser corrected, $\eta^2 = 0.68$, Switch*Task: $F(5,29) = 4.24$, $p = .043$, Greenhouse-Geisser corrected, $\eta^2 = 0.28$, see Figure 10.5 left). Participants were faster to react when the visual alerts were active. Searching for vehicles was faster than searching for humans. External switch trials were significantly slower as compared to repeat and internal switch trials (Repeat*Internal: $t(15) = 1.28, p = .64$; Repeat*External: $t(15) = -5.15, p < .001$; Internal*External: $t(15) = -6.43, p < .001$).

F1 The F1 accuracy score was significantly influenced by the factor switch and the interaction Switch*Task (switch: $F(1,14) = 38.2$, $p = .002$, $\eta^2 = 0.91$, Switch*Task: $F(5,14) = 19.89$, $p < .001$, Greenhouse-Geisser corrected, $\eta^2 = 0.83$, see Figure 10.5 center). Repeat trials were significantly more accurate as compared to switch trials (Repeat*Internal: $t(15) = 7.9, p < .001$; Repeat*External: $t(15) = 7.19, p < .001$; Internal*External: $t(15) = -0.70, p = 1$).

Misses The frequency of misses in the Search task was significantly influenced by the factors task and switch (Task: $F(1,14) = 18.17$, $p < .001$, $\eta^2 = 0.05$, Switch: $F(2,14) = 8.34$, $p = .002$, Greenhouse-Geisser corrected, $\eta^2 = 0.02$, see Figure 10.5 right). Searching humans resulted in significantly more misses. Internal switches had significantly less misses as compared to repeat trials (Repeat*Internal: $t(15) = 4.08, p = .001$; Repeat*External: $t(15) = 1.88, p = .21$; Internal*External: $t(15) = -2.2, p = .11$).

Navigation Task - Heading Mode

Between Condition Analysis

Reaction Time Reaction times in the navigation task heading mode depended significantly on the factors block, switch, and their interaction (Block: $F(3,29) = 4.16$, $p = .018$,

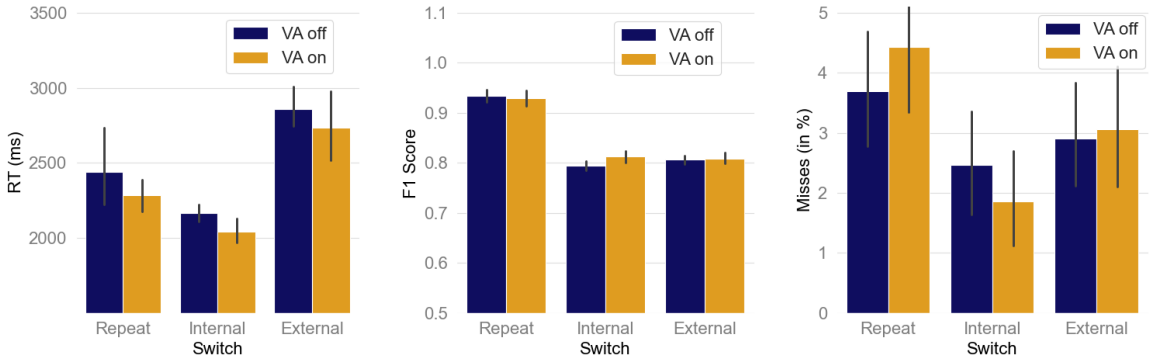

Fig. 10.5 Results of the within-subjects analysis of the search task

Greenhouse-Geisser corrected, $\eta^2 = 0.13$; Switch: $F(2,29) = 16.25$, $p < .001$, Greenhouse-Geisser corrected, $\eta^2 = 0.38$; Block*Switch: $F(11,29) = 12.37$, $p < .001$, Greenhouse-Geisser corrected, $\eta^2 = 0.314$, see Figure 10.6 top left). Reaction times decreased over time. Internal switch trials were significantly slower than both external and internal switch trials (Repeat*Internal: $t(30) = -3.31, p = .005$; Repeat*External: $t(30) = 2.37, p = .064$; Internal*External: $t(30) = 5.68, p < .001$).

Correct Direction Heading deviation was influenced by the switching ($F(3,29) = 5.82$, $p = .006$, Greenhouse-Geisser corrected, $\eta^2 = 0.17$, see Figure 10.6 bottom left). Repeat trials were significantly less correct as compared to both switch trials (Repeat*Internal: $t(30) = -3.05, p = .01$; Repeat*External: $t(30) = -2.87, p = .017$; Internal*External: $t(30) = .184, = 1$).

Deviation Heading deviation was influenced by the factor Block ($F(3,29) = 3.93$, $p = .016$, Greenhouse-Geisser corrected, $\eta^2 = 0.12$, see Figure 10.6 bottom left). Deviation decreased over time.

Misses Heading deviation was influenced by switches ($F(2,29) = 4.11, p = .025$, Greenhouse-Geisser corrected, $\eta^2 = 0.13$, see Figure 10.6 bottom right). External switches were significantly more frequent as compared to repeat trials (Repeat*Internal: $t(30) = -0.72, p = 1$; Repeat*External: $t(30) = -2.763, p = .023$; Internal*External: $t(30) = -2.05, p = .136$).

Within Condition Analysis

Reaction Time Reaction times in the navigation task heading mode depended significantly on the factors visual alert and switch (visual alert: $F(1,14) = 9.49$, $p = .008$, $\eta^2 = 0.4$; Switch: $F(2,14) = 17.79$, $p < .001$, Greenhouse-Geisser corrected, $\eta^2 = 0.56$, see Figure 10.7 top left). Trials with visual alerts were significantly faster. Repeat trials were significantly faster than both types of switch trials (Repeat*Internal: $t(15) = -467, p < .001$; Repeat*External: $t(15) = -5.56, p < .001$; Internal*External: $t(15) = -088, p = 1$).

Correct Direction Heading deviation was not significantly influenced by any independent variable.

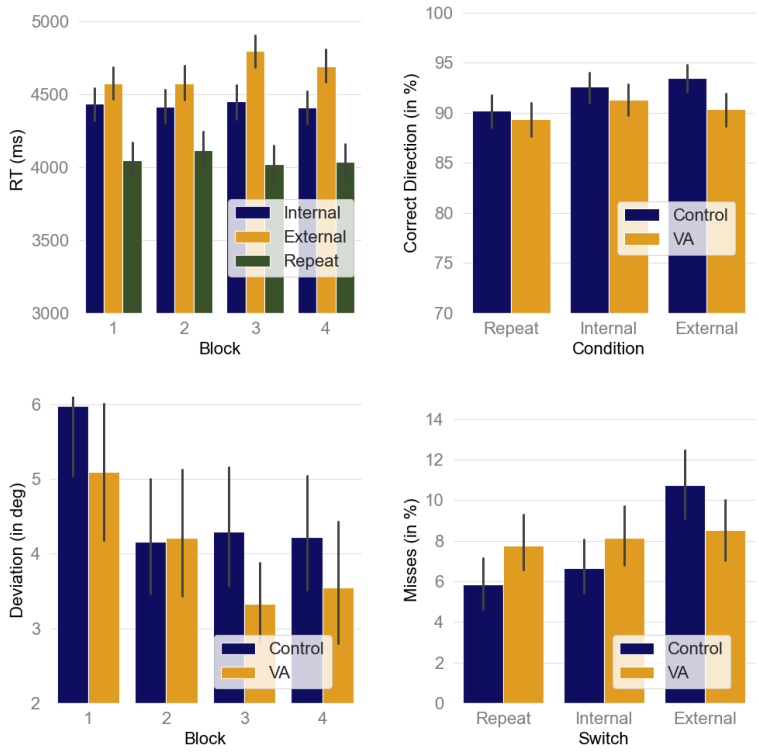

Fig. 10.6 Results of the between subjects analysis of the Navigation task, Heading Mode

Deviation Heading deviation was also not significantly influenced by any independent variable.

Misses Misses in the heading task were influenced by the visual alerts ($F(1,14) = 9.6$, $p = .008$, $\eta^2 = 0.02$, see Figure 10.7 top right). Misses were significantly less likely with the visual alerts.

Navigation Task - Waypoint Mode

Between Condition Analysis

Reaction Time Reaction times for choosing a waypoint in the Navigation task depended on both switches and the interaction switch*block (Switch: $F(2,29) = 43.63$, $p < .001$, Greenhouse-Geisser corrected, $\eta^2 = 0.61$; Block*Switch: $F(11,29) = 3.02$, $p = .015$, Greenhouse-Geisser corrected, $\eta^2 = 0.1$, see Figure 10.8 left). Repeat trials were the fastest, followed by internal and then external switch trials. All types differed from each other significantly External switches were significantly more frequent as compared to repeat trials (Repeat*Internal: $t(30) = -3.61, p = .002$; Repeat*External: $t(30) = -9.27, p < .001$; Internal*External: $t(30) = -5.65, p < .001$).

Accuracy No factor influenced the accuracy of Waypoint choices.

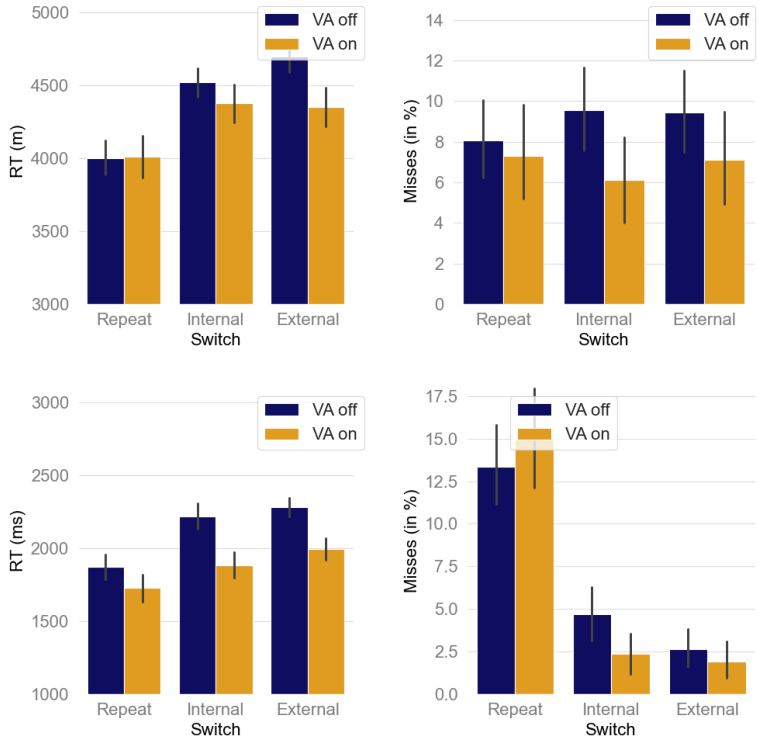

Fig. 10.7 Results of the within-subjects analysis of the Navigation task; Top: Heading Mode. Bottom: Waypoint Mode

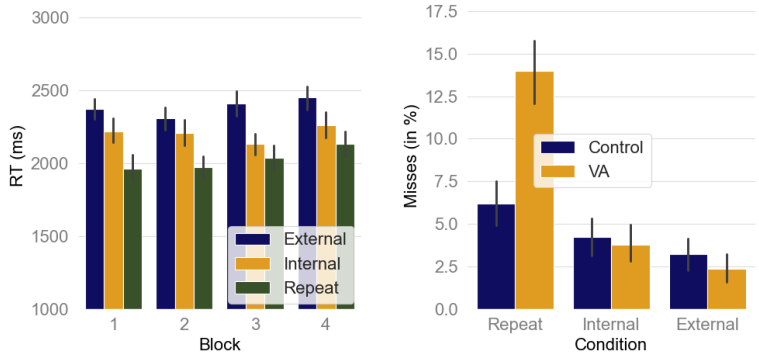

Fig. 10.8 Results of the statistical analysis of the Navigation task, Waypoint Mode

Misses Misses were influenced significantly by switches and the interaction switch*condition (Switch: $F(2,29) = 16.43$, $p < .001$, Greenhouse-Geisser corrected, $\eta^2 = 0.37$; Condition*Switch: $F(5,29) = 6.2$, $p = .015$, Greenhouse-Geisser corrected, $\eta^2 = 0.18$, see Figure 10.8 right). Misses were significantly more likely in repeat trials of the visual alerts condition ($t(30) = -3.3, p = 0.023$). Repeat trials had significantly more misses than switch trials (Repeat*Internal: $t(30) = 4.49, p < .001$; Repeat*External: $t(30) = 5.33, p < .001$; Internal*External: $t(30) = .84, p = 1$).

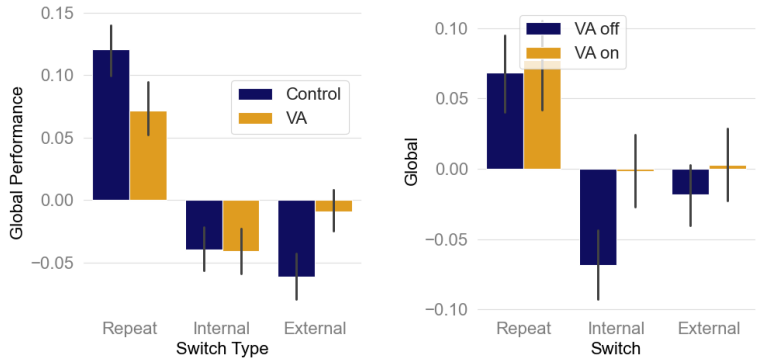

Fig. 10.9 Results of the Between Subjects analysis of the global performance score. Left: Between subjects. Right: Within-subject

Within Condition Analysis

Reaction Time Reaction times for choosing a waypoint in the Navigation task depended on the visual alerts, switches and the interaction switch*visual alert (visual alert: $F(1,14) = 44.01$, $p < .001$, $\eta^2 = 0.8$; switch: $F(2,14) = 11.29$, $p = .002$, Greenhouse-Geisser corrected, $\eta^2 = 0.51$; visual alert*switch: $F(5,14) = 7.422$, $p = .005$, Greenhouse-Geisser corrected, $\eta^2 = 0.4$, see Figure 10.7 bottom left). Reaction times were significantly faster with the visual alerts activated.

Accuracy No factor influenced the accuracy of Waypoint choices.

Misses Misses were influenced significantly by switches($F(2,14) = 13.33$, $p = .002$, Greenhouse-Geisser corrected, $\eta^2 = 0.53$, see Figure 10.7 bottom right). Overall, misses were less likely with the visual alerts activated. Particularly for both types of switch trials, misses were less likely due to the visual alerts (internal switch Visual alerts: $t(15) = 5.16, p < .001$; external switch Visual alerts: $t(15) = 5.49, p < .001$;).

Global performance score

Between Condition Analysis The accuracy and miss combined global performance score was significantly influenced by switches and the interaction switch*condition (Switch: $F(2,29) = 65.77$, $p < .001$, Greenhouse-Geisser corrected, $\eta^2 = 0.7$; Condition*Switch: $F(5,29) = 7.78$, $p = .002$, Greenhouse-Geisser corrected, $\eta^2 = 0.22$, see Figure 10.9, right). External switch trials performed significantly better with the visual alerts than without them ($t(30) = -3.74, p = .006$).

Within Condition Analysis The accuracy and miss combined global performance score was significantly influenced by switches ($F(2,14) = 17.86$, $p < .001$, Greenhouse-Geisser corrected, $\eta^2 = 0.6$, see Figure 10.9, right). Repeat trials had the best overall score, followed by external switch trials and then internal switch trials (Repeat*Internal: $t(15) = 5.93, p < .001$; Repeat*External: $t(15) = 3.58, p = .003$; Internal*External: $t(15) = -236, p = .027$).

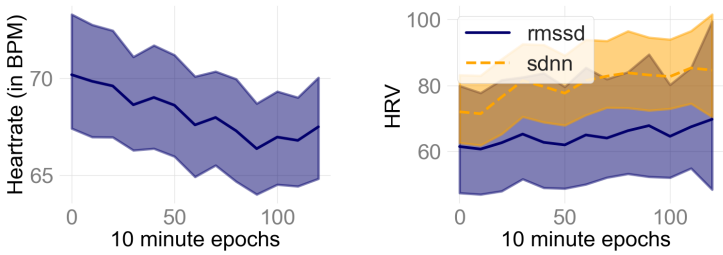

Fig. 10.10 Results of the cardiac analysis: Left: Heart Rate over time; Right: HRV over time.

II.10.3.2 ECG

HR (measured in BPM), as well as HRV (measured with the RMSSD and SDNN), were all sensitive to changes over time (BPM: $F(3,27) = 17.68$, $p < .001$, Greenhouse-Geisser corrected, $\eta^2 = 0.029$, RMSSD: $F(3,27) = 6.683$, $p = .003$, Greenhouse-Geisser corrected, $\eta^2 = 0.002$, SDNN: $F(3,27) = 18.36$, $p < .001$, Greenhouse-Geisser corrected, $\eta^2 = 0.021$). Over time, HR decreased while both SDNN and RMSSD increased (see Figure 10.10). Between conditions, there were no significant differences in HR or HRV.

II.10.3.3 EEG

Between subjects frequency analysis

Alpha Band Alpha power did not differ significantly between conditions but increased over time (factor blocks) in all 4 electrode clusters (Frontal: $F(3,29) = 5.33$, $p = .016$, Greenhouse-Geisser corrected, $\eta^2 = 0.17$; Central: $F(3,29) = 9.48$, $p < .001$, Greenhouse-Geisser corrected, $\eta^2 = 0.27$; Parietal: $F(3,29) = 11.24$, $p < .001$, Greenhouse-Geisser corrected, $\eta^2 = 0.3$; Occipital: $F(3,29) = 5.03$, $p = .021$, Greenhouse-Geisser corrected, $\eta^2 = 0.162$, see Figure 10.11).

Theta Band The theta band did not change in power across conditions. The central cluster of electrodes showed a significant increase in theta power over time (Central: $F(3,29) = 3.88$, $p = .045$, Greenhouse-Geisser corrected, $\eta^2 = 0.13$, see Figure 10.11 top right).

Gamma Band The gamma band showed no significant change over time or across conditions.

Beta Band Beta frequency power did not change over time or across conditions.

Exploratory Time of Day Analysis

The exploratory analysis highlights the effect of the Time of Day (TOD) on some of the results obtained in this experiment. We focussed only on the between-subjects analysis for the behavioural data as the 15-participant subset would likely be too biased by the unbalanced TOD factor.

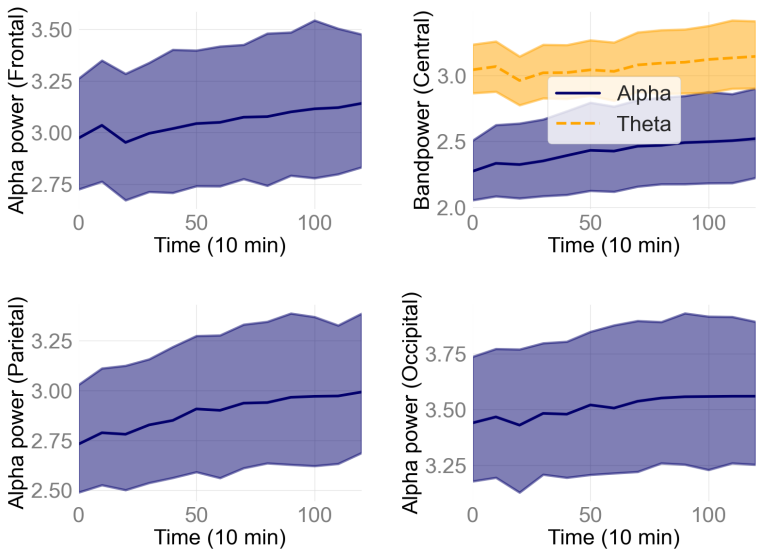

Fig. 10.11 Results of the Between Subjects analysis of the EEG frequency bands.

Subjective

Including the TOD factor resulted in the KSS no longer showing any significant effects. The results of the other subjective questionnaires remained unchanged following the inclusion of the SPS scale.

Behavioral Between Groups

Adding the TOD factor as a covariate on the dependent variable of the F1 score in the Search Task resulted in TOD, Condition and the interaction Switch*Condition being significant (TOD: $F(1,29) = 9.2, p = .006, \eta^2 = .269$; condition: $F(1,29) = 5.84, p = .023, \eta^2 = .189$; Switch*Condition: $F(5,29) = 6.547., p = .005$, Greenhouse-Geisser corrected, $\eta^2 = .208$). Visual Alerts improved performance, and participants in the afternoon performed better. Multiple comparisons showed that performance with the visual alerts improved, particularly in external switch trials (Control External * Visual Alerts External: $t(30) = -3.248, p = .03$). Contrary, afternoon participants choose the correct heading direction less often ($F(1,29) = 5.073, p = .033, \eta^2 = .158$).

Inclusion of the TOD factor also resulted in the factor Condition becoming significant on the dependent variable reaction time in the waypoint mode ($F(1,29) = 11.122, p = .002, \eta^2 = .292$). Visual Alerts significantly improved the reaction times (365.14 ms difference). For the global performance score, the inclusion of TOD factor revealed a significant improvement in performance with the visual alerts for external switch trials while resulting in a decrease in performance in repeat trials (control Repeat * Visual Alerts Repeat: $t(30) = 3.308, p = .025$; control External * Visual Alerts External: $t(30) = -3.89, p = .004$). TOD itself did not influence the global performance score.

EEG & ECG

Including the TOD factor did not significantly alter any of the EEG results. The heart rate of participants in the afternoon was significantly higher ($F(1,27) = 4.288, p = .049, \eta^2 = .132$).

II.10.4 Discussion

This experiment aimed to evaluate the effectiveness of visual alerts (VA) as an adaptive interface to support the cognitive flexibility of operators in the context of long-endurance missions. Two groups of participants performed a two-hour search and navigation task. In one group, the visual alert adaptation was activated pseudorandomly, mimicking an adaptive interface.

Subjective results showed that fatigue and workload increased in both groups over time. Surprisingly, as measured by the SPS measure, mental fatigue showed a significantly higher increase in fatigue in the control condition than in the visual alerts condition, suggesting that the visual alerts may help suppress mental fatigue. Alternatively, this may indicate that the visual alerts render the task more engaging. Notably, this result remained significant following the inclusion of the TOD factor in the exploratory analysis. Behavioural results were divided into a within and a between-condition analysis. In the between-condition analysis, the visual alerts showed little effect. Notably, misses were more likely in the VA condition for repeat trials in the Navigation task Waypoint mode. On the other hand, the global performance score showed a better performance for external switch trials with the visual alerts without any significant performance decrement in the other trials. The inclusion of the TOD factor in the exploratory analysis revealed more factors. Here, the visual alerts were shown to improve accuracy and decrease reaction time while showing reduced performance in the correctly chosen direction in the heading task.

The absence of more significant effects between the groups is likely attributable to four factors. The visual alerts were only active for around 30-35 % of the overall 2-hour period, but the overall behaviour is compared in the between-condition analysis. Furthermore, participants in the VA condition may improve their behaviour when the visual alerts are active but disengage when they are not active. In addition, the schedule for the visual alerts was not based on the actual mental state of the participants but was randomly assigned. Finally, participants might require more training with visual alerts. In a future study, TOD should be a balanced factor with equal numbers of participants arriving in the morning and in the afternoon, or with the inclusion of other factors controlling for the influence of said factor.

Another aspect to highlight is that internal switches in the heading mode are slower than external switches in the same mode. For the Waypoint mode, the opposite is observed. This may indicate that participants tend to rest their hands on the joystick. Therefore, they do not need to move their hand in an external switch and can react quickly. In an internal switch, however, they must move their hand to the trackball. The opposite is true for the waypoint mode, where both an external and an internal switch require hand movement.

The within-subjects analysis of behaviour focused only on the visual alerts group and compared the trials in which the visual alerts were active vs. those in which the visual alerts were inactive. This analysis shows a general improvement in reaction time when the visual alerts are active. This shows that the VA support the operators. Still, in a realistic scenario, they would not improve overall reaction time due to the delay caused by the visual alerts themselves (see Figure 10.2). However, the visual alerts also significantly reduced the misses in the heading mode of the navigation task.

EEG and ECG analysis showed no difference between groups but supported the results of the subjective analysis, indicating an increase in mental fatigue.

Overall, the behavioural analysis shows that visual alerts impact behaviour, and despite the tendency to indicate performance improvement, the results are not conclusive enough to state that visual alerts in this form improve performance. This is likely attributable to the

analysis, sample size (15 participants per group), training, and the fact that the visual alerts were independent of the operator. Brain-computer interfaces and Physiological computing systems may likely improve this by activating the visual alerts only when the operator is in a degraded mental state and requires support. The results also align with the hypothesis that an adaptive interface improves performance best when the operator is in a mental state requiring support (see Section I.2.2.3).

Currently we are preparing a contribution to an international conference, to present the results of this work.

Part III
Discussion & Afterword

Chapter 11

Discussion

The here presented work investigates UAS operators, the mental fatigue-related human factors issues, and possible ways to mitigate these issues with passive Brain-Computer Interfaces and Adaptive Interfaces. To investigate six Research objectives (RO), we conducted 5 experiments ranging from fundamental studies investigating cognitive flexibility (see Chapter 6) to more ecological studies testing countermeasures in UAS simulations (see Chapter 10). The following discussion is structured along the 6 ROs with open questions limitations and a future research section at the end of each group of ROs (UAS operations, mental fatigue & cognitive flexibility).

III.11.1 RO1) UAS Operations & Interfaces

Central Ideas

Relevant Contributions:

- UASOS (see Chapter 9)
  - The UASOS experiment was a more realistic implementation of our cognitive flexibility protocol, allowing for investigating cognitive flexibility and mental fatigue in UAS operations.
  - Participants alternated between a search and a navigation task, each with two submodes of operation.
- UASOS II (see Chapter 10): The UASOS II experiment used the UASOS protocol but added visual alerts for mitigating task-switching costs to one group of participants. Learning from the DTS-II study, the visual alerts were explained to the participants. Visual Alerts improved performance on certain metrics.

Key Points:

- Besides mental fatigue, operators likely encounter operational difficulties due to HMI issues.
- The limited availability of operators is a major constraint for the research on UAS operators.

One future question: **Can visual alerts support UAS operators under more realistic conditions?**

III.11.1.1 RO1 a) Why do Uncrewed Aerial Systems (UAS) suffer from a high rate of human factor-related mishaps?

Due to their long endurance, versatility, and cheap operational costs, UAS military and civilian operations are increasing in popularity and scope (Cauchard et al., 2021)[1]. Despite high levels of automation in UAS, human factor mishaps are still common (see Section I.2.1.4, Waraich et al., 2013). Factors such as retasking and the necessity to continuously observe the visual feed of the UAS necessitate operators to work long hours with little breaks (see Section I.2.1.5). As a result, operators experience mental fatigue, resulting in attentional errors, difficulty switching tasks, and decreased situational awareness (see Section I.3.1.4). Shift work and 24-hour operations add to fatigue issues and circadian rhythm disruption. Beyond fatigue, mishaps may also be attributed to poor interface design. Mishaps include an operator mistaking the engine off switch with the gear down switch or operators thinking the UAV is already on the ground while still airborne (see Section I.2.1.4). These examples showcase the potential for improvement in interface design. Finally, procedures and checklists present a final risk factor frequently causing UAS mishaps.

There are likely to be many more issues in UAS operations, and they are also likely to be more complex, than presented here. Unfortunately, access to operators is limited. Operators

[1] https://www.icao.int/annual-report-2021/Pages/emerging-and-cross-cutting-aviation-issues-increased-use-of-unmanned-aircraft-systems-uas.aspx

are often either in training or in actual deployment. Furthermore, the French air force currently only operates 12 Reapers[2], further limiting the number of pilots that could be available for research.

III.11.1.2 RO1 b) Which modifications in interface design can improve the safety of UAS operations?

A simple solution to address the issue of mental fatigue would be more frequent rotations between teams operating a single UAS. The problem is that this would require many more operators to operate one UAS, negating one of the great advantages of UAS (cost). Another problem is that hand-off procedures between teams are a large risk factor during operations (see Section I.2.1.4).

The other obvious solution would be to mitigate mental fatigue. However, effective methods of doing this mostly revolve around taking breaks. A viable alternative to mitigating mental fatigue is accommodating its effects (see Section I.3.1.4). Our research identified cognitive flexibility as a dangerous intersection of several factors. Cognitive flexibility is required during UAS operations (several different tasks to perform at all times, also retasking) and has been indicated by operators to pose problems when fatigued. It is also been shown that mental fatigue negatively affects cognitive flexibility.

As stated in the previous section, mental fatigue is not the only issue that UAS operators face. Interface design will likely be at fault for some UAS mishaps, and improvements may easily result in safer operations.

Adaptive User Interfaces (AUI) present a possible solution to improve the user interface. AUI adapts based on input data, including physiological measures (see Section I.2.2.2). As shown in Section I.2.2.3 AUI present an advantage over static modifications of the interface, as they can adapt to the modified capacities of a user in specific situations. AUIs may mitigate the effects caused by mental fatigue rather than reducing mental fatigue.

One example of AUI are the visual alerts that were tested in the DTS-II and UASOS-II study (see Chapters 7 & 10 for the studies and Section III.11.3.2 for a more detailed discussion on visual alerts and their effectiveness). Visual alerts can be implemented in several ways within UAS operations. They could be linked to communications (and instructions) that operators receive, flight parameters and aspects that require regular surveillance.

For several reasons, adaptive interfaces using visual alerts are an interesting concept for HMI. Adaptive User Interfaces are often confronted with the question, "If there is a way to improve the interface, why not use it all the time?" One argument is the optimal workload theory, which states that participants perform best when task complexity matches their current mental state (see Figure 2.7). Along those lines, an improved interface may result in the underload of the participant. The flaw of this argument is that if a participant is experiencing cognitive underload, then making the task more 'difficult' is not the only way to resolve this issue. One might as well add a task, increasing the cognitive load to match the mental state of the operator. In that way, the participant would profit from an improved interface and simultaneously increase their work. Visual Alerts as AUIs have another justification. Evaluating performance on a multidimensional scale may only significantly improve when the participant is in a suboptimal state (see Figure 2.8).

[2] https://lignesdedefense.blogs.ouest-france.fr/archive/2020/01/15/un-nouveau-systeme-de-drone-mq-9-block-5-livre-la-semaine-de-20798.html

III.11.1.3 Open questions, limitations & future research

During this PhD, access to UAS operators and understanding their operations was complicated because of their limited availability and because many aspects of UAS OPS remain classified as military secrets. Ideally, to design interactions for UAS operations, more access to UAS operators is needed to apply a user-centred design methodology (Beaudouin-Lafon & Mackay, 2007). Interviews, observations and other means of obtaining information are important for understanding the particularities of UAS operations. Future studies measuring the physiological activity of operators during training missions would provide much-needed insight into crucial aspects of the Human-Machine Interaction (HMI) and the degrees of mental fatigue that operators experience (Tvaryanas et al., 2008).

More operator contact would likely also have improved the realism of the UASOS protocol concerning how the operator is portrayed in the study. Here, the roles of the sensor operator and the pilot are mixed, while in reality, this work is executed by two different persons.

Getting feedback from operators on the visual alerts would also be valuable. Likely, there are constraints in UAS operations that we are unaware of.

Moreover, one limitation of the experimental environments developed during this PhD is that adaptations focus on visual alerts. In the future, studying other types of adaptive interfaces would be interesting. The literature also highlights the importance of trust in systems (see Section I.2.2.6). This aspect was considered in the design, but future studies should test to what degree participants trusted the system and how trust could be increased. Furthermore, some studies highlight the advantage of giving operators control over the adaptation (to turn it on or off at any given time). It would be interesting to see if this increases performance with the visual alerts.

The next step in lab-based research would be to move the UASOS protocol towards more realism. It could be possible to increase the complexity and create scenarios or investigate how operators could interact. Future studies might investigate how visual alerts may aid in contexts other than UAS OPS such as nuclear powerplant supervision or single pilot operations.

III.11.2 RO2) Mental fatigue

Central Ideas

Relevant Contributions:

- Double Task Switching (DTS) (see Chapter 6): The DTS study tested a fundamental task-switching protocol examining similarity and mental fatigue. Behavioral results concerning mental fatigue showed little to no effect on participants. This may be partially attributable to learning effects.
- Remot3e (see Chapter 8): In the Remote experiment, participants had to perform a visual search task for 60 minutes while equipped with an EEG, ECG and an eye-tracker
  - While mental fatigue was shown to impact some metrics of behaviour, there was a small global effect of TOT on behaviour.
  - Performance and TOT estimation of mental fatigue showed that good classifications were easily achievable with TOT labels but not with performance metrics.
- UASOS (see Chapter 9): The UASOS experiment recreated the DTS protocol in the context of UAS OPS. As with the Remot3e experiment, some but not all parameters were influenced by mental fatigue (TOT). Task engagement and learning effects may have played large roles in this result.
- UASOS II (see Chapter 10): The added visual alerts in the experiments resulted in a smaller increase in subjective mental fatigue.

Key Points:

- Issues of assigning fatigue labels to data feed into the larger question of how mental fatigue should be defined.
- Besides performance estimation and TOT, fatigue indexes allow for another approach of labelling mental fatigue.

One future question: **What definition of mental fatigue allows for generalizable and accurate predictions of factors such as behaviour and subjective feeling?**

III.11.2.1 RO2 a) What is a useful definition of mental fatigue within complex operations?

Scientific constructs such as mental fatigue are not directly observable, meaning there is no way to measure said construct accurately. Their value lies in understanding interactions between relevant variables and predicting outcomes. However, it is difficult to determine if a false prediction is due to an incomplete or faulty construct. For example, if an intelligence test does not reliably predict professional success, is that because it doesn't measure intelligence very well, intelligence and professional success are unrelated, professional success is related to more factors than intelligence or because of a combination of one or more reasons?

In its simplest form, mental fatigue is defined as changes in subjective feelings, behaviour and physiology resulting from changes in motivation and time-on-task (see Sections I.3.1.1 & I.3.1.2).

While our experiments have repeatedly been shown to induce mental fatigue on the subjective level, and we were also able to observe physiological changes, the behaviour was often more robust and did not always change as expected.

In Section I.3.1.5, we already stated that changes in motivation, learning and novelty effects, and compensatory strategies are possible explanations for the absence of mental fatigue effects.

Except for the DTS experiment (see Chapter 6), all of our experiments were conducted within one session. It is reasonable to expect that learning effects may reduce the impact of mental fatigue. Especially the UASOS and UASOS II experiments (see Chapters 9 & 10) involved very complex tasks. While the training may have been sufficient to allow for a complete understanding of the task, participants may not have been able to create appropriate strategies. This is likely not the case in realistic environments, where operators often have extensive experience.

The absence of behavioural effects may also be attributable to the incomplete measuring of behaviour. Especially in a complex task such as UASOS, behaviour may not be adequately captured by looking at accuracies and reaction times but rather by metrics that capture more general aspects such as situational awareness or planning.

These are reasonable explanations for the unexpected effects of mental fatigue. However, it's also worth considering that our current understanding of mental fatigue is incomplete. As explored in Section I.3.1.4, mental fatigue should not be seen as directly resulting in behavioural changes but rather as influencing attention, cognitive flexibility and situational awareness. So, while little changes were observed behaviorally, it is possible that participants had, e.g. reduced situational awareness, but the behavioural measures did not pick up this metric.

A better understanding of mental fatigue and the covariate and confounder variables that drive subjective, behavioural, and physiological outcomes will result in a more complete and accurate construct.

III.11.2.2 RO2 b) How do markers of mental fatigue (physiological, behavioural & subjective) perform at predicting mental fatigue depending on different definitions of the construct?

Performance Estimation

A more complete construct of mental fatigue promises to understand how a person becomes fatigued and how this would affect said person. It also has very practical implications. As explored in Section I.5.1, supervised machine learning algorithms required labelled data with a *'ground truth'*. In research and practical applications of mental fatigue, this is often either TOT or performance (see Section I.5.2.2 for a discussion on the issue). The differences between the prediction of performance and TOT that we observed in the Remot3e and UASOS experiments (see Chapters 8 & 9), along with the behavioural results discussed above, highlight the difference between both approaches.

Performance estimation can be interpreted in at least two ways. The argument can be made that any performance change is at least partly due to mental fatigue, making these changes a marker of mental fatigue. Performance estimation is then also mental fatigue estimation.

Seeing performance as a marker of mental fatigue should also incorporate aspects of attention, situational awareness and cognitive flexibility. Behavioural markers should, therefore, be

markers that are congruent with changes in said mental state. In this case, the argument goes along the lines of: "Mental fatigue is expected to reduce top-down attention. Changes in top-down attention are measurable with behavioural metric X but not with behavioural metric Y. Therefore, measuring metric X may also be correlated with mental fatigue."

Alternatively, performance estimation may also be used directly, forgoing the theoretical construct of mental fatigue (and its link to TOT), by aiming to predict performance regardless of the mental state that causes performance changes.

This is an appealing idea, but there are further issues to consider beyond practical difficulties in achieving strong classification performances. First and foremost, performance needs to be defined (Brouwer et al., 2015). As seen the UASOS experiment (see Chapter 9) and discussed in Section I.5.2.2, defining performance becomes ever more complex as observed metrics increase. Perhaps, rather than looking at simple metrics in complex tasks, it is a more robust approach to develop more abstract metrics of behaviour that objectify strategic thinking or planning. Another issue of performance estimation is that it is not sure to generalize to other tasks, as there is no underlying assumption as to why performance changes occur. The mental state of performance may also be very subject-dependent. A participant focusing on speed may perceive and thus influence their mental state and physiology if they believe they are doing well.

These issues need to be addressed by further research aiming to improve the accuracy and dependability of performance estimation using pBCIs and physiological computing systems. While many improvements in both domains have often focused on optimizing the machine learning aspects (e.g. the use of Riemannian geometry see Congedo et al., 2017), a better understanding of performance's cognitive and neuroscientific mechanisms may be the more urgent direction. This is because other than, for example, Motor Imagery, TOT or SSVEP protocols, performance has many, or depending on how it is seen, no ground truth. That is not to say that there is no research on performance estimation. However, this mostly focuses on fundamental tasks or uses event-related designs to investigate ERPs (Alhazmi et al., 2015; Blaha et al., 2016; Ko et al., 2015; Müller et al., 2008). The applicability of ERPs for performance estimation is limited when moving towards more realistic tasks as events may occur less frequently or tasks may become less defined by events (Chavarriaga & Millan, 2010).

Fatigue Indexes

Beyond TOT and performance estimation, other approaches to estimating mental fatigue should be considered.

Using subjective questionnaire results may improve TOT metrics, but due to the issues of subjective questionnaires (see Section I.3.1.5), these metrics are likely to resemble those of TOT closely. One possibility is the idea of creating physiology-based indexes of mental fatigue (Monseigne et al., 2019; Wang et al., 2018). The idea is that the ground truth of mental fatigue is based on one or more physiological metrics. This has the advantage that it allows for non-linear effects of mental fatigue (unlike TOT) but might also be generalizable across different tasks (unlike performance estimation).

Two points need to be considered here. If, for example, some EEG metrics are used to create an index of mental fatigue, and the goal is to predict that index, using the EEG data would be confounded. The other issue with indexes is how an index is initially validated. If an index was validated using TOT or performance, a pBCI predicting the index is essentially a system predicting TOT or performance.

In the end, mental fatigue remains a construct, and our predictions will always reflect the assumptions made about the structure of said construct. For research, this means that there should be continued efforts to understand the construct, what drives fatigue and what the results are. For practical applications, the ground truth is that TOT, performance or an Index of Fatigue may be of more importance than the approach used to predict this. The leading question should be: "What do we gain from predicting X?"

Alternatively, the issue could be framed as a neuroscientific issue. Assuming that the underlying neural mechanisms that drive, e.g. motor imagery, mental fatigue and other target states, are the same across participants with noise due to anatomy, current state, etc., we could assume that markers of this true mental state should generalize across participants. This argument, at its most extreme, thus sees mental state detection as a purely neuroscientific problem. The truth likely is that BCIs and PCS need both neuroscience and machine learning. However, changing the perspective highlights that for a long time, the trend has been to test new algorithms, classifiers etc. and to evaluate them on their outcome and not based on the underlying hypotheses why X particular algorithm should be particularly adapt at detecting a mental state and what we can learn about the brain and behaviour with said algorithm.

Furthermore, BCIs and PCS also present a phenomenal way of hypothesis testing. A hypothesis can be transformed into a machine-learning model and tested on participants in real-time. For example, the hypothesis may be that frontal theta power increases with mental workload. If this hypothesis is injected a priori into a machine learning algorithm that may then be calibrated to a participant to detect changes in frontal theta power, we can use this BCI to test our hypothesis in real time.

III.11.2.3 Open questions, limitations & future research

As explored in Section I.3.1, mental fatigue and fatigue are significant issues in many domains. Still, it is a concept that is somewhat poorly understood. Many questions on the interactions between TOT, motivation, task complexity and performance remain to be answered. Further investigating the links between physiological markers and mental fatigue would also be interesting. Does increased alpha band power drive fatigue, or does the power increase result from mental fatigue? Transcranial magnetic stimulation and combined EEG fMRI studies may pave the way towards a more comprehensive understanding of the origins of mental fatigue and its physiological markers (Peters et al., 2020 as an example of a combined EEG, fMRI, TMS study & Fiene et al., 2018 as an example of how cerebral stimulation can effect mental fatigue).

The other big question is how performance estimation can be improved, especially when tasks have multiple dependent measures. Most often, performance is measured in terms of accuracy and reaction times. Evaluating performance multidimensionally as proposed in Section I.2.2.3 may result in better performance estimation (due to better performance labels) and further our understanding of performance, but has, to the best of our knowledge, not been done before. The approach of utility cures allows for both a single digit of performance, expressed as the utility, but it also allows for the evaluation of how each single metric plays into the overall scores. It furthermore allows the use of individual utility curves for each task. Creating ways to objectively measure situational awareness or strategic thinking in complex tasks may also be interesting. Eye-tracking could support this, as it may allow the extraction of movement patterns that can be identified as more or less strategic. In future steps it may also be of interest to evaluate the capability to estimate performance using a regression approach rather than classification (for an example for an example of this see Bose et al., 2020).

The UASOS study (see Chapter 9) used eye-tracking glasses. Unfortunately, hardware and software constraints from the actual eye-tracking product made the data unusable. The data would likely have allowed for a more precise performance estimation. Simple metrics such as saccade velocity and fixation times can be a reliable marker of fatigue (Di Stasi et al., 2013). Furthermore, it would have been interesting to extract more complex gaze patterns. Is the participant's gaze different before an error? Do misses occur because the participant did not look at the task or did not process the information? These and other questions may be best addressed by redesigning the UASOS task onto one screen and using a fixed eye tracker. Alternatively, or additionally, the inclusion of Electrooculography (EOG) electrodes for the use of extracting eye blink metrics may present a simple approach to gaining valuable eye-movement information.

Similarly, metrics related to using the joystick or the trackball/mouse may provide valuable information for fatigue detection.

Assessing the influence of personal beliefs on performance or even manipulating these beliefs could also be a path towards a better understanding of performance estimation. Finally, studies should investigate if clustering could allow for the detection of behavioural strategies in response to mental fatigue. Perhaps people group into different response archetypes (Wood & Jennings, 1976). In addition, future studies should take more precautions in accounting for the effects of caffeine, circadian rhythm and sleep. The DTS (see Chapter 6) study found no impact of these factors on performance. To shorten subsequent studies, these questionnaires were partially omitted. However, the exploratory analysis of the UASOS-II study (see Chapter 10) showed a significant impact of time-of-day on several performance metrics. Fatigue modelling as done with fatigue management systems (Cabon et al., 2012) would likely improve the modelling of mental fatigue.

Mental fatigue is fascinating, not least because everybody has experienced it. For now, there is no universal solution to mitigate fatigue, but a better scientific understanding of mental fatigue may likely reduce the negative impact in the future.

III.11.3 RO3) Cognitive flexibility

Central Ideas

Relevant Contributions:

- Double Task Switching (DTS, see Chapter 6)
  - The DTS study tested a fundamental protocol to validate whether task-switching is different when similar or dissimilar tasks are concerned.
  - The study showed that similarity impacts reaction times and accuracies differently. Switching between dissimilar tasks results in longer reaction times but higher accuracy than switching between similar tasks.
- Double Task Switching-II (DTS-II, see Chapter 7): The DTS-II study tested if different types of visual alerts could reduce task-switching costs. The study's results could not show any impact of visual alerts on cognitive flexibility.
- UASOS (see Chapter 9): The UASOS experiment recreated the double task switching protocol in the context of UAS OPS and overall showed to generally recreate the task-switching effects.
- UASOS II (see Chapter 10): The added visual alerts in the UASOS-II experiment improved reaction times as well as some accuracy-related performance metrics.

Key Points:

- Cognitive flexibility is a central point of many complex operations, but it is not systematically researched using more realistic tasks.
- Visual Alerts, as implemented in the UASOS II task, significantly improved performance.

Future Questions: **How can we best understand cognitive flexibility in realistic contexts?**

III.11.3.1 RO3 a) What do we need to understand about cognitive flexibility to move from fundamental tasks to more realistic environments?

Many studies investigating task-switching focus on fundamental protocols that use tasks such as judging if a number is smaller or bigger than 5 and whether a number is odd or even (Koch et al., 2018, see Section I.3.2.2). In realistic contexts, UAS operations operators switch between very different tasks (see Section I.2.1.3). The goal of the DTS study was to investigate whether switching between similar tasks (as done in the fundamental protocols, switching between a high/low and odd/even task) is different than switching between dissimilar tasks (as in real-world applications; switching from navigation to communication or mission planning). While studies such as Arrington et al. (2003) and Cooper et al. (2015) have previously examined switching between more than two tasks, our study presents a systematic effort to investigate the factor of similarity in cognitive flexibility. We were able to show that there is a difference when switching between tasks depending on similarity (see Chapter 6). Participants take longer switching between dissimilar tasks, but errors are more frequent when switching be-

tween similar tasks. With the task-rule convergence effect, we also discovered that distracting information had a stronger impact on performance when congruent with the task participants had just switched to.

The more realistic UASOS protocol was strongly inspired by the DTS protocol with 4 different tasks, again arranged in two pairs of two tasks (see Chapter 9). Globally, the effects of the DTS study were recreated; however, several aspects need to be considered here. The task required participants to switch between modalities for the navigation task (using either a joystick or a trackball). Participants also had to use two screens and thus shift their gaze when switching tasks but not when repeating tasks. This effect is perhaps less strong as the task instructions were located in the centre portion of the screen. Therefore, participants had to direct their attention irrespective of trial type. This study shows that the fundamental aspects of task-switching can be studied in more realistic contexts and that task-switching effects also exist in these contexts. To the best of our knowledge, task-switching has not been investigated in such realistic conditions before.

III.11.3.2 RO3 b) Are there ways to reduce the effects of task-switching?

In two studies, we examined the effect of visual alerts on performance (see Chapters 7 & 10). Visual Alerts are supposed to let participants know that they have to switch tasks in an upcoming trial, thus reducing the task-switching costs. Furthermore, tasks with distributed information allow the participant to focus on the relevant task as they do not have to check whether they have to switch tasks at any given point. The DTS-II study aimed to examine whether the mere presence of visual flashes can reduce task-switching costs. The results showed no significant performance improvement due to the flashes. This was likely due to a combination of a complex study design and lacking power due to a small sample size. In the UASOS-II study, we examined the effects of visual alerts on a more ecological setting.

There was no overall improvement in performance. However, we observed that the visual alerts improved performance in reaction time and some non-reaction time-related metrics. In the form that the visual alerts were implemented here, that is by appearing instead of actual task instruction and delaying said instruction by 700 ms, improved reaction times would not be helpful, as the 700 ms delay would negate any positive effect. However, the effect still highlights that the visual alerts positively impacted performance. Furthermore, the reaction time differences were relatively large (more than 350 ms for the Waypoint mode for example). Reducing the delay of the visual alerts may allow for an improvement in reaction time.

The exploratory analysis also revealed many more significant changes and improvements in performance that likely have been masked by TOD effects. Subjective fatigue did not increase as much for participants who performed the task with the visual alerts as it did for those without them. This may be due to the visual alerts rendering the task less fatiguing, but it could also be due to the task being more engaging with visual alerts (Hopstaken et al., 2015a).

The absence of more global performance improvement is likely attributable to the fact that the visual alerts were not activated and deactivated based on the users' mental state but rather pseudorandomly. Coupling this adaptive interface with a BCI or physiological computing system would likely improve the performance. Furthermore, the study might require a larger sample size. Several sessions to allow participants to explore the usefulness of the adaptation fully may also help (Lee and See, 2004, see Section I.2.2.6). Finally, future studies should investigate whether participants disengage once the visual alerts are deactivated. If that is the case, solutions to this issue need to be explored. One possibility would be to have the BCI or Physio-

logical system suggest whether or not to activate the visual alerts but let the participants make the final choice themselves (see Section I.2.2.6 for further information on how understanding an adaptation and Trust influence the efficacy of adaptive interfaces).

To the best of our knowledge, mitigating the negative impact of mental fatigue by targeting its effects (foremost decreases in cognitive flexibility) presents a novel approach to the problem of mental fatigue. Further studies need to investigate the extent to which this particular solution (the visual alerts) is successful. Yet, the approach itself may be applicable to other effects of mental fatigue, such as situational awareness and attention, and may even be useful in successfully counteracting mental overload or other suboptimal mental states.

III.11.3.3 Open questions, limitations & future research

Questions and future research on cognitive flexibility due to the DTS and UASOS studies go into three directions: (1) Investigating fundamental aspects of CF. (2) Investigating the effects of visual alerts on CF. (3) CF in realistic contexts.

The DTS study was novel as it used four instead of two studies but was simultaneously capable of controlling for congruency between stimuli. The results merit further investigation. The task congruency convergence showed that the previous task interacts with the presented stimuli. A dedicated study to replicate this effect may allow a better understanding of CF. We saw strong learning effects across all of our studies. Studies with more sessions to better account for this would likely also allow for a better investigation of the effects of mental fatigue. More extensive pre-testing and a dedicated pilot study would likely have significantly improved the studies. This would have allowed for detecting the extent to which both pauses and learning effects influence mental fatigue. Learning effects were especially underestimated in our experiments. Dealing with these would, on the other hand, have required a much larger experimental design. The DTS study (see Chapter 6) showed that one training session is insufficient to eliminate learning effects. So, to avoid a long and fatiguing training session before the actual task, participants would likely have to come to the lab for two or even more training sessions. This risks a much higher rate of attrition.

More training and fewer conditions may also have improved the investigation of the visual alerts. Fewer conditions and a simple two-task switching protocol rather than the 4 task protocol would allow for a more rigorous investigation of the effects of visual alerts on cognitive flexibility.

The next important step should be to develop a robust way to estimate performance within the UASOS task. Once achieved, this could then be linked to the visual alerts, creating a true adaptive interface. Moving towards this goal some preliminary tests using transfer learning algorithms have been conducted (Kobler et al., 2022). The results are detailed in Appendix V.14.6. A working adaptive interface would allow for a complete evaluation of the effectiveness of the visual alerts and their feasibility as a support to UAS operators.

Finally, to move CF towards more realistic protocols, the next step would likely be to recreate the UASOS protocol, accounting for aspects such as gaze shifting and modality changes. This could be achieved by presenting the task on one screen and having the Numpad register all responses. As alluded to in the section on UAS operations, the step after that should be to look at 'larger' less frequent switches. This way, one could investigate whether the issue of retasking is directly related to cognitive flexibility.

III.11.4 Conclusion

The work presented in this manuscript contributes to the scientific community in several ways.

To our knowledge, the idea of multidimensional performance evaluation as an argument for using adaptive interfaces has not yet been suggested. Furthermore, this work shows the importance of incorporating several behavioural measures to understand a person's behaviour.

The work on passive Brain-Computer Interfaces in this manuscript highlights an issue with mental fatigue. The definition of mental fatigue often revolves around time-on-task. While producing strong results in terms of classification accuracy, their real-world application is limited. On the other hand, performance estimation is difficult, especially when performing more realistic tasks.

One central bottleneck for the research on passive Brain-Computer Interfaces is data availability (Hinss, Somon, et al., 2021). The process of testing new hypotheses or algorithms can be time and resource-consuming if a new dataset needs to be acquired for each test. Open datasets present a simple but underused opportunity to alleviate this issue. To work towards more open data in the passive Brain-Computer Interface community, we published a dataset of several cognitive tasks to do so with a specific focus on transfer learning. To this extent the dataset allows for cross-participant, cross-session and cross-task mental state estimation (Hinss, Jahanpour, Somon, et al., 2023,see Appendix V.14.7).

The work on mental fatigue also highlights several important aspects. Mental fatigue should be seen as a mental state that influences cognition, thus resulting in behavioural changes rather than influencing behaviour and performance directly. Cognitive flexibility is one of the aspects influenced by mental fatigue. Here, we could show that switching between tasks depends on similarity. This aspect is important as cognitive flexibility is mostly studied in fundamental contexts with similar tasks, assuming these results would map to more realistic contexts with dissimilar tasks.

The fact that mental fatigue influences cognitive flexibility and attention presents an understudied opportunity to mitigate the negative effects of mental fatigue beyond resting. We showed that visual alerts can reduce task-switching costs when alternating between tasks. Beyond this, we found that the presence of visual alerts results in lower subjective mental fatigue.

While the visual alerts as they were tested here are not yet usable for UAS operations, refinement and optimization of the protocol likely present an opportunity for interface improvement to support operators suffering from mental fatigue.

Chapter 12

Afterword

III.12.1 Some thoughts on the ethics of military UAS operations

Uncrewed Aerial system (UAS, Drones) warfare is controversial. Its opponents criticize the lack of oversight, the civilian harm and the imbalance in risk for operators and enemy combatants. Its proponents argue that UAS allow for reliable, cheap and safe operations that do not risk the lives of their operators [1].

This section will not try to find a definitive answer to the question of whether UAS are ethically acceptable or not. It should be seen as a collection of thoughts and arguments that may allow every reader to make up their own opinion more informedly.

The horror of war is evident. Even at a high cost in terms of freedom, political and economic power, diplomacy should always be the first, second, third and fourth option. Unfortunately, war is still a reality in our daily lives, as seen with the Russian attack on Ukraine in 2022 or the Israel-Hamas war following the attacks on October 7th 2023 [2].

There are many ways to decide what is and is not ethically and legally acceptable during an armed conflict. Teleology, for example, argues that the end justifies the means, while Kantian ethics argue that the motivation for action needs to be motivated by a sense of duty. Here, we will approach this problem from the perspective of the just war theory (Glazebrook, 2019; Moseley, 2024), as this theory applies specificily to armed conflict and separates several aspects of war. It also is the basis of international humanitarian law[3]. The theory focuses on three distinct parts: *jus Ad Bellum* ethics of starting a war, *jus In Bellum* ethics during the war and *jus Post Bellum* ethics after the war.

[1] https://www.defense.gov/News/News-Stories/Article/Article/2780257/dod-august-29-strike-in-kabul-tragic-mistake-kills-10-civilians/, https://theweek.com/news/defence/960548/the-ethics-of-military-drones

[2] https://www.bbc.com/news/war-in-ukraine, https://www.reuters.com/world/israel-hamas/

[3] https://www.icrc.org/en/war-and-law

We will focus mostly on *jus In bellum*, synonymous with International Humanitarian Law and the Genava Conventions of 1949,1977 & 2005. By doing this, we can avoid arguments such as *"The defending party has the right to use any system at their disposition"* (*jus Ad bellum*) and focus on UAS as a system only. According to the just war theory, action in war needs to meet two characteristics: *distinction* and *proportionality*. Distinction refers to the extent to which any action or weapon is likely to harm non-combatants[4]. Are UAS inherently more likely to harm civilians as compared to other weapons? The many news articles of civilian deaths to UAS drone strikes come to mind that might instinctively trigger a 'yes' response to this question[5]. Still, we need to consider further points before making that assessment. Except for the Russio-Ukranian war, UAS have been disproportionately used in asymmetrical conflicts, that is, conflicts in which the military capabilities of the belligerent powers are so different that they cannot attack each other in the same way (for example, the US Vietnam War, the US Afghanistan War or the US Iraq War)[6]. As combatants in these conflicts often wear no uniform, distinction is much more difficult and hence is likely to result in high civilian casualties. A manned air vehicle (fighter jet or helicopter) may have the same difficulty in determining the status of a person as an Uncrewed vehicle (Sexton, 2024; Shane, 2012). So higher casualty rates are not directly attributable to the use of UAS but rather to the asymmetrical nature of the conflict.

Furthermore, UAS missions often plan their strikes weeks in advance, verifying over a long period that the person they are observing is their target. This is partially attributable to the low operating costs of UAS in both financial and military terms. Manned Surveillance using scout soldiers, helicopters or other means is associated with the cost of increased risk to the operators. Furthermore UAS may execute surveillance missions not otherwise feasible with manned operations[7]. A study comparing the ratio of civilian casualties in Pakistan due to US drone strikes found that the percentage of civilian casualties was between 4 and 20%. Operations by the Pakistani military, not using UAS, against militant tribes resulted in 40% casualties (Plaw & Fricker, 2012).

The idea of a highly precise weapon executing targeted killings is horrifying, but discounting this as unethical neglects the fact that the set of alternatives to a drone strike is not limited to no action but may also mean a less precise strike [8].

We also need to consider damages other than casualties. UAS are difficult to detect and very silent. The psychological harm inflicted on a civilian population living in fear of being attacked 'out-of-the-blue' (in this case, literally the blue sky) at all times should not be understated and is unique to UAS[9] (Hijazi et al., 2019). To summarize, UAS are not an inherently less distinctive weapon system, and their low cost of operating may actually, through better surveillance, allow for better distinction. Nevertheless, UAS, by seeding fear, result in psychological harm to the population.

The second characteristic of just war is proportionality, an adequate response to the situation that minimizes the overall suffering. UAS again allow for very precise and targeted strikes

[4] https://legal.un.org/repertory/art51.shtml

[5] https://www.defense.gov/News/News-Stories/Article/Article/2780257/dod-august-29-strike-in-kabul-tragic-mistake-kills-10-civilians/

[6] https://www.britannica.com/topic/asymmetrical-warfare

[7] https://www.nytimes.com/2012/07/15/sunday-review/the-moral-case-for-drones.html

[8] For an introduction into the history of the ongoing struggle for precision within the US Air Force, please refer to Gladwell (2021)

[9] https://www.theatlantic.com/politics/archive/2013/10/saddest-words-congresss-briefing-drone-strikes/354548/

on military leaders, which, in theory, makes UAS an asset for a proportional response. Furthermore, the use of UAS has low political risk. If a manned aircraft is shot down and a pilot is killed, there is political pressure to retaliate. With UAS, we have seen that that is not necessarily the case as with the 2023 Black Seea incident, where a Russian fighter jet damaged an American UAS [10].

However, the lower risk of retaliation may have also caused the surge of targeted killings done by drones[11] (Carroll, 2012), as these assets can be used knowing that no one will counter attack.

So, have UAS resulted in fewer casualties due to their high precision, or have they allowed governments to have assassins at their hand with little risk of consequences? As with many things, the truth likely lies somewhere in between and largely depends on how a country or armed force employs these systems.

When discussing proportionality, one point that may be brought up is the fact that the UAS costs several millions of dollars and shoots missiles that cost more than 100.000 $ per piece for attacking combatants that have substantially fewer means at their disposition (Beno & Adamcik, 2014). While certainly a point of interest, this is much more of an argument of *jus Ad Bellum* ("Should a much stronger nation attack a weaker one?"), and it is not unique to UAS systems.

To summarize, UAS can be used with distinction and proportionality, minimizing civilian suffering. Unlike a weapon such as poisonous gas, landmines or expanding bullets where the argument could be made that any use would likely result in disproportionate suffering of both civilians and combatants, UAS may be used in a way that is by the *jus in Bellum* principles. However, much like any other weapon, they may be used unjustly.

Still, concluding that UAS are ethically acceptable weapons requires some further consideration.

One argument that has been repeatedly raised is that UAS operators face no risk to their lives while engaging in war. If that is an argument against using UAS, then a fighter jet dropping bombs on undefended targets should be held to the same standard, as their risk is very low. The same may go for the operators of long-range artillery or cruise missiles.

Research also shows that the operators of UAS suffer disproportionate psychological damage in the form of PTSD, risk of accidents and moral disengagement (Chappelle et al., 2010; Hijazi et al., 2019). This may be attributable to long working hours. In the case of US military personnel, the fact that the operators are stationed within the United States means that their military activity is much less removed from their everyday lives. However, nations and their operators choose to employ these systems. So, these actors (either the nations deciding to employ strike UAS or the soldiers enlisting, knowing there is a possibility to become a UAS operator) decide to accept the associated risks of using UAS.

For ethics in war, combatants need to be held accountable for their actions to limit an individual's power. This is likely an issue in UAS operations, as an operator of the MQ-9 Reaper in the USAF explained in an interview [12]. According to the USAF operator UAS teams often have the liberty to decide whether or not to strike a target depending on their risk assessment. This is not unlike other operations. Remarkably, the same USAF operators executing the strike are

[10] https://www.reuters.com/world/europe/us-says-reaper-drone-crashes-into-black-sea-after-russian-intercept-2023-03-14/

[11] https://www.newamerica.org/future-security/reports/americas-counterterrorism-wars/the-war-in-yemen/

[12] https://www.youtube.com/watch?v=26CgiDqWzd0

responsible for determining whether the target they striked was an actual combatant following a strike. This may somewhat incentivize a very 'liberal' attack policy for operators. In the case of a false positive, one can claim that the person killed was a combatant; this reduces the risk of false negatives, which would put friendly troops at risk.

Another critical aspect to consider is the increasing autonomy of UAS. Systems that have the authority to decide whether or not to strike a target independently are not a far-fetched thing of the future but have been extensively studied[13]. Giving a machine the capability to decide whether to kill a person is much more ethically questionable. Unfortunately, it would likely present a major strategic advantage to the military employing said weapon.

History also shows that there is a trend that weapons that have a strategic advantage will be used despite ethical objections. Carpet bombing, the practice of sending large bomber fleets over major industrial cities to cripple the enemies, was largely regarded as unethical before and during the start of the Second World War (Gladwell, 2021). Its strategic advantage quickly resulted in all major countries engaging in the practice. Even weapons that have been used at some point but are no longer used may fit that pattern. Poisonous gas, one of the horrors of World War I, was not used in the Second World War. A popular narrative is that poisonous gas has not been used since because it is an unethical weapon. While perhaps partially attributable to ethical reasons, the weapon had lost much of its strategic advantage during the Second World War. Large or incendiary bombs proved more effective and rendered poisonous gas obsolete in a battlefield no longer characterised by trenches but by mechanized infantry [14]. In addition, poisonous agents have been used since World War I. One example was the Iran-Iraq war, partially defined by trench warfare (like the First World War). Here, the Iraqi military used poisonous gas to break through the Iranian lines.

The same argument could be made for atomic weapons. These weapons of unimaginable destructive power may be most useful as a deterrent but not as an actual weapon [15]. Intuitively the recent announcements of Russia to start drills of their nonstrategic nuclear missile forces would seem like a counterexample to this argument. While certainly also an escalation, sabre-rattling is also part of the nuclear deterrent. If a nuclear-armed nation cannot project readiness to use its arsenal, then the deterrent is no longer effective.

These examples show that weapons considered unethical are likely to be used if they have a tactical or strategic advantage. The danger of autonomous UAS is that, unlike poisonous gas or atomic weapons, they might see extensive use due to a large strategic advantage.

There are likely many further points to consider, but to summarize: UAS as compared to fighter jets, helicopters etc. may induce disproportionately more psychological harm in civilians, their boundary of use is lower due to a low risk of retaliation, the policies put in place by armed forces may incentivize an aggressive use of the system, and they are at risk of being a gateway to being used as autonomous killing systems. On the other hand, they present a possibility of reducing civilian harm due to high precision and very targeted striking.

UAS are here to stay, so keeping an ongoing discourse on the topic is important. While advancements in science and technology may lead towards autonomous systems with a large ethical risk, work on human factors, HMI, and neuroergonomics may also improve the way UAS are used.

[13] https://www.newsweek.com/ai-drone-facial-recognition-ethic-issues-1784088
[14] https://www.sciencehistory.org/stories/magazine/a-brief-history-of-chemical-war/
[15] https://www.icj-cij.org/case/95

As this was only a very brief introduction on ethics in UAS warfare, there are several sources that we can recommend for diving further into the topic:

- On the ethics of war and the just war theory
  - https://legal.un.org/repertory/art51.shtml
  - https://www.icrc.org/en/war-and-law
  - https://iep.utm.edu/justwar/
  - https://www.icrc.org/en/war-and-law/treaties-customary-law/geneva-conventions
  - https://casebook.icrc.org/atoz/glossary/hague-conventions
- On the ethics of UAS as weapons
  - https://www.nytimes.com/2012/07/15/sunday-review/the-moral-case-for-drones.html
  - https://www.theatlantic.com/politics/archive/2013/10/saddest-words-congresss-briefing-drone-strikes/354548/
  - https://rabble.ca/human-rights/moral-buffer-death-missy-cummings-on-military-drones/
- Case studies of UAS warfare
  - https://www.youtube.com/watch?v=26CgiDqWzd0
  - https://www.defense.gov/News/News-Stories/Article/Article/2780257/dod-august-29-strike-in-kabul-tragic-mistake-kills-10-civilians/
- On Game Theory and US military doctrine
  - https://www.dancarlin.com/hardcore-history-59-the-destroyer-of-worlds/
  - https://www.lawfaremedia.org/article/pentagon-releases-first-ever-policy-on-civilian-harm-reduction
  - https://www.nytimes.com/2021/04/25/books/review/the-bomber-mafia-malcolm-gladwell.htm

Part IV
References

Adrian, E. D., & Matthews, B. H. (1934). The Berger rhythm: potential changes from the occipital lobes in man [Publisher: Oxford University Press]. *Brain*, *57*(4), 355–385 (cit. on p. 79).

Ahn, M., & Jun, S. C. (2015). Performance variation in motor imagery brain–computer interface: A brief review. *Journal of Neuroscience Methods*, *243*, 103–110. https://doi.org/10.1016/j.jneumeth.2015.01.033 (cit. on p. 116).

Ahn, S., Nguyen, T., Jang, H., Kim, J. G., & Jun, S. C. (2016). Exploring Neuro-Physiological Correlates of Drivers' Mental Fatigue Caused by Sleep Deprivation Using Simultaneous EEG, ECG, and fNIRS Data [Place: Lausanne, Switzerland Publisher: Frontiers Research Foundation Section: Original Research ARTICLE]. *Frontiers in Human Neuroscience*. https://doi.org/http://dx.doi.org/10.3389/fnhum.2016.00219 (cit. on pp. 113, 115)

Place: Lausanne, Switzerland Publisher: Frontiers Research Foundation Section: Original Research ARTICLE.

Air Force, S. C. (2024). Mishap Investigation Process. Retrieved March 21, 2024, from https://www.safety.af.mil/Home/Mishap-Investigation-Process/ (cit. on p. 25).

Airbus. (2008). A320 Flight Crew Operating Manual. (cit. on pp. 39, 44).

Åkerstedt, T., & Gillberg, M. (1990). Subjective and Objective Sleepiness in the Active Individual. *International Journal of Neuroscience*, *52*(1-2), 29–37. https://doi.org/10.3109/00207459008994241 (cit. on pp. 153, 161).

Alhazmi, S., Saini, M. K., & El Saddik, A. (2015). Multimedia Fatigue Detection for Adaptive Infotainment User Interface. *Proceedings of the 2nd Workshop on Computational Models of Social Interactions: Human-Computer-Media Communication*, 15–24. https://doi.org/10.1145/2810397.2810400 (cit. on pp. 44, 45, 110, 113, 207).

Allport, D., Styles, E., & Hsieh, S. (1994). Shifting intentional set: Exploring the dynamic control of tasks. *Attention and performance 15: Conscious and nonconscious information processing. Attention and performance series*, 421–452 (cit. on pp. 65, 66).

Amaro, E., & Barker, G. J. (2006). Study design in fMRI: Basic principles. *Brain and Cognition*, *60*(3), 220–232. https://doi.org/10.1016/j.bandc.2005.11.009 (cit. on p. 64).

André, N., Audiffren, M., & Baumeister, R. F. (2019). An Integrative Model of Effortful Control. *Frontiers in Systems Neuroscience*, *13*. Retrieved November 9, 2023, from https://www.frontiersin.org/articles/10.3389/fnsys.2019.00079 (cit. on p. 55).

Ansari, S., Naghdy, F., Du, H., & Pahnwar, Y. N. (2021). Driver Mental Fatigue Detection Based on Head Posture Using New Modified reLU-BiLSTM Deep Neural Network [Conference Name: IEEE Transactions on Intelligent Transportation Systems]. *IEEE Transactions on Intelligent Transportation Systems*, 1–13. https://doi.org/10.1109/TITS.2021.3098309 (cit. on pp. 113, 115)

Conference Name: IEEE Transactions on Intelligent Transportation Systems.

APA. (2018). APA Dictionary of Psychology. Retrieved November 23, 2023, from https://dictionary.apa.org/ (cit. on p. 51).

Arai, T. (1912). *Mental fatigue*. Teachers College, Columbia University. (Cit. on pp. 49, 63).

Aricò, P., Borghini, G., Di Flumeri, G., Colosimo, A., Bonelli, S., Golfetti, A., Pozzi, S., Imbert, J.-P., Granger, G., Benhacene, R., & Babiloni, F. (2016). Adaptive Automation Triggered by EEG-Based Mental Workload Index: A Passive Brain-Computer Interface Application in Realistic Air Traffic Control Environment. *Frontiers in Human Neuroscience*, *10*, 539. https://doi.org/10.3389/fnhum.2016.00539 (cit. on pp. 38, 42, 108, 292).

Aricò, P., Borghini, G., Di Flumeri, G., Sciaraffa, N., Colosimo, A., & Babiloni, F. (2017). Passive BCI in Operational Environments: Insights, Recent Advances, and Future Trends [Conference Name: IEEE Transactions on Biomedical Engineering]. *IEEE Transactions on Biomedical Engineering*, *64*(7), 1431–1436. https://doi.org/10.1109/TBME.2017.2694856 (cit. on pp. 108, 110)
Conference Name: IEEE Transactions on Biomedical Engineering.

Armour, C., & Ross, J. (2017). The Health and Well-Being of Military Drone Operators and Intelligence Analysts: A Systematic Review. *Military Psychology*, *29*(2), 83–98. https://doi.org/10.1037/mil0000149 (cit. on p. 26).

Arnould-Taylor, W. (1998). *A Textbook of Anatomy and Physiology*. Nelson Thornes. (Cit. on p. 72).

Arrabito, G. R., Ho, G., Lambert, A., Rutley, M., Keillor, J., Chiu, A., Au, H., & Hou, M. (2010). Human Factors Issues for Controlling Uninhabited Aerial Vehicles: 107 (cit. on pp. 24, 27).

Arrington, C. M., Altmann, E. M., & Carr, T. H. (2003). Tasks of a feather flock together: Similarity effects in task switching. *Memory & Cognition*, *31*(5), 781–789. https://doi.org/10.3758/BF03196116 (cit. on pp. 65, 68, 210).

Arrington, C. M., & Logan, G. D. (2004). The Cost of a Voluntary Task Switch. *Psychological Science*, *15*(9), 610–615. https://doi.org/10.1111/j.0956-7976.2004.00728.x (cit. on pp. 39, 43, 66–68, 126).

Arrington, C. M., & Yates, M. M. (2009). The role of attentional networks in voluntary task switching [Place: US Publisher: Psychonomic Society]. *Psychonomic Bulletin & Review*, *16*(4), 660–665. https://doi.org/10.3758/PBR.16.4.660 (cit. on p. 39)
Place: US Publisher: Psychonomic Society.

Austin. (2011, September). *Unmanned Aircraft Systems: UAVS Design, Development and Deployment* [Google-Books-ID: 03gdqhU61C0C]. John Wiley & Sons. (Cit. on pp. 13, 20, 23, 24, 27).

Austin, L. (2023). Civilian Harm Mitigation and Response Action Plan. (Cit. on p. 26).

Bank, T. W. (2022). *GDP* (tech. rep.). Retrieved November 23, 2023, from https://data.worldbank.org (cit. on p. 50).

Barachant, A., Bonnet, S., Congedo, M., & Jutten, C. (2010). Riemannian Geometry Applied to BCI Classification. In V. Vigneron, V. Zarzoso, E. Moreau, R. Gribonval, & E. Vincent (Eds.), *Latent Variable Analysis and Signal Separation* (pp. 629–636). Springer. https://doi.org/10.1007/978-3-642-15995-4_78 (cit. on pp. 114, 303).

Barachant, A., Bonnet, S., Congedo, M., & Jutten, C. (2012). Multiclass Brain–Computer Interface Classification by Riemannian Geometry [Conference Name: IEEE Transactions on Biomedical Engineering]. *IEEE Transactions on Biomedical Engineering*, *59*(4), 920–928. https://doi.org/10.1109/TBME.2011.2172210 (cit. on pp. 114, 303).

Barber, D., Reinerman-Jones, L., & Hidalgo, M. (2019). Optimizing Military Human-Robot Teaming: An Evaluation of Task Load and Modality Switch Cost [Publisher: SAGE Publications Inc]. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *63*(1), 1751–1755. https://doi.org/10.1177/1071181319631352 (cit. on p. 68)
Publisher: SAGE Publications Inc.

Baumeister, R. F. (2003). Ego Depletion and Self-Regulation Failure: A Resource Model of Self-Control. *Alcoholism: Clinical and Experimental Research*, *27*(2), 281–284. https://doi.org/10.1097/01.ALC.0000060879.61384.A4 (cit. on p. 52).

Bear, M. F., Connors, B. W., & Paradiso, M. A. (2007). *Neuroscience: exploring the brain* (3. ed). Lippincott Williams & Wilkins. (Cit. on p. 90).
Literaturangaben. - Literaturverz. S. [817] - 836.

Beatty, J., Lucero-Wagoner, B., Cacioppo, J., Tassinary, L., & Berntson, G. (2000). Handbook of psychophysiology [Publisher: Cambridge University Press Cambridge]. *The pupillary system.*, 142–162 (cit. on p. 93).

Beaudouin-Lafon, M., & Mackay, W. E. (2007). Prototyping Tools and Techniques [Num Pages: 24]. In *The Human-Computer Interaction Handbook* (2nd ed.). CRC Press. (Cit. on p. 204).

Belo, J., Clerc, M., & Schön, D. (2021). EEG-Based Auditory Attention Detection and Its Possible Future Applications for Passive BCI. *Frontiers in Computer Science*, *3*, 661178. https://doi.org/10.3389/fcomp.2021.661178 (cit. on p. 96).

Benaroch, C., Sadatnejad, K., Roc, A., Appriou, A., Monseigne, T., Pramij, S., Mladenovic, J., Pillette, L., Jeunet, C., & Lotte, F. (2021). Long-term BCI training of a tetraplegic user: Adaptive Riemannian classifiers and user training [Publisher: Frontiers Media SA]. *Frontiers in Human Neuroscience*, *15*, 635653. https://doi.org/https://doi.org/10.3389/fnhum.2021.635653 (cit. on p. 291).

Benchetrit, Y., Banville, H., & King, J.-R. (2023, October). Brain decoding: toward real-time reconstruction of visual perception [arXiv:2310.19812 [cs, eess, q-bio]]. https://doi.org/10.48550/arXiv.2310.19812 (cit. on p. 83).
Comment: 16 pages, 8 figures, preprint.

Bendak, S., & Rashid, H. S. (2020). Fatigue in aviation: A systematic review of the literature. *International Journal of Industrial Ergonomics*, *76*, 102928. https://doi.org/10.1016/j.ergon.2020.102928 (cit. on p. 61).

Beno, V., & Adamcik, F. (2014). Unmanned combat air vehicle: MQ-9 Reaper. *Scientific Research & Education in the Air Force-AFASES*, *1* (cit. on pp. 22, 24, 217).

Benton, D., & Parker, P. (1998). Breakfast, blood glucose, and cognition. *The American Journal of Clinical Nutrition*, *67*(4), 772S–778S. https://doi.org/10.1093/ajcn/67.4.772S (cit. on p. 53).

Benton, D., & Owens, D. S. (1993). Blood glucose and human memory. *Psychopharmacology*, *113*(1), 83–88. https://doi.org/10.1007/BF02244338 (cit. on p. 53).

Beppi, C., Ribeiro Violante, I., Scott, G., & Sandrone, S. (2021). EEG, MEG and neuromodulatory approaches to explore cognition: Current status and future directions. *Brain and Cognition*, *148*, 105677. https://doi.org/10.1016/j.bandc.2020.105677 (cit. on p. 79).

Berg, E. A. (1948). A Simple Objective Technique for Measuring Flexibility in Thinking [Publisher: Taylor & Francis Group]. *The Journal of General Psychology*. Retrieved January 23, 2024, from https://www.tandfonline.com/doi/abs/10.1080/00221309.1948.9918159 (cit. on p. 67).

Berger, H. (1929). Über das Elektrenkephalogramm des Menschen. *Archiv für Psychiatrie und Nervenkrankheiten*, *87*(1), 527–570. https://doi.org/10.1007/BF01797193 (cit. on pp. 79, 106).

Berka, C., Levendowski, D. J., Lumicao, M. N., Yau, A., Davis, G., Zivkovic, V. T., Olmstead, R. E., Tremoulet, P. D., & Craven, P. L. (2007). EEG Correlates of Task Engagement and Mental Workload in Vigilance, Learning, and Memory Tasks. *Aviation, space, and environmental medicine*, *78*(5), 14 (cit. on p. 317).

Bertrand, M., & Mullainathan, S. (2001). Do People Mean What They Say? Implications for Subjective Survey Data. *American Economic Review*, *91*(2), 67–72. https://doi.org/10.1257/aer.91.2.67 (cit. on p. 62).

Bigdely-Shamlo, N., Mullen, T., Kothe, C., Su, K.-M., & Robbins, K. A. (2015). The PREP pipeline: standardized preprocessing for large-scale EEG analysis. *Frontiers in Neuroinformatics*, *9*. https://doi.org/http://journal.frontiersin.org/Article/10.3389/fninf.2015.00016/abstract (cit. on p. 301).

Bigger, J. T., Fleiss, J. L., Steinman, R. C., Rolnitzky, L. M., Kleiger, R. E., & Rottman, J. N. (1992). Frequency domain measures of heart period variability and mortality after myocardial infarction. [Publisher: American Heart Association]. *Circulation*, *85*(1), 164–171. https://doi.org/10.1161/01.CIR.85.1.164 (cit. on p. 89).

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

Birbaumer, N., Ghanayim, N., Hinterberger, T., Iversen, I., Kotchoubey, B., Kübler, A., Perelmouter, J., Taub, E., & Flor, H. (1999). A spelling device for the paralysed. *Nature*, *398*(6725), 297–298. https://doi.org/10.1038/18581 (cit. on p. 106).

Blackmon, R. B., & Gramopadhye, A. K. (1995). Improving Construction Safety by Providing Positive Feedback on Backup Alarms. *Journal of Construction Engineering and Management*, *121*(2), 166–171. https://doi.org/10.1061/(ASCE)0733-9364(1995)121:2(166) (cit. on pp. 57, 60).

Blaha, L. M., Fisher, C. R., Walsh, M. M., Veksler, B. Z., & Gunzelmann, G. (2016). Real-Time Fatigue Monitoring with Computational Cognitive Models. In D. D. Schmorrow & C. M. Fidopiastis (Eds.), *Foundations of Augmented Cognition: Neuroergonomics and Operational Neuroscience* (pp. 299–310). Springer International Publishing. https://doi.org/10.1007/978-3-319-39955-3_28 (cit. on pp. 110, 207).

Blankertz, B., Dornhege, G., Krauledat, M., Müller, K.-R., & Curio, G. (2007). The non-invasive Berlin Brain–Computer Interface: Fast acquisition of effective performance in untrained subjects. *NeuroImage*, *37*(2), 539–550. https://doi.org/10.1016/j.neuroimage.2007.01.051 (cit. on p. 117).

Bliss, J. P., Gilson, R. D., & Deaton, J. E. (1995). Human probability matching behaviour in response to alarms of varying reliability. *Ergonomics*, *38*(11), 2300–2312. https://doi.org/10.1080/00140139508925269 (cit. on pp. 57, 60)
Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/00140139508925269.

Boardman, J. D., & Robert, S. A. (2000). Neighborhood Socioeconomic Status and Perceptions of Self-Efficacy [Publisher: SAGE Publications Inc]. *Sociological Perspectives*, *43*(1), 117–136. https://doi.org/10.2307/1389785 (cit. on p. 97).

Boksem, M. A., Meijman, T. F., & Lorist, M. M. (2005). Effects of mental fatigue on attention: An ERP study. *Cognitive Brain Research*, *25*(1), 107–116. https://doi.org/10.1016/j.cogbrainres.2005.04.011 (cit. on p. 182).

Boksem, M. A., Meijman, T. F., & Lorist, M. M. (2006). Mental fatigue, motivation and action monitoring. *Biological Psychology*, *72*(2), 123–132. https://doi.org/10.1016/j.biopsycho.2005.08.007 (cit. on pp. 52, 53).

Boksem, M. A., & Tops, M. (2008). Mental fatigue: Costs and benefits. *Brain Research Reviews*, *59*(1), 125–139. https://doi.org/10.1016/j.brainresrev.2008.07.001 (cit. on p. 53).

Bollimunta, A., Mo, J., Schroeder, C. E., & Ding, M. (2011). Neuronal mechanisms and attentional modulation of corticothalamic oscillations. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *31*(13), 4935–4943. https://doi.org/10.1523/JNEUROSCI.5580-10.2011 (cit. on p. 79).

Borghini, G., Astolfi, L., Vecchiato, G., Mattia, D., & Babiloni, F. (2014). Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. *Neuroscience & Biobehavioral Reviews*, *44*, 58–75. https://doi.org/10.1016/j.neubiorev.2012.10.003 (cit. on pp. 93, 112, 113, 115).

Borragán, G., Guerrero-Mosquera, C., Guillaume, C., Slama, H., & Peigneux, P. (2019). Decreased prefrontal connectivity parallels cognitive fatigue-related performance decline after sleep deprivation. An optical imaging study. *Biological Psychology*, *144*, 115–124. https://doi.org/10.1016/j.biopsycho.2019.03.004 (cit. on p. 52).

Bose, R., Wang, H., Dragomir, A., Thakor, N. V., Bezerianos, A., & Li, J. (2020). Regression-Based Continuous Driving Fatigue Estimation: Toward Practical Implementation [Conference Name: IEEE Transactions on Cognitive and Developmental Systems]. *IEEE Transactions on Cognitive and Developmental Systems*, *12*(2), 323–331. https://doi.org/10.1109/TCDS.2019.2929858 (cit. on pp. 112–114, 208).

Bougrain, L., Rimbert, S., Rodrigues, P. L. C., Canron, G., & Lotte, F. (2021). Guidelines to use Transfer Learning for Motor Imagery Detection: an experimental study [ISSN: 1948-3554]. *2021 10th International IEEE/EMBS Conference on Neural Engineering (NER)*, 5–8. https://doi.org/10.1109/NER49283.2021.9441254 (cit. on p. 114).

Bradski, G. (2000). The OpenCV Library. *Dr. Dobb's Journal of Software Tools* (cit. on p. 171).

Breznitz, S. (2013, May). *Cry Wolf: The Psychology of False Alarms*. Psychology Press. (Cit. on p. 38).
Google-Books-ID: lxwLVy16wqoC.

Britannica, T. E. o. E. (2023, November). Edward L. Thorndike | American Psychologist, Educational Theorist | Britannica. Retrieved November 22, 2023, from https://www.britannica.com/biography/Edward-L-Thorndike (cit. on p. 49).

Bronkhorst, A. (2000). The Cocktail Party Phenomenon: A Review of Research on Speech Intelligibility in Multiple-Talker Conditions. *Acta Acustica united with Acustica*, *86*, 117–128 (cit. on p. 57).

Brouwer, A.-M., Hogervorst, M. A., Van Erp, J. B. F., Heffelaar, T., Zimmerman, P. H., & Oostenveld, R. (2012). Estimating workload using EEG spectral power and ERPs in the n-back task. *Journal of Neural Engineering*, *9*(4), 045008. https://doi.org/10.1088/1741-2560/9/4/045008 (cit. on pp. 108, 291, 295, 317).

Brouwer, A.-M., Zander, T. O., van Erp, J. B. F., Korteling, J. E., & Bronkhorst, A. W. (2015). Using neurophysiological signals that reflect cognitive or affective state: six recommendations to avoid common pitfalls. *Frontiers in Neuroscience*, *9*. https://doi.org/https://www.frontiersin.org/article/10.3389/fnins.2015.00136 (cit. on pp. 110, 207).

Brown, I. (2000). Coping With Driver Fatigue: Is the Long Journey Nearly Over? In *Stress, workload, and fatigue* (pp. 596–606). CRC Press. (Cit. on pp. 52, 62).

Browne, D. (2016, March). *Adaptive User Interfaces*. Elsevier. (Cit. on p. 31).
Google-Books-ID: zMTSBQAAQBAJ.

Cabon, P., Deharvengt, S., Grau, J. Y., Maille, N., Berechet, I., & Mollard, R. (2012). Research and guidelines for implementing Fatigue Risk Management Systems for the French regional airlines. *Accident Analysis & Prevention*, *45*, 41–44. https://doi.org/10.1016/j.aap.2011.09.024 (cit. on pp. 61, 209).

Cabrera Castillos, K., Ladouce, S., Darmet, L., & Dehais, F. (2023). Burst c-VEP Based BCI: Optimizing stimulus design for enhanced classification with minimal calibration data and improved user experience. *NeuroImage*, *284*, 120446. https://doi.org/10.1016/j.neuroimage.2023.120446 (cit. on p. 106).

Cacioppo, J. T., & Tassinary, L. G. (1990). Inferring psychological significance from physiological signals [Place: US Publisher: American Psychological Association]. *American Psychologist*, *45*(1), 16–28. https://doi.org/10.1037/0003-066X.45.1.16 (cit. on p. 110).

Caid, S., Hauret, D., Wolff, M., & Mollard, R. (2016). Fatigue study and discourse analysis of french uninhabited aerial vehicle (UAV) operators to understand operational issues. *Proceedings of the 15th Ergo'IA "Ergonomie Et Informatique Avancé" Conference on - Ergo'IA '16*, 1–8. https://doi.org/10.1145/3050385.3050399 (cit. on pp. 26, 61).

Cakra, Y. E., & Distiawan Trisedya, B. (2015). Stock price prediction using linear regression based on sentiment analysis. *2015 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, 147–154. https://doi.org/10.1109/ICACSIS.2015.7415179 (cit. on p. 96).

Cam, L. M. L., & Neyman, J. (1967). *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* [Google-Books-ID: IC4Ku_7dBFUC]. University of California Press. (Cit. on p. 96).

Cannon, J., McCarthy, M. M., Lee, S., Lee, J., Börgers, C., Whittington, M. A., & Kopell, N. (2014). Neurosystems: brain rhythms and cognitive processing. *European Journal of Neuroscience*, *39*(5), 705–719. https://doi.org/10.1111/ejn.12453 (cit. on p. 79).

Cansiz, S. (2023). Guide To Covariance, Covariance Matrix & Eigenvalues | Built In. Retrieved April 6, 2024, from https://builtin.com/data-science/covariance-matrix (cit. on p. 114).

Cao, L., Li, J., Xu, Y., Zhu, H., & Jiang, C. (2016). A Hybrid Vigilance Monitoring Study for Mental Fatigue and Its Neural Activities. *Cognitive Computation*, *8*(2), 228–236. https://doi.org/10.1007/s12559-015-9351-y (cit. on pp. 112, 113).

Carlisle, R. (2022, February). *Afghanistan War, Third Edition*. Infobase Holdings, Inc. (Cit. on p. 25).

Carroll, R. (2012). The philosopher making the moral case for US drones. *The Guardian*. Retrieved April 25, 2024, from https://www.theguardian.com/world/2012/aug/02/philosopher-moral-case-drones (cit. on p. 217).

Cash, J. J. (2009). Alert fatigue. *American Journal of Health-System Pharmacy*, *66*(23), 2098–2101. https://doi.org/10.2146/ajhp090181 (cit. on pp. 57, 60).

Casso, M.-I., Jeunet, C., & Roy, R. N. (2021). Heading for motor imagery brain-computer interfaces (MI-BCIs) usable out-of-the-lab: Impact of dry electrode setup on classification accuracy [ISSN: 1948-3554]. *2021 10th International IEEE/EMBS Conference on Neural Engineering (NER)*, 690–693. https://doi.org/10.1109/NER49283.2021.9441410 (cit. on p. 157).

Cauchard, J. R., Khamis, M., Garcia, J., Kljun, M., & Brock, A. M. (2021). Toward a Roadmap for Human-Drone Interaction [Place: New York, NY, USA Publisher: Association for Computing Machinery]. *Interactions*, *28*(2), 76–81. https://doi.org/10.1145/3447889 (cit. on p. 202).

Cavanagh, J. F., & Frank, M. J. (2014). Frontal theta as a mechanism for cognitive control. *Trends in Cognitive Sciences*, *18*(8), 414–421. https://doi.org/10.1016/j.tics.2014.04.012 (cit. on p. 79).

Ceynar, M. L., Nkwonta, A. M., Cline, D. C. L., & Grahe, J. (2023, January). Studying Mental Fatigue: Dr. Tsuruko Arai Haraguchi Inspires from the Past. In *Early Psychological Research Contributions from Women of Color, Volume 1* (1st ed., pp. 12–23). Routledge. https://doi.org/10.4324/9781003246183-2 (cit. on p. 49).

Chadwick, R. A., Gillan, D. J., Simon, D., & Pazuchanics, S. (2004). Cognitive Analysis Methods for Control of Multiple Robots: Robotics on $5 a Day [Publisher: SAGE Publications

Inc]. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *48*(3), 688–692. https://doi.org/10.1177/154193120404800388 (cit. on p. 70).

Chang, C.-Y., Hsu, S.-H., Pion-Tonachini, L., & Jung, T.-P. (2018). Evaluation of Artifact Subspace Reconstruction for Automatic EEG Artifact Removal. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference*, *2018*, 1242–1245. https://doi.org/10.1109/EMBC.2018.8512547 (cit. on p. 82).

Chang-Hwan, C.-H. (Ed.). (2018). *Computational EEG Analysis: Methods and Applications*. Springer Singapore. https://doi.org/10.1007/978-981-13-0908-3 (cit. on pp. 75, 76, 83, 84).

Chappelle, W., McDonald, K., & King, R. E. (2010). Psychological attributes critical to the performance of MQ-1 Predator and MQ-9 Reaper US Air Force sensor operators. *Air Force Research Laboratory, Brooks City-Base, TX* (cit. on pp. 22, 23, 217).

Charbonnier, S., Roy, R. N., Bonnet, S., & Campagne, A. (2016). EEG index for control operators' mental fatigue monitoring using interactions between brain regions. *Expert Systems with Applications*, *52*, 91–98. https://doi.org/10.1016/j.eswa.2016.01.013 (cit. on pp. 113, 115).

Chavarriaga, R., & Millan, J. D. R. (2010). Learning From EEG Error-Related Potentials in Noninvasive Brain-Computer Interfaces. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *18*(4), 381–388. https://doi.org/10.1109/TNSRE.2010.2053387 (cit. on pp. 112, 114, 207, 318).

Chavarriaga, R., Sobolewski, A., & Millán, J. d. R. (2014). Errare machinale est: the use of error-related potentials in brain-machine interfaces. *Frontiers in Neuroscience*, *8*. https://doi.org/https://www.frontiersin.org/article/10.3389/fnins.2014.00208 (cit. on pp. 302, 318).

Chen, J., Barnes, M., & Harper-Sciarini, M. (2010). Supervisory Control of Unmanned Vehicles, 60 (cit. on pp. 23, 68).

Chen, J. Y. C., & Barnes, M. J. (2014). Human–Agent Teaming for Multirobot Control: A Review of Human Factors Issues. *IEEE Transactions on Human-Machine Systems*, *44*(1), 13–29. https://doi.org/10.1109/THMS.2013.2293535 (cit. on pp. 40, 46, 68).

Chen, Y., Wiesel, A., Eldar, Y. C., & Hero, A. O. (2010). Shrinkage Algorithms for MMSE Covariance Estimation [Conference Name: IEEE Transactions on Signal Processing]. *IEEE Transactions on Signal Processing*, *58*(10), 5016–5029. https://doi.org/10.1109/TSP.2010.2053029 (cit. on p. 157).

Chen, Y., Wu, J., & Li, F. (2022). Reconfiguration of response rule is more difficult than that of task goal: Behavior and electrophysiological evidence. *Neuroscience Letters*, *774*, 136517. https://doi.org/10.1016/j.neulet.2022.136517 (cit. on p. 68).

Cheng, M., Lu, Z., & Wang, H. (2017). Regularized common spatial patterns with subject-to-subject transfer of EEG signals. *Cognitive Neurodynamics*, *11*(2), 173–181. https://doi.org/10.1007/s11571-016-9417-x (cit. on p. 117).

Chenot, Q., Hinss, M. F., Roy, R. N., & Scannella, S. (2023). EEG-based Machine Learning to detect executive functions engagement during complex tasks. https://doi.org/https://enac.hal.science/hal-04638715 (cit. on p. 268).

Chew, J. Y., Kawamoto, M., Okuma, T., Yoshida, E., & Kato, N. (2021). Adaptive attention-based human machine interface system for teleoperation of industrial vehicle. *Scientific Reports*, *11*(1), 17284. https://doi.org/10.1038/s41598-021-96682-0 (cit. on pp. 38, 42).

Chikhi, S., Matton, N., & Blanchet, S. (2022). EEG power spectral measures of cognitive workload: A meta-analysis. *Psychophysiology*, *59*(6), e14009. https://doi.org/10.1111/psyp.

14009 (cit. on p. 84)
e14009 PsyP-2021-0459.R1.

Christie, S. T., & Schrater, P. (2015). Cognitive cost as dynamic allocation of energetic resources. *Frontiers in Neuroscience*, *9*. https://doi.org/10.3389/fnins.2015.00289 (cit. on pp. 37, 53, 54).

Clerc, M., Bougrain, L., & Lotte, F. (Eds.). (2016). *Brain-computer interfaces 2: technology and applications*. ISTE ; Wiley. (Cit. on pp. 106, 108).

Cobb, C. W., & Douglas, P. H. (1928). A Theory of Production [Publisher: American Economic Association]. *The American Economic Review*, *18*(1), 139–165. Retrieved June 20, 2024, from https://www.jstor.org/stable/1811556 (cit. on p. 36).

Combrisson, E., & Jerbi, K. (2015). Exceeding chance level by chance: The caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy. *Journal of Neuroscience Methods*, *250*, 126–136. https://doi.org/10.1016/j.jneumeth.2015.01.010 (cit. on p. 103).

Congedo, M., Barachant, A., & Bhatia, R. (2017). Riemannian geometry for EEG-based brain-computer interfaces; a primer and a review. *Brain-Computer Interfaces*, *4*(3), 155–174. https://doi.org/10.1080/2326263X.2017.1297192 (cit. on pp. 114, 207).

Congedo, M., & Lopes da Silva, F. (2018, January). Event-related potentials: General aspects of methodology and quantification. In D. L. Schomer & F. H. L. d. Silva (Eds.), *Niedermeyer's Electroencephalography: Basic Principles, Clinical Applications, and Related Fields, 7th edition*. Oxford University Press. Retrieved March 8, 2024, from https://hal.science/hal-01953600 (cit. on p. 83).

Cooke, N. (2006). *Human factors of remotely operated vehicles* (1st ed). Elsevier JAI. (Cit. on p. 26). Print version record.

Cooper, P. S., Garrett, P. M., Rennie, J. L., & Karayanidis, F. (2015). Task Uncertainty Can Account for Mixing and Switch Costs in Task-Switching [Publisher: Public Library of Science]. *PLOS ONE*, *10*(6), e0131556. https://doi.org/10.1371/journal.pone.0131556 (cit. on pp. 65, 126, 210)
Publisher: Public Library of Science.

Cooper, R. P. (2010). Cognitive Control: Componential or Emergent?: Topics in Cognitive Science(2010). *Topics in Cognitive Science*, *2*(4), 598–613. https://doi.org/10.1111/j.1756-8765.2010.01110.x (cit. on p. 66).

Craig, A., Tran, Y., Wijesuriya, N., & Nguyen, H. (2012). Regional brain wave activity changes associated with fatigue. *Psychophysiology*, *49*(4), 574–582. https://doi.org/https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1469-8986.2011.01329.x (cit. on p. 317).

Croft, R. J., & Barry, R. J. (2000). Removal of ocular artifact from the EEG: a review. *Neurophysiologie Clinique/Clinical Neurophysiology*, *30*(1), 5–19. https://doi.org/10.1016/S0987-7053(00)00055-1 (cit. on pp. 80, 82).

Csathó, Á., Van der Linden, D., & Matuz, A. (2024). Change in heart rate variability with increasing time-on-task as a marker for mental fatigue: A systematic review. *Biological Psychology*, *185*, 108727. https://doi.org/10.1016/j.biopsycho.2023.108727 (cit. on p. 115).

Csathó, Á., van der Linden, D., Hernádi, I., Buzás, P., & Kalmár, G. (2012). Effects of mental fatigue on the capacity limits of visual attention. *Journal of Cognitive Psychology*, *24*(5), 511–524. https://doi.org/10.1080/20445911.2012.658039 (cit. on pp. 55–58, 182).

Cuevas, H., Fiore, S., Caldwell, B., & Strater, L. (2007). Augmenting team cognition in human-automation teams performing in complex operational environments. *Aviation, Space, and Environmental Medicine*, *78*(5 Suppl), B63–70 (cit. on p. 23).

Cuevas, J. (2019, January). Neurotransmitters and Their Life Cycle. In *Reference Module in Biomedical Sciences*. Elsevier. https://doi.org/10.1016/B978-0-12-801238-3.11318-2 (cit. on p. 75).

Cullen, T. M. (2011). *The MQ-9 Reaper remotely piloted aircraft : humans and machines in action* [Thesis]. Massachusetts Institute of Technology [Accepted: 2013-08-22T19:01:38Z]. Retrieved March 22, 2024, from https://dspace.mit.edu/handle/1721.1/80249 (cit. on p. 22).

Danielmeier, C., & Ullsperger, M. (2011). Post-Error Adjustments. *Frontiers in Psychology*, *2*. Retrieved February 13, 2024, from https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2011.00233 (cit. on pp. 112, 281, 286).

Dawson, D., Searle, A. K., & Paterson, J. L. (2014). Look before you (s)leep: Evaluating the use of fatigue detection technologies within a fatigue risk management system for the road transport industry. *Sleep Medicine Reviews*, *18*(2), 141–152. https://doi.org/10.1016/j.smrv.2013.03.003 (cit. on p. 93).

De Jong, R. (1995). The Role of Preparation in Overlapping-task Performance. *The Quarterly Journal of Experimental Psychology Section A*, *48*(1), 2–25. https://doi.org/10.1080/14640749508401372 (cit. on p. 58).

de Carvalho, J., da Rocha, A., de Oliveira Nascimento, F., Neto, J., & Junqueira, L. (2002). Development of a Matlab software for analysis of heart rate variability. *6th International Conference on Signal Processing, 2002.*, *2*, 1488–1491 vol.2. https://doi.org/10.1109/ICOSP.2002.1180076 (cit. on p. 155).

de Cheveigné, A., & Arzounian, D. (2018). Robust detrending, rereferencing, outlier detection, and inpainting for multichannel data. *Neuroimage*, *172*, 903–912. https://doi.org/10.1016/j.neuroimage.2018.01.035 (cit. on p. 81).

de Cheveigné, A., & Nelken, I. (2019). Filters: When, Why, and How (Not) to Use Them. *Neuron*, *102*(2), 280–293. https://doi.org/10.1016/j.neuron.2019.02.039 (cit. on pp. 81, 82, 84).

Défossez, A., Caucheteux, C., Rapin, J., Kabeli, O., & King, J.-R. (2023). Decoding speech perception from non-invasive brain recordings [Number: 10 Publisher: Nature Publishing Group]. *Nature Machine Intelligence*, *5*(10), 1097–1107. https://doi.org/10.1038/s42256-023-00714-5 (cit. on pp. 83, 174).

Dehaene, S., & Changeux, J.-P. (2011). Experimental and Theoretical Approaches to Conscious Processing. *Neuron*, *70*(2), 200–227. https://doi.org/10.1016/j.neuron.2011.03.018 (cit. on p. 62).

Dehaene, S., & Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. *Cognition*, *79*(1), 1–37. https://doi.org/10.1016/S0010-0277(00)00123-2 (cit. on p. 54).

Dehais, F., Dupres, A., Di Flumeri, G., Verdiere, K., Borghini, G., Babiloni, F., & Roy, R. (2018). Monitoring pilot's cognitive fatigue with engagement features in simulated and actual flight conditions using an hybrid fNIRS-EEG passive BCI [ISSN: 2577-1655]. *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 544–549. https://doi.org/10.1109/SMC.2018.00102 (cit. on pp. 109, 113, 115)
ISSN: 2577-1655.

Dehais, F., Duprès, A., Blum, S., Drougard, N., Scannella, S., Roy, R. N., & Lotte, F. (2019). Monitoring Pilot's Mental Workload Using ERPs and Spectral Power with a Six-Dry-

Electrode EEG System in Real Flight Conditions [Number: 6 Publisher: Multidisciplinary Digital Publishing Institute]. *Sensors*, *19*(6), 1324. https://doi.org/10.3390/s19061324 (cit. on p. 83).

Dehais, F., Ladouce, S., Darmet, L., Nong, T.-V., Ferraro, G., Torre Tresols, J., Velut, S., & Labedan, P. (2022). Dual Passive Reactive Brain-Computer Interface: A Novel Approach to Human-Machine Symbiosis [Publisher: Frontiers]. *Frontiers in Neuroergonomics*, *3*. https://doi.org/10.3389/fnrgo.2022.824780 (cit. on pp. 108, 114).

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009 (cit. on p. 157).

DeVellis, R. F., & Thorpe, C. T. (2021, September). *Scale Development: Theory and Applications* [Google-Books-ID: QddDEAAAQBAJ]. SAGE Publications. (Cit. on p. 62).

de Visser, E. J., Beatty, P. J., Estepp, J. R., Kohn, S., Abubshait, A., Fedota, J. R., & McDonald, C. G. (2018). Learning From the Slips of Others: Neural Correlates of Trust in Automated Agents. *Frontiers in Human Neuroscience*, *12*, 309. https://doi.org/10.3389/fnhum.2018.00309 (cit. on p. 46).

Deza, A., Peters, J. R., Taylor, G. S., Surana, A., & Eckstein, M. P. (2017). Attention Allocation Aid for Visual Search. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 220–231. https://doi.org/10.1145/3025453.3025834 (cit. on pp. 38, 42).

Di Stasi, L. L., Catena, A., Cañas, J. J., Macknik, S. L., & Martinez-Conde, S. (2013). Saccadic velocity as an arousal index in naturalistic tasks. *Neuroscience & Biobehavioral Reviews*, *37*(5), 968–975. https://doi.org/10.1016/j.neubiorev.2013.03.011 (cit. on pp. 92, 93, 209).

Di Stasi, L. L., Diaz-Piedra, C., Suárez, J., McCamy, M. B., Martinez-Conde, S., Roca-Dorda, J., & Catena, A. (2015). Task complexity modulates pilot electroencephalographic activity during real flights. *Psychophysiology*, *52*(7), 951–956. https://doi.org/10.1111/psyp.12419 (cit. on pp. 113, 115).

Di Stasi, L. L., McCamy, M. B., Martinez-Conde, S., Gayles, E., Hoare, C., Foster, M., Catena, A., & Macknik, S. L. (2016). Effects of long and short simulated flights on the saccadic eye movement velocity of aviators. *Physiology & Behavior*, *153*, 91–96. https://doi.org/10.1016/j.physbeh.2015.10.024 (cit. on pp. 113, 115).

Di Stasi, L. L., Renner, R., Catena, A., Cañas, J. J., Velichkovsky, B. M., & Pannasch, S. (2012). Towards a driver fatigue test based on the saccadic main sequence: A partial validation by subjective report data. *Transportation Research Part C: Emerging Technologies*, *21*(1), 122–133. https://doi.org/10.1016/j.trc.2011.07.002 (cit. on pp. 113, 115, 281).

Diamond, A. (2013). Executive Functions [_eprint: https://doi.org/10.1146/annurev-psych-113011-143750]. *Annual Review of Psychology*, *64*(1), 135–168. https://doi.org/10.1146/annurev-psych-113011-143750 (cit. on p. 65).

Diarra, M., Marchitto, M., Bressolle, M.-C., Baccino, T., & Drai-Zerbib, V. (2023). A narrative review of the interconnection between pilot acute stress, startle, and surprise effects in the aviation context: Contribution of physiological measurements [Publisher: Frontiers]. *Frontiers in Neuroergonomics*, *4*. https://doi.org/10.3389/fnrgo.2023.1059476 (cit. on p. 40).

Dictionary, O. (2023). Fatigue. *Oxford Dictionary*. https://www.oxfordlearnersdictionaries.com/definition/american_english/fatigue (cit. on p. 51).

Dimitrakopoulos, G. N., Kakkos, I., Thakor, N. V., Bezerianos, A., & Sun, Y. (2017). A mental fatigue index based on regression using mulitband EEG features with application in simulated driving [ISSN: 1558-4615]. *2017 39th Annual International Conference of the*

IEEE Engineering in Medicine and Biology Society (EMBC), 3220–3223. https://doi.org/10.1109/EMBC.2017.8037542 (cit. on pp. 112, 113).

Domingos, P. (2012). A few useful things to know about machine learning. *Communications of the ACM*, *55*(10), 78–87. https://doi.org/10.1145/2347736.2347755 (cit. on pp. 95, 97, 103, 104).

Dorneich, M., Ververs, P., Mathan, S., & Whitlow, S. (2005). A Joint Human-Automation Cognitive System to Support Rapid Decision-Making in Hostile Environments, 2390–2395 Vol. 3. https://doi.org/10.1109/ICSMC.2005.1571506 (cit. on pp. 31–33).

Dorrian, J., Roach, G. D., Fletcher, A., & Dawson, D. (2007). Simulated train driving: Fatigue, self-awareness and cognitive disengagement. *Applied Ergonomics*, *38*(2), 155–166. https://doi.org/10.1016/j.apergo.2006.03.006 (cit. on pp. 55, 56, 60).

Duarte, J. M., & Berton, L. (2023). A review of semi-supervised learning for text classification. *Artificial Intelligence Review*, *56*(9), 9401–9469. https://doi.org/10.1007/s10462-023-10393-8 (cit. on p. 96).

Dubey, P. (1986). Inefficiency of Nash Equilibria [Publisher: INFORMS]. *Mathematics of Operations Research*. https://doi.org/10.1287/moor.11.1.1 (cit. on p. 37)
Publisher: INFORMS.

Durham, J. C. ((2020). Literature Review and Annotated Bibliography (1990 – 2019): Duty Time, Shift Work, and Operator Fatigue for Consideration of Unmanned Aircraft Systems in Air Carrier Operations, 70 (cit. on p. 26).

Dustman, R. E., Boswell, R. S., & Porter, P. B. (1962). Beta Brain Waves as an Index of Alertness. *Science*, *137*(3529), 533–534. https://doi.org/10.1126/science.137.3529.533 (cit. on p. 79).

Ellis, B. W., Johns, M. W., Lancaster, R., Raptopoulos, P., Angelopoulos, N., & Priest, R. G. (1981). The St. Mary's Hospital Sleep Questionnaire: A Study of Reliability. *Sleep*, *4*(1), 93–97. https://doi.org/10.1093/sleep/4.1.93 (cit. on pp. 130, 280).

Endsley, M. R., & Kaber, D. B. (1999). Level of automation effects on performance, situation awareness and workload in a dynamic control task. *Ergonomics*, *42*(3), 462–492. https://doi.org/10.1080/001401399185595 (cit. on p. 23).

Endsley, M. R. (1995). Toward a Theory of Situation Awareness in Dynamic Systems [Publisher: SAGE Publications Inc]. *Human Factors*, *37*(1), 32–64. https://doi.org/10.1518/001872095779049543 (cit. on p. 60)
Publisher: SAGE Publications Inc.

Endsley, M. R. (1996). Automation and Situation Awareness [Num Pages: 19]. In *Automation and Human Performance*. CRC Press. (Cit. on pp. 23, 40).

Engel, A. K., & Fries, P. (2010). Beta-band oscillations — signalling the status quo? *Current Opinion in Neurobiology*, *20*(2), 156–165. https://doi.org/10.1016/j.conb.2010.02.015 (cit. on p. 79).

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, *16*(1), 143–149. https://doi.org/10.3758/BF03203267 (cit. on pp. 58, 112, 291, 292, 296, 318).

Eurocontrol. (2012). Some Perspectives on Fatigue Risk Management Systems. https://skybrary.aero/sites/default/files/bookshelf/4595.pdf (cit. on p. 61).

Ewing, K. C., Fairclough, S. H., & Gilleade, K. (2016). Evaluation of an Adaptive Game that Uses EEG Measures Validated during the Design Process as Inputs to a Biocybernetic Loop [Publisher: Frontiers]. *Frontiers in Human Neuroscience*, *10*. https://doi.org/10.3389/fnhum.2016.00223 (cit. on p. 108).

FAA. (2012). *Advisory Circular* (tech. rep. No. 117-3). FAA. (Cit. on p. 51).

Fabiani, M., Buckley, J., Gratton, G., Coles, M. G. H., Donchin, E., & Logie, R. (1989). The training of complex task performance. *Acta Psychologica*, *71*(1), 259–299. https://doi.org/10.1016/0001-6918(89)90012-7 (cit. on p. 23).

Fairclough, S. H. (2009). Fundamentals of physiological computing [Conference Name: Interacting with Computers]. *Interacting with Computers*, *21*(1-2), 133–145. https://doi.org/10.1016/j.intcom.2008.10.011 (cit. on pp. 47, 105, 110).

Fairclough, S. H. (2017). Physiological Computing and Intelligent Adaptation. In *Emotions and Affect in Human Factors and Human-Computer Interaction* (pp. 539–556). Elsevier. https://doi.org/10.1016/B978-0-12-801851-4.00020-3 (cit. on pp. 14, 105).

Fairclough, S. H., & Lotte, F. (2020). Grand challenges in neurotechnology and system neuroergonomics [Publisher: Frontiers]. *Frontiers in Neuroergonomics*, 2. https://doi.org/https://doi.org/10.3389/fnrgo.2020.602504 (cit. on p. 291).

Fan, X., Zhou, Q., Liu, Z., & Xie, F. (2015). Electroencephalogram assessment of mental fatigue in visual search (F. Liu, D.-H. Lee, R. Lagoa, & S. Kumar, Eds.) [Publisher: IOS Press]. *Bio-Medical Materials and Engineering*, *26*(s1), S1455–S1463. https://doi.org/10.3233/BME-151444 (cit. on pp. 112–115, 152).

Fananas, M. A., & Soria, X. (2024). *Online mental fatigue estimation for long endurance Unmanned Aerial Systems (UAS) operators*. (Cit. on p. 102).

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses [Publisher: Springer]. *Behavior research methods*, *41*(4), 1149–1160 (cit. on p. 126).

Feigh, K. M., Dorneich, M. C., & Hayes, C. C. (2012). Toward a Characterization of Adaptive Systems: A Framework for Researchers and System Designers. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *54*(6), 1008–1024. https://doi.org/10.1177/0018720812443983 (cit. on pp. 32, 33).

Fiene, M., Rufener, K. S., Kuehne, M., Matzke, M., Heinze, H.-J., & Zaehle, T. (2018). Electrophysiological and behavioral effects of frontal transcranial direct current stimulation on cognitive fatigue in multiple sclerosis. *Journal of Neurology*, *265*(3), 607–617. https://doi.org/10.1007/s00415-018-8754-6 (cit. on p. 208).

Figalová, N., Chuang, L. L., Pichen, J., Baumann, M., & Pollatos, O. (2022). Ambient Light Conveying Reliability Improves Drivers' Takeover Performance without Increasing Mental Workload. *Multimodal Technologies and Interaction*, *6*(9), 73. https://doi.org/10.3390/mti6090073 (cit. on p. 37).

Findlater, L., & McGrenere, J. (2004). A comparison of static, adaptive, and adaptable menus. *Proceedings of the 2004 conference on Human factors in computing systems - CHI '04*, 89–96. https://doi.org/10.1145/985692.985704 (cit. on p. 31).

Fine, E. J., Ionita, C. C., & Lohr, L. (2002). The History of the Development of the Cerebellar Examination. *Seminars in Neurology*, *22*(4), 375–384. https://doi.org/10.1055/s-2002-36759 (cit. on p. 72).

Frey, J., Mühl, C., Lotte, F., & Hachet, M. (2014). Review of the Use of Electroencephalography as an Evaluation Method for Human-Computer Interaction (cit. on p. 108).

Fries, P., Nikolić, D., & Singer, W. (2007). The gamma cycle. *Trends in Neurosciences*, *30*(7), 309–316. https://doi.org/10.1016/j.tins.2007.05.005 (cit. on p. 79).

Furogori, S., Yoshizawa, N., Iname, C., & Mirua, Y. (2005). Estimation of Driver Fatigue by Pressure Distribution on Seat in Long Term Driving. *Review of automotive engineering*, *26*(1), 053–058. Retrieved February 5, 2022, from https://www.safetylit.org/citations/

index.php?fuseaction=citations.viewdetails&citationIds[]=citjournalarticle_185440_5 (cit. on pp. 113, 115).

Gabbard, J. (2023). How Air Force drone pilots train to fly the $32 million MQ-9 Reaper. Retrieved March 19, 2024, from https://www.businessinsider.com/how-drone-pilots-train-to-fly-the-mq-9-reaper-2023-8 (cit. on p. 23).

Gajos, K. Z., Czerwinski, M., Tan, D. S., & Weld, D. S. (2006). Exploring the design space for adaptive graphical user interfaces. *Proceedings of the working conference on Advanced visual interfaces*, 201–208. https://doi.org/10.1145/1133265.1133306 (cit. on p. 47).

Gander, P. H., Mangie, J., van den Berg, M. J., Smith, A. A. T., Mulrine, H. M., & Signal, T. L. (2014). Crew Fatigue Safety Performance Indicators for Fatigue Risk Management Systems. *Aviation, Space, and Environmental Medicine*, *85*(2). Retrieved April 26, 2022, from https://trid.trb.org/view/1290640 (cit. on p. 277).

Gardner, A. B., Worrell, G. A., Marsh, E., Dlugos, D., & Litt, B. (2007). Human and automated detection of high-frequency oscillations in clinical intracranial EEG recordings. *Clinical Neurophysiology*, *118*(5), 1134–1143. https://doi.org/10.1016/j.clinph.2006.12.019 (cit. on p. 76).

Geanakoplos, J. (1989). Arrow-Debreu Model of General Equilibrium. In J. Eatwell, M. Milgate, & P. Newman (Eds.), *General Equilibrium* (pp. 43–61). Palgrave Macmillan UK. https://doi.org/10.1007/978-1-349-19802-3_3 (cit. on p. 35).

Gergelyfi, M., Sanz-Arigita, E. J., Solopchuk, O., Dricot, L., Jacob, B., & Zénon, A. (2021). Mental fatigue correlates with depression of task-related network and augmented DMN activity but spares the reward circuit. *NeuroImage*, *243*, 118532. https://doi.org/10.1016/j.neuroimage.2021.118532 (cit. on p. 53).

Geurts, H. M., Corbett, B., & Solomon, M. (2009). The paradox of cognitive flexibility in autism. *Trends in cognitive sciences*, *13*(2), 74–82. https://doi.org/10.1016/j.tics.2008.11.006 (cit. on p. 67).

Ghanbary, A., Ashnagar, M., Habibi, E., & Sadeghi, S. (2016). Evaluation of Rating Scale Mental Effort (RSME) effectiveness for mental workload assessment in nurses [Publisher: Journal of Occupational Health and Epidemiology]. *Journal of Occupational Health and Epidemiology*, *5*(4), 211–217. https://doi.org/10.18869/acadpub.johe.5.4.211 (cit. on pp. 173, 279).

Ghanbary Sartang, A., Ashnagar, M., Habibi, E., & Sadeghi, S. (2016). Evaluation of Rating Scale Mental Effort (RSME) effectiveness for mental workload assessment in nurses. *Journal of Occupational Health and Epidemiology*, *5*(4), 211–217. https://doi.org/10.18869/acadpub.johe.5.4.211 (cit. on pp. 292, 317)
Publisher: Journal of Occupational Health and Epidemiology.

Ghimire, N. P. (2023). Final Report on the accident of ATR72/212A. https://www.tourism.gov.np/files/1/9N-ANC%20FINAL%20Report.pdf (cit. on p. 27).

Giese, S., Carr, D., & Chahl, J. (2013). Implications for unmanned systems research of military UAV mishap statistics. *2013 IEEE Intelligent Vehicles Symposium (IV)*, 1191–1196. https://doi.org/10.1109/IVS.2013.6629628 (cit. on p. 26).

Gladwell, M. (2021, April). *The Bomber Mafia: A Dream, a Temptation, and the Longest Night of the Second World War*. Little, Brown; Company. (Cit. on pp. 216, 218).

Glazebrook, T. (2019, November). Are drone strikes ever ethical? Retrieved April 25, 2024, from https://www.defense.gov/News/News-Stories/Article/Article/2780257/dod-august-29-strike-in-kabul-tragic-mistake-kills-10-civilians/ (cit. on p. 215).

Goldberger, A. L., & Gold-berger, E. (2013). Clinical electrocardiography, a simplified approach [Publisher: LWW]. *Critical Care Medicine*, *9*(12), 891–892 (cit. on pp. 86, 87).

Goncharova, I., McFarland, D., Vaughan, T., & Wolpaw, J. (2003). EMG contamination of EEG: spectral and topographical characteristics. *Clinical Neurophysiology*, *114*(9), 1580–1593. https://doi.org/10.1016/S1388-2457(03)00093-2 (cit. on pp. 77, 80).

Goschke, T. (2000). Intentional reconfiguration and involuntary persistence in task-set switching. *Control of Cognitive Processes: Attention and Performance XVIII, 331-355 (2000)*, *18* (cit. on pp. 66, 68, 69).

Goyal, A., Miller, J., Qasim, S. E., Watrous, A. J., Zhang, H., Stein, J. M., Inman, C. S., Gross, R. E., Willie, J. T., Lega, B., Lin, J.-J., Sharan, A., Wu, C., Sperling, M. R., Sheth, S. A., McKhann, G. M., Smith, E. H., Schevon, C., & Jacobs, J. (2020). Functionally distinct high and low theta oscillations in the human hippocampus. *Nature Communications*, *11*(1), 2469. https://doi.org/10.1038/s41467-020-15670-6 (cit. on p. 79).

Graefe, J., Engelhardt, D., & Bengler, K. (2021). What does Well-Designed Adaptivity Mean for Drivers? A Research Approach to Develop Recommendations for Adaptive In-Vehicle User Interfaces that are Understandable, Transparent and Controllable. *13th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 43–46. https://doi.org/10.1145/3473682.3480261 (cit. on pp. 45, 46).

Grosse-Wentrup, M., & Schölkopf, B. (2013). A review of performance variations in SMR-based Brain- Computer interfaces (BCIs) [Publisher: Springer]. *Brain-Computer Interface Research*, 39–51. https://doi.org/10.1007/978-3-642-36083-1_5 (cit. on p. 291).

Guo, Z., Chen, R., Zhang, K., Pan, Y., & Wu, J. (2016). The Impairing Effect of Mental Fatigue on Visual Sustained Attention under Monotonous Multi-Object Visual Attention Task in Long Durations: An Event-Related Potential Based Study [Publisher: Public Library of Science]. *PLOS ONE*, *11*(9), e0163360. https://doi.org/10.1371/journal.pone.0163360 (cit. on pp. 55–58)
Publisher: Public Library of Science.

Gupta, S. G., Ghonge, D. M., & Jawandhiya, P. M. (2013). Review of Unmanned Aircraft System (UAS). https://doi.org/10.2139/ssrn.3451039 (cit. on p. 24).

Haans, A. (2019). Contrast Analysis: A Tutorial. *Practical Assessment, Research, and Evaluation*, *23*(1). https://doi.org/https://doi.org/10.7275/7dey-zd62 (cit. on p. 300).

Haider, A. (2021, January). *Introduction - Joint Air Power Competence Centre*. Retrieved March 13, 2024, from https://www.japcc.org/chapters/c-uas-introduction/ (cit. on pp. 13, 20).

Haider, A. (2022, August). It's a Question of Gender! - Joint Air Power Competence Centre. Retrieved March 14, 2024, from https://www.japcc.org/online-feature/its-a-question-of-gender/ (cit. on p. 20).

Hajinoroozi, M., Zhang, J., & Huang, Y. (2017). Prediction of fatigue-related driver performance from EEG data by deep Riemannian model [ISSN: 1558-4615]. *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 4167–4170. https://doi.org/10.1109/EMBC.2017.8037774 (cit. on pp. 113, 114).

Hamann, A., & Carstengerdes, N. (2023). Assessing the development of mental fatigue during simulated flights with concurrent EEG-fNIRS measurement [Number: 1 Publisher: Nature Publishing Group]. *Scientific Reports*, *13*(1), 4738. https://doi.org/10.1038/s41598-023-31264-w (cit. on pp. 63, 110, 136).

Han, J., Kamber, M., & Pei, J. (2012). *Data mining: concepts and techniques* (3rd ed). Elsevier/Morgan Kaufmann. (Cit. on p. 95).

Harbour, S., Christensen, J., Estepp, J., & Gray, T. (2013). An Integrated Neuroergonomic Assessment of In-Flight Pilot Workload. *17th International Symposium on Aviation Psychology*, 621–626. https://corescholar.libraries.wright.edu/isap_2013/11 (cit. on p. 301).

Hardy, L. H., Rand, G., & Rittler, M. C. (1945). Tests for the Detection and Analysis of Color-Blindness. I. The Ishihara Test: An Evaluation [Publisher: Optica Publishing Group]. *JOSA*, *35*(4), 268–275. https://doi.org/10.1364/JOSA.35.000268 (cit. on pp. 128, 130, 142, 275)
Publisher: Optica Publishing Group.

Hari, R., & Puce, A. (2023). *MEG-EEG Primer* [Google-Books-ID: wRrOEAAAQBAJ]. Oxford University Press. (Cit. on pp. 76, 79).

Harmony, T. (2013). The functional significance of delta oscillations in cognitive processing. *Frontiers in Integrative Neuroscience*, *7*, 83. https://doi.org/10.3389/fnint.2013.00083 (cit. on p. 79).

Hart, S. G. (1986, January). NASA Task Load Index (TLX) [NTRS Author Affiliations: NASA Ames Research Center NTRS Document ID: 20000021488 NTRS Research Center: Ames Research Center (ARC)]. Retrieved June 6, 2024, from https://ntrs.nasa.gov/citations/20000021488 (cit. on p. 280).

Hassib, M., Braun, M., Pfleging, B., & Alt, F. (2019). Detecting and Influencing Driver Emotions Using Psycho-Physiological Sensors and Ambient Light [Series Title: Lecture Notes in Computer Science]. In D. Lamas, F. Loizides, L. Nacke, H. Petrie, M. Winckler, & P. Zaphiris (Eds.), *Human-Computer Interaction – INTERACT 2019* (pp. 721–742, Vol. 11746). Springer International Publishing. https://doi.org/10.1007/978-3-030-29381-9_43 (cit. on p. 37).

He, J. (2013). Fatigue Detection using Smartphones. *Journal of Ergonomics*, *03*(03). https://doi.org/10.4172/2165-7556.1000120 (cit. on pp. 113, 115).

Heitz, R. P. (2014). The speed-accuracy tradeoff: history, physiology, methodology, and behavior. *Frontiers in Neuroscience*, *8*. Retrieved February 20, 2024, from https://www.frontiersin.org/journals/neuroscience/articles/10.3389/fnins.2014.00150 (cit. on pp. 110, 111).

Helton, W. S., & Wen, J. (2023). Will the real resource theory please stand up! Vigilance is a renewable resource and should be modeled as such. *Experimental Brain Research*, *241*(5), 1263–1270. https://doi.org/10.1007/s00221-023-06604-x (cit. on p. 52).

Hewett, T. T., Baecker, R., Card, S., Carey, T., Gasen, J., Mantei, M., Perlman, G., Strong, G., & Verplank, W. (1992). *ACM SIGCHI Curricula for Human-Computer Interaction* (Technical Report) (Num Pages: 173 ISBN-10: 0897914740). Association for Computing Machinery. New York, NY, USA. (Cit. on p. 31).

Hijazi, A., Ferguson, C. J., Richard Ferraro, F., Hall, H., Hovee, M., & Wilcox, S. (2019). Psychological Dimensions of Drone Warfare. *Current Psychology*, *38*(5), 1285–1296. https://doi.org/10.1007/s12144-017-9684-7 (cit. on pp. 216, 217).

Hinss, M., Jahanpour, E., Brock, A., Imbert, J.-P., & Roy, R. (2022). Long-endurance UAS operations translated to neuroergonomics. *ICCAS22-International Conference on Cognitive Aircraft Systems*. https://doi.org/hal-04149728 (cit. on p. 268).

Hinss, M. F., Brock, A. M., & Roy, R. N. (2022a). Cognitive effects of prolonged continuous human-machine interaction: The case for mental state-based adaptive interfaces. *Frontiers in Neuroergonomics*, *3*. https://doi.org/10.3389/fnrgo.2022.935092 (cit. on pp. 13, 32, 42, 43, 267).

Hinss, M. F., Brock, A. M., & Roy, R. N. (2023a). The double task-switching protocol: An investigation into the effects of similarity and conflict on cognitive flexibility in the context of mental fatigue (G. G. De La Torre, Ed.). *PLOS ONE*, *18*(2), e0279021. https://doi.org/10.1371/journal.pone.0279021 (cit. on pp. 137, 171, 173, 267).

Hinss, M. F., Brock, A. M., & Roy, R. N. (2022b). Is it time for passive Brain Computer Interfaces in UAV Operations. https://doi.org/https://enac.hal.science/hal-04638704 (cit. on p. 268).

Hinss, M. F., Brock, A. M., & Roy, R. N. (2024). Double Task Switching: an investigation into the effects of similarity and task-rule congruency on cognitive flexibility. *PLOS ONE* (cit. on pp. 137, 267).

Hinss, M. F., Darmet, L., Somon, B., Jahanpour, E., Lotte, F., Ladouce, S., & Roy, R. N. (2021a, July). An EEG dataset for cross-session mental workload estimation: Passive BCI competition of the Neuroergonomics Conference 2021. https://doi.org/10.5281/zenodo.5055046 (cit. on p. 291).
The project was validated by the local ethical committee of the University of Toulouse (CER number 2021-342).

Hinss, M. F., Darmet, L., Somon, B., Jahanpour, E., Lotte, F., Ladouce, S., & Roy, R. N. (2021b). An EEG dataset for cross-session mental workload estimation: Passive BCI competition of the Neuroergonomics Conference 2021 (Version 2)[Data set]. *Neuroergonomics Conference, Munich, Germany. Zenodo*. https://doi.org/10.5281/zenodo.5055046 (cit. on p. 268).

Hinss, M. F., Jahanpour, E. S., Brock, A. M., & Roy, R. N. (2024). A passive Brain-Computer Interface for operator mental fatigue estimation in monotonous surveillance operations: time-on-task and performance labeling issues. *Journal of Neural Engineering* (cit. on pp. 167, 267).

Hinss, M. F., Jahanpour, E. S., Somon, B., Pluchon, L., Dehais, F., & Roy, R. N. (2022a, July). COG-BCI database: A multi-session and multi-task EEG cognitive dataset for passive brain-computer interfaces. https://doi.org/10.5281/zenodo.6923638 (cit. on pp. 116, 268, 291, 292).
Type: dataset.

Hinss, M. F., Jahanpour, E. S., Somon, B., Pluchon, L., Dehais, F., & Roy, R. N. (2022b, July). COG-BCI database: A multi-session and multi-task EEG cognitive dataset for passive brain-computer interfaces. https://doi.org/10.5281/ZENODO.6874129 (cit. on p. 318).
Other This database has been created as part of a project validated by the local ethical committee of the University of Toulouse (CER number 2021-342).

Hinss, M. F., Jahanpour, E. S., Somon, B., Pluchon, L., Dehais, F., & Roy, R. N. (2023). Open multi-session and multi-task EEG cognitive Dataset for passive brain-computer Interface Applications. *Scientific Data*, *10*(1), 85. https://doi.org/10.1038/s41597-022-01898-y (cit. on pp. 213, 268).

Hinss, M. F., Somon, B., Dehais, F., & Roy, R. N. (2021). Open EEG Datasets for Passive Brain-Computer Interface Applications: Lacks and Perspectives. *10th International IEEE/EMBS Conference on Neural Engineering*, 686–689. https://doi.org/10.1109/NER49283.2021.9441214 (cit. on pp. 213, 269, 291).

Hinss, M. F., Vitale, V. M., Brock, A. M., & Roy, R. N. (2024). EEG-based performance estimation during a realistic drone piloting task. https://doi.org/hal-04596992v1 (cit. on pp. 182, 268).

Hinss, M. F., Vitale, V. M., Phan, N. T., Roy, R. N., & Brock, A. M. (2024). UASOS: An Experimental Environment For Assessing Mental Fatigue & Cognitive Flexibility During Drone Operations. https://doi.org/https://doi.org/10.1145/3610977.3637484 (cit. on pp. 182, 268).

Hinss, M. F., Brock, A., & Roy, R. (2023b). Etude et caractérisation de la fatigue mentale et de la flexibilité cognitive dans le contexte des opérations drone longue endurance, 100. https://doi.org/hal-04362545v1 (cit. on p. 267).

Hinss, M. F., Jahanpour, E. S., Brock, A. M., & Roy, R. N. (2023). Labeling mental fatigue for passive BCI applications: Accuracy vs applicability tradeoff. *BCI meeting*. https://doi.org/hal-04146193 (cit. on pp. 167, 268).

Hoang, M.-T. O., Grøntved, K. A. R., van Berkel, N., Skov, M. B., Christensen, A. L., & Merritt, T. (2023). Drone Swarms to Support Search and Rescue Operations: Opportunities and Challenges. In B. J. Dunstan, J. T. K. V. Koh, D. Turnbull Tillman, & S. A. Brown (Eds.), *Cultural Robotics: Social Robots and Their Emergent Cultural Ecologies* (pp. 163–176). Springer International Publishing. https://doi.org/10.1007/978-3-031-28138-9_11 (cit. on p. 25).

Hoang, M.-T. O., van Berkel, N., Skov, M. B., & Merritt, T. R. (2023). Challenges and Requirements in Multi-Drone Interfaces. *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–9. https://doi.org/10.1145/3544549.3585673 (cit. on p. 27).

Hockey, G. R. J. (2011). A motivational control theory of cognitive fatigue. In *Cognitive fatigue: Multidisciplinary perspectives on current research and future applications* (pp. 167–187). American Psychological Association. https://doi.org/10.1037/12343-008 (cit. on pp. 52, 53).

Hockey, R. (2013). *The Psychology of Fatigue: Work, Effort and Control*. Cambridge University Press. https://doi.org/10.1017/CBO9781139015394 (cit. on p. 62).

Hodgkin, A. L., & Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of Physiology*, *117*(4), 500–544. Retrieved February 2, 2024, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1392413/ (cit. on p. 73).

Hogervorst, M. A., Brouwer, A.-M., & van Erp, J. B. F. (2014). Combining and comparing EEG, peripheral physiology and eye-related measures for the assessment of mental workload [Publisher: Frontiers]. *Frontiers in Neuroscience*, *8*. https://doi.org/10.3389/fnins.2014.00322 (cit. on p. 108).

Holgado, A. J. N., Terry, J. R., & Bogacz, R. (2010). Conditions for the Generation of Beta Oscillations in the Subthalamic Nucleus–Globus Pallidus Network [Publisher: Society for Neuroscience Section: Articles]. *Journal of Neuroscience*, *30*(37), 12340–12352. https://doi.org/10.1523/JNEUROSCI.0817-10.2010 (cit. on p. 79).

Holtzer, R., Shuman, M., Mahoney, J. R., Lipton, R., & Verghese, J. (2010). Cognitive Fatigue Defined in the Context of Attention Networks. *Aging, Neuropsychology, and Cognition*, *18*(1), 108–128. https://doi.org/10.1080/13825585.2010.517826 (cit. on pp. 57, 58) Publisher: Routledge _eprint: https://doi.org/10.1080/13825585.2010.517826.

Homölle, S., & Oostenveld, R. (2019). Using a structured-light 3D scanner to improve EEG source modeling with more accurate electrode positions. *Journal of Neuroscience Methods*, *326*, 108378. https://doi.org/https://www.sciencedirect.com/science/article/pii/S0165027019302353 (cit. on p. 298).

Höök, K. (2000). Steps to take before intelligent user interfaces become real. *Interacting with Computers*, *12*(4), 409–426. https://doi.org/10.1016/S0953-5438(99)00006-5 (cit. on p. 48)
Conference Name: Interacting with Computers.

Hopstaken, J. F., van der Linden, D., Bakker, A. B., & Kompier, M. A. J. (2015a). A multifaceted investigation of the link between mental fatigue and task disengagement: Mental fatigue and task disengagement. *Psychophysiology*, *52*(3), 305–315. https://doi.org/10.1111/psyp.12339 (cit. on pp. 37, 211).

Hopstaken, J. F., van der Linden, D., Bakker, A. B., & Kompier, M. A. J. (2015b). The window of my eyes: Task disengagement and mental fatigue covary with pupil dynamics. *Biological Psychology*, *110*, 100–106. https://doi.org/10.1016/j.biopsycho.2015.06.013 (cit. on pp. 52–54, 114).

Hopstaken, J. F., van der Linden, D., Bakker, A. B., Kompier, M. A. J., & Leung, Y. K. (2016). Shifts in attention during mental fatigue: Evidence from subjective, behavioral, physiological, and eye-tracking data. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(6), 878–889. https://doi.org/10.1037/xhp0000189 (cit. on pp. 52, 53, 93, 113–115).

Hou, M., Kobierski, R. D., & Brown, M. (2007). Intelligent Adaptive Interfaces for the Control of Multiple UAVs. *Journal of Cognitive Engineering and Decision Making*, *1*(3), 327–362. https://doi.org/10.1518/155534307X255654 (cit. on pp. 46, 48).

Hu, J., & Min, J. (2018). Automated detection of driver fatigue based on EEG signals using gradient boosting decision tree model. *Cognitive Neurodynamics*, *12*(4), 431–440. https://doi.org/10.1007/s11571-018-9485-1 (cit. on p. 159).

Huang, K.-C., Huang, T.-Y., Chuang, C.-H., King, J.-T., Wang, Y.-K., Lin, C.-T., & Jung, T.-P. (2016). An EEG-Based Fatigue Detection and Mitigation System. *International Journal of Neural Systems*, *26*(04), 1650018. https://doi.org/10.1142/S0129065716500180 (cit. on pp. 112, 113).

Huang, S., Li, J., Zhang, P., & Zhang, W. (2018). Detection of mental fatigue state with wearable ECG devices. *International Journal of Medical Informatics*, *119*, 39–46. https://doi.org/10.1016/j.ijmedinf.2018.08.010 (cit. on pp. 110, 113, 115).

Huddlestone, J., Sears, R., & Harris, D. (2017). The use of operational event sequence diagrams and work domain analysis techniques for the specification of the crewing configuration of a single-pilot commercial aircraft. *Cognition, Technology & Work*, *19*(2), 289–302. https://doi.org/10.1007/s10111-017-0423-5 (cit. on p. 29).

Hursh, S. R., Redmond, D. P., Johnson, M. L., Thorne, D. R., Belenky, G., Balkin, T. J., Storm, W. F., Miller, J. C., & Eddy, D. R. (2004). Fatigue Models for Applied Research in Warfighting. *Aviation, Space, and Environmental Medicine*, *75*(3), A44–A53 (cit. on p. 61).

Husta, C., Dalmaijer, E., Belopolsky, A., & Mathôt, S. (2019). The Pupillary Light Response Reflects Visual Working Memory Content. *Journal of Experimental Psychology: Human Perception and Performance*, *45*. https://doi.org/10.1037/xhp0000689 (cit. on p. 93).

Hutton, S. (2022). What are the technical challenges involved in eye tracking? (Cit. on pp. 91, 92).

ICAO. (2016). *Manual for the Oversight of fatigue Management Approaches* (tech. rep. No. Doc 9966). https://www.icao.int/safety/fatiguemanagement/FRMS%20Tools/Doc%209966.FRMS.2016%20Edition.en.pdf (cit. on p. 51).

Ionescu, T. (2012). Exploring the nature of cognitive flexibility. *New Ideas in Psychology*, *30*(2), 190–200. https://doi.org/10.1016/j.newideapsych.2011.11.001 (cit. on p. 59).

Islam, M. K., Rastegarnia, A., & Yang, Z. (2016). Methods for artifact detection and removal from scalp EEG: A review. *Neurophysiologie Clinique/Clinical Neurophysiology*, *46*(4), 287–305. https://doi.org/10.1016/j.neucli.2016.07.002 (cit. on p. 80).

ISO 9241-110:2020 Part 110: Interaction principles ["the extent to which a product can be used by specified end-users to achieve specified goal with effectiveness, efficiency and satisfaction in a specified context of use"]. (2021). https://www.iso.org/obp/ui/#iso:std:iso:9241:-110:ed-2:v1:en (cit. on p. 31).

Jackson, C. A., & Earl, L. (2006). Prevalence of fatigue among commercial pilots. *Occupational Medicine*, *56*(4), 263–268. https://doi.org/10.1093/occmed/kql021 (cit. on p. 50).

Jacquet, T., Poulin-Charronnat, B., Bard, P., Perra, J., & Lepers, R. (2021). Physical Activity and Music to Counteract Mental Fatigue. *Neuroscience*, *478*, 75–88. https://doi.org/10.1016/j.neuroscience.2021.09.019 (cit. on p. 37).

Jaeggi, S. M., Buschkuehl, M., Perrig, W. J., & Meier, B. (2010). The concurrent validity of the N-back task as a working memory measure. *Memory*, *18*(4), 394–412. https://doi.org/https://doi.org/10.1080/09658211003702171 (cit. on pp. 295, 317).

Jahanpour, E. S., Berberian, B., Imbert, J.-P., & Roy, R. N. (2020). Cognitive fatigue assessment in operational settings: a review and UAS implications**This work was supported by the Defense Innovation Agency (AID) of the French Ministry of Defense (research project CONCORDE N°2019 65 0090004707501). *IFAC-PapersOnLine*, *53*(5), 330–337. https://doi.org/10.1016/j.ifacol.2021.04.188 (cit. on pp. 112, 113).

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2017). *An introduction to statistical learning: with applications in R* (Corrected at 8th printing). Springer. https://doi.org/10.1007/978-1-4614-7138-7 (cit. on p. 104).
©Springer ... 2013 (Corrected at 8th printing 2017).

Jansen, B. H., Agarwal, G., Hegde, A., & Boutros, N. N. (2003). Phase synchronization of the ongoing EEG and auditory EP generation. *Clinical Neurophysiology*, *114*(1), 79–85. https://doi.org/10.1016/S1388-2457(02)00327-9 (cit. on p. 83).

Jasper, H. (1958). The ten-twenty electrode system of the international federation \textbar CiNii Research. *Electroencephalogr. Clin. Neurophysio*, 10, 370–375. https://doi.org/DOI:10.1016/0013-4694(58)90053-1 (cit. on pp. 77, 154, 157, 174, 298, 300).

Jenkins, A. (2013). Feminist Voices - Tsuruko Haraguchi. Retrieved November 22, 2023, from https://feministvoices.com/profiles/tsuruko-haraguchi (cit. on p. 49).

Jersild, A. T. (1927). *Mental set and shift*. Columbia university. (Cit. on p. 66).

Ji, Q., Zhu, Z., & Lan, P. (2004). Real-time nonintrusive monitoring and prediction of driver fatigue [Conference Name: IEEE Transactions on Vehicular Technology]. *IEEE Transactions on Vehicular Technology*, *53*(4), 1052–1068. https://doi.org/10.1109/TVT.2004.830974 (cit. on pp. 113, 115).

John, S. M., Smallman, H. S., & Manes, D. I. (2005). Recovery from Interruptions to a Dynamic Monitoring Task: The Beguiling Utility of Instant Replay. *th ANNUAL MEETING*, 5 (cit. on pp. 40, 42, 70).

Joseph, A. W., & Murugesh, R. (2020). Potential Eye Tracking Metrics and Indicators to Measure Cognitive Load in Human-Computer Interaction Research. *Journal of scientific research*, *64*(01), 168–175. https://doi.org/10.37398/JSR.2020.640137 (cit. on p. 93).

Joynt, R. (2024, January). Brainstem | Definition, Structure, & Function | Britannica. Retrieved February 2, 2024, from https://www.britannica.com/science/brainstem (cit. on p. 72).

Kahneman, D. (1973). *Attention and effort*. Prentice-Hall. (Cit. on pp. 34, 52).

Kaida, K., Takahashi, M., Åkerstedt, T., Nakata, A., Otsuka, Y., Haratani, T., & Fukasawa, K. (2006). Validation of the Karolinska sleepiness scale against performance and EEG variables. *Clinical Neurophysiology*, *117*(7), 1574–1581. https://doi.org/10.1016/j.clinph.2006.03.011 (cit. on pp. 62, 113, 130, 276, 292, 317).

Kamzanova, A. T., Matthews, G., Kustubayeva, A. M., & Jakupov, S. M. (2011). EEG Indices to Time-On-Task Effects and to a Workload Manipulation (Cueing). *International Journal of Psychological and Behavioral Sciences*, *5*(8), 928–931. Retrieved February 26, 2024, from https://publications.waset.org/8925/eeg-indices-to-time-on-task-effects-and-to-a-workload-manipulation-cueing (cit. on pp. 79, 182).

Kaplan Berkaya, S., Uysal, A. K., Sora Gunal, E., Ergin, S., Gunal, S., & Gulmezoglu, M. B. (2018). A survey on ECG analysis. *Biomedical Signal Processing and Control*, *43*, 216–235. https://doi.org/10.1016/j.bspc.2018.03.003 (cit. on p. 87).

Kappenman, E. S., & Luck, S. J. (2010). The effects of electrode impedance on data quality and statistical significance in ERP recordings. *Psychophysiology*, *47*(5), 888–904. https://doi.org/10.1111/j.1469-8986.2010.01009.x (cit. on p. 81).

Kar, A., & Corcoran, P. (2017). A Review and Analysis of Eye-Gaze Estimation Systems, Algorithms and Performance Evaluation Methods in Consumer Platforms [Conference Name: IEEE Access]. *IEEE Access*, *5*, 16495–16519. https://doi.org/10.1109/ACCESS.2017.2735633 (cit. on p. 92).

Katsuki, F., & Constantinidis, C. (2014). Bottom-up and top-down attention: different processes and overlapping neural systems. *The Neuroscientist*, *20*(5), 509–521. https://doi.org/10.1177/1073858413514136 (cit. on p. 58).

Kaya, I. (2019). A brief summary of EEG artifact handling [Publisher: in Artificial Intelligence]. *Brain-computer interface*, (9) (cit. on pp. 80, 82, 167).

Kesting, A., Treiber, M., Schönhof, M., & Helbing, D. (2007). Extending Adaptive Cruise Control to Adaptive Driving Strategies. *Transportation Research Record: Journal of the Transportation Research Board*, *2000*(1), 16–24. https://doi.org/10.3141/2000-03 (cit. on p. 44).

Khan, M. J., Hong, K.-S., Naseer, N., & Bhutta, M. R. (2014). A hybrid EEG-fNIRS BCI: Motor imagery for EEG and mental arithmetic for fNIRS [ISSN: 2093-7121]. *2014 14th International Conference on Control, Automation and Systems (ICCAS 2014)*, 275–278. https://doi.org/10.1109/ICCAS.2014.6988001 (cit. on p. 106).

Kiefer, P., Giannopoulos, I., Athanasios Anagnostopoulos, V., Schöning, J., & Raubal, M. (2017). Controllability matters: The user experience of adaptive maps. *GeoInformatica*, *21*(3), 619–641. https://doi.org/10.1007/s10707-016-0282-x (cit. on pp. 45, 46).

Kiesel, A., Steinhauser, M., Wendt, M., Falkenstein, M., Jost, K., Philipp, A. M., & Koch, I. (2010). Control and interference in task switching—A review. *Psychological Bulletin*, *136*(5), 849–874. https://doi.org/10.1037/a0019842 (cit. on pp. 14, 69).

Kiesel, A., Wendt, M., & Peters, A. (2007). Task switching: On the origin of response congruency effects [Place: Germany Publisher: Springer]. *Psychological Research*, *71*(2), 117–125. https://doi.org/10.1007/s00426-005-0004-8 (cit. on p. 285).

Kim, H., Luo, J., Chu, S., Cannard, C., Hoffmann, S., & Miyakoshi, M. (2023). ICA's bug: How ghost ICs emerge from effective rank deficiency caused by EEG electrode interpolation and incorrect re-referencing [Publisher: Frontiers]. *Frontiers in Signal Processing*, *3*. https://doi.org/10.3389/frsip.2023.1064138 (cit. on p. 81).

Kirchner, W. K. (1958). Age differences in short-term retention of rapidly changing information. [Publisher: American Psychological Association]. *Journal of experimental psychology*, *55*(4), 352 (cit. on pp. 110, 111).

Kizilcec, R. F. (2016). How Much Information?: Effects of Transparency on Trust in an Algorithmic Interface. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 2390–2395. https://doi.org/10.1145/2858036.2858402 (cit. on p. 47).

Kleinsorge, T. (2004). Hierarchical Switching with Two Types of Judgment and Two Stimulus Dimensions. *Experimental Psychology*, *51*(2), 145–149. https://doi.org/10.1027/1618-3169.51.2.145 (cit. on p. 68)
Publisher: Hogrefe Publishing.

Kleinsorge, T., & Heuer, H. (1999). Hierarchical switching in a multi-dimensional task space. *Psychological Research*, *62*(4), 300–312. https://doi.org/10.1007/s004260050060 (cit. on p. 68).

Knyazev, G. G. (2012). EEG delta oscillations as a correlate of basic homeostatic and motivational processes. *Neuroscience & Biobehavioral Reviews*, *36*(1), 677–695. https://doi.org/10.1016/j.neubiorev.2011.10.002 (cit. on p. 79).

Ko, H.-k., Snodderly, D. M., & Poletti, M. (2016). Eye movements between saccades: Measuring ocular drift and tremor. *Vision Research*, *122*, 93–104. https://doi.org/10.1016/j.visres.2016.03.006 (cit. on pp. 93, 112, 113).

Ko, L.-W., Komarov, O., Lai, W.-K., Liang, W.-G., & Jung, T.-P. (2020). Eyeblink recognition improves fatigue prediction from single-channel forehead EEG in a realistic sustained attention task [Publisher: IOP Publishing]. *Journal of Neural Engineering*, *17*(3), 036015. Retrieved October 6, 2022, from https://doi.org/10.1088/1741-2552/ab909f (cit. on pp. 113, 115).

Ko, L.-W., Lai, W.-K., Liang, W.-G., Chuang, C.-H., Lu, S.-W., Lu, Y.-C., Hsiung, T.-Y., Wu, H.-H., & Lin, C.-T. (2015). Single channel wireless EEG device for real-time fatigue level detection [ISSN: 2161-4407]. *2015 International Joint Conference on Neural Networks (IJCNN)*, 1–5. https://doi.org/10.1109/IJCNN.2015.7280817 (cit. on pp. 111–113, 207).

Kobler, R., Hirayama, J.-i., Zhao, Q., & Kawanabe, M. (2022). SPD domain-specific batch normalization to crack interpretable unsupervised domain adaptation in EEG. *Advances in Neural Information Processing Systems*, *35*, 6219–6235 (cit. on pp. 212, 289).

Koch, I. (2003). The role of external cues for endogenous advance reconfiguration in task switching. *Psychonomic Bulletin & Review*, *10*(2), 488–492. https://doi.org/10.3758/BF03196511 (cit. on p. 66).

Koch, I., Poljac, E., Müller, H., & Kiesel, A. (2018). Cognitive structure, flexibility, and plasticity in human multitasking—An integrative review of dual-task and task-switching research. *Psychological Bulletin*, *144*(6), 557–583. https://doi.org/10.1037/bul0000144 (cit. on pp. 13, 39, 59, 66, 210).

Koelstra, S., Muhl, C., Soleymani, M., Lee, J.-S., Yazdani, A., Ebrahimi, T., Pun, T., Nijholt, A., & Patras, I. (2012). DEAP: A Database for Emotion Analysis ;Using Physiological Signals. *IEEE Transactions on Affective Computing*, *3*(1), 18–31. https://doi.org/10.1109/T-AFFC.2011.15 (cit. on p. 292).

Kok, A. (2022). Cognitive control, motivation and fatigue: A cognitive neuroscience perspective. *Brain and Cognition*, *160*, 105880. https://doi.org/10.1016/j.bandc.2022.105880 (cit. on p. 55).

Kothe, C. (2015). Sccn/labstreaminglayer [original-date: 2018-02-28T10:50:12Z]. Retrieved January 11, 2023, from https://github.com/sccn/labstreaminglayer (cit. on p. 174).

Kothe, C., Medine, D., Boulay, C., Grivich, M., & Stenner, T. (2014). Lab streaming layer. *URL https://github. com/sccn/labstreaminglayer* (cit. on pp. 153, 154, 171).

Krishnan, A., Kumar, R., Etienne, A., Robinson, A., Kelly, S. K., Behrmann, M., Tarr, M. J., & Grover, P. (2018). Challenges and Opportunities in Instrumentation and Use of High-Density EEG for Underserved Regions. In C. M. Kebe, A. Gueye, A. Ndiaye, & A. Garba (Eds.), *Innovations and Interdisciplinary Solutions for Underserved Areas* (pp. 72–82). Springer International Publishing. https://doi.org/10.1007/978-3-319-98878-8_7 (cit. on p. 77).

Krol, Laurens R. & Zander, Thorsten O. (2017). Passive Bci-Based Neuroadaptive Systems [ISBN: 9783851255331 Publisher: Verlag der Technischen Universität Graz]. https://doi.org/10.3217/978-3-85125-533-1-46 (cit. on pp. 108, 110)
SeriesInformation Proceedings Of The 7Th Graz Brain-Computer Interface Conference 2017, From Vision To Reality, September 18-22, 2017 Graz University Of Technology, Austria.

Kübler, A. (2020). The history of BCI: From a vision for the future to real support for personhood in people with locked-in syndrome. *Neuroethics*, *13*(2), 163–180. https://doi.org/10.1007/s12152-019-09409-4 (cit. on p. 106).

Kubrick, S. (1968). 2001: A Space Odyssey. (Cit. on p. 105).

Kudesia, R. S., Pandey, A., & Reina, C. S. (2022). Doing More With Less: Interactive Effects of Cognitive Resources and Mindfulness Training in Coping With Mental Fatigue From Multitasking [Publisher: SAGE Publications Inc]. *Journal of Management*, *48*(2), 410–439. https://doi.org/10.1177/0149206320964570 (cit. on p. 59)
Publisher: SAGE Publications Inc.

Labedan, P., Darodes-De-Tailly, N., Dehais, F., & Peysakhovich, V. (2021). Virtual reality for pilot training: study of cardiac activity, 81–88. Retrieved July 25, 2022, from https://doi.org/10.5220/0010296700810088 (cit. on p. 301).

Ladouce, S., & Dehais, F. (2024, March). Frequency-tagging of spatial attention using perilimiminal flickers. https://doi.org/10.1101/2024.02.29.582725 (cit. on p. 106).

Lajos, B. (1995). Adaptive human-computer interfaces for man-machine interaction in computer-integrated systems. *Computer Integrated Manufacturing Systems*, *2*(8), 133–142. Retrieved January 29, 2022, from https://www.infona.pl//resource/bwmeta1.element.elsevier-58a7f3f6-267e-314a-a2fd-3a9ee663a620 (cit. on p. 48).

Lal, S. K., & Craig, A. (2001). A critical review of the psychophysiology of driver fatigue. *Biological Psychology*, *55*(3), 173–194. https://doi.org/10.1016/s0301-0511(00)00085-5 (cit. on pp. 113–115).

Lal, S. K. L., & Craig, A. (2002). Driver fatigue: electroencephalography and psychological assessment. *Psychophysiology*, *39*(3), 313–321. https://doi.org/10.1017/s0048577201393095 (cit. on p. 114).

Lamond, N., Dawson, D., & Roach, G. D. (2005). Fatigue Assessment in the Field: Validation of a Hand-Held Electronic Psychomotor Vigilance Task. *Aviation, Space, and Environmental Medicine*, *76*(5), 486–489 (cit. on pp. 292, 294, 317).

Land, M. F. (1999). Motion and vision: why animals move their eyes. *Journal of Comparative Physiology A*, *185*(4), 341–352. https://doi.org/10.1007/s003590050393 (cit. on p. 91).

Lau, N., Skraaning, G., Jamieson, G. A., & Burns, C. M. (2008). Enhancing Operator Task Performance during Monitoring for Unanticipated Events through Ecological Interface Design [Publisher: SAGE Publications Inc]. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *52*(4), 448–452. https://doi.org/10.1177/154193120805200452 (cit. on p. 23).

Lavie, T., & Meyer, J. (2010). Benefits and costs of adaptive user interfaces. *International Journal of Human-Computer Studies*, *68*(8), 508–524. https://doi.org/10.1016/j.ijhcs.2010.01.004 (cit. on p. 46).

LeCun, Y., Huang, F. J., & Bottou, L. (2004). Learning methods for generic object recognition with invariance to pose and lighting. *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, *2*, II–104 (cit. on pp. 154, 155, 170).

Lee, B.-G., Lee, B.-L., & Chung, W.-Y. (2014). Mobile Healthcare for Automatic Driving Sleep-Onset Detection Using Wavelet-Based EEG and Respiration Signals [Number: 10 Publisher: Multidisciplinary Digital Publishing Institute]. *Sensors*, *14*(10), 17915–17936. https://doi.org/10.3390/s141017915 (cit. on pp. 112, 113, 115)
Number: 10 Publisher: Multidisciplinary Digital Publishing Institute.

Lee, J. D., & See, K. A. (2004). Trust in Automation: Designing for Appropriate Reliance. *Human Factors*, 31 (cit. on pp. 46, 47, 211).

Lencer, R., & Trillenberg, P. (2022). Smooth Pursuit Eye Movements. (Cit. on p. 93).

Lepage, K. Q., Kramer, M. A., & Chu, C. J. (2014). A statistically robust EEG re-referencing procedure to mitigate reference effect. *Journal of Neuroscience Methods*, *235*, 101–116. https://doi.org/10.1016/j.jneumeth.2014.05.008 (cit. on p. 81).

Leske, S., & Dalal, S. S. (2019). Reducing power line noise in EEG and MEG data via spectrum interpolation. *NeuroImage*, *189*, 763–776. https://doi.org/10.1016/j.neuroimage.2019.01.026 (cit. on pp. 81, 113, 115).

Li, G., Li, B., Wang, G., Zhang, J., & Wang, J. (2017). A New Method for Human Mental Fatigue Detection with Several EEG Channels. *Journal of Medical and Biological Engineering*, *37*(2), 240–247. https://doi.org/10.1007/s40846-017-0224-6 (cit. on p. 45).

Li, P. (2015, August). *Safety analysis using a Smart Safety Helmet embedded with IMU and EEG sensors applied in industrial facility* [Master's thesis, Université du Québec à Chicoutimi]. Retrieved October 11, 2022, from https://constellation.uqac.ca/id/eprint/3315/ (cit. on pp. 113, 115).

Li, Z., Li, S. E., Li, R., Cheng, B., & Shi, J. (2017). Online Detection of Driver Fatigue Using Steering Wheel Angles for Real Driving Conditions [Number: 3 Publisher: Multidisciplinary Digital Publishing Institute]. *Sensors*, *17*(3), 495. https://doi.org/10.3390/s17030495 (cit. on p. 113)
Number: 3 Publisher: Multidisciplinary Digital Publishing Institute.

Liefooghe, B., Barrouillet, P., Vandierendonck, A., & Camos, V. (2008). Working memory costs of task switching [Number: 3]. *JOURNAL OF EXPERIMENTAL PSYCHOLOGY-LEARNING MEMORY AND COGNITION*, *34*(3), 478–494. Retrieved June 30, 2022, from http://hdl.handle.net/1854/LU-417381 (cit. on pp. 66, 68, 69)
Number: 3.

Lim, J.-H., Lee, J.-H., Hwang, H.-J., Kim, D. H., & Im, C.-H. (2015). Development of a hybrid mental spelling system combining SSVEP-based brain–computer interface and webcam-based eye tracking. *Biomedical Signal Processing and Control*, *21*, 99–104. https://doi.org/10.1016/j.bspc.2015.05.012 (cit. on pp. 106, 108).

Lim, Y., Gardi, A., Sabatini, R., Ramasamy, S., Kistan, T., Ezer, N., Vince, J., & Bolia, R. (2018). Avionics Human-Machine Interfaces and Interactions for Manned and Unmanned Aircraft. *Progress in Aerospace Sciences*, *102*, 1–46. https://doi.org/10.1016/j.paerosci.2018.05.002 (cit. on pp. 30, 58).

Lim, Y., Liu, J., Ramasamy, S., & Sabatini, R. (2016). Cognitive Remote Pilot-Aircraft Interface for UAS Operations. *Proceedings of the 2016 International Conference on Intelligent Unmanned Systems (ICIUS 2016)*, 5 (cit. on p. 129).

Liu, C. C., Hosking, S. G., & Lenné, M. G. (2009). Predicting driver drowsiness using vehicle measures: Recent insights and future challenges. *Journal of Safety Research*, *40*(4), 239–245. https://doi.org/10.1016/j.jsr.2009.04.005 (cit. on p. 45).

Liu, J., Li, R., Ling, S. L., Zhou, J. R., Lim, K. W., Srinivasan, R., Tan, K. C., Ho, W. K., & Tay, A. (2001). Intelligent Alarm Management Through Suppressing Nuisance Alarms and Providing Operator Advice. *IFAC Proceedings Volumes*, *34*(27), 81–86. https://doi.org/10.1016/S1474-6670(17)33572-3 (cit. on p. 38).

Liu, Y., Lan, Z., Cui, J., Sourina, O., & Müller-Wittig, W. (2020). Inter-subject transfer learning for EEG-based mental fatigue recognition. *Advanced Engineering Informatics*, *46*, 101157. https://doi.org/10.1016/j.aei.2020.101157 (cit. on p. 116).

Logan, G. D., & Gordon, R. D. (2001). Executive control of visual attention in dual-task situations. *Psychological Review*, *108*(2), 393–434. https://doi.org/10.1037/0033-295x.108.2.393 (cit. on pp. 65, 126).

Logan, G. D. (2004). Working memory, task switching, and executive control in the task span procedure. *Journal of experimental psychology General*, *133*(2), 218–236. https://doi.org/10.1037/0096-3445.133.2.218 (cit. on p. 66).

Lopes da Silva, F. (2013). EEG and MEG: Relevance to Neuroscience. *Neuron*, *80*(5), 1112–1128. https://doi.org/10.1016/j.neuron.2013.10.017 (cit. on p. 79).

Lorist, M. M., Bezdan, E., ten Caat, M., Span, M. M., Roerdink, J. B. T. M., & Maurits, N. M. (2009). The influence of mental fatigue and motivation on neural network dynamics; an EEG coherence study. *Brain Research*, *1270*, 95–106. https://doi.org/10.1016/j.brainres.2009.03.015 (cit. on pp. 55–57, 59, 112, 113).

Lorist, M. M., Boksem, M. A., & Ridderinkhof, K. R. (2005). Impaired cognitive control and reduced cingulate activity during mental fatigue. *Cognitive Brain Research*, *24*(2), 199–205. https://doi.org/10.1016/j.cogbrainres.2005.01.018 (cit. on p. 58).

Lorist, M. M., Klein, M., Nieuwenhuis, S., Jong, R. D., Mulder, G., & Meijman, T. F. (2000). Mental fatigue and task control: Planning and preparation, 12 (cit. on pp. 39, 43, 59).

Lotte, F., & Roy, R. N. (2019, January). Chapter 7 - Brain–Computer Interface Contributions to Neuroergonomics. In H. Ayaz & F. Dehais (Eds.), *Neuroergonomics* (pp. 43–48). Academic Press. https://doi.org/10.1016/B978-0-12-811926-6.00007-5 (cit. on pp. 14, 106, 108).

Lu, Z., Coster, X., & De Winter, J. (2017). How much time do drivers need to obtain situation awareness? A laboratory-based study of automated driving. *Applied Ergonomics*, *60*, 293–304. https://doi.org/10.1016/j.apergo.2016.12.003 (cit. on p. 40).

Luck, S. J. (2005, January). *Introduction to the Event-related Potential Technique* (1st edition). MIT Press. (Cit. on p. 83).

Luck, S. J., & Kappenman, E. S. (2011, December). *The Oxford Handbook of Event-Related Potential Components* [Google-Books-ID: gItoAgAAQBAJ]. Oxford University Press. (Cit. on p. 112).

Lunn, E. (2022, March). What Is Dark Mode – And Should You Be Using It? [Publication Title: Forbes Advisor UK]. Retrieved June 29, 2022, from https://www.forbes.com/uk/advisor/mobile-phones/what-is-dark-mode-and-should-you-be-using-it/ (cit. on p. 44).
Section: Mobile Phones.

Maffei, A., & Angrilli, A. (2019). Spontaneous blink rate as an index of attention and emotion during film clips viewing. *Physiology & Behavior*, *204*, 256–263. https://doi.org/10.1016/j.physbeh.2019.02.037 (cit. on p. 93).

Maglogiannis, I. G. (2007). *Emerging Artificial Intelligence Applications in Computer Engineering: Real Word AI Systems with Applications in EHealth, HCI, Information Retrieval and Pervasive Technologies* [Google-Books-ID: vLiTXDHr_sYC]. IOS Press. (Cit. on p. 98).

Malik, M. (1996). Heart Rate Variability. *Annals of Noninvasive Electrocardiology*, *1*(2), 151–181. https://doi.org/10.1111/j.1542-474X.1996.tb00275.x (cit. on p. 89).

Manning, C., Hall, J., & Gold, P. (1990). Glucose Effects on Memory and Other Neuropsychological Tests in Elderly Humans [Publisher: SAGE Publications Inc]. *Psychological Science*, *1*(5), 307–311. https://doi.org/10.1111/j.1467-9280.1990.tb00223.x (cit. on p. 53).

Manning, S. D., Rash, C. E., LeDuc, P. A., Noback, R. K., & McKeon, J. (2004, March). *The Role of Human Causal Factors in U.S. Army Unmanned Aerial Vehicle Accidents:* (tech. rep.). Defense Technical Information Center. Fort Belvoir, VA. https://doi.org/10.21236/ADA421592 (cit. on p. 26).

Marquart, G., Cabrall, C., & De Winter, J. (2015). Review of Eye-related Measures of Drivers' Mental Workload. *Procedia Manufacturing*, *3*, 2854–2861. https://doi.org/10.1016/j.promfg.2015.07.783 (cit. on p. 115).

Marquez, J. J., & Cummings, M. L. (2008). Design and Evaluation of Path Planning Decision Support for Planetary Surface Exploration. *Journal of Aerospace Computing, Information, and Communication*, *5*(3), 57–71. https://doi.org/10.2514/1.26248 (cit. on pp. 40, 43).

Martin, P. Y., & Benton, D. (1999). The Influence of a Glucose Drink on a Demanding Working Memory Task. *Physiology & Behavior*, *67*(1), 69–74. https://doi.org/10.1016/S0031-9384(99)00040-2 (cit. on p. 53).

Martinez-Conde, S. (2022). Fixations. (Cit. on p. 93).

Martinez-Conde, S., Otero-Millan, J., & Macknik, S. L. (2013). The impact of microsaccades on vision: towards a unified theory of saccadic function. *Nature Reviews Neuroscience*, *14*(2), 83–96. https://doi.org/10.1038/nrn3405 (cit. on p. 93).

Mathôt, S. (2022). Pupillometry: Psychology, Physiology, and Function. (Cit. on p. 90).

Matthews, G. (Ed.). (2012). *The handbook of operator fatigue*. Ashgate Pub. Company. (Cit. on pp. 51, 52).

Mayr, U., & Kliegl, R. (2000). Task-set switching and long-term memory retrieval. *Journal of experimental psychology Learning, memory, and cognition*, *26*(5), 1124–1140. https://doi.org/10.1037//0278-7393.26.5.1124 (cit. on pp. 65, 66).

Mayr, U., & Kliegl, R. (2003). Differential effects of cue changes and task changes on task-set selection costs. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(3), 362–372. https://doi.org/10.1037/0278-7393.29.3.362 (cit. on p. 69).

McDonald, A. D., Schwarz, C., Lee, J. D., & Brown, T. L. (2012). Real-Time Detection of Drowsiness Related Lane Departures Using Steering Wheel Angle [Publisher: SAGE Publications Inc]. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *56*(1), 2201–2205. https://doi.org/10.1177/1071181312561464 (cit. on pp. 45, 113)
Publisher: SAGE Publications Inc.

McIntire, L. K., McKinley, R. A., Goodyear, C., & McIntire, J. P. (2014). Detection of vigilance performance using eye blinks. *Applied Ergonomics*, *45*(2, Part B), 354–362. https://doi.org/10.1016/j.apergo.2013.04.020 (cit. on p. 93).

Meiran, N. (1996). Reconfiguration of processing mode prior to task performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(6), 1423–1442. https://doi.org/10.1037/0278-7393.22.6.1423 (cit. on p. 66).

Meiran, N. (2008). The Dual Implication of Dual Affordance: Stimulus-Task Binding and Attentional Focus Changing During Task Preparation. *Experimental Psychology*, *55*(4), 251–259. https://doi.org/10.1027/1618-3169.55.4.251 (cit. on p. 68).

Melo, H. M., Nascimento, L. M., & Takase, E. (2017). Mental fatigue and heart rate variability (HRV): The time-on-task effect. *Psychology & Neuroscience*, *10*(4), 428–436. https://doi.org/10.1037/pne0000110 (cit. on pp. 113, 115)
Place: US Publisher: Educational Publishing Foundation.

Meyer, M., Endedijk, H. M., Van Ede, F., & Hunnius, S. (2019). Theta oscillations in 4-year-olds are sensitive to task engagement and task demands. *Scientific Reports*, *9*(1), 6049. https://doi.org/10.1038/s41598-019-42615-x (cit. on p. 79).

Miley, A. Å., Kecklund, G., & Åkerstedt, T. (2016). Comparing two versions of the Karolinska Sleepiness Scale (KSS). *Sleep and Biological Rhythms*, *14*(3), 257–260. https://doi.org/10.1007/s41105-016-0048-8 (cit. on p. 279).

Minguillon, J., Lopez-Gordo, M. A., & Pelayo, F. (2017). Trends in EEG-BCI for daily-life: Requirements for artifact removal. *Biomedical Signal Processing and Control*, *31*, 407–418. https://doi.org/10.1016/j.bspc.2016.09.005 (cit. on pp. 82, 83).

Miraz, M. H., Ali, M., & Excell, P. S. (2021). Adaptive user interfaces and universal usability through plasticity of user interface design. *Computer Science Review*, *40*, 100363. https://doi.org/10.1016/j.cosrev.2021.100363 (cit. on pp. 32, 37, 44, 105, 109).

Mitchell, T. M. (1997). *Machine Learning*. McGraw-Hill. (Cit. on p. 95).

Mizuno, K., Tanaka, M., Tajima, K., Okada, N., Rokushima, K., & Watanabe, Y. (2009). Effects of mild-stream bathing on recovery from mental fatigue. *Medical Science Monitor*, *16*(1), CR8–CR14. Retrieved January 26, 2022, from https://www.medscimonit.com/abstract/index/idArt/878306 (cit. on p. 37)
Publisher: International Scientific Information, Inc.

Möckel, T., Beste, C., & Wascher, E. (2015). The Effects of Time on Task in Response Selection - An ERP Study of Mental Fatigue. *Scientific Reports*, *5*(1), 10113. https://doi.org/10.1038/srep10113 (cit. on pp. 113, 114).

Monchi, O., Petrides, M., Petre, V., Worsley, K., & Dagher, A. (2001). Wisconsin Card Sorting Revisited: Distinct Neural Circuits Participating in Different Stages of the Task Identified by Event-Related Functional Magnetic Resonance Imaging [Publisher: Society for Neuroscience Section: ARTICLE]. *Journal of Neuroscience*, *21*(19), 7733–7741. https://doi.org/10.1523/JNEUROSCI.21-19-07733.2001 (cit. on p. 60)
Publisher: Society for Neuroscience Section: ARTICLE.

Monseigne, T., Lotte, F., Bioulac, S., Philip, P., & Micoulaud-Franchi, J.-A. (2019). Design and preliminary study of a neurofeedback protocol to self-regulate an EEG marker of drowsiness. Retrieved May 22, 2024, from https://inria.hal.science/hal-02130755 (cit. on pp. 113, 114, 207).

Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences*, *7*(3), 134–140. https://doi.org/10.1016/S1364-6613(03)00028-7 (cit. on pp. 65, 67).

Monsell, S., & Mizon, G. A. (2006). Can the task-cuing paradigm measure an endogenous task-set reconfiguration process. *Journal of experimental psychology Human perception and performance*, *32*(3), 493–516. https://doi.org/10.1037/0096-1523.32.3.493 (cit. on p. 66).

Monsell, S., Sumner, P., & Waters, H. (2003). Task-set reconfiguration with predictable and unpredictable task switches. *Memory & Cognition*, *31*(3), 327–342. https://doi.org/10.3758/BF03194391 (cit. on pp. 39, 55, 56, 59, 65).

Monsell, S., Yeung, N., & Azuma, R. (2000). Reconfiguration of task-set: Is it easier to switch to the weaker task? *Psychological Research*, *63*(3), 250–264. https://doi.org/10.1007/s004269900005 (cit. on p. 136).

Moreno-Ocampo, L. (2006). Allegations concerning War Crimes. https://www.icc-cpi.int/sites/default/files/NR/rdonlyres/04D143C8-19FB-466C-AB77-4CDB2FDEBEF7/143682/OTP_letter_to_senders_re_Iraq_9_February_2006.pdf (cit. on p. 25).

Morris, T. L., & Miller, J. C. (1996). Electrooculographic and performance indices of fatigue during simulated flight. *Biological Psychology*, *42*(3), 343–360. https://doi.org/10.1016/0301-0511(95)05166-X (cit. on pp. 113, 115).

Moseley, A. (2024). Just War Theory | Internet Encyclopedia of Philosophy. Retrieved April 25, 2024, from https://iep.utm.edu/justwar/ (cit. on p. 215).

Mueller-Putz, G., Scherer, R., Brunner, C., Leeb, R., & Pfurtscheller, G. (2008). Better than random: A closer look on BCI results. *International Journal of Bioelectromagnetism* (cit. on pp. 103, 158, 174, 303).

Mühl, C., Jeunet, C., & Lotte, F. (2014). EEG-based Workload Estimation across Affective Contexts. *Frontiers in Neuroscience*, *8*. https://doi.org/10.3389/fnins.2014.00114 (cit. on p. 291).

Müller, K.-R., Tangermann, M., Dornhege, G., Krauledat, M., Curio, G., & Blankertz, B. (2008). Machine learning for real-time single-trial EEG-analysis: From brain–computer interfacing to mental state monitoring. *Journal of Neuroscience Methods*, *167*(1), 82–90. https://doi.org/10.1016/j.jneumeth.2007.09.022 (cit. on pp. 111, 181, 207).

Murphy, A. H. (1993). What Is a Good Forecast? An Essay on the Nature of Goodness in Weather Forecasting [Publisher: American Meteorological Society Section: Weather and Forecasting]. *Weather and Forecasting*, *8*(2), 281–293. https://doi.org/10.1175/1520-0434(1993)008<0281:WIAGFA>2.0.CO;2 (cit. on pp. 61, 63).

Nieuwenhuis, S., Holroyd, C. B., Mol, N., & Coles, M. G. H. (2004). Reinforcement-related brain potentials from medial frontal cortex: origins and functional significance. *Neuroscience & Biobehavioral Reviews*, *28*(4), 441–448. https://doi.org/https://www.sciencedirect.com/science/article/pii/S0149763404000521 (cit. on p. 302).

Nieuwenhuis, S., Ridderinkhof, K. R., Blom, J., Band, G. P., & Kok, A. (2001). Error-related brain potentials are differentially related to awareness of response errors: Evidence from an antisaccade task. *Psychophysiology*, *38*(5), 752–760. https://doi.org/10.1111/1469-8986.3850752 (cit. on p. 112).

Niso, G., Krol, L. R., Combrisson, E., Dubarry, A. S., Elliott, M. A., François, C., Héjja-Brichard, Y., Herbst, S. K., Jerbi, K., Kovic, V., Lehongre, K., Luck, S. J., Mercier, M., Mosher, J. C., Pavlov, Y. G., Puce, A., Schettino, A., Schön, D., Sinnott-Armstrong, W., … Chaumon, M. (2022). Good scientific practice in EEG and MEG research: Progress and perspectives. *NeuroImage*, *257*, 119056. https://doi.org/10.1016/j.neuroimage.2022.119056 (cit. on p. 81).

NTSB. (1990). *Grounding of the U.S. Tankship EXXON VALDEZ on Bligh Reef, Prince William sound Bear Valdez, Alaska* (tech. rep. No. NTSB/MAR-90/04). (Cit. on p. 50).

NTSB. (1999). *Evaluation of US Department of Transportation efforts in the 1990s to address operator fatigue* (tech. rep.) (Publication Title: Safety Report NTSB/SR-99/01). National Transportation Safety Board. Washington, DC, US. (Cit. on p. 50).

NTSB. (2006). *Collision with Trees and Crash Short of the Runway, Corporate Airlines Flight 5966 BAE Systems BAE-J3201, N875JX Kirksville, Missouri October 19, 2004* (tech. rep.). (Cit. on p. 50).

OACI, O. D. L. C. I. (2016). *Manuel Pour La Supervision Des Approches de Gestion de La Fatigue* (tech. rep. No. Doc 9966. Deuxième édition). OACI. (Cit. on p. 276).

Oberauer, K., Wilhelm, O., Schulze, R., & Süss, H.-M. (2005). Working memory and intelligence - their correlation and their relation: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin*, *131*(1), 61–65. https://doi.org/https://bibliographie.ub.uni-due.de/servlets/DozBibEntryServlet?id=ubo_mods_00031441 (cit. on p. 296).

O'Brien, J., & Bull Schaefer, R. A. (2020). Deadly distraction – Eastern Air 401: the accident that changed aviation forever. *The CASE Journal*, *16*(3), 345–368. https://doi.org/10.1108/TCJ-10-2019-0095 (cit. on p. 60).

Oken, B. S. (1986). Filtering and aliasing of muscle activity in EEG frequency analysis. *Electroencephalography and Clinical Neurophysiology*, *64*(1), 77–80. https://doi.org/10.1016/0013-4694(86)90045-3 (cit. on p. 81).

on science and technology House of Representatives, C. (1986). *Invesitgation of the Challenger accident* (tech. rep.). Washington, DC, US. (Cit. on p. 50).

Oncu, M., & Yildiz, S. (2014, December). *An Analysis of Human Causal Factors in Unmanned Aerial Vehicle (UAV) Accidents:* (tech. rep.). Defense Technical Information Center. Fort Belvoir, VA. https://doi.org/10.21236/ADA620843 (cit. on pp. 23, 24).

Onnasch, L., Ruff, S., & Manzey, D. (2014). Operators adaptation to imperfect automation – Impact of miss-prone alarm systems on attention allocation and performance. *International Journal of Human-Computer Studies*, *72*(10), 772–782. https://doi.org/10.1016/j.ijhcs.2014.05.001 (cit. on p. 41).

Onnasch, L., Wickens, C., Li, H., & Manzey, D. (2013). Human Performance Consequences of Stages and Levels of Automation An Integrated Meta-Analysis. *Human Factors The Journal of the Human Factors and Ergonomics Society*. https://doi.org/10.1177/0018720813501549 (cit. on pp. 38, 40).

Otero-Millan, J., Troncoso, X. G., Macknik, S. L., Serrano-Pedraza, I., & Martinez-Conde, S. (2008). Saccades and microsaccades during visual fixation, exploration, and search: Foundations for a common saccadic generator. *Journal of Vision*, *8*(14), 21. https://doi.org/10.1167/8.14.21 (cit. on p. 93).

Ousler, G. W. I., Hagberg, K. W., Schindelar, M., Welch, D., & Abelson, M. B. (2008). The Ocular Protection Index. *Cornea*, *27*(5), 509. https://doi.org/10.1097/ICO.0b013e31816583f6 (cit. on p. 93).

Pan, S. J., & Yang, Q. (2010). A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering*, *22*(10), 1345–1359. https://doi.org/10.1109/TKDE.2009.191 (cit. on p. 291).

Parasuraman, R., Cosenzo, K., & de Visser, E. (2009). Adaptive Automation for Human Supervision of Multiple Uninhabited Vehicles: Effects on Change Detection, Situation Awareness, and Mental Workload. *Military Psychology - MIL PSYCHOL*, *21*, 270–297. https://doi.org/10.1080/08995600902768800 (cit. on p. 42).

Parasuraman, R., Galster, S., Squire, P., Furukawa, H., & Miller, C. (2005). A Flexible Delegation-Type Interface Enhances System Performance in Human Supervision of Multiple Robots: Empirical Studies With RoboFlag. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, *35*(4), 481–493. https://doi.org/10.1109/TSMCA.2005.850598 (cit. on p. 38).

Parasuraman, R., Sheridan, T., & Wickens, C. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, *30*(3), 286–297. https://doi.org/10.1109/3468.844354 (cit. on p. 32) Conference Name: IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans.

Park, J., Park, J., Shin, D., & Choi, Y. (2021). A BCI Based Alerting System for Attention Recovery of UAV Operators. *Sensors*, *21*(7), 2447. https://doi.org/10.3390/s21072447 (cit. on pp. 38, 42).

Partala, T., Luimula, M., & Saukko, O. (2006). Automatic rotation and zooming in mobile road maps. *Proceedings of the 8th conference on Human-computer interaction with mobile devices and services*, 255–258. https://doi.org/10.1145/1152215.1152269 (cit. on pp. 31–33, 47).

Paterno, F. (2014). *User Interface Design Adaptation*. Interaction Design Foundation - IxDF. Retrieved July 4, 2024, from https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/user-interface-design-adaptation (cit. on pp. 31, 32).

Paul, M. A., Hursh, S. R., & Love, R. J. (2020). The Importance of Validating Sleep Behavior Models for Fatigue Management Software in Military Aviation. *Military Medicine*, *185*(11-12), e1986–e1991. https://doi.org/10.1093/milmed/usaa210 (cit. on p. 61).

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy [Publisher: Springer]. *Behavior research methods*, *51*, 195–203 (cit. on pp. 171, 287).

Peiris, M. T. R., Jones, R. D., Davidson, P. R., Carroll, G. J., & Bones, P. J. (2006). Frequent lapses of responsiveness during an extended visuomotor tracking task in non-sleep-deprived subjects. *Journal of Sleep Research*, *15*(3), 291–300. https://doi.org/10.1111/j.1365-2869.2006.00545.x (cit. on p. 182).

Peters, J. C., Reithler, J., Graaf, T. A. d., Schuhmann, T., Goebel, R., & Sack, A. T. (2020). Concurrent human TMS-EEG-fMRI enables monitoring of oscillatory brain state-dependent gating of cortico-subcortical network activity [Publisher: Nature Publishing Group]. *Communications Biology*, *3*(1), 1–11. https://doi.org/10.1038/s42003-020-0764-0 (cit. on p. 208).

Petruo, V. A., Mückschel, M., & Beste, C. (2018). On the role of the prefrontal cortex in fatigue effects on cognitive flexibility - a system neurophysiological approach [Bandiera_abtest: a Cc_license_type: cc_by Cg_type: Nature Research Journals Number: 1 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Cognitive control;Human behaviour Subject_term_id: cognitive-control;human-behaviour]. *Scientific Reports*, *8*(1), 6395. https://doi.org/10.1038/s41598-018-24834-w (cit. on pp. 40, 43, 57, 59) Bandiera_abtest: a Cc_license_type: cc_by Cg_type: Nature Research Journals Number: 1 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Cognitive control;Human behaviour Subject_term_id: cognitive-control;human-behaviour.

Peysakhovich, V., Lefrançois, O., Dehais, F., & Causse, M. (2018). The Neuroergonomics of Aircraft Cockpits: The Four Stages of Eye-Tracking Integration to Enhance Flight Safety. *Safety*, *4*(1), 8. https://doi.org/10.3390/safety4010008 (cit. on pp. 41, 43).

Phillips, R. O. (2015). A review of definitions of fatigue – And a step towards a whole definition. *Transportation Research Part F: Traffic Psychology and Behaviour*, *29*, 48–56. https://doi.org/10.1016/j.trf.2015.01.003 (cit. on pp. 51, 52, 110).

Piechulla, W., Mayser, C., Gehrke, H., & König, W. (2003). Reducing drivers' mental workload by means of an adaptive man–machine interface. *Transportation Research Part F: Traffic*

Psychology and Behaviour, *6*(4), 233–248. https://doi.org/10.1016/j.trf.2003.08.001 (cit. on pp. 38, 39, 42).

Pimenta, A., Carneiro, D. R., Novais, P., & Neves, J. (2013). Monitoring mental fatigue through the analysis of keyboard and mouse interaction patterns. https://doi.org/10.1007/978-3-642-40846-5_23 (cit. on pp. 45, 113)
Accepted: 2013-12-03T11:30:58Z ISSN: 0302-9743.

Pinel, J. P. J. (2008, January). *Biopsychology: International Edition* (7e édition). Pearson. (Cit. on pp. 72, 76).

Pinti, P., Tachtsidis, I., Hamilton, A., Hirsch, J., Aichelburg, C., Gilbert, S., & Burgess, P. W. (2020). The present and future use of functional near-infrared spectroscopy (fNIRS) for cognitive neuroscience. *Annals of the New York Academy of Sciences*, *1464*(1), 5–29. https://doi.org/10.1111/nyas.13948 (cit. on p. 64).

Pion-Tonachini, L., Kreutz-Delgado, K., & Makeig, S. (2019). ICLabel: An automated electroencephalographic independent component classifier, dataset, and website. *NeuroImage*, *198*, 181–197. https://doi.org/10.1016/j.neuroimage.2019.05.026 (cit. on pp. 301, 303).

Plaw, A., & Fricker, M. S. (2012). Tracking the Predators: Evaluating the US Drone Campaign in Pakistan [Publisher: Oxford University Press]. *International Studies Perspectives*, *13*(4), 344–365. Retrieved April 29, 2024, from https://www.jstor.org/stable/44218719 (cit. on p. 216).

Ploughman, M. (2008). Exercise is brain food: The effects of physical activity on cognitive function. *Developmental Neurorehabilitation*, *11*(3), 236–240. https://doi.org/10.1080/17518420801997007 (cit. on p. 52).

Plukaard, S., Huizinga, M., Krabbendam, L., & Jolles, J. (2015). Cognitive flexibility in healthy students is affected by fatigue: An experimental study. *Learning and Individual Differences*, *38*, 18–25. https://doi.org/10.1016/j.lindif.2015.01.003 (cit. on pp. 55–59).

Posner, M. I. (1980). Orienting of Attention. *Quarterly Journal of Experimental Psychology*, *32*(1), 3–25. https://doi.org/10.1080/00335558008248231 (cit. on p. 122).

Pouget, P. (2022a). Introduction ot the Field of Eye Movement Research. (Cit. on pp. 90–94).

Pouget, P. (2022b). Saccades. (Cit. on p. 92).

Powell, D. M. C., Spencer, M. B., & Petrie, K. J. (2011). Automated Collection of Fatigue Ratings at the Top of Descent: A Practical Commercial Airline Tool. *Aviation, Space, and Environmental Medicine*, *82*(11), 1037–1041. https://doi.org/10.3357/ASEM.3115.2011 (cit. on p. 61).

Pradeep, P., & Krishnamoorthy, S. (2019). The MOM of context-aware systems: A survey. *Computer Communications*, *137*, 44–69. https://doi.org/10.1016/j.comcom.2019.02.002 (cit. on pp. 31, 32).

Pritchett, A. R., Kim, S. Y., & Feigh, K. M. (2014). Measuring human-automation function allocation [Place: US Publisher: Sage Publications]. *Journal of Cognitive Engineering and Decision Making*, *8*(1), 52–77. https://doi.org/10.1177/1555343413490166 (cit. on pp. 46, 48)
Place: US Publisher: Sage Publications.

Qi, P., Ru, H., Gao, L., Zhang, X., Zhou, T., Tian, Y., Thakor, N., Bezerianos, A., Li, J., & Sun, Y. (2019). Neural Mechanisms of Mental Fatigue Revisited: New Insights from the Brain Connectome. *Engineering*, *5*(2), 276–286. https://doi.org/10.1016/j.eng.2018.11.025 (cit. on pp. 53, 54, 63).

Radovic, M. (2019, March). Tech Talk: Identify Drone Autonomy - Drone Industry Insights [Section: Market]. Retrieved March 14, 2024, from https://droneii.com/drone-autonomy (cit. on pp. 20, 21).

Rahman, M. M., Sarkar, A. K., Hossain, M. A., Hossain, M. S., Islam, M. R., Hossain, M. B., Quinn, J. M., & Moni, M. A. (2021). Recognition of human emotions using EEG signals: A review. *Computers in Biology and Medicine*, *136*, 104696. https://doi.org/10.1016/j.compbiomed.2021.104696 (cit. on p. 84).

Ratwani, R. M., Andrews, A. E., McCurry, M., Trafton, J. G., & Peterson, M. S. (2007). Using Peripheral Processing and Spatial Memory to Facilitate Task Resumption. *st ANNUAL MEETING*, 5 (cit. on p. 40).

Raufi, B., & Longo, L. (2022). An Evaluation of the EEG Alpha-to-Theta and Theta-to-Alpha Band Ratios as Indexes of Mental Workload [Publisher: Frontiers]. *Frontiers in Neuroinformatics*, *16*. https://doi.org/10.3389/fninf.2022.861967 (cit. on p. 84).

Reifman, J., Kumar, K., Khitrov, M. Y., Liu, J., & Ramakrishnan, S. (2018). PC-PVT 2.0: An updated platform for psychomotor vigilance task testing, analysis, prediction, and visualization. *Journal of Neuroscience Methods*, *304*, 39–45. https://doi.org/https://www.sciencedirect.com/science/article/pii/S0165027018301109 (cit. on pp. 294, 317).

Reinerman-Jones, L., Taylor, G., Sprouse, K., Barber, D., & Hudson, I. (2011). Adaptive Automation as a Task Switching and Task Congruence Challenge. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *55*(1), 197–201. https://doi.org/10.1177/1071181311551041 (cit. on pp. 40, 68).

Rezaei, Z. (2021). A review on image-based approaches for breast cancer detection, segmentation, and classification. *Expert Systems with Applications*, *182*, 115204. https://doi.org/10.1016/j.eswa.2021.115204 (cit. on p. 96).

Robert J. Hockey, G. (1997). Compensatory control in the regulation of human performance under stress and high workload: A cognitive-energetical framework. *Biological Psychology*, *45*(1), 73–93. https://doi.org/10.1016/S0301-0511(96)05223-4 (cit. on pp. 63, 110).

Roda, C. (2011). Human attention and its implications for human-computer interaction. In *Human attention in digital environments* (pp. 11–62). Cambridge University Press. https://doi.org/10.1017/CBO9780511974519.002 (cit. on p. 57).

Rodríguez-Fernández, V., Menéndez, H. D., & Camacho, D. (2017). Analysing temporal performance profiles of UAV operators using time series clustering. *Expert Systems with Applications*, *70*, 103–118. https://doi.org/10.1016/j.eswa.2016.10.044 (cit. on p. 113).

Rogers, R. D., & Monsell, S. (1995). Costs of a predictible switch between simple cognitive tasks. [Place: US Publisher: American Psychological Association]. *Journal of Experimental Psychology: General*, *124*(2), 207–231. https://doi.org/10.1037/0096-3445.124.2.207 (cit. on pp. 65–67, 69)
Place: US Publisher: American Psychological Association.

Roldán, J. J., Peña-Tapia, E., Martín-Barrio, A., Olivares-Méndez, M. A., Del Cerro, J., & Barrientos, A. (2017). Multi-Robot Interfaces and Operator Situational Awareness: Study of the Impact of Immersion and Prediction [Number: 8 Publisher: Multidisciplinary Digital Publishing Institute]. *Sensors*, *17*(8), 1720. https://doi.org/10.3390/s17081720 (cit. on pp. 40, 42, 44)
Number: 8 Publisher: Multidisciplinary Digital Publishing Institute.

Rönnlund, M., Lövdén, M., & Nilsson, L.-G. (2001). Adult Age Differences in Tower of Hanoi Performance: Influence From Demographic and Cognitive Variables. *Aging, Neuropsy-*

chology, and Cognition, *8*(4), 269–283. https://doi.org/10.1076/anec.8.4.269.5641 (cit. on p. 60).

Rousseeuw, P. J., & Driessen, K. V. (1999). A Fast Algorithm for the Minimum Covariance Determinant Estimator. *Technometrics*, *41*(3), 212–223. https://doi.org/10.1080/00401706.1999.10485670 (cit. on p. 157).

Roy, R., Hinss, M., Darmet, L., Ladouce, S., Jahanpour, E., Somon, B., Xu, X., Drougard, N., Dehais, F., & Lotte, F. (2022). Retrospective on the First Passive Brain-Computer Interface Competition on Cross-Session Workload Estimation. *Frontiers in Neuroergonomics*, *3*, 838342. https://doi.org/10.3389/fnrgo.2022.838342 (cit. on pp. 114, 268).

Roy, R. N., Bonnet, S., Charbonnier, S., & Campagne, A. (2013). Mental fatigue and working memory load estimation: Interaction and implications for EEG-based passive BCI [ISSN: 1558-4615]. *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 6607–6610. https://doi.org/10.1109/EMBC.2013.6611070 (cit. on pp. 112, 113, 159, 291).

Roy, R. N., Bovo, A., Gateau, T., Dehais, F., & Carvalho Chanel, C. P. (2016). Operator Engagement During Prolonged Simulated UAV Operation. *IFAC-PapersOnLine*, *49*(32), 171–176. https://doi.org/10.1016/j.ifacol.2016.12.209 (cit. on pp. 93, 113, 115).

Roy, R. N., Charbonnier, S., & Bonnet, S. (2014a). Detection of mental fatigue using an active BCI inspired signal processing chain. *IFAC Proceedings Volumes*, *47*(3), 2963–2968. https://doi.org/10.3182/20140824-6-ZA-1003.00897 (cit. on pp. 112–114).

Roy, R. N., Charbonnier, S., & Bonnet, S. (2014b). Eye blink characterization from frontal EEG electrodes using source separation and pattern recognition algorithms [Publisher: IOP Publishing]. *Biomedical Signal Processing and Control*, *47*(3), 2963–2968. https://doi.org/10.1088/1741-2560/13/2/026019/meta (cit. on pp. 113, 115)
Publisher: IOP Publishing.

Roy, R. N., Drougard, N., Gateau, T., Dehais, F., & Chanel, C. P. C. (2020). How Can Physiological Computing Benefit Human-Robot Interaction? *Robotics*, *9*(4), 100. https://doi.org/10.3390/robotics9040100 (cit. on pp. 44, 45)
Number: 4 Publisher: Multidisciplinary Digital Publishing Institute.

Rubinstein, J. S., Meyer, D. E., & Evans, J. E. (2001). Executive control of cognitive processes in task switching. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(4), 763–797. https://doi.org/10.1037/0096-1523.27.4.763 (cit. on pp. 39, 43, 69, 136).

Rucci, M., & Poletti, M. (2015). Control and Functions of Fixational Eye Movements. *Annual Review of Vision Science*, *1*(1), 499–518. https://doi.org/10.1146/annurev-vision-082114-035742 (cit. on p. 93).

Ruskin, K. J., Corvin, C., Rice, S., Richards, G., Winter, S. R., & Clebone Ruskin, A. (2021). Alarms, alerts, and warnings in air traffic control: An analysis of reports from the Aviation Safety Reporting System. *Transportation Research Interdisciplinary Perspectives*, *12*, 100502. https://doi.org/10.1016/j.trip.2021.100502 (cit. on p. 38).

Saint-Lot, J., Imbert, J.-P., & Dehais, F. (2020). Red Alert: A Cognitive Countermeasure to Mitigate Attentional Tunneling. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–6. https://doi.org/10.1145/3313831.3376709 (cit. on pp. 38, 39, 43, 46).

Salahuddin, L., Cho, J., Jeong, M. G., & Kim, D. (2007). Ultra Short Term Analysis of Heart Rate Variability for Monitoring Mental Stress in Mobile Settings [ISSN: 1558-4615]. *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 4656–4659. https://doi.org/10.1109/IEMBS.2007.4353378 (cit. on pp. 89, 157, 167).

Salomone, M., Audiffren, M., & Berberian, B. (2023). Overview and Perspectives on the Assessment and Mitigation of Cognitive Fatigue in Operational Settings. In *Cyber Physical Human Systems* (pp. 401–428). John Wiley & Sons, Ltd. (Cit. on p. 52).

Salvucci, D. D., Taatgen, N. A., & Borst, J. P. (2009). Toward a Unified Theory of the Multitasking Continuum: From Concurrent Performance to Task Switching, Interruption, and Resumption, 10 (cit. on pp. 59, 66).

Samn, S. W., & Perelli, L. P. (1982). *Estimating Aircrew Fatigue: A Technique with Application to Airlift Operations* (Technical Report No. SAM-TR-82-21). Brooks AFB, TX: USAF School of Aerospace Medicine. (Cit. on pp. 62, 113, 128, 130, 153, 161, 277).

Santana, R., Marti, L., & Zhang, M. (2019). GP-based methods for domain adaptation: using brain decoding across subjects as a test-case. *Genetic Programming and Evolvable Machines*, *20*(3), 385–411. https://doi.org/10.1007/s10710-019-09352-6 (cit. on p. 116).

Santiago-Espada, Y., Myer, R. R., Latorella, K. A., & Comstock, J. R. (2011, July). *The Multi-Attribute Task Battery II (MATB-II) Software for Human Performance and Workload Research: A User's Guide* (tech. rep. No. L-20031). Retrieved July 6, 2022, from https://ntrs.nasa.gov/citations/20110014456 (cit. on pp. 297, 318).
NTRS Author Affiliations: NASA Langley Research Center, Stinger Ghaffarian Technologies, Inc. (SGT, Inc.) NTRS Document ID: 20110014456 NTRS Research Center: Langley Research Center (LaRC).

Sarker, I. H. (2021). Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Computer Science*, *2*(3), 160. https://doi.org/10.1007/s42979-021-00592-x (cit. on pp. 95–98, 104).

Sauseng, P., Griesmayr, B., Freunberger, R., & Klimesch, W. (2010). Control mechanisms in working memory: A possible function of EEG theta oscillations. *Neuroscience & Biobehavioral Reviews*, *34*(7), 1015–1022. https://doi.org/10.1016/j.neubiorev.2009.12.006 (cit. on p. 79).

Sazgar, M., & Young, M. G. (2019). EEG Artifacts. In *Absolute Epilepsy and EEG Rotation Review: Essentials for Trainees* (pp. 149–162). Springer International Publishing. https://doi.org/10.1007/978-3-030-03511-2_8 (cit. on p. 80).

Schaefer, K. E., Chen, J. Y. C., Szalma, J. L., & Hancock, P. A. (2016). A Meta-Analysis of Factors Influencing the Development of Trust in Automation: Implications for Understanding Autonomy in Future Systems [Publisher: SAGE Publications Inc]. *Human Factors*, *58*(3), 377–400. https://doi.org/10.1177/0018720816634228 (cit. on p. 47).

Scheffers, M. K., & Coles, M. G. H. (2000). Performance monitoring in a confusing world: Error-related brain activity, judgments of response accuracy, and types of errors. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(1), 141–151. https://doi.org/10.1037/0096-1523.26.1.141 (cit. on p. 112).

Scheiman, T., Chappelle, W., & Sanford, E. (2017). U.S. Air Force Special Operations Command Remotely Piloted Aircraft Operator Fatigue Levels and Compensatory Strategies, 49 (cit. on pp. 26, 27, 70).

Schumacher, E. H., Seymour, T. L., Glass, J. M., Fencsik, D. E., Lauber, E. J., Kieras, D. E., & Meyer, D. E. (2001). Virtually Perfect Time Sharing in Dual-Task Performance: Uncorking the Central Cognitive Bottleneck. *Psychological Science*, *12*(2), 101–108. https://doi.org/10.1111/1467-9280.00318 (cit. on p. 66).

Scott, S. D., Mercier, S., Cummings, M. L., & Wang, E. (2006). Assisting Interruption Recovery in Supervisory Control of Multiple Uavs. *th ANNUAL MEETING*, 5 (cit. on pp. 40, 42, 46).

Semmens, R., Martelaro, N., Kaveti, P., Stent, S., & Ju, W. (2019). Is Now A Good Time?: An Empirical Study of Vehicle-Driver Communication Timing. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–12. https://doi.org/10.1145/3290605.3300867 (cit. on p. 38).

Senoussi, M., Verdiere, K. J., Bovo, A., Cavalho Chanel, C. P., Dehais, F., & Roy, R. N. (2017). Prestimulus antero-posterior EEG connectivity predicts performance in a UAV monitoring task. *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 1167–1172. https://doi.org/10.1109/SMC.2017.8122770 (cit. on p. 292).

Sexton, E. (2024, April). Asymmetrical warfare | Unconventional Tactics & Strategies | Britannica. Retrieved April 29, 2024, from https://www.britannica.com/topic/asymmetrical-warfare (cit. on p. 216).

Shaffer, F., & Ginsberg, J. P. (2017). An Overview of Heart Rate Variability Metrics and Norms. *Frontiers in Public Health*, *5*, 258. https://doi.org/10.3389/fpubh.2017.00258 (cit. on pp. 89, 167, 301).

Shaffer, F., McCraty, R., & Zerr, C. L. (2014). A healthy heart is not a metronome: an integrative review of the heart's anatomy and heart rate variability [Publisher: Frontiers]. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.01040 (cit. on p. 89).

Shah, A. J., Lampert, R., Goldberg, J., Veledar, E., Bremner, J. D., & Vaccarino, V. (2013). Posttraumatic Stress Disorder and Impaired Autonomic Modulation in Male Twins. *Biological Psychiatry*, *73*(11), 1103–1110. https://doi.org/10.1016/j.biopsych.2013.01.019 (cit. on p. 89).

Shah, S., Dey, D., Lovett, C., & Kapoor, A. (2017). AirSim: High-Fidelity Visual and Physical Simulation for Autonomous Vehicles. *Field and Service Robotics*. https://arxiv.org/abs/1705.05065 (cit. on p. 170).

Shahbakhti, M., Beiramvand, M., Rejer, I., Augustyniak, P., Broniec-Wójcik, A., Wierzchon, M., & Marozas, V. (2022). Simultaneous Eye Blink Characterization and Elimination From Low-Channel Prefrontal EEG Signals Enhances Driver Drowsiness Detection [Conference Name: IEEE Journal of Biomedical and Health Informatics]. *IEEE Journal of Biomedical and Health Informatics*, *26*(3), 1001–1012. https://doi.org/10.1109/JBHI.2021.3096984 (cit. on pp. 110, 112, 113, 115).

Shane, S. (2012). The Moral Case for Drones. *The New York Times*. Retrieved April 29, 2024, from https://www.nytimes.com/2012/07/15/sunday-review/the-moral-case-for-drones.html (cit. on p. 216).

Shappell, S., & Wiegmann, D. (2000). The Human Factors Analysis and Classification System–HFACS. *Publications*. https://commons.erau.edu/publication/737 (cit. on p. 26).

Shen, K.-Q., Li, X.-P., Ong, C.-J., Shao, S.-Y., & Wilder-Smith, E. P. (2008). EEG-based mental fatigue measurement using multi-class support vector machines with confidence estimate. *Clinical Neurophysiology*, *119*(7), 1524–1533. https://doi.org/10.1016/j.clinph.2008.03.012 (cit. on p. 110).

Sheng, Y., Huang, K., Wang, L., & Wei, P. (2018). The Optimal Number and Distribution of Channels in Mental Fatigue Classification Based on GA-SVM. *Proceedings of the 2018 5th International Conference on Bioinformatics Research and Applications*, 34–39. https://doi.org/10.1145/3309129.3309140 (cit. on pp. 112–115).

Sherbino, J., & Norman, G. (2021). Task Switching, Multitasking, and Errors: A Psychologic Perspective on the Impact of Interruptions. *Annals of Emergency Medicine*, *78*(3), 425–428. https://doi.org/10.1016/j.annemergmed.2021.07.120 (cit. on p. 70).

Shin, J., von Lühmann, A., Kim, D.-W., Mehnert, J., Hwang, H.-J., & Müller, K.-R. (2018). Simultaneous acquisition of EEG and NIRS during cognitive tasks for an open access dataset. *Scientific Data*, *5*(1), 180003. https://doi.org/http://www.nature.com/articles/sdata20183 (cit. on p. 291).

Signal, T. L., Ratieta, D., & Gander, P. H. (2008). Flight crew fatigue management in a more flexible regulatory environment: an overview of the New Zealand aviation industry. *Chronobiology International*, *25*(2), 373–388. https://doi.org/10.1080/07420520802118202 (cit. on p. 61).

Sikander, G., & Anwar, S. (2019). Driver Fatigue Detection Systems: A Review [Conference Name: IEEE Transactions on Intelligent Transportation Systems]. *IEEE Transactions on Intelligent Transportation Systems*, *20*(6), 2339–2352. https://doi.org/10.1109/TITS.2018.2868499 (cit. on pp. 44, 45, 113)
Conference Name: IEEE Transactions on Intelligent Transportation Systems.

Simon, M., Schmidt, E. A., Kincses, W. E., Fritzsche, M., Bruns, A., Aufmuth, C., Bogdan, M., Rosenstiel, W., & Schrauf, M. (2011). EEG Alpha Spindle Measures as Indicators of Driver Fatigue under Real Traffic Conditions. *Clinical Neurophysiology*, *122*(6), 1168–1178. https://doi.org/10.1016/j.clinph.2010.10.044 (cit. on p. 301).

Singh, A., Hussain, A. A., Lal, S., & Guesgen, H. W. (2021). A Comprehensive Review on Critical Issues and Possible Solutions of Motor Imagery Based Electroencephalography Brain-Computer Interface [Number: 6 Publisher: Multidisciplinary Digital Publishing Institute]. *Sensors*, *21*(6), 2173. https://doi.org/10.3390/s21062173 (cit. on p. 98).

Singh, G., Chanel, C. P. C., & Roy, R. N. (2021). Mental Workload Estimation Based on Physiological Features for Pilot-UAV Teaming Applications. *Frontiers in Human Neuroscience*, *15*. Retrieved April 12, 2023, from https://www.frontiersin.org/articles/10.3389/fnhum.2021.692878 (cit. on p. 154).

Singh, G., Roy, R. N., & Chanel, C. P. (2022). POMDP-Based Adaptive Interaction Through Physiological Computing. *Frontiers in Artificial Intelligence and Applications*. https://doi.org/10.3233/FAIA220188 (cit. on p. 115).

Siqi-Liu, A., & Egner, T. (2023). Task sets define boundaries of learned cognitive flexibility in list-wide proportion switch manipulations. [Publisher: American Psychological Association]. *Journal of Experimental Psychology: Human Perception and Performance* (cit. on p. 67).

Skaramagkas, V., Giannakakis, G., Ktistakis, E., Manousos, D., Karatzanis, I., Tachos, N., Tripoliti, E., Marias, K., Fotiadis, D. I., & Tsiknakis, M. (2023). Review of Eye Tracking Metrics Involved in Emotional and Cognitive Processes. *IEEE Reviews in Biomedical Engineering*, *16*, 260–277. https://doi.org/10.1109/RBME.2021.3066072 (cit. on pp. 92, 93).

Smallman, H. S., & John, M. S. (2003). Chex (Change History Explicit): New HCI Concepts for Change Awareness [Publisher: SAGE Publications Inc]. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *47*(3), 528–532. https://doi.org/10.1177/154193120304700358 (cit. on pp. 40, 42)
Publisher: SAGE Publications Inc.

Smith, M. E., Gevins, A., Brown, H., Karnik, A., & Du, R. (2001). Monitoring Task Loading with Multivariate EEG Measures during Complex Forms of Human-Computer Interaction [Publisher: SAGE Publications Inc]. *Human Factors*, *43*(3), 366–380. https://doi.org/10.1518/001872001775898287 (cit. on p. 318).

Smith, M. R., Chai, R., Nguyen, H. T., Marcora, S. M., & Coutts, A. J. (2019). Comparing the Effects of Three Cognitive Tasks on Indicators of Mental Fatigue. *The Journal of Psychology*, *153*(8), 759–783. https://doi.org/10.1080/00223980.2019.1611530 (cit. on p. 301).

Smolders, K. C., & de Kort, Y. A. (2014). Bright light and mental fatigue: Effects on alertness, vitality, performance and physiological arousal. *Journal of Environmental Psychology*, *39*, 77–91. https://doi.org/10.1016/j.jenvp.2013.12.010 (cit. on p. 37).

Sörnmo, L., & Laguna, P. (2005, June). *Bioelectrical Signal Processing in Cardiac and Neurological Applications* [Google-Books-ID: RQv7tFFXYyIC]. Academic Press. (Cit. on p. 80).

Spector, A., & Biederman, I. (1976). Mental Set and Mental Shift Revisited [Publisher: University of Illinois Press]. *The American Journal of Psychology*, *89*(4), 669–679. https://doi.org/10.2307/1421465 (cit. on p. 66).

Sporns, O. (2012, August). *Discovering the Human Connectome*. MIT Press. (Cit. on p. 54).

Squire, P. N., & Parasuraman, R. (2010). Effects of automation and task load on task switching during human supervision of multiple semi-autonomous robots in a dynamic environment. *Ergonomics*, *53*(8), 951–961. https://doi.org/10.1080/00140139.2010.489969 (cit. on pp. 27, 59)
Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/00140139.2010.489969.

Squire, P., Trafton, G., & Parasuraman, R. (2006). Human Control of Multiple Unmanned Vehicles: Effects of Interface Type on Execution and Task Switching Times, 7 (cit. on pp. 27, 40, 68, 70).

Squires, N. K., Squires, K. C., & Hillyard, S. A. (1975). Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalography and Clinical Neurophysiology*, *38*(4), 387–401. https://doi.org/10.1016/0013-4694(75)90263-1 (cit. on p. 83).

Srivastava, R., Kumar, S., & Kumar, B. (2023, January). 7 - Classification model of machine learning for medical data analysis. In T. Goswami & G. R. Sinha (Eds.), *Statistical Modeling in Machine Learning* (pp. 111–132). Academic Press. https://doi.org/10.1016/B978-0-323-91776-6.00017-8 (cit. on pp. 100, 103).

Stanton, N., Chambers, P., & Piggott, J. (2001). Situational awareness and safety. *Safety Science*, *39*(3), 189–204. https://doi.org/10.1016/S0925-7535(01)00010-8 (cit. on p. 60).

Stokes, A., & Wickens, C. (1990). Display technoloy: Human Factors Conepts. *Society of Automotive Engineers Inc* (cit. on p. 35).

Storr, W. (2022, January). Voltmeter - Measurement of Voltage. Retrieved May 31, 2024, from https://www.electronics-tutorials.ws/blog/voltmeter.html (cit. on p. 75).

Straetmans, L., Holtze, B., Debener, S., Jaeger, M., & Mirkovic, B. (2022). Neural tracking to go: auditory attention decoding and saliency detection with mobile EEG [Publisher: IOP Publishing]. *Journal of Neural Engineering*, *18*(6), 066054. https://doi.org/10.1088/1741-2552/ac42b5 (cit. on p. 76).

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. [Publisher: Psychological Review Company]. *Journal of experimental psychology*, *18*(6), 643 (cit. on pp. 110, 111).

Subha, D. P., Joseph, P. K., Acharya U, R., & Lim, C. M. (2010). EEG signal analysis: a survey. *Journal of Medical Systems*, *34*(2), 195–212. https://doi.org/10.1007/s10916-008-9231-z (cit. on p. 116).

Sun, Y., Lim, J., Kwok, K., & Bezerianos, A. (2014). Functional cortical connectivity analysis of mental fatigue unmasks hemispheric asymmetry and changes in small-world networks. *Brain and Cognition*, *85*, 220–230. https://doi.org/10.1016/j.bandc.2013.12.011 (cit. on pp. 57, 58, 108, 113, 114).

Swanson, R. A., Morton, M. M., Sagar, S. M., & Sharp, F. R. (1992). Sensory stimulation induces local cerebral glycogenolysis: Demonstration by autoradiography. *Neuroscience*, *51*(2), 451–461. https://doi.org/10.1016/0306-4522(92)90329-Z (cit. on p. 53).

Takasuna, M. (2013). Tsuruko Haraguchi (1886 - 1915). Retrieved November 22, 2023, from https://www.apadivisions.org/division-35/about/heritage/tsuruko-haraguchi-biography (cit. on p. 49).

Tallon-Baudry, C., Kreiter, A., & Bertrand, O. (1999). Sustained and transient oscillatory responses in the gamma and beta bands in a visual short-term memory task in humans. *Visual Neuroscience*, *16*(3), 449–459. https://doi.org/10.1017/s0952523899163065 (cit. on p. 79).

Tallon-Baudry, C., Bertrand, O., Delpuech, C., & Pernier, J. (1996). Stimulus Specificity of Phase-Locked and Non-Phase-Locked 40 Hz Visual Responses in Human [Publisher: Society for Neuroscience Section: Articles]. *Journal of Neuroscience*, *16*(13), 4240–4249. https://doi.org/10.1523/JNEUROSCI.16-13-04240.1996 (cit. on p. 83).

Tamburro, G., Croce, P., Zappasodi, F., & Comani, S. (2021). Automated Detection and Removal of Cardiac and Pulse Interferences from Neonatal EEG Signals. *Sensors (Basel, Switzerland)*, *21*(19), 6364. https://doi.org/10.3390/s21196364 (cit. on p. 80).

Tao, D., Tan, H., Wang, H., Zhang, X., Qu, X., & Zhang, T. (2019). A Systematic Review of Physiological Measures of Mental Workload [Number: 15 Publisher: Multidisciplinary Digital Publishing Institute]. *International Journal of Environmental Research and Public Health*, *16*(15), 2716. https://doi.org/10.3390/ijerph16152716 (cit. on p. 115).

Taylor, G. S., Reinerman-Jones, L. E., Szalma, J. L., Mouloua, M., & Hancock, P. A. (2013). What to Automate: Addressing the Multidimensionality of Cognitive Resources Through System Design [Publisher: SAGE Publications]. *Journal of Cognitive Engineering and Decision Making*, *7*(4), 311–329. https://doi.org/10.1177/1555343413495396 (cit. on p. 43).

Theeuwes, J. (1991). Exogenous and endogenous control of attention: The effect of visual onsets and offsets. *Perception & Psychophysics*, *49*(1), 83–90. https://doi.org/10.3758/BF03211619 (cit. on p. 57).

Thompson, W. T., Lopez, N., Hickey, P., DaLuz, C., & Cadwell, J. L. (2006). DTIC ADA443145: Effects of Shift Work and Sustained Operations: Operator Performance in Remotely Piloted Aircraft (OP-REPAIR). Retrieved January 13, 2023, from http://archive.org/details/DTIC_ADA443145 (cit. on p. 26).

Ting, C.-H., Mahfouf, M., Nassef, A., Linkens, D. A., Panoutsos, G., Nickel, P., Roberts, A. C., & Hockey, G. R. J. (2010). Real-Time Adaptive Automation System Based on Identification of Operator Functional State in Simulated Process Control Operations [Conference Name: IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans]. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, *40*(2), 251–262. https://doi.org/10.1109/TSMCA.2009.2035301 (cit. on pp. 112, 113, 115).

Tingstad, A., Goldfeld, D. A., Menthe, L., Guffey, R. A., Haldeman, Z., Langeland, K. S., Cordova, A., Waina, E. M., & Gintautas, B. (2021). *Assessing the Value of Intelligence Collected by US Air Force Airborne Intelligence, Surveillance, and Reconnaissance Platforms*. (Cit. on p. 26).

Todi, K., Bailly, G., Leiva, L., & Oulasvirta, A. (2021). Adapting User Interfaces with Model-based Reinforcement Learning. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–13. https://doi.org/10.1145/3411764.3445497 (cit. on p. 44).

Tohka, J., & van Gils, M. (2021). Evaluation of machine learning algorithms for Health and Wellness applications: a tutorial [arXiv:2008.13690 [cs, stat]]. *Computers in Biology and Medicine*, *132*, 104324. https://doi.org/10.1016/j.compbiomed.2021.104324 (cit. on p. 103)
Comment: To be published in Computers in Biology and Medicine.

Tran, Y., Craig, A., Craig, R., Chai, R., & Nguyen, H. (2020). The influence of mental fatigue on brain activity: Evidence from a systematic review with meta-analyses. *Psychophysiology*, *57*(5), e13554. https://doi.org/10.1111/psyp.13554 (cit. on pp. 84, 112, 113).

Trejo, L. J., Kubitz, K., Rosipal, R., Kochavi, R. L., & Montgomery, L. D. (2015). EEG-Based Estimation and Classification of Mental Fatigue [Number: 05 Publisher: Scientific Research Publishing]. *Psychology*, *06*(05), 572–589. https://doi.org/10.4236/psych.2015.65055 (cit. on pp. 112, 113)
Number: 5 Publisher: Scientific Research Publishing.

Tvaryanas, A. P., Platte, W., Swigart, C., Colebank, J., & Miller, N. L. (2008). A Resurvey of Shift Work-Related Fatigue in MQ-1 Predator Unmanned Aircraft System Crewmembers [Number: 1 Publisher: Nature Publishing Group]. *npj Digital Medicine*, *4*(1), 1–5. https://doi.org/10.1038/s41746-021-00415-6 (cit. on pp. 26, 204)
Number: 1 Publisher: Nature Publishing Group.

Urigüen, J. A., & Garcia-Zapirain, B. (2015). EEG artifact removal—state-of-the-art and guidelines. *Journal of Neural Engineering*, *12*(3), 031001. https://doi.org/10.1088/1741-2560/12/3/031001 (cit. on pp. 80, 82, 83, 167, 182).

USAspending. (2022). *Federal Agency Spending Profiles | USAspending* (tech. rep.). Retrieved November 23, 2023, from https://usaspending.gov/agency (cit. on p. 50).

Vagia, M., Transeth, A. A., & Fjerdingen, S. A. (2016). A literature review on the levels of automation during the years. What are the different taxonomies that have been proposed? *Applied Ergonomics*, *53*, 190–202. https://doi.org/10.1016/j.apergo.2015.09.013 (cit. on p. 20).

Van Benthem, K., & Herdman, C. M. (2016). Cognitive Factors Mediate the Relation Between Age and Flight Path Maintenance in General Aviation [Publisher: Hogrefe Publishing]. *Aviation Psychology and Applied Human Factors*, *6*(2), 81–90. https://doi.org/10.1027/2192-0923/a000102 (cit. on p. 70).

Van Cutsem, J., De Pauw, K., Marcora, S., Meeusen, R., & Roelands, B. (2018). A caffeine-maltodextrin mouth rinse counters mental fatigue. *Psychopharmacology*, *235*(4), 947–958. https://doi.org/10.1007/s00213-017-4809-0 (cit. on p. 37).

van de Merwe, F., Kähler, N., & Securius, P. (2016). Crew-centred Design of Ships – The CyClaDes Project. *Transportation Research Procedia*, *14*, 1611–1620. https://doi.org/10.1016/j.trpro.2016.05.126 (cit. on p. 46).

van der Linden, D. (2011). The urge to stop: The cognitive and biological nature of acute mental fatigue. In *Cognitive fatigue: Multidisciplinary perspectives on current research and future applications* (pp. 149–164). American Psychological Association. https://doi.org/10.1037/12343-007 (cit. on pp. 63, 110).

van der Linden, D., Frese, M., & Meijman, T. F. (2003). Mental fatigue and the control of cognitive processes: effects on perseveration and planning. *Acta Psychologica*, *113*(1), 45–65. https://doi.org/10.1016/S0001-6918(02)00150-6 (cit. on pp. 51, 55–57, 59).

van der Linden, D., Frese, M., & Sonnentag, S. (2003). The Impact of Mental Fatigue on Exploration in a Complex Computer Task: Rigidity and Loss of Systematic Strategies. *Human*

Factors: The Journal of the Human Factors and Ergonomics Society, *45*(3), 483–494. https://doi.org/10.1518/hfes.45.3.483.27256 (cit. on pp. 55–58, 60).

Vandierendonck, A., Liefooghe, B., & Verbruggen, F. (2010). Task switching: Interplay of reconfiguration and interference control. *Psychological Bulletin*, *136*(4), 601–626. https://doi.org/10.1037/a0019791 (cit. on pp. 39, 65, 66, 68, 69).

Veale, J. F. (2014). Edinburgh Handedness Inventory - Short Form: a revised version based on confirmatory factor analysis. *Laterality*, *19*(2), 164–177. https://doi.org/10.1080/1357650X.2013.783045 (cit. on pp. 127, 130, 274, 292).

Verdière, K. J., Roy, R. N., & Dehais, F. (2018). Detecting Pilot's Engagement Using fNIRS Connectivity Features in an Automated vs. Manual Landing Scenario. *Frontiers in Human Neuroscience*, *12*, 6. https://doi.org/http://journal.frontiersin.org/article/10.3389/fnhum.2018.00006/full (cit. on pp. 292, 297).

Vidal, T., & Schiffer, M. (2020, July). Born-Again Tree Ensembles. In H. D. III & A. Singh (Eds.), *Proceedings of the 37th International Conference on Machine Learning* (pp. 9743–9753, Vol. 119). PMLR. https://proceedings.mlr.press/v119/vidal20a.html (cit. on p. 100).

Vijaya K, K., Deepika, D., & Lei, D. (2014). EEG Correlates of Fluctuation in Cognitive Performance in an Air Traffic Control Task, 30 (cit. on pp. 55–57, 60, 112, 113).

Waggoner, P. D. (2020). Unsupervised Machine Learning for Clustering in Political and Social Research [ISBN: 9781108883955 9781108793384 Publisher: Cambridge University Press]. *Elements in Quantitative and Computational Methods for the Social Sciences*. https://doi.org/10.1017/9781108883955 (cit. on p. 96).

Wakita, T., Ozawa, K., Miyajima, C., Igarashi, K., Itou, K., Takeda, K., & Itakura, F. (2006). Driver Identification Using Driving Behavior Signals [Publisher: The Institute of Electronics, Information and Communication Engineers]. *IEICE TRANSACTIONS on Information and Systems*, *E89-D*(3), 1188–1194. https://doi.org/10.1093/ietisy/e89-d.3.1188 (cit. on p. 45)
Publisher: The Institute of Electronics, Information and Communication Engineers.

Wan, Z., Yang, R., Huang, M., Zeng, N., & Liu, X. (2021). A review on transfer learning in EEG signal analysis. *Neurocomputing*, *421*, 1–14. https://doi.org/10.1016/j.neucom.2020.09.017 (cit. on p. 116).

Wang, C., Trongnetrpunya, A., Samuel, I. B. H., Ding, M., & Kluger, B. M. (2016). Compensatory Neural Activity in Response to Cognitive Fatigue. *The Journal of Neuroscience*, *36*(14), 3919–3924. https://doi.org/10.1523/JNEUROSCI.3652-15.2016 (cit. on pp. 113–115).

Wang, F., Wang, H., & Fu, R. (2018). Real-Time ECG-Based Detection of Fatigue Driving Using Sample Entropy [Number: 3 Publisher: Multidisciplinary Digital Publishing Institute]. *Entropy*, *20*(3), 196. https://doi.org/10.3390/e20030196 (cit. on pp. 112–115, 207).

Wang, J., & Lewis, M. (2007). Human control for cooperating robot teams. *Proceedings of the ACM/IEEE international conference on Human-robot interaction*, 9–16. https://doi.org/10.1145/1228716.1228719 (cit. on p. 70).

Wang, X., & Xu, C. (2016). Driver drowsiness detection based on non-intrusive metrics considering individual specifics. *Accident Analysis & Prevention*, *95*, 350–357. https://doi.org/10.1016/j.aap.2015.09.002 (cit. on pp. 93, 113, 115).

Waraich, Q. R. (, Mazzuchi, T. A., Sarkani, S., & Rico, D. F. (2013). Minimizing Human Factors Mishaps in Unmanned Aircraft Systems. *Ergonomics in Design: The Quarterly of Human Factors Applications*, *21*(1), 25–32. https://doi.org/10.1177/1064804612463215 (cit. on pp. 25–27, 202).

Wascher, E., Rasch, B., Sänger, J., Hoffmann, S., Schneider, D., Rinkenauer, G., Heuer, H., & Gutberlet, I. (2014). Frontal theta activity reflects distinct aspects of mental fatigue. *Biological Psychology*, *96*, 57–65. https://doi.org/10.1016/j.biopsycho.2013.11.010 (cit. on pp. 79, 112, 113, 174, 182).

Waser, M., & Garn, H. (2013). Removing cardiac interference from the electroencephalogram using a modified Pan-Tompkins algorithm and linear regression [ISSN: 1558-4615]. *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2028–2031. https://doi.org/10.1109/EMBC.2013.6609929 (cit. on p. 82).

Welch, J. (2011). An Evidence-Based Approach to Reduce Nuisance Alarms and Alarm Fatigue. *Biomedical Instrumentation & Technology*, *45*(s1), 46–52. https://doi.org/10.2345/0899-8205-45.s1.46 (cit. on p. 39).

West, R., & Alain, C. (2000). Evidence for the Transient Nature of a Neural System Supporting Goal-directed Action. *Cerebral Cortex*, *10*(8), 748–752. https://doi.org/10.1093/cercor/10.8.748 (cit. on p. 63).

Wickens, C. D., & McCarley, J. S. (2008). *Applied attention theory*. CRC Press. (Cit. on p. 58).

Wickens, C. D., & McCarley, J. S. (2019, April). *Applied Attention Theory*. CRC Press. https://doi.org/10.1201/9780429059261 (cit. on pp. 38, 58).

Williams, K. (2004). A Summary of Unmanned Aircraft Accident/Incident Data: Human Factors Implications, 17 (cit. on p. 27).

Williams, K. (2006). Human Factors Implications of Unmanned Aircraft Accidents: Flight Control Problems. *Advances in Human Performance and Cognitive Engineering Research*, *7*. https://doi.org/10.1016/S1479-3601(05)07008-6 (cit. on p. 27).

Wohleber, R. W., Matthews, G., Lin, J., Szalma, J. L., Calhoun, G. L., Funke, G. J., Chiu, C.-Y. P., & Ruff, H. A. (2019). Vigilance and Automation Dependence in Operation of Multiple Unmanned Aerial Systems (UAS): A Simulation Study [Publisher: SAGE Publications Inc]. *Human Factors*, *61*(3), 488–505. https://doi.org/10.1177/0018720818799468 (cit. on p. 26).

Wong, I. (2021). Center for Work and Fatigue Research. cdc.gov/niosh/topics/fatigue/center.html (cit. on p. 50).

Wood, C. C., & Jennings, J. R. (1976). Speed-accuracy tradeoff functions in choice reaction time: Experimental designs and computational procedures. *Perception & Psychophysics*, *19*(1), 92–102. https://doi.org/10.3758/BF03199392 (cit. on pp. 62, 182, 209).

Wriessnegger, S. C., Raggam, P., Kostoglou, K., & Müller-Putz, G. R. (2021). Mental State Detection Using Riemannian Geometry on Electroencephalogram Brain Signals [Publisher: Frontiers]. *Frontiers in Human Neuroscience*, *15*. https://doi.org/10.3389/fnhum.2021.746081 (cit. on pp. 113, 114).

Wu, J., Chen, X.-Y., Zhang, H., Xiong, L.-D., Lei, H., & Deng, S.-H. (2019). Hyperparameter Optimization for Machine Learning Models Based on Bayesian Optimizationb. *Journal of Electronic Science and Technology*, *17*(1), 26–40. https://doi.org/10.11989/JEST.1674-862X.80904120 (cit. on pp. 102, 103).

Wu, X., Wanyan, X., & Zhuang, D. (2015). Pilot's visual attention allocation modeling under fatigue [Publisher: IOS Press]. *Technology and Health Care*, *23*(s2), S373–S381. https://doi.org/10.3233/THC-150974 (cit. on pp. 93, 113, 115).

Xenou, K., Chalkiadakis, G., & Afantenos, S. (2019). Deep Reinforcement Learning in Strategic Board Game Environments. In M. Slavkovik (Ed.), *Multi-Agent Systems* (pp. 233–248). Springer International Publishing. https://doi.org/10.1007/978-3-030-14174-5_16 (cit. on p. 97).

Xing, Y., Lv, C., Cao, D., & Hang, P. (2021). Toward human-vehicle collaboration: Review and perspectives on human-centered collaborative automated driving. *Transportation Research Part C: Emerging Technologies*, *128*, 103199. https://doi.org/10.1016/j.trc.2021.103199 (cit. on p. 46).

Yamano, E., Tanaka, M., Ishii, A., Tsuruoka, N., Abe, K., & Watanabe, Y. (2013). Effects of chicken essence on recovery from mental fatigue in healthy males. *Medical Science Monitor : International Medical Journal of Experimental and Clinical Research*, *19*, 540–547. https://doi.org/10.12659/MSM.883971 (cit. on p. 37).

Yerushalmy, J. (2023). MQ-9 Reaper: what is the US drone that collided with a Russian jet and how is it used? *The Guardian*. Retrieved March 22, 2024, from https://www.theguardian.com/world/2023/mar/15/mq-9-reaper-what-is-the-us-drone-that-collided-with-a-russian-jet-and-how-is-it-used (cit. on p. 22).

Yger, F., Berar, M., & Lotte, F. (2017). Riemannian Approaches in Brain-Computer Interfaces: A Review. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *25*(10), 1753–1762. https://doi.org/10.1109/TNSRE.2016.2627016 (cit. on p. 114).

Ying, X. (2019). An Overview of Overfitting and its Solutions [Publisher: IOP Publishing]. *Journal of Physics: Conference Series*, *1168*(2), 022022. https://doi.org/10.1088/1742-6596/1168/2/022022 (cit. on p. 104).

Young, M. S., & Stanton, N. A. (2002a). Attention and automation: New perspectives on mental underload and performance. *Theoretical Issues in Ergonomics Science*, *3*(2), 178–194. https://doi.org/10.1080/14639220210123789 (cit. on pp. 34, 35).

Young, M. S., & Stanton, N. A. (2002b). Malleable attentional resources theory: a new explanation for the effects of mental underload on performance. *Human Factors*, *44*(3), 365–375. https://doi.org/10.1518/0018720024497709 (cit. on p. 34).

Yu, D., Antonik, C. W., Webber, F., Watz, E., & Bennett, W. (2021). Correction to: Multi-modal physiological sensing approach for distinguishing high workload events in remotely piloted aircraft simulation. *Human-Intelligent Systems Integration*, *3*(2), 201–211. https://doi.org/10.1007/s42454-021-00033-3 (cit. on pp. 112, 113).

Yu, S., Mückschel, M., & Beste, C. (2021). Event-related synchronization/desynchronization and functional neuroanatomical regions associated with fatigue effects on cognitive flexibility. *Journal of Neurophysiology*, *126*(2), 383–397. https://doi.org/10.1152/jn.00228.2021 (cit. on pp. 55–59).

Yuval-Greenberg, S., Tomer, O., Keren, A. S., Nelken, I., & Deouell, L. Y. (2008). Transient induced gamma-band response in EEG as a manifestation of miniature saccades. *Neuron*, *58*(3), 429–441. https://doi.org/10.1016/j.neuron.2008.03.027 (cit. on p. 79).

Zander, Brönstrup, Lorenz, & Krol. (2014, March). Towards BCI-Based Implicit Control in Human–Computer Interaction. https://doi.org/10.1007/978-1-4471-6392-3_4 (cit. on pp. 108, 109).

Zander, T. O., Shetty, K., Lorenz, R., Leff, D. R., Krol, L. R., Darzi, A. W., Gramann, K., & Yang, G.-Z. (2017). Automated Task Load Detection with Electroencephalography: Towards Passive Brain–Computer Interfacing in Robotic Surgery. *Journal of Medical Robotics Research*, *02*(01), 1750003. https://doi.org/10.1142/S2424905X17500039 (cit. on pp. 108, 110).

Zeng, H., Li, X., Borghini, G., Zhao, Y., Aricò, P., Di Flumeri, G., Sciaraffa, N., Zakaria, W., Kong, W., & Babiloni, F. (2021). An EEG-Based Transfer Learning Method for Cross-Subject Fatigue Mental State Prediction. *Sensors*, *21*(7), 2369. https://doi.org/https://www.mdpi.com/1424-8220/21/7/2369 (cit. on p. 292).

Zhang, X., Sun, Y., Zhang, Y., & Su, S. (2021). Multi-agent modelling and situational awareness analysis of human-computer interaction in the aircraft cockpit: A case study [Conference Name: IEEE Sensors Journal]. *Simulation Modelling Practice and Theory*, *111*(15), 16826–16838. https://doi.org/10.1016/j.simpat.2021.102355 (cit. on p. 40).

Zhao, C., Zhao, M., Liu, J., & Zheng, C. (2012). Electroencephalogram and electrocardiograph assessment of mental fatigue in a driving simulator. *Accident Analysis & Prevention*, *45*, 83–90. https://doi.org/10.1016/j.aap.2011.11.019 (cit. on pp. 113, 115).

Zhao, C., Zheng, C., Zhao, M., Liu, J., & Tu, Y. (2011). Automatic classification of driving mental fatigue with EEG by wavelet packet energy and KPCA-SVM. *International Journal of Innovative Computing, Information and Control*, *7* (cit. on p. 159).

Zhao, W., Van Someren, E. J. W., Li, C., Chen, X., Gui, W., Tian, Y., Liu, Y., & Lei, X. (2021). EEG spectral analysis in insomnia disorder: A systematic review and meta-analysis. *Sleep Medicine Reviews*, *59*, 101457. https://doi.org/10.1016/j.smrv.2021.101457 (cit. on p. 84).

Zheng, W.-L., & Lu, B.-L. (2015). Investigating Critical Frequency Bands and Channels for EEG-Based Emotion Recognition with Deep Neural Networks. *IEEE Transactions on Autonomous Mental Development*, *7*(3), 162–175. https://doi.org/10.1109/TAMD.2015.2431497 (cit. on p. 292).

Zhu, C., & Han, J. (2022). The Higher, More Complicated: The Neural Mechanism of Hierarchical Task Switching on Prefrontal Cortex [Number: 5 Publisher: Multidisciplinary Digital Publishing Institute]. *Brain Sciences*, *12*(5), 645. https://doi.org/10.3390/brainsci12050645 (cit. on p. 68)
Number: 5 Publisher: Multidisciplinary Digital Publishing Institute.

Part V
Scientific Output & Appendices

Chapter 13

Scientific Output

V.13.1 Journal Articles

- **Published**: Hinss, M. F., Brock, A. M., & Roy, R. N. (2022a). Cognitive effects of prolonged continuous human-machine interaction: The case for mental state-based adaptive interfaces. *Frontiers in Neuroergonomics*, *3*. https://doi.org/10.3389/fnrgo.2022.935092

- **Preregistered**: Hinss, M. F., Brock, A. M., & Roy, R. N. (2023a). The double task-switching protocol: An investigation into the effects of similarity and conflict on cognitive flexibility in the context of mental fatigue (G. G. De La Torre, Ed.). *PLOS ONE*, *18*(2), e0279021. https://doi.org/10.1371/journal.pone.0279021

- **Accepted**: Hinss, M. F., Brock, A. M., & Roy, R. N. (2024). Double Task Switching: an investigation into the effects of similarity and task-rule congruency on cognitive flexibility. *PLOS ONE*

- **Accepted**: Hinss, M. F., Jahanpour, E. S., Brock, A. M., & Roy, R. N. (2024). A passive Brain-Computer Interface for operator mental fatigue estimation in monotonous surveillance operations: time-on-task and performance labeling issues. *Journal of Neural Engineering*

V.13.2 Conference Articles

- **ÉPIQUE 2023 - Paris, France**: Hinss, M. F., Brock, A., & Roy, R. (2023b). Etude et caractérisation de la fatigue mentale et de la flexibilité cognitive dans le contexte des opérations drone longue endurance, 100. https://doi.org/hal-04362545v1

- **Human-Robot Interaction 2024 - Boulder, United States**: Hinss, M. F., Vitale, V. M., Phan, N. T., Roy, R. N., & Brock, A. M. (2024). UASOS: An Experimental Environment For Assessing Mental Fatigue & Cognitive Flexibility During Drone Operations. https://doi.org/https://doi.org/10.1145/3610977.3637484

- **Graz BCI Meeting 2024 - Graz, Austria (Accepted)**: Hinss, M. F., Vitale, V. M., Brock, A. M., & Roy, R. N. (2024). EEG-based performance estimation during a realistic drone piloting task. https://doi.org/hal-04596992v1

V.13.3 Conference Abstracts

- **CORTICO Days 2022 - Autrans, France**: Hinss, M. F., Brock, A. M., & Roy, R. N. (2022b). Is it time for passive Brain Computer Interfaces in UAV Operations. https://doi.org/https://enac.hal.science/hal-04638704

- **ICCAS 2022 - Toulouse, France**: Hinss, M., Jahanpour, E., Brock, A., Imbert, J.-P., & Roy, R. (2022). Long-endurance UAS operations translated to neuroergonomics. *ICCAS22-International Conference on Cognitive Aircraft Systems*. https://doi.org/hal-04149728

- **CORTICO Days 2023 - Paris, France**: Chenot, Q., Hinss, M. F., Roy, R. N., & Scannella, S. (2023). EEG-based Machine Learning to detect executive functions engagement during complex tasks. https://doi.org/https://enac.hal.science/hal-04638715

- **BCI Meeting 2023 - Brussels, Belgium**: Hinss, M. F., Jahanpour, E. S., Brock, A. M., & Roy, R. N. (2023). Labeling mental fatigue for passive BCI applications: Accuracy vs applicability tradeoff. *BCI meeting*. https://doi.org/hal-04146193

V.13.4 Internship

- **Journal Article**: Hinss, M. F., Jahanpour, E. S., Somon, B., Pluchon, L., Dehais, F., & Roy, R. N. (2023). Open multi-session and multi-task EEG cognitive Dataset for passive brain-computer Interface Applications. *Scientific Data*, *10*(1), 85. https://doi.org/10.1038/s41597-022-01898-y

- **Journal Article**: Roy, R., Hinss, M., Darmet, L., Ladouce, S., Jahanpour, E., Somon, B., Xu, X., Drougard, N., Dehais, F., & Lotte, F. (2022). Retrospective on the First Passive Brain-Computer Interface Competition on Cross-Session Workload Estimation. *Frontiers in Neuroergonomics*, *3*, 838342. https://doi.org/10.3389/fnrgo.2022.838342

- **Dataset**: Hinss, M. F., Darmet, L., Somon, B., Jahanpour, E., Lotte, F., Ladouce, S., & Roy, R. N. (2021b). An EEG dataset for cross-session mental workload estimation: Passive BCI competition of the Neuroergonomics Conference 2021 (Version 2)[Data set]. *Neuroergonomics Conference, Munich, Germany. Zenodo*. https://doi.org/10.5281/zenodo.5055046

- **Dataset**: Hinss, M. F., Jahanpour, E. S., Somon, B., Pluchon, L., Dehais, F., & Roy, R. N. (2022a, July). COG-BCI database: A multi-session and multi-task EEG cognitive dataset for passive brain-computer interfaces. https://doi.org/10.5281/zenodo.6923638
  Type: dataset

- **IEEE virtual Conference on Neural Engineering 2021 - Virtual**: Hinss, M. F., Somon, B., Dehais, F., & Roy, R. N. (2021). Open EEG Datasets for Passive Brain-Computer Interface Applications: Lacks and Perspectives. *10th International IEEE/EMBS Conference on Neural Engineering*, 686–689. https://doi.org/10.1109/NER49283.2021.9441214

Chapter 14

Appendices

V.14.1 Picture Sources

Figure	Image	Source	Figure	Image	Source
2.1	RQ-4	Link	4.3	Action Potential	own (BioRender)
2.1	MQ-9	Link	4.4	Meninies	own (BioRender)
2.1	Watchkeeper	Link	4.5	EEG	own Image
2.1	ScanEagle	Link	4.5	EEG signal	own Image
2.1	Raven	Link	4.5	ECG	Link
2.1	Black Hornet	Link	4.5	ECG signal	own Image
2.1	MQ-9	Link	4.5	Eye-Tracking	Link
2.1	MQ-9 Size	Link	4.5	Eye-Tracking data	own Image
2.3	MQ-9 GCS	Link	4.5	fMRI	Link
2.4	Pup	Link	4.5	fMRI data	Link
2.4	Pup Cockpit	Link	4.5	MEG	Link
2.4	P-38	Link	4.5	MEG signal	Link
2.4	P-38 Cockpit	Link	4.5	fNIRS	Link
2.4	L1011	Link	4.5	fNIRS data	Link
2.4	L1011 Cockpit	Link	4.6	10-20 System	own (BioRender)
2.4	A320	Link	4.7	Frequency bands	own (BioRender)
2.4	A320 Cockpit	Link	4.8	Filters	de Cheveigne, 2019
2.4	MQ-9	Link	4.12	Heart	own (BioRender)
2.4	MQ-9 Cockpit	Link	4.13	ECG	own (BioRender)
2.4	Ukrainian Drone	Link	4.14	Retina anatomy	own (BioRender)
2.4	Drone VR	Link	4.15	Eye-tracking	own (BioRender)
2.5	Map	Partala, 2006	5.2	LDA	own(Python)
2.5	AugCog	Dorneich, 2005	5.3	SVM	own(Python)
2.6	Adaptive Taxonomy	Hinss, 2022	5.4	Random Forest	own(Python)
2.7	MART	Young, 2002	5.6	Motor Imagery	own (BioRender)
2.9	Red Alert	Saint-Lot 2020	5.7	P300	own (BioRender)
4.1	Brain anatomy	own (BioRender)	5.8	passive BCI	own (BioRender)
4.2	Neuron	own (BioRender)			

Tab. 14.1 Sources of Images used in the Manuscript. The authors created any figures that are not referenced.

V.14.2 Questionnaires

V.14.2.1 Demographics

The Demographics questionnaire includes questions about age, education, gender and employment (see Figure 14.1).

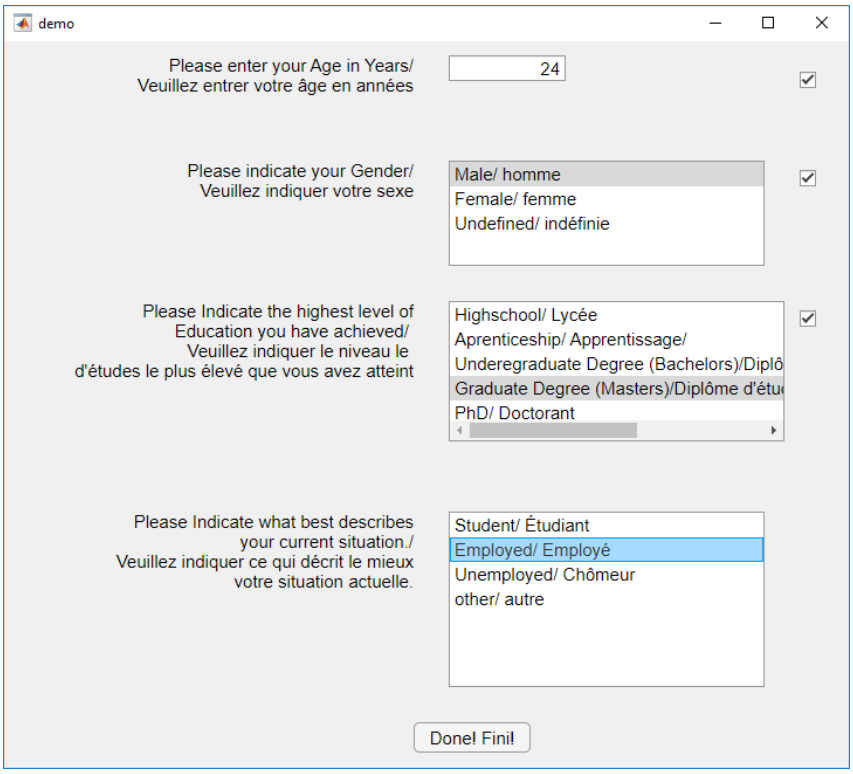

Fig. 14.1 The demographics questionnaire

V.14.2.2 Edinburgh Handedness Questionnaire

The shortened version of the Edinburgh Handedness Inventory encompasses 4 items (e.g. writing) scored on a 5-point scale ranging from:" Always right; Usually right; Both equally; Usually left; Always left". This shortened version is a faster measure while maintaining its reliability (see Figure 14.2, Veale (2014)).

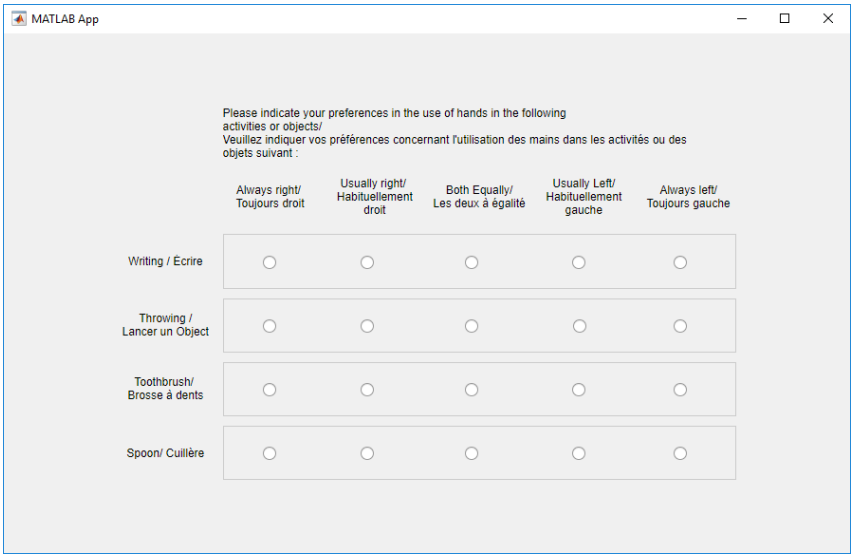

Fig. 14.2 Edinburgh Handedness Questionnaire

V.14.2.3 Ishihara Colorblindness Test

The Ishihara test for colourblindness (Hardy et al., 1945) encompasses stimuli that can determine the degree and type of colourblindness a person has. It was presented in printed form, using a validated version of the test made available by the US Department of Infectious Diseases (see figure 14.3). Participants were informed that a licensed physician is required for a medically accurate test. Should the suspicion of colourblindness arise during the assessment, participants were notified and advised to seek a medical professional.

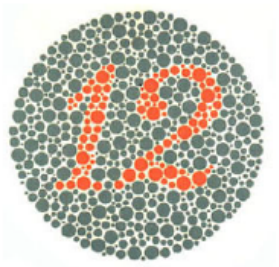

Plate 1

Everyone should see number 12.

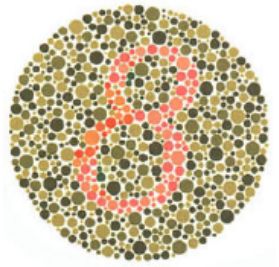

Plate 2

Normal view: 8

Red-green deficiency: 3

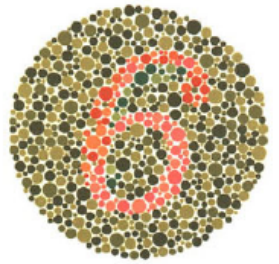

Plate 3

Normal view: 6

Red-green deficiency: 5

Fig. 14.3 Ishihara Colorblindness Test

V.14.2.4 Karolinska Sleepiness Scale

The Karolinska Sleepiness Scale (KSS) is a 9-point Likert scale, ranging from "extremely 1 = alert" to "9 = extremely sleepy–fighting sleep" (see Figure 14.4, Kaida et al., 2006). The translation into french is from the International Civil Aviation Organization ICAO, but not psychometrically validated (OACI, 2016).

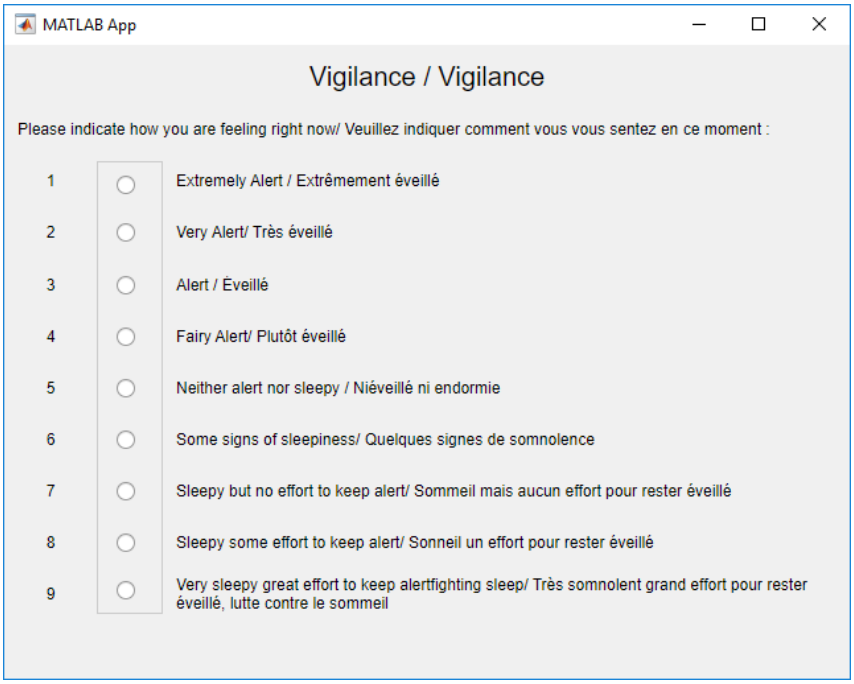

Fig. 14.4 Karolinska Sleepiness Scale

V.14.2.5 Samn-Perelli Scale

The Samn-Perelli Scale is a 7-point Likert scale used to assess fatigue (Gander et al., 2014; Samn and Perelli, 1982, see figure 14.5).

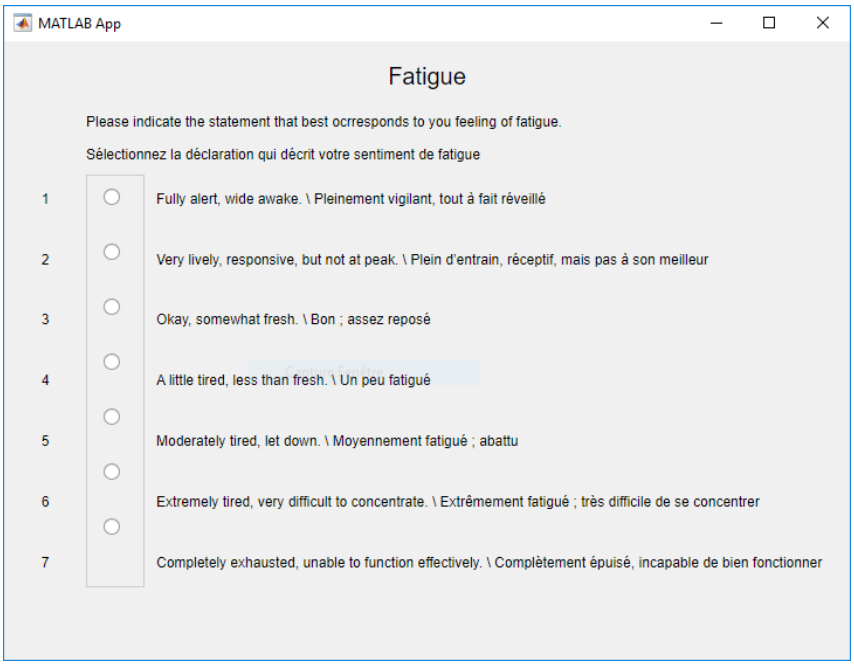

Fig. 14.5 Samn-Pernelli Scale

V.14.2.6 Visual Analogue Scale

Two visual analogue scales in the form of digitally presented sliders were used to assess fatigue. One scale assessed cognitive fatigue (VAS-F) and one for sleepiness (VAS-S). The following definitions were given to distinguish the two scales: "*Cognitive fatigue refers to the need to stop an ongoing cognitive/mental effort that has become too demanding in a situation. Sleepiness fatigue refers to significant sleep pressure leading to the need to close the eyes and fall asleep.*" Two VAS scales were presented as a 10 cm horizontal line with the ends labelled as follows: left, 0 cm = "no fatigue"; right, 10 cm = "complete exhaustion". Participants were asked to place a slider on this horizontal line that best described their state of fatigue(see Figure 14.6).

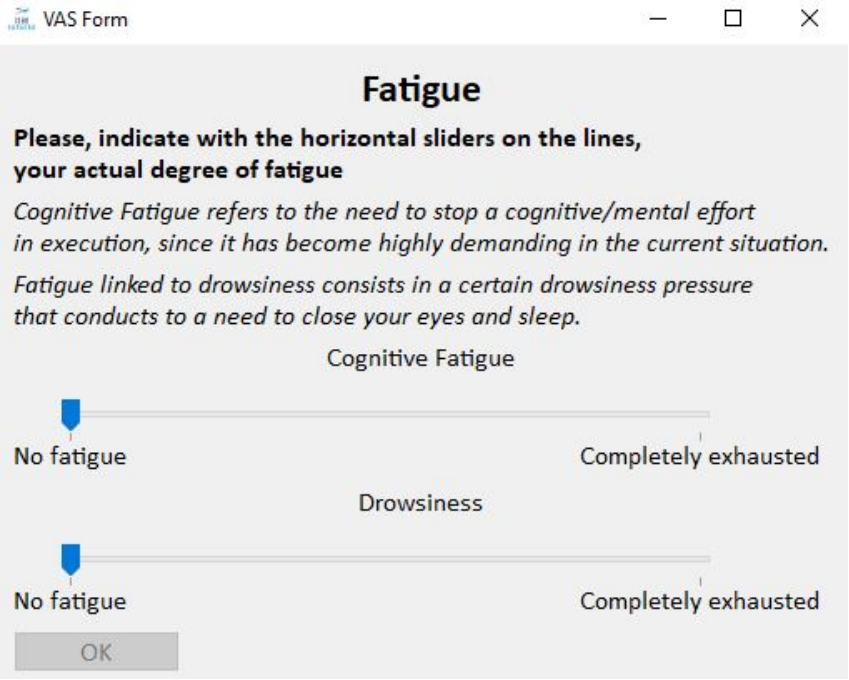

Fig. 14.6 Visual Analogue Scale

V.14.2.7 Rating Scale Mental Effort

The RSME scale Ghanbary et al. (2016) evaluates participants' mental effort invested in the task. The versions in which all items are labelled was used in all experiments see figure 14.7, Miley et al., 2016).

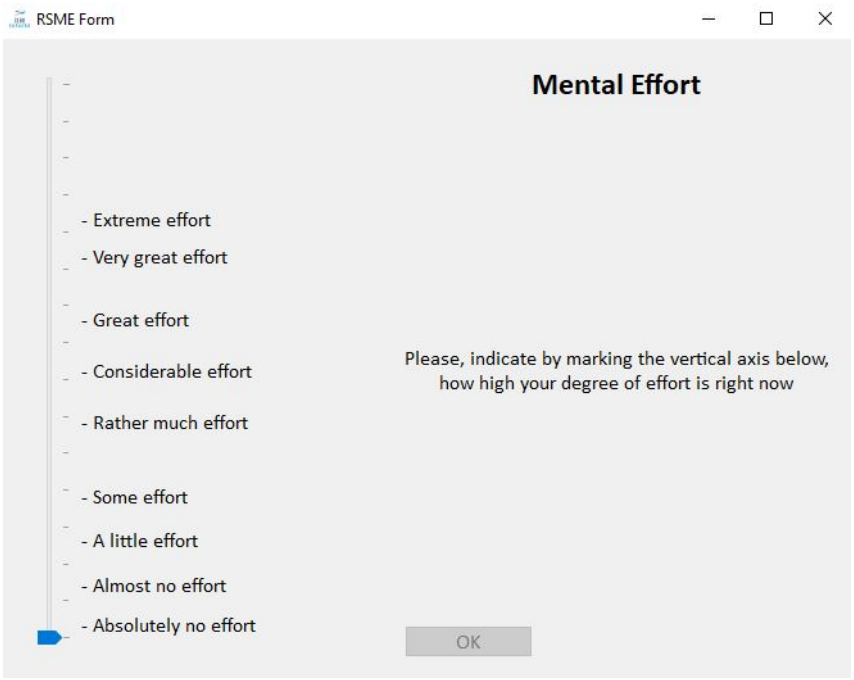

Fig. 14.7 Rating Scale Mental Effort

V.14.2.8 NASA Task Load Index

The NASA Task Load Index (NASA-TLX) is a multidimensional assessment tool for the perceived workload. It includes 6 subjective VAS assessing Mental, Physical and Temporal demand as well as Performance, Effort, and Frustration (Hart, 1986).

V.14.2.9 The St Mary's Hospital Sleep Questionnaire

The St Mary's Hospital Sleep Questionnaire (QSN) was used to estimate the sleep quality on the previous night. A translated version of this questionnaire has been produced (see Appendix V.14.2.9, Ellis et al., 1981). The scoring of this questionnaire is not standardized due to the use of Likert scales and free responses. In addition, two questions regarding caffeine consumption have been added to this questionnaire.

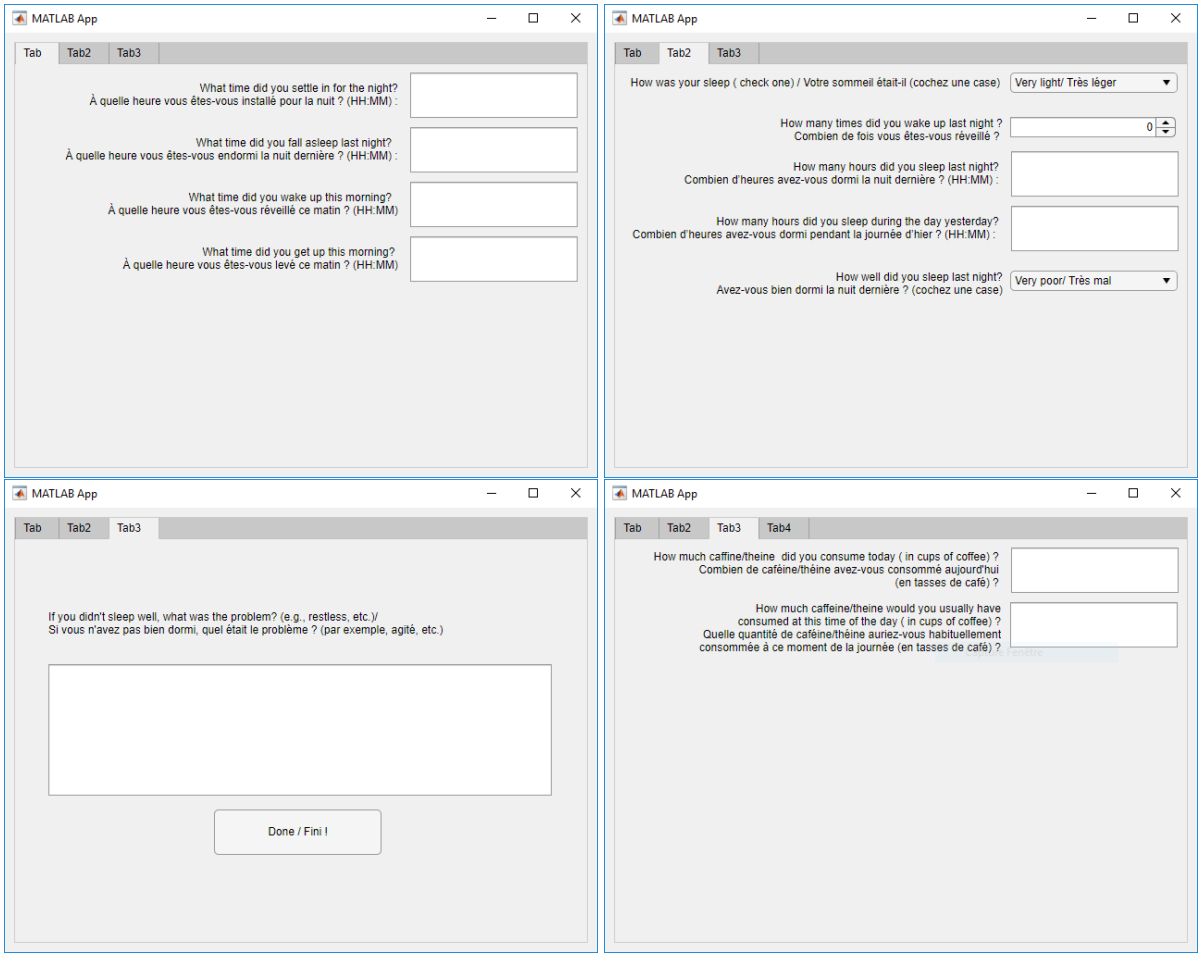

Fig. 14.8 The St Mary's Hospital Sleep Questionnaire

V.14.3 Remot3e - Saccade detection

With an acquisition frequency of around 50 Hz (data analysis revealed some inconsistencies in the acquisition frequency of the hardware that was accounted for during the data processing), robust estimation of saccade velocity metrics was impossible. Instead of using the common threshold of saccades of 300 degrees/sec and a time of 30-120 ms (Di Stasi et al., 2012), eye movements for which the velocity exceeded 1.75 standard deviations from the average eye movement speed were defined as saccades. Upon inspection of the data, this method was chosen as the best solution for detecting saccades. This metric was used only for mental fatigue and performance estimation, not for the validation of the database. This was the best possible approximation, given the low acquisition frequency. The validity of this metric is debatable. The threshold was chosen following the visual inspection of the eye movements of several subjects at different time points during the experiment.

V.14.4 Doulbe Task Switching - Exploratory Analysis

V.14.4.1 Methods

Rebound Effect

Rebound effects, the increase in error probability and RT following an error, are well-studied phenomena (Danielmeier & Ullsperger, 2011). In the exploratory analysis, we investigated this using a Kruskal-Wallis one-way analysis to account for unequal sample sizes. The dependent variables were RT and ACC.

Task-Rule Congruency Convergence

We were interested in whether task-rule incongruent trials were influenced by the fact that the incongruent information displayed was from the task from which the participants had just switched. For example, during trial n-1, the participant was performing the high/low task and was now instructed to switch to the horizontal/vertical task. Is there a significant difference in behaviour if the task-rule incongruent information presented in trial n corresponds to the task of trial n-1 (in our example, high/low)? To investigate this, we limited the sample to switch trials and compared trials where this occurred to all other types of trials. Four possible scenarios could occur:

1. Internal switch trial with internal task-rule incongruent information
2. Internal switch trial with double task-rule task-rule
3. External switch trial with external task-rule incongruent information
4. External switch trial with double task-rule incongruent information

A Kruskal-Wallis one-way analysis was performed to account for unequal sample sizes. The dependent variables were RT and ACC. These trials were compared to the most similar subset: internal and external switches with task-rule incongruent information presented simultaneously.

Timepoint Mixed Model

Initially, the trial number was planned to be included in the mixed model analyses on RT and ACC. However, the number of parameters to estimate by SPSS would have been computationally impossible, resulting in a singular fit. We recalculated the primary analyses with the additional factor timepoint to mitigate this. Timepoint was a division of each block into three separate points. This calculation was performed on both dependent variables ACC and RT.

Simple Task-Rule Congruency Models

In total 7 different conditions of task-rule congruency were explored. In a simplified analysis, we recalculated the same models but regrouped these 7 types of task-rule congruency into 3: Neutral trials, Task-rule congruent trials, and Task-Rule incongruent trials. This calculation was performed on both dependent variables ACC and RT. The sample size was unequal because the number of neutral trials was significantly lower than the task-rule congruent or incongruent trials.

Questionnaires

To explore possible relations between the subjective questionnaires and the dependent variables, we correlated the SPS and KSS scores with RT and ACC. Correlations with Age and EDI scores were excluded due two skewed data (young population and very few left-handed participants). We also performed a Mann-Whitney U test on both dependent variables to explore possible gender effects. Furthermore, we performed repeated measures of ANOVA with Timepoint and Session on both the KSS and SPS scores.

V.14.4.2 Results

Rebound Effects

The Mann-Whitney U tests on RT and ACC for trials following an error were both significant (RT: $z = 14.66, p < 0.001$, ACC $z = -19.98, p < 0.001$). Following an error, participants were 71 ms slower, with the likelihood of making an error increasing by almost 10% (see Graph 14.9).

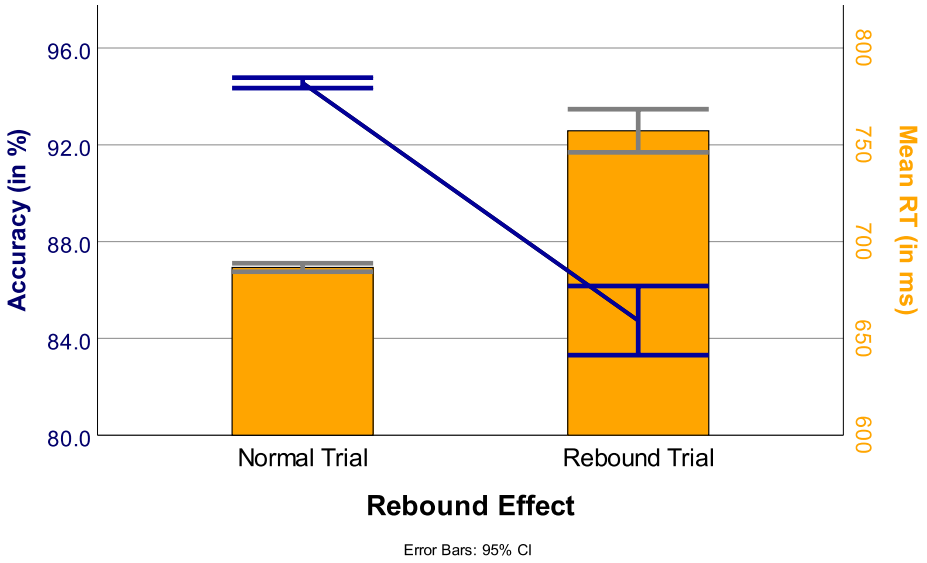

Fig. 14.9 RTs (in ms, Bars) and ACC (in %, Line) for the rebound effect, with 95% confidence interval.

Task-rule-congruent convergence

The Mann-Whitney U tests on RT and ACC for trials in which the task-rule incongruent information that was presented was coherent with the task of the previous trial were both significant (RT: $z = 14.50 p > 0.001$, ACC $z = -5.73, p < 0.001$). The task-rule congruency convergent trials resulted in an increase of RT of 48 ms and an increased likelihood of making an error by 2%. Results are summarized in Graph 14.10.

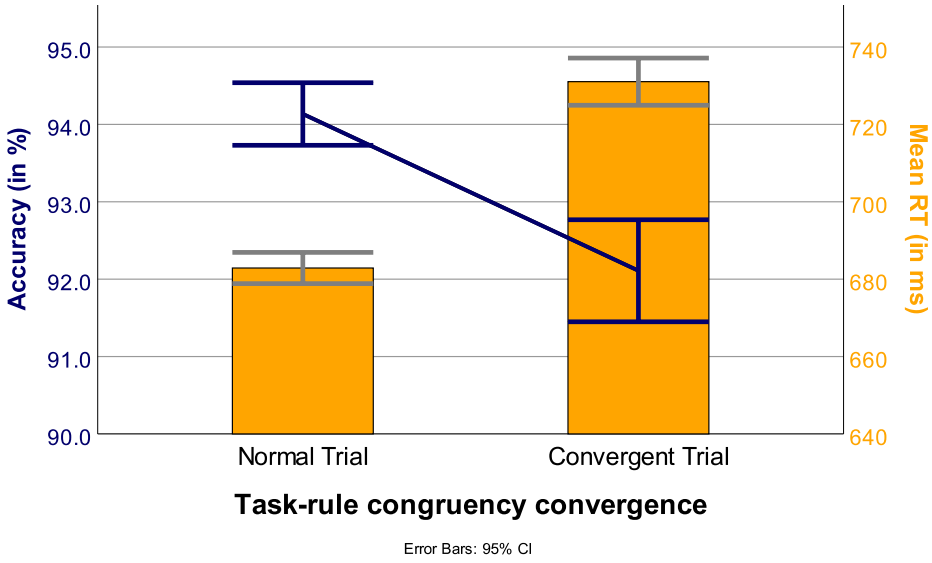

Fig. 14.10 RT (in ms, Bars) and ACC (in %, Line) for the effect of congruency task convergence, with 95% confidence interval.

Timepoint Mixed Model

The effect of timepoint on RT was significant (Timepoint: $F(45123, 2) = 4.83 p < .01$) in the analysis as well as its interaction with both the factors Switch and Block (Switch: $F(45124, 4) = 2.79, p < .05$, Block: $F(45123, 8) = 4.20, p < .001$). Pairwise comparisons between time points showed a significant difference between the first and the last time point (1-112 vs 225-336: $t(45123), p < 0.01$) and the second and third time points (113-224 vs 225-336: $t(45123), p < 0.05$). RTs decreased across time points. With regard to ACC the main effect of timepoint was significant (Timepoint: $F(45123, 2) = .06, p < .01$). The interaction between task and time point was also significant (Task & Timepint: $F(45123, 6) = 3.09, p < .01$). Pairwise comparisons between timepoints showed a significant decrease (94.4% to 93.4%) in ACC from timepoint 1 to timepoint 3 (113-224 vs 225-336: $t(45123), p < 0.01$). Results are summarized in Graph 14.11.

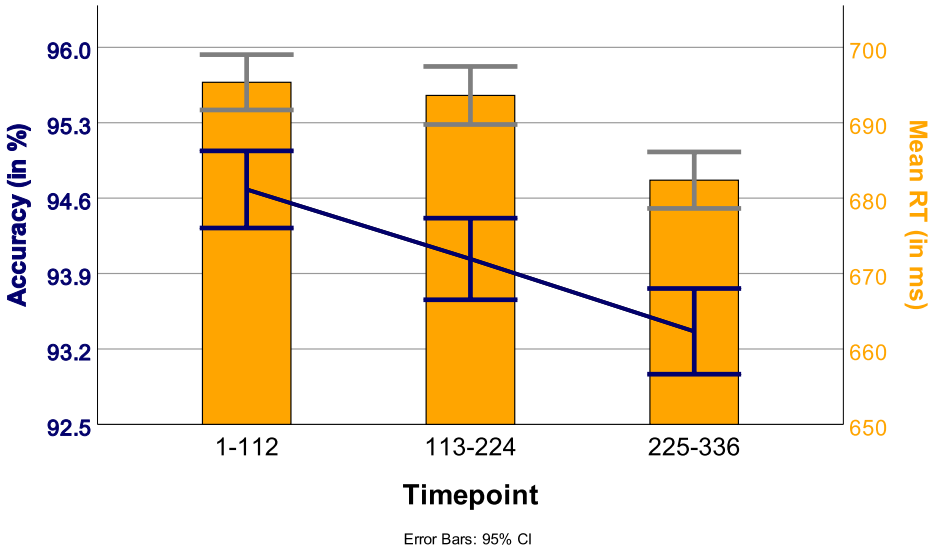

Fig. 14.11 RTs (in ms, Bars) and ACC (in%, Line) across time points, with 95% confidence interval.

Simplified Task-Rule Congruency Models

The simple task-rule congruency was a significant factor in the mixed model (simplified task-rule congruency: $F(300041, 2) = 32.23, p < .001$) with RT as a dependent variable. The Interactions between simplified task-rule congruency and Switch, Block as well as Task were also significant (Switch: $F(30041, 8) = 3.83, p < .005$, Block: $F(30041, 8) = 4.78, p < .001$, Task: $F(30041, 6) = 5.65, p < .001$). All pairwise comparisons were significant. Task-rule neutral trials (679 ms) were the fastest followed by task-rule congruent (693 ms) and finally task-rule incongruent trials (702 ms) (task-rule neutral vs task-rule congruent: $t(45123), p < .001$, task-rule neutral vs task-rule incongruent $t(45123), p < .001$, task-rule congruent vs task-rule incongruent: $t(45123), p < .001$). A similar main effect was also found in the analysis with ACC as a dependent variable (simplified task-rule congruency: $F(45123, 2) = 9.46, p < .001$). No interactions were significant. The pairwise comparisons showed a significant difference between task-rule incongruent trials with both task-rule neutral and task-rule congruent trials (task-rule neutral vs task-rule congruent: $t(45123), p < .005$, task-rule congruent vs task-rule incongruent:

$t(45123), p < .001$). However, task-rule congruent and task-rule neutral trials did not differ significantly (task-rule neutral vs task-rule incongruent $t(45123), p = 1$). Results are summarized in Graph 14.12.

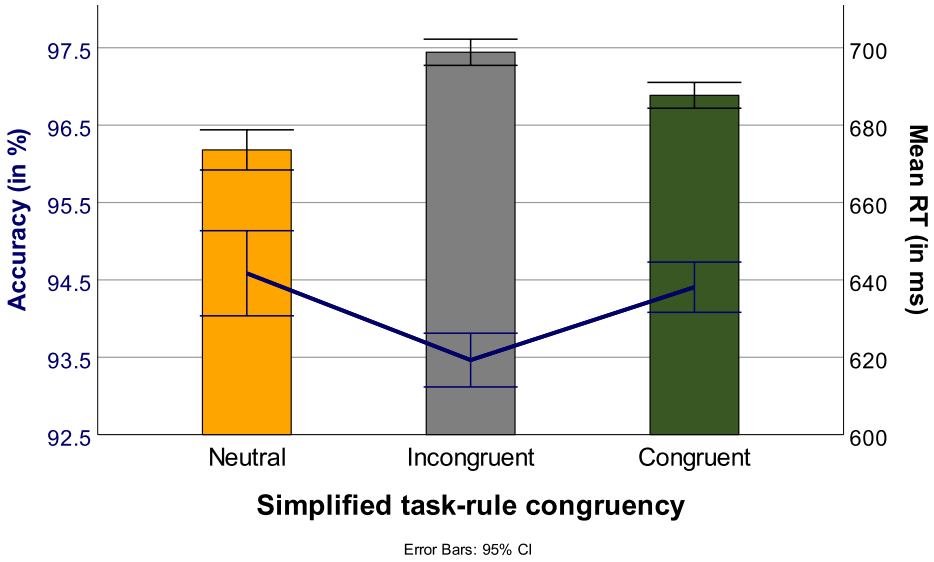

Fig. 14.12 RTs (in ms, Bars) and ACC (in %, Line) for each simplified task-rule congruency condition, with 95% confidence interval.

Questionnaires

The Mann-Whitney U tests of gender on RT and ACC were both significant (RT: $z = 14.00, p < 0.001$, ACC $z = 4.64, p < 0.001$, see Figure 14.13 Top). Female participants were less accurate but responded faster.

As expected, the repeated measures ANOVA showed an increase in SPS scores from pre-task measurement to post-task measurement only during the testing session ($F(119, 1) = 13.20, p < .001$, $t(119) = -5.430, p < .001$).

As with the SPS, as expected, the repeated measures ANOVA showed that reported fatigue increased from pre-task measurement to post-task measurement only during the testing session. ($F(119, 1) = 17.88, p < .001$, $t(119) = -6.121, p < .001$).

V.14.4.3 Discussion

The exploratory analysis of the task-rule congruency convergence shows a novel effect. Suppose a participant is presented with a task and task-rule incongruent information. In that case, that information has a significantly higher impact on the performance if that incongruent information is coherent with the task the participant had to perform in the previous trial. The results show a significantly higher RT and a 2% increase in error rate. This finding results from an exploratory analysis; therefore, the validity is limited. Further studies should test whether this effect is reproducible. To the best of our knowledge, this effect has, in this form, not been observed or at least not been documented before.

The argument could be made that the observed effect is a case of the congruency-frequency effect observed by (Kiesel et al., 2007). The researchers showed that congruency effects were

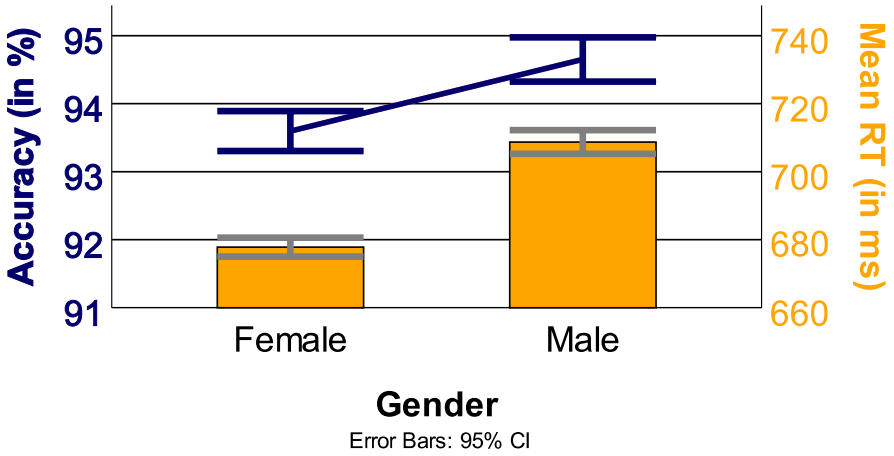

Fig. 14.13 Summary of effects of gender on RT (in ms, bars) and ACC (in %, line), with 95% confidence interval.

larger if the presented incongruent stimulus had previously occurred more frequently. Our observation differs in two significant ways. The frequency of all stimuli is constant across blocks. Furthermore, the effect is observed across all stimuli of one task, highlighting that the effect is dependent on the task set and not the stimulus.

These findings may be interpreted in two ways. The increase in TSC may be attributable to confusion among the participants as they believe they must repeat the previous task. The results may also support the interference view of cognitive flexibility. Stimuli relevant to the last task reinforce the activation of the previous task set.

Another surprising effect found in the post hoc analysis was the strength of the rebound effect. Participants were around twice as likely to commit another error following an error. Interestingly, this supports the notion that participants were very aware when making errors, as they received no feedback following each trial. Rebound effects are a well-documented phenomenon. For further information, please refer to (Danielmeier & Ullsperger, 2011).

V.14.5 UASOS - Open Source Code

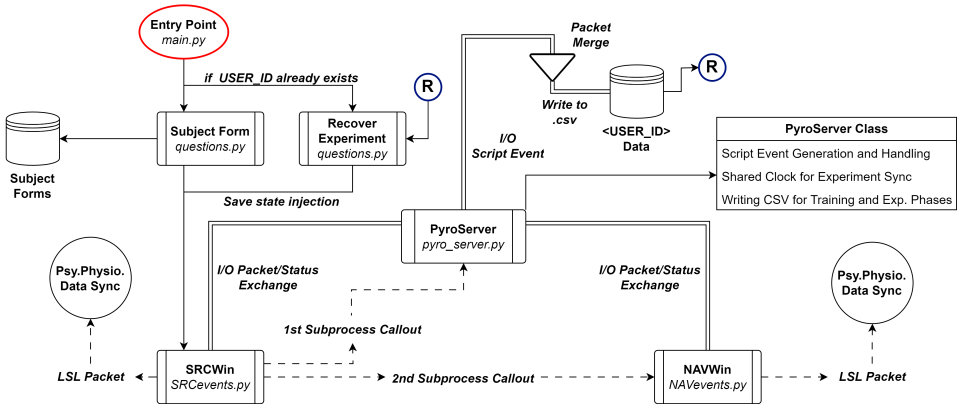

Fig. 14.14 UASOS Architecture: The Driver/Driven Gear architecture is reproduced here. `SRCWin` manages the lives of its child instances (`PyroServer` and `NAVWin`) with which it interacts during the experiment. Each Graphical has its own LSL Channel. Two separate channels allow for a higher temporal precision.

Fig. 14.15 Architecture of the UASOS environment.

V.14.5.1 Characteristics

UASOS was built with Python 3.8.10. A key requirement for the task is PsychoPy[1] (Peirce et al., 2019), which can be installed using the wizard from the project page. The installer includes Python 3.8.10 within the installation folder. The configuration of the IDE interpreter should also be set to this folder. Once PsychoPy is added, UASOS can easily be installed. The UASOS repository must be downloaded to the desktop, and the IDE interpreter must be set to the executable `python.exe` hosted within the PsychoPy installation folder. Then, UASOS can be opened.

Following the installation, the experiment's parameters can be adjusted to the requirements of the experiment using the `scriptgen4HPC` module. This module generates a dataset of scripts, with each script dictating the order and duration of all events during the experiment. We also implemented a version of said module that uses multiprocessing to create a set of scripts.

A novel aspect of the UASOS is the Driver/Driven Gear architecture, allowing for a synchronized exchange of status and data packets between both the Search (`SRCWin`) and the Navigation Task Windows (`NAVWin`). Two separate instances communicate through a Remote Object `PyroServer` in the middle exchanging 2 types of packets (see Figure 14.15):

- The **Status()** object transfers the actual experimental status, integrating the graphical mode to draw, timestamp & pause the actual state of each graphical instance.
- The **Packet()** object transfers each iteration's instructions to record the measured behavioural metrics.

[1] https://www.psychopy.org/download.html

Time synchronization is based on a shared timestamp value released by a PsychoPy `Clock()` object hosted by the server. UASOS is available on GitHub[2], and this release is suitable for screen resolutions up to 4K. We recommend a dual monitor setup with equal resolutions.

V.14.5.2 Code/Software

The behavioural data from the Search and Navigation task are collected by PyroServer, merged and saved as a `.csv` file. Each Participant (with their corresponding ID) will have their output saved in a separate folder under `./results/IDxxx`. The folder contains six files:

- `setup_IDxxx.dat`: contains all parameters for restarting in case of interruptions
- `data_IDxxx.csv & *_end.csv`: two CSV files with the pre and post-questionnaire data.
- `res_ddmmYYYY_HHmmss_IDxxx.csv & *_t.csv`: files containing the behavioural data for the main and the training phases respectively.
- `VAS_IDxxx.csv`: file with the questionnaire data, asked periodically during the main phase.

Detailed information on all aspects of the code can be found in the following section (V.14.5.3) and in the `README.md` file.

V.14.5.3 Usage Notes

Researchers may customize the experimental parameters using the scriptgen or scriptgen4HPC tools. The scriptgen4HPC creates a database of scripts that comply with a set of previously determined experimental parameters. Users can adapt parameters such as:

- `max_size_dset: int`: overall number of event scripts to generate
- `max_time: int`: maximum time to let the scriptgen4HPC script run
- `phase_gen: string`: Selection of the type of dataset to build (e.g. training or main phases)
- `exp_time_main: int`: duration for main phase scripts
- `exp_time_train: int`: duration for single task training scripts
- `it_time: int`: average time for updating instructions (new trial)
- `jitter: int`: jitter for instructions
- `treshold: float`: maximal deviation from a balanced design (e.g. equal amount of trials in each task) in the main phase
- `treshold_train: float`: maximal deviation from a balanced design in the training phase

[2] https://github.com/Marcels-2-Neurons/uasos-project

Should the script be run on a personal computer instead of an HPC Node, the number of threads should be reduced to (`num_threads = mp.cpu_count() - 1`). This should allow the PC to manage the essential processes. Beyond this, users may add a separate dataset of images for the search task, which requires labelling all images.

To run the experiment, users will need a trackball (or a mouse), a keyboard, two screens and a joystick. The current version of UASOS uses the `pysticks`[3] toolbox for the joystick implementation. Should no joystick be available, users can alternatively use several game controllers.

V.14.6 UASOS - Transfer Learning

V.14.6.1 Motivation

As expressed in Section I.5.2.4, Transfer Learning is a large problem for EEG-based Physiological Computing Systems and BCIs. Other than, for example, motor imagery-based BCIs, training a model before testing it all within one session is difficult in the context of mental fatigue. For an adequate training the participant would be required to be fatigued at some point during the training, therefore biasing the testing session. So, the only solution to implement a machine learning-based BCI is to transfer information across sessions or participants.

V.14.6.2 Methods

Using the data from the UASOS experiment (see Chapter 9), we tested three transfer learning pipelines for inter-participant fatigue detection using either TOT labels or performance labels. Results were compared using a repeated measures ANOVA with the dependent variable of Accuracy. The test was performed using leave one out analysis on a set of 24 participants.

Support Vector Machine

The data was preprocessed and labelled as described in Section II.9.3.4. Next, the covariance matrices for each epoch were calculated and normalized based on the mean semi-positive definite matrix for all epochs of one subject. The data size is then reduced to the upper triangle of the data, as the Covariance matrices are symmetric along the diagonal. The data is then used to train a Support Vector Machine with an RBF kernel and validated using Leave One Out Analysis.

TSM-Net

The TSM-Net[4] was built to tackle transfer learning for BCIs by combining Riemannian and Deep learning approaches. This model's major advantage is that it can work without calibration data, using unsupervised domain adaptation. This means that, theoretically, the model can classify the EEG data of a new participant in real-time without having prior access to that participant's data (Kobler et al., 2022). Here, we will compare the model's performance using an offline approach (where the model is calibrated to the target domain) and a simulated online approach where the model calibrates during the actual testing.

[3] https://github.com/simondlevy/PySticks

[4] This work was done in collaboration Reinmar Kobler at ATR (Kyoto, Japan), who built the initial TSM-Net.

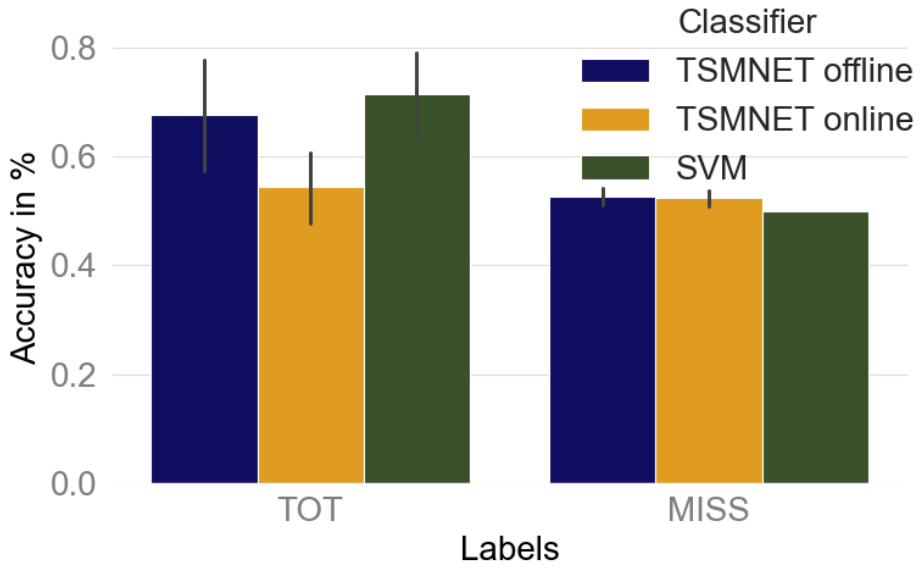

Fig. 14.16 Summary of the results of the TOT and performance estimation.

V.14.6.3 Results

In addition to the interaction effect, both the main effects of label-type and classifier were significant (label: $(1, 23) = 12.323$, $p = .0002$ $\eta^2 = .349$; classifier: $(2, 23) = 5.811$, $p = .015$, Greenhouse-Geisser corrected, $\eta^2 = .284$; label*classifier:$(5, 23) = 9.103$, $p = .003$, Greenhouse-Geisser corrected, $\eta^2 = .284$). All classifiers using the performance labels performed at around 50% accuracy. For the TOT, both the SVM and the offline TSMNET performed significantly better at 71.5 % (SVM) & 67.6 % (TSMNET offline) respectively. The online implementation of the TSMNET had an average accuracy of 54.4% (see Figure 14.16).

V.14.7 COG-BCI - An Open EEG Database for passive BCIs

Central Ideas

Brain-computer interfaces and especially passive Brain-Computer interfaces (pBCI), with their ability to estimate and monitor user mental states, are receiving increasing attention from both the fundamental research and the applied research and development communities.
Testing new pipelines, benchmarking classifiers, and feature extraction algorithms is central to further research within this domain. Unfortunately, data sharing in pBCI research is still scarce. The COG-BCI database encompasses the recordings of 29 participants over 3 separate sessions with 4 different tasks (MATB, N-Back, PVT, Flanker) designed to elicit different mental states for a total of over 100 hours of open EEG data. This dataset was validated on a subjective, behavioural and physiological level to ensure its usefulness to the pBCI community. Furthermore, a proof of concept is given with an example of mental workload estimation pipeline and results to ensure the data can be used to design and evaluate pBCI pipelines. This body of work presents a large effort to promote using pBCIs in an open science framework.

V.14.7.1 Introduction

Development and use in real life of passive Brain-Computer Interfaces (pBCI) is unfortunately hindered by a decrease in performance due to strong variability that originates both between and within user factors (Fairclough & Lotte, 2020). In particular, EEG signals and BCI performance have been reported to change across days significantly, tasks, users' mental states and contexts (Benaroch et al., 2021; Grosse-Wentrup & Schölkopf, 2013; Hinss, Darmet, et al., 2021a; Hinss, Somon, et al., 2021; Mühl et al., 2014; Roy et al., 2013). A relevant solution to tackle this issue is to develop transfer learning methods. Transfer learning is an overarching term for a group of problems that refer to the decrease in effectiveness/accuracy of machine-learning algorithms and pipelines when applying the algorithm to slightly different but related data (Pan & Yang, 2010). To that end, open datasets present a cost-effective alternative to collecting data individually for each experiment. They enable researchers to validate their algorithms and compare their performances on a similar data set. Nonetheless, the number of freely available datasets is, with exceptions (Shin et al., 2018), very limited (see Hinss, Somon, et al., 2021 for further details). The advantages of data sharing and the simultaneous lack of free data were the motivating factors for creating the COG-BCI database.

Hence, this database was designed to propose a sufficient amount of fully informed data in various conditions representing various cognitive states, thus allowing researchers to investigate transfer learning (Hinss, Jahanpour, Somon, et al., 2022a). Four tasks were carefully selected to allow the assessment of various mental states. The *Flanker task* is used for decision-making and conflict evaluation (Eriksen & Eriksen, 1974). The task consists of deciding an arrow's direction while simultaneously presenting distracting (incongruent) or non-distracting (congruent) stimuli. Afterwards, participants receive feedback, allowing researchers to investigate the effects of errors and trial-based feedback.

The *N-Back task* is known for taxing working memory and has, therefore been frequently used to elicit different levels of mental workload (Brouwer et al., 2012). Here, participants have to retain numbers that appear on the screen and decide if the currently presented letter is the same as the Nth letter before (N equal to 0, 1 or 2). Another ecological way to measure mental

workload is the MATB-II, in which participants must simultaneously perform four aviation-related subtasks. Lastly, the Psychomotor Vigilance Task (PVT) was selected as a simple and well-established measure for robust estimation of vigilance and fatigue (Lamond et al., 2005). In this task, participants must react as fast as possible by pressing a button to a stimulus appearing pseudo-randomly on a computer screen. The described data is publicly available (Hinss, Jahanpour, Somon, et al., 2022a). The upcoming sections will detail the data collection and analysis for data validation. The results (subjective, behavioural and physiological) are further interpreted to determine if the tasks were correctly implemented and elicited the expected mental states and their physiological markers. These results prove their possible use for pBCI development with a three-class machine-learning estimation pipeline example. Lastly, usage notes are provided.

V.14.7.2 Methods

Participants

Ethical Approval for the data collection and subsequent distribution was obtained from the Comité d'Éthique de la Recherche - CER at Université de Toulouse (CER number 2021-342). Based on the average number of participants presented in the pBCI literature (around 15 participants per experiment (Aricò et al., 2016; Verdière et al., 2018)), cognitive science literature (around 20 participants per experiment (Senoussi et al., 2017; Zeng et al., 2021)) and similar databases (SEED Zheng and Lu, 2015 15 participants and DEAP (Koelstra et al., 2012) 32 participants) we initially recruited 35 participants to perform this experiment. Due to participant dropout and technical issues with the data collection, a total of 29 participants were included in the final experimental setup. Participants (11 Female, 18 Male) were on average 23.9 (Std. 3.20) y.o. All but 4 participants were students, while the others were employees. 14 participants had obtained a bachelor's degree, while 13 had obtained a master's degree. Based on the Edinburgh Handedness scale, 2 participants were left-handed.

Experimental Paradigm

The experiment was conducted over 3 sessions spaced one week apart during which participants had to perform 4 different tasks: the N-back task, the MATB-II, the PVT and the arrow-based Eriksen flanker task (Eriksen & Eriksen, 1974). Upon their arrival for the first session, participants were informed about the study procedure and asked to read and sign the informed consent.Participants consented to the collection and subsequent distribution of their subjective, behavioral and physiological (cardiac and EEG) data for the duration of the experiment. Participants also consented to having a 3D picture taken. Then they were asked to fill in the demographic questionnaires, as well as the Edinburgh Handedness Inventory, (Veale, 2014). The EEG was set up while participants received task instructions. To ensure precise electrode location for within and between-subjects variability, a 3D picture of their head with the electrode cap was recorded at the beginning of each session. Then, participants completed a short training sequence for all 4 tasks. After training, they filled in the Karolinska Sleepiness Scale (Kaida et al., 2006) (KSS, see Appendix V.14.2.4) and performed a two-minute resting state (one-minute eyes open and one minute eyes closed) while their brain activity was recorded. Participants then started performing the tasks, which were presented twice in a pseudorandom order followed systematically by a Rating Scale Mental Effort (Ghanbary Sartang et al., 2016) (RSME). An additional KSS was also completed following the PVT only. Participants

were allowed to take short breaks in between tasks. After all tasks were fully completed, participants performed another resting-state period before filling in the KSS scale once more. The entire experimental protocol is detailed in Figure 14.18. Completing the tasks took from 65 to 80 minutes, depending on the number of breaks the participants took. With setting up the electrodes and training, each session lasted about two hours, with the first session often lasting a bit longer due to more detailed instructions. Due to the ongoing pandemic at the time of acquisition, a special sanitary protocol was adopted.



Fig. 14.17 Experimental setup with a participant equipped with the 64-electrode EEG system performing the MATB-II task. Consent for the publication of this picture was obtained from the participants.

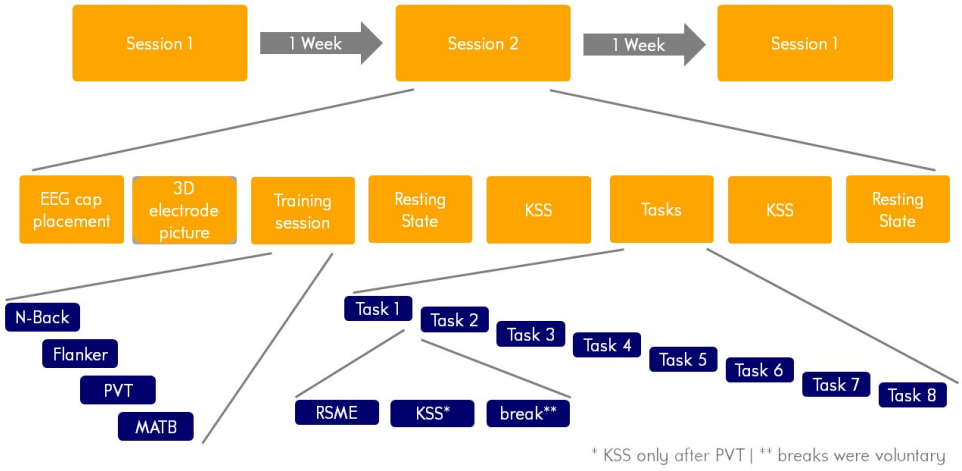

Fig. 14.18 Experimental Procedure, detailing each step for all three sessions.

Materials

The database was created by acquiring participants' subjective, behavioural and physiological data using the experimental protocol presented above. Details on the questionnaires, tasks and acquisition devices used are given in this section.

Subjective Questionnaires All questionnaires were coded, administered with MATLAB v.2021a (The Mathworks Inc.), and presented on a $60Hz$ LCD computer screen. Questionnaires included a demographics questionnaire (see Appendix V.14.2.1), the Edinburgh Handedness Questionnaire (EDI, see Appendix V.14.2.2), the Karolinska Sleepeniess Scale (KSS, see Appendix V.14.2.4) and the Rating Scale Mental Effort (RSME, see Section V.14.2.7).

Tasks

Psychomotor Vigilance Task The PVT is a 10-minute task allowing to measure vigilance (Lamond et al., 2005; Reifman et al., 2018). Participants are asked to react as fast as possible to the appearance of a timer on the computer screen by pressing the keyboard's space bar. This task was designed to closely resemble the PC-PVT 2.0, an established computer version of the PVT (Reifman et al., 2018). Each trial starts with an interstimulus interval (ISI) between two and ten seconds. Then, the timer is displayed on the screen until participants respond. Following a response, the timer stops and displays the participant's reaction time on the screen for another $500ms$. Participants have to perform 90 trials per session, leading the task to roughly 10 minutes. Participants' reaction times and responses are collected throughout the whole task. Figure 14.19 illustrates the task.

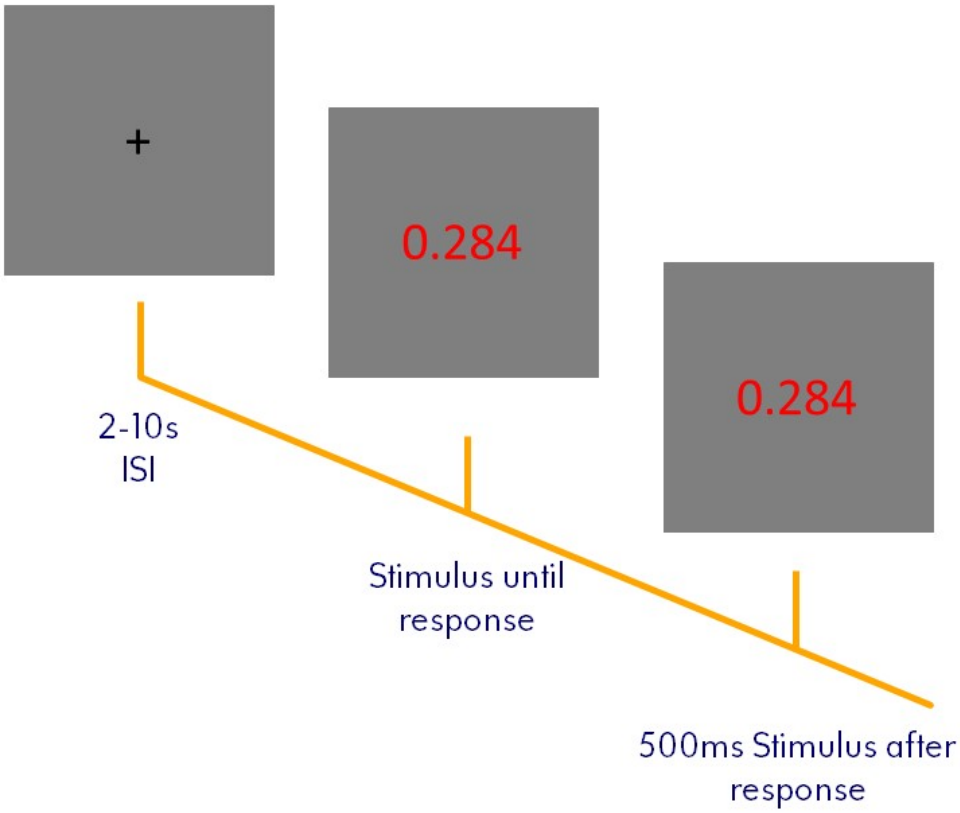

Fig. 14.19 The PVT task: Following an ISI of 2-10 seconds, a red timer shows the time elapsed from trial onset until reaction. Once the participant reacts with a button push, the timer stops and shows the reaction time for 500 ms before the next trial starts.

N-Back Task The N-Back task is a widely used measure of both working memory and mental workload (Brouwer et al., 2012; Jaeggi et al., 2010). This task is selected to elicit different levels of mental workload across several conditions without changing the amount of visual information presented or the amount of motor responses required thus avoiding EEG data contamination.

On a computer screen, participants are presented with single letters appearing for a short period of time. Participants are instructed to remember the order in which the numbers appear and to react with a button press if the presented number is the same as the Nth number presented before. The N here is the variable that determines the difficulty of a particular block. With the increasing size of N, the difficulty of this task increases, as more numbers need to be retained. Here, 3 different conditions are chosen: 0-Back, 1-Back and 2-Back, corresponding to easy, medium and high workload levels. Trials begin with the presentation of a number (between 1 and 9) for $500ms$, followed by a blank screen for $1500ms$ (Figure 14.20). If the number presented is a hit number participants are instructed to respond by hitting the space bar. A hit number corresponds to the number 3 in the 0-back condition, the same number as the previous

one in the 1-back condition, and the same number as 2 trials before in the 2-back condition. Each block consists of 48 trials and lasts approximately 2 minutes. The display frequency of hit numbers was fixed at $\frac{1}{3}$ (16 trials per block) in all three conditions. Participants completed 3 blocks ($\sim$ 6 minutes) of each condition for 9 total blocks.
Additionally, in the 2-back condition, 5 conflict trials are added per block. These trials are characterized by a number being followed immediately by the same number again. Participants are asked not to react to these trials, but this has been suggested to result in eliciting conflict (Oberauer et al., 2005). As these trials would also occur during full randomization of the numbers, they have no adverse effect on the participant's performance. Before the onset of each block, the participant was informed about the condition of the block and given short instructions on what to do. Participants' responses and reaction times are recorded throughout the entire task for each session.

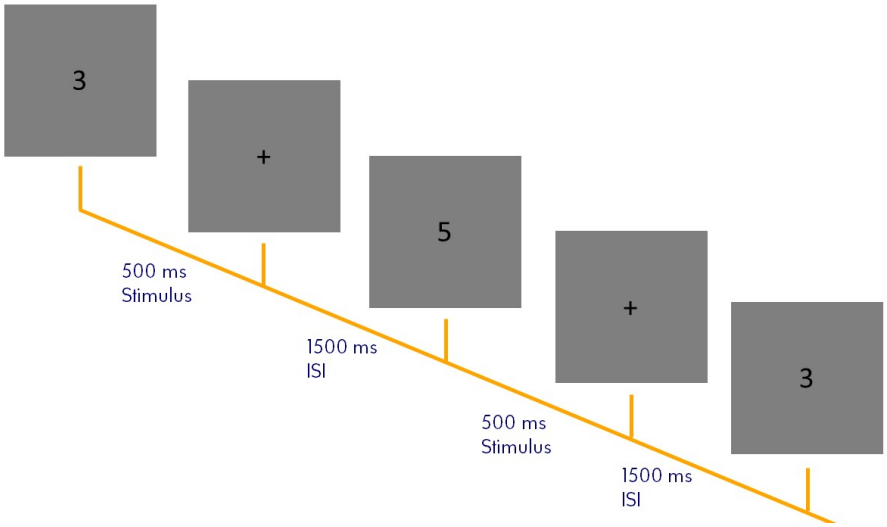

Fig. 14.20 The N-Back task: Single digits are presented for 500 ms with a 1500 ms ISI. If in the 2-back condition, the last stimuli (3) would be a hit trial and require a response from the participant, as the digit shown 2 trials before was the same.

Flanker Task The Flanker task is a simple choice reaction task to elicit errors and conflict during a binary decision (Eriksen & Eriksen, 1974). In its arrowhead version, participants are presented with stimuli composed of 5 horizontal arrows. They are instructed to react to the middle arrow and ignore the flanking arrows on either side. These flanker stimuli can either point towards the same direction (congruent condition) or the other direction (incongruent condition) as the target, central arrow. A typical stimulus may therefore look like '< < > < <' or '< < < < <', where the respective targets are '>' and '<'. Each trial begins with an ISI of $2000m$, followed by the stimulus display for $16ms$ (Figure 14.21). This display time was determined in a pilot study based on changes in error rates. Each of the four possible stimuli ('> > < > >' '< < < < <' '< < > < <' '> > > > >') are presented equally frequently (25% of all trials) in a pseudorandom order. Following the stimulus presentation, participants are required to respond by stating the target direction with the "S" and "L" keys of the keyboard. During this time a blank screen with a fixation cross is shown for $2500 \pm 250ms$. At the end of each trial, participants received feedback about the outcome (correct, incorrect, miss) of their trial for $500ms$ In total, 120 trials were performed (30 for each type), with a complete run taking around 10 minutes. Before

the onset of the run, participants received instructions. Participants' responses, error rates and reaction times to each type of stimulus are recorded throughout the entire task for each session.

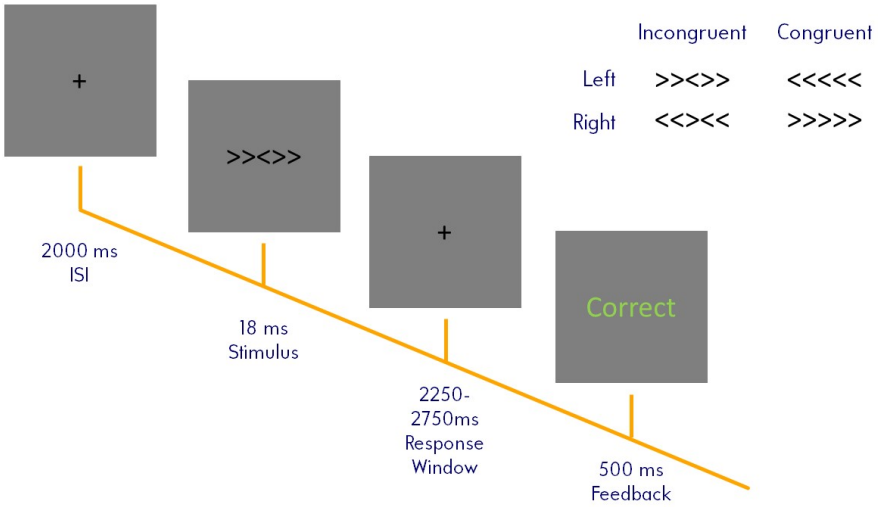

Fig. 14.21 The Flanker task: Following a 2000 ms ISI one of the 4 possible stimuli (bottom left) is shown for 18 ms. Participants then have between 2250-2750 ms to react before they receive feedback for 500 ms.

MATB-II Task NASA developed the MATB-II task to assess task-switching and mental workload capacities in more realistic environments (Santiago-Espada et al., 2011) 14.17. As the MATB-II also taxes mental workload the inclusion of this task not only allows for more ecologically valid taxation of mental workload, it also allows researchers to investigate transfer learning between the N-Back task and a more realistic measures of mental workload. Participants are presented with up to 4 different tasks that they have to complete simultaneously. This provides a highly realistic environment of operational systems that researchers can control to create different degrees of cognitive workload and difficulty. For this study, combinations of four of the available subtasks of the MATB-II were used and presented with an adapted version of the MATB-II software (Figure 14.22). This version is coded in MATLAB v.2021a (The MathWorks Inc.) and provides the same measures as the original MATB-II task (Verdière et al., 2018). The four subtasks selected from the MTAB-II are tracking (TRACK), system monitoring (SYSMON), communication (COMM) and resource management (RESMAN); leaving task scheduling unused. In the tracking task (TRACK – top right corner), participants are presented with a moving target inside a window and have to keep the target within the central square area by controlling a joystick. The degree of difficulty can be adapted by modifying the degree and the speed at which the target moves. Participants must monitor gauges and warning lights for the System monitoring task (SYSMON – top left corner). Action is required in the absence of green lights, the presence of red lights and deviations of four moving pointers dials from a midpoint. Participants must press keys F5, F6 and F1 to F4 in those cases. The degree of difficulty can be adapted by increasing the number of events the participant has to react to. In the communication task (COMM – bottom left corner). Participants must listen to radio messages and select specific radio channels and change frequencies accordingly when the messages are directed to him/her while ignoring messages not directed to them. Participants' responses

are recorded with joystick-based button presses. Workload level can be adapted by including fewer or more messages directed or not to the participant. The last task used is the resource management task (RESMAN – bottom right corner). Participants are presented with an interface displaying two main fuel tanks (A and B) and four subsidiary tanks (C to F) interconnected via eight pumps (numbers 1 to 8, Type keys). The goal is to maintain a specific fuel level in both main tanks. Participants can do this by activating or deactivating the pumps. Events such as pump failures can be introduced to increase the task's difficulty. Pump activation/deactivation can be done by pressing buttons (F1-F6). Whenever a pump fails, it is inoperable for a short time.

In the current study, participants are asked to perform three independent runs of 5 minutes each corresponding to a different degree of difficulty. Participants only engaged in the system monitoring and tracking tasks for the easy condition. For the medium condition, participants engaged in both tasks and the resource management task. For the difficult condition, the communications task was added, and the tracking task was made more difficult. Before each run, the participants also received a short instruction.

Data Acquisition

Subjective, behavioral and physiological data are recorded throughout the entire session for each session independently and with each participant. Accurate data synchronization and stimuli recording is performed through the LabRecorder of the LabStreaming Layer software and its related API (https://labstreaminglayer.readthedocs.io/info/intro.html).

Stimulus Presentation and Response Acquisition All tasks were coded in MATLAB v.2021a (The Mathworks Inc.), with the PsychToolBox-3 (http://psychtoolbox.org/). To display the stimuli, a desktop computer with a 60 Hz screen was used. Response acquisition occurred with a Keyboard and an Extreme 3D Pro Logitec Joystick for the MATB task. Participants were seated approximately 50 cm away from the screen.

EEG System In this experimental campaign, we used an EEG system (electroencephalography) with 64 active Ag-AgCl electrodes (ActiCap, Brain Products Gmbh) and an ActiCHamp amplifier (Brain Products, Gmbh). Electrode locations followed the standard 10-20 system (Jasper, 1958). Electrode used here names can be found in Figure 14.23. For participants 1-9 the electrode Cz was not recorded. In addition, one electrode dedicated to recording peripheral electrocardiographic (ECG) activity was placed on the left fifth intercostal. To obtain the precise location of the electrodes on the scalp of each participant at each session, a 3D scanning camera by STRUCTURE (https://structure.io/ and the get_chanlocs plug-in developed specifically for electrode localisation purposes was used (GitHub.com/sccn/get_chanlocs/wiki Homölle and Oostenveld, 2019). For this, a 3D camera was located on an IPad (Apple Inc.) and calibrated once for the entire experimental campaign. Then, after equipping the participant with the electrode cap, and right before the experiment, a 3D scanning of their head was realised as described on the Structure Sensor website and recommended specifically for EEG on the get_chanlocs wiki.

Data was recorded continuously during each session. Reference was located at Fpz (Figure 14.23). Before acquisition, impedances were improved such that acquisition started with all electrode impedances lower than $25k\Omega$. The signal was amplified, digitized at a 24-bit rate, and sampled at 500 Hz with a $0.05\mu V$ resolution. No filtering was applied during the acquisition.

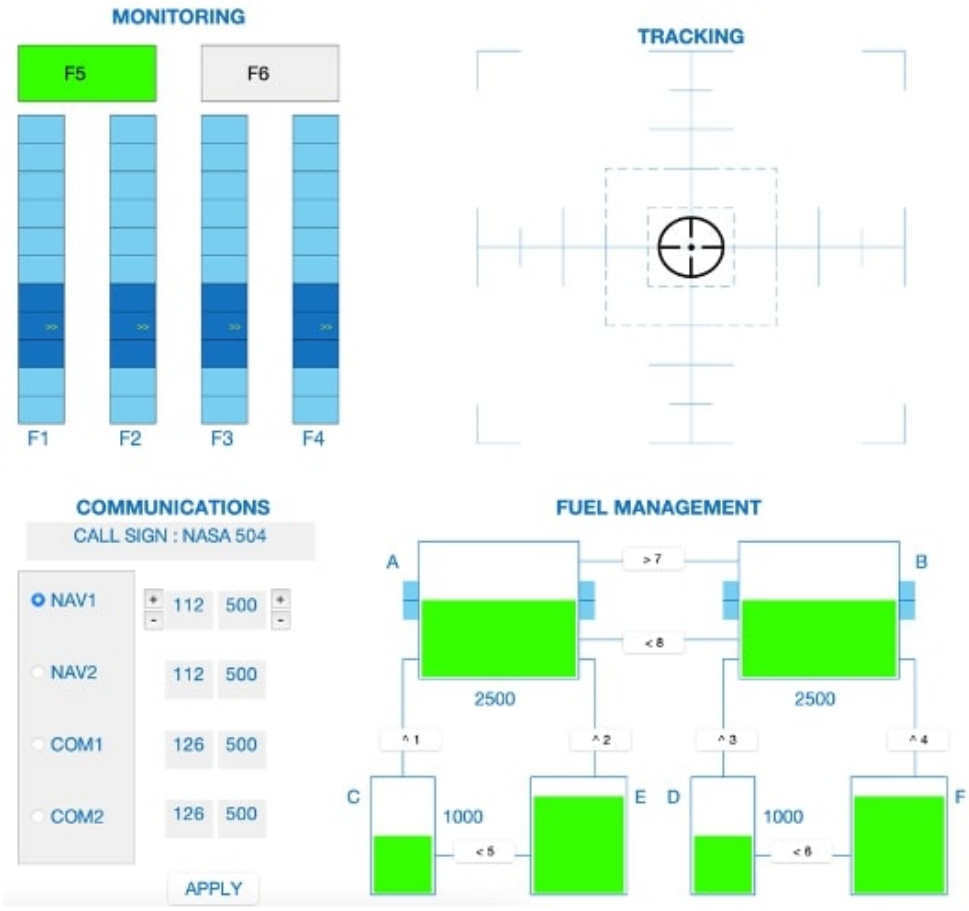

Fig. 14.22 The MATB task: Top Left: Monitoring sub-task (SYSMON), where participants have to react to alarms with button presses if any of the gauges indicate abnormal values (change in color for F5 and F6, extremely high or low values for F1 - F4). Top Right: Tracking task (TRACK) requires the participants to keep the circle in the middle of the target by controlling a joystick. Bottom Left: Communications sub-task (COMM), requiring participants to react to radio messages and to change frequencies according to the message by using the Joystick. Bottom Right: The Fuel Management task (RESMAN), where participants need to keep the reservoir level of the two main tanks at a certain level by activating and deactivating 8 pumps.

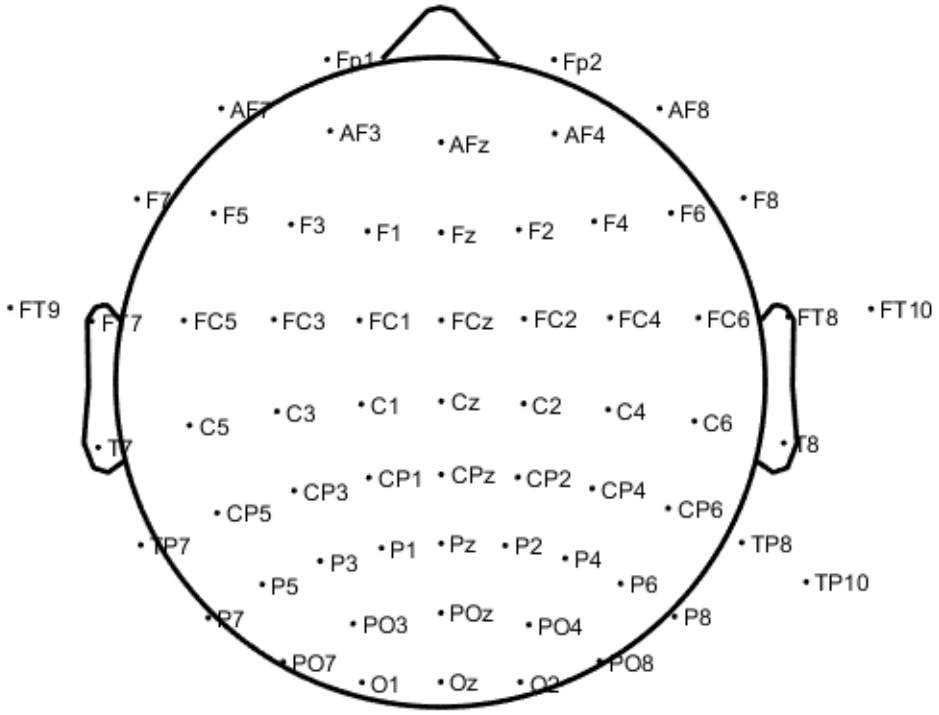

Fig. 14.23 EEG channel locations following the extended 10-20 system (Jasper, 1958). Electrode TP9 was sacrificed to record peripheral ECG.

Data Processing

To allow the community to perform their pipeline design and tests, the data that we provide is raw. Hence, only the structure and format of these data is detailed below.

Data Processing and Analysis

This section presents the analyses performed to ensure that the experimental protocol elicited the mental states of interest and that the pBCI community is aware of the relevance of the dataset. First, the processing applied to the data to enable statistical analysis of the results is presented. Next, the results obtained are presented, thanks to the statistical analyses. Lastly, proof of usability for pBCI is provided with a simple pipeline and estimation results, followed by a conclusion on this technical validation. For all reported statistics, a p-value of $<.05$ was considered significant. The Bonferroni method was applied for multiple corrections, and the reported p-values were multiplied by the factor of tests. Following the GLM, contrast analyses (t-test) were used to report statistical significance (Haans, 2019). Data Processing and Analysis were performed using MATLAB v.2021a (The Mathworks Inc.) with the EEGlab toolbox v.2022.0 (https://sccn.ucsd.edu/eeglab/index.php) and JASP v. 0.16.3 (https://jasp-stats.org/).

Cardiac activity The data from the cardiac electrode was first down-sampled to 250 Hz before a frequency band pass FIR filter was applied (1-40 Hz). The data were then epoched into

10-second segments. Using R peak detection, the heart rate (HR) as well as heart rate variability (HRV in the time domain) were computed for each epoch. HRV was measured with the Standard deviation of NN (i.e., normal R-R) intervals (SDNN) and the root mean square of the successive differences (RMSSD). Both measures have been used in neuroergonomic experiments for mental workload and vigilance evaluation (Harbour et al., 2013; Labedan et al., 2021). The RMSSD is a measure that is often used for shorter time-windows as it reflects beat-to-beat variations while the SDNN is a metric more suited for long-term analysis, reflecting overall cardiac variation (Shaffer & Ginsberg, 2017).

Cerebral activity The EEG data were processed with a usual pre-processing pipeline. Data were first down-sampled to $250Hz$. Next, a $1Hz$ high-pass FIR filter was applied. An automatic channel rejection and interpolation, based on a two standard deviation criterion was then used. On average, 0.34 channels were interpolated per task. Using the cleanline (Bigdely-Shamlo et al., 2015) toolbox, the $50Hz$ line noise was then filtered. The data were then epoched into 0.5-second segments and an automatic epoch rejection using a 2 standard deviation criterion was applied. On average, 16 epochs (8 seconds) were rejected per task. Finally, an independent component analysis (ICA) and subsequent component rejection were performed using the standard runica extended function in EEGLAB v. 2022.0 (https://sccn.ucsd.edu/eeglab/index.php) and the IClabel toolbox (Pion-Tonachini et al., 2019). The thresholds for rejecting eye, heart and muscle components were all set to >90%. On average seven components were rejected per participant and session. After the pre-processing, data was merged again into a continuous signal.

When the data was used for machine learning (i.e. proof of usability for pBCI, see Section V.14.7.2), an adapted version of this pipeline was applied. Here, the data was first epoch before the pre-processing was applied to each epoch individually. The automatic epoch rejection was, therefore, not used.

For the EEG analysis, the power in the theta ($4-8Hz$) and alpha ($8-13Hz$) bands were extracted at each electrode of the continuous data using the spectopo EEGLAB function and the absolute mean power was computed for each frequency band. Electrode clusters were selected to calculate the power by brain areas, identical to those of Simon et al. (Simon et al., 2011). For the frontal area, a cluster of 10 electrodes was averaged: F3, F1, Fz, F2, F4, FC3, FC1, FCz, FC2, FC4; for the central area, 10 electrodes were averaged: C3, C1, Cz, C2, C4, CP3, CP1, CPz, CP2, CP4; and for the parieto-occipital area, a cluster of 11 electrodes was averaged: P3, P1, Pz, P2, P4, PO3, POz, PO4, O1, Oz, O2.

Specific per-task analyses

Psychomotor Vigilance Task The behavioral analysis of the PVT focused on the reaction times of participants. It was analyzed whether there was a significant effect of time-on-task (TOT; measured per trial), Session and ISI. Cardiac analysis of the PVT focused on the normalized values of HR and HRV. A General Linear Model (GLM) with TOT (measured in the 10-second epochs) as a covariate and session as a fixed factor on the dependent variables of HRV (SDNN or RMSSD) andHR (beats per minute; BPM) was performed. For the EEG analysis, three 1-minute time windows were selected within the signal: one at the beginning, one in the middle and one at the end of the PVT (Smith et al., 2019). A Greenhouse-Geisser correction was applied when the assumption of sphericity condition was violated. A GLM with Session

(Session: 1, 2, 3) and Time window (Time window: beginning, middle, end) as fixed factors was performed on the alpha power on each Region of Interest (ROI: frontal, central and posterior).

MATB-II Task To analyze the subjective RSME scores, a $3x3$ (Session: 1, 2, 3; and Difficulty: Easy, Medium, Difficult) repeated-measures analysis of variance (RMANOVA) was performed.
As with the subjective scores RMANOVAS were performed on the performance of the tracking and the system monitoring tasks. This was done as these were the only two tasks assessed across all conditions. For the tracking task, the average absolute distance from the center of the square (root mean squared distance) was used as a performance measure. The average reaction time to alarms was used for the system monitoring task.
For all analyses, a Greenhouse-Geisser correction was applied when the assumption of sphericity was violated. Cardiac data was analyzed using a GLM (Fixed factors: Session x Condition, TOT as a covariate) on the dependent variables of HRV (SDNN or RMSS) andHR (BPM).

For the EEG analysis, the data were analyzed according to the difficulty of the task (Easy; Medium; Difficult), the cortical area (frontal and posterior) and the theta and alpha bands. A GLM (Fixed factors: Session x Difficulty) was performed on each ROI (frontal, central and posterior).

N-Back Task A $3x3x3$ RMANOVA (Session x Condition x Block) was performed on participants' error rates. A Greenhouse-Geisser correction was applied when the assumption of sphericity was violated.
The data was analyzed using the same factors for reaction times, using a GLM (dependent variable: reaction time).
Cardiac data were analyzed in a similar way as was the cardiac data of the MATB task. A GLM (Fixed factors: Session x Condition x Block) on the dependent variables of HRV (SDNN or RMSSD) andHR (BPM) was performed. Here Block was added as an additional factor. The EEG data were analyzed according to the difficulty (Easy, Medium, Difficult) of the N-back task, with a GLM (fixed factors: Session x Difficulty) on each ROI (frontal, central and posterior) and for both the alpha and theta band, respectively.

Flanker Task The behavioral results of the Flanker task were analyzed using GLMs and RMANOVAs. The first analysis (GLM) focused on reaction times as a dependent variable with congruency (congruent vs incongruent trial), session (1,2 or 3) and response accuracy (error vs correct response) as fixed factors. Furthermore, it was investigated if session or congruency affected accuracy (RMANOVA).
To analyze the EEG data, ERPs were computed for both the moment of response (when the participant reacted to the stimulus) and for the feedback (when participants received positive or negative feedback regarding the current trial). At the moment of the response, the analysis focused on the ERN amplitude, which is sensitive to accuracy (Chavarriaga et al., 2014). In the feedback time window, the analysis focused on the FRN and P300, which are sensitive to feedback valence (Nieuwenhuis et al., 2004). To determine the peak amplitude of the three components, the same procedure was applied: the most negative (for the ERN and FRN) or positive (for the P300) point in the $0-150ms$ post-response, and in the $175-325ms$ and $250-450ms$ post-feedback time-windows respectively for the ERN, FRN and P300 were identified for each epoch. Then the average around this point was computed (± 2 points for a 5-point average). To analyze the results a GLM with the Fixed factors of congruency, correctness, and

the session was performed on the dependent variables of peak amplitudes of the P300, FRN and ERN components.

Mental Workload estimation Mental Workload estimation was performed on both the N-Back and the MATB tasks, with three different mental workload levels each (0-back, 1-back, 2-back; and MATB Easy, MATB Medium, MATB Difficult). For both tasks, the same preprocessing, feature extraction and classification pipeline was used. To preprocess the data, the following steps were executed: First, 5-second non-overlapping epochs were created from the continuous data, which were then downsampled to 250 Hz. Next, FIR filters (eegfiltnew function from EEGLAB with default parameters) were applied to filter the data either in the theta or the alpha band (i.e., $[4-8]Hz$ and $[8-12]Hz$ respectively). Automatic channel rejection with subsequent interpolation of the rejected electrodes was performed. If the standard deviation of a channel was more than two standard deviations larger than that of the other channels, the channel was interpolated, using linear interpolation. A full-rank average referencing was performed to keep full rank for the following independent component analysis (ICA). Finally, an ICA with IClabel component rejection was performed (Pion-Tonachini et al., 2019). Eye, muscle and heart components with over 90% confidence were removed. Figure 14.24 a) summarizes how many components on average, were removed per task. Figure 14.24 b) details how many channels were interpolated per task. Differences between tasks concerning the amounts of components removed and channels interpolated are most likely due to differences in motor activity across conditions.

To classify the data, the covariance matrix of a subset of 10 channels (F3, Fz, F4, FCz, C3, C4, CPz, P3, Pz, P4) was calculated for each epoch. The data were then divided into training and testing sets using a 5-fold cross-validation procedure. The data were trained and tested on a Riemannian Minimum Distance to Mean (MDM) classifier (Barachant et al., 2010, 2012). Based on at least 60 trials in each of the 3 classes, the upper limit for the 33% chance level is 41.7% (Mueller-Putz et al., 2008). Accuracy was then evaluated using a RMANOVA with the factors of session (1 vs. 2 vs. 3), difficulty, and bandwidth (theta vs. alpha).

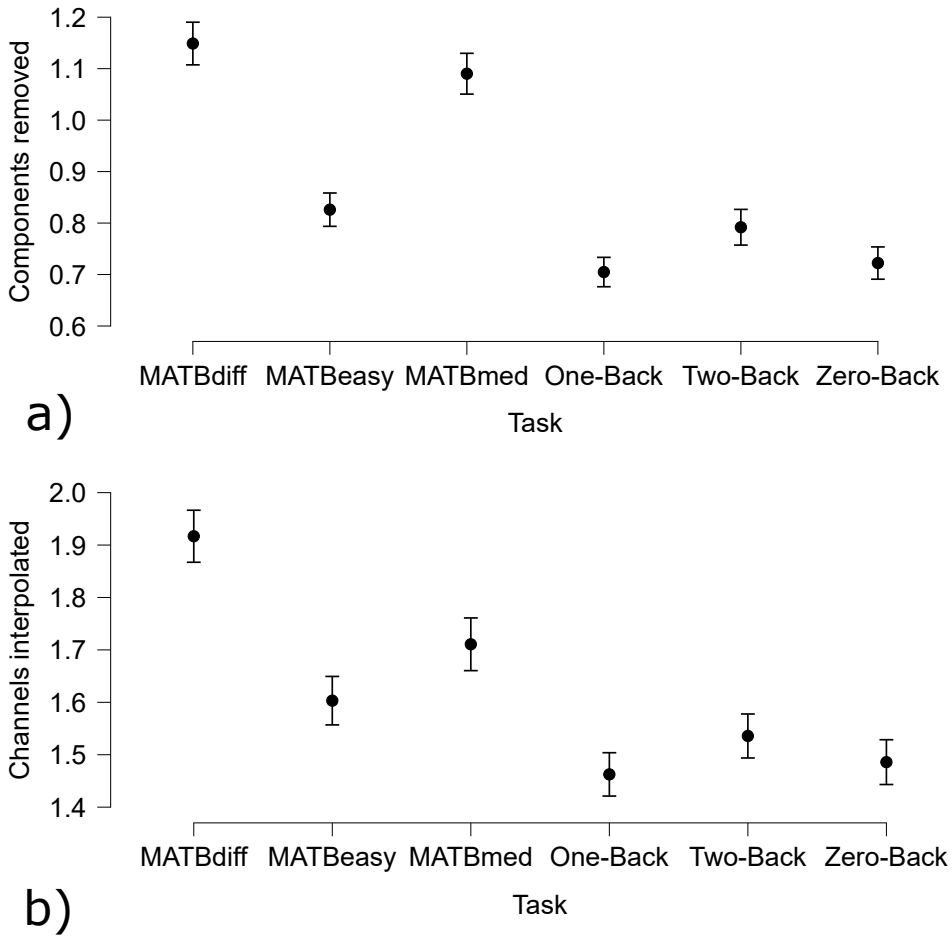

Fig. 14.24 EEG data preprocessing for the mental state estimation example: a) Average amount of removed components; b) Average amount of interpolated channels.

V.14.7.3 Validation results

Here, we present the results obtained for each task independently and the results for mental workload classification. RSME, as well as KSS scores, did not significantly change across sessions.

Psychomotor Vigilance Task

- **Behavior**: The GLM showed that ISI, TOT and session all had significant effects on the reaction time (TOT: t(28)=12.87 p<.005 ; ISI: t(28)=-19.32 p<.005; Session: t(28)=-2.31 p<.05; Figure 14.25 a). With increasing TOT, participants took significantly longer to respond to stimuli, while longer ISI resulted in quicker responses. Contrast analysis showed significantly lower reaction times during session 3 as compared to session 2 (t(28)=-2.31 p<.05).

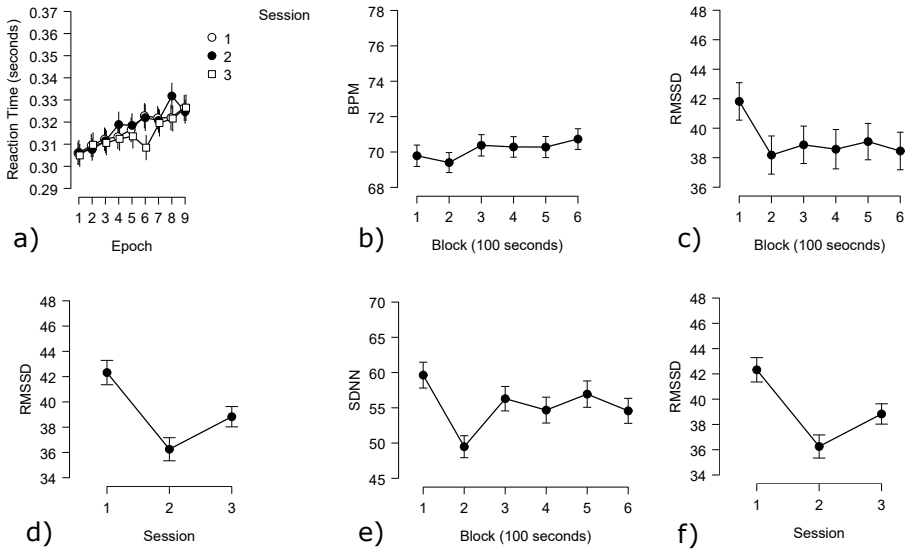

Fig. 14.25 Behavioral and cardiac PVT results: a) Average reaction time over trials (10 trials each) per session. A consistent increase in RT over sessions can be observed; b) Average BPM over blocks of 100 seconds; c) Average RMSSD (measure of HRV) over blocks; d) Average SDNN (measure of HRV) over sessions.

- **Cardiac activity**: Both TOT and Session influenced participants'HR. It increased over sessions, as well as throughout each session (Epoch: t(4946)=2.88 p<.005; Session 1 - Session 2: t(4946)=9.86 p<.001; Session 1 - Session 3: t(4946)=7.12 p<.001; Session 2 - Session 3: t(4946)=2.67 p<.01; Figure 14.25 b).

  HRV however, significantly decreased as time increased (RMSSD: Epoch: t(4945)=-2.6 p<.01). The SDNN was not sensitive to changes within sessions. As observed with the N-Back task, the lowest HRV was observed during Session 2 (RMSSD: Session 1 - Session 2: t(4945)=-9.72 p<.001; Session 2 - Session 3: t(4945)=-5.50 p<.001; Session 1 - Session 3: t(4945)=-4.19 p<.001; Figure 14.25; SDNN Session 1 - Session 2: t(4945)=-6.15 p<.001; Session 2 - Session 3: t(4945)=-3.45 p<.001; Session 1 - Session 3: t(4945)=-2.67 p<.01; Figure 14.25).

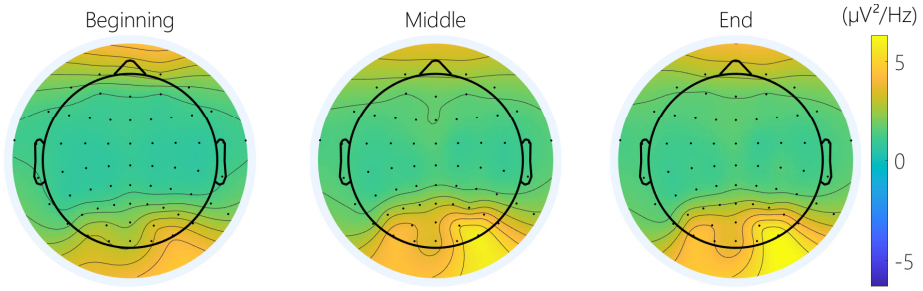

Fig. 14.26 Topography of the average alpha power over three-time windows of 1 minute, in the beginning, the middle, and at the end of the PVT.

- **Cerebral activity**: There was a significant effect of TOT, as indicated by the time win-

dows, on the frontal, central and posterior alpha power (Figure 14.26). A contrast analysis showed significantly lower alpha power at the beginning of the task, compared to the middle and the end of the task on all ROI (Beginning-Middle: frontal: t(256)= -2.51 p<.05; central: t(256)= -2.47, p<.05; posterior: t(256)= -2.47 p<.05; Beginning-End: frontal: t(256)= -2.51 p<.05; central: t(256)= -2.47, p<.05; posterior: t(256)= -2.47 p<.05; Figure 14.27).

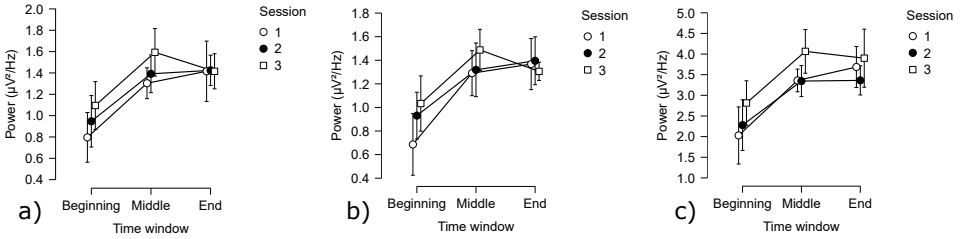

Fig. 14.27 Cerebral PVT results: a) Average alpha power in the frontal area; b) Average alpha power in the central area; c) Average alpha power in the posterior area

MATB-II

- **RSME questionnaire**: RMSE scores were dependent on both session and difficulty (Session: F(2,104)=17.15 p<.005; Difficulty: F(2,104)=92.73 p<.005; Figure 14.28 a). Across sessions, RSME scores decreased, while increasing difficulty was reflected in higher scores.

- **Behavior**: The TRACK scores (average absolute distance from midpoint) showed significant effects of difficulty as well as session (Session: F(2,56)= 27.63 p<.005; Difficulty: F(2,56)=421.99 p<.005). The interaction was also significant (Interaction F(4,112)=23.12 p<.005; Figure 14.28 b). In general, the difficult condition showed much higher deviation scores as compared to the other conditions
. However, an increase in performance was observed across sessions.
Analysis of the SYSMON scores (average reaction times to alarms) showed significant effects of difficulty as well as session (Session: F(2,56)= 15.17 p<.005; Difficulty: F(2,56)=62.19 p<.005). The interaction was also significant (Interaction F(4,112)=6.86 p<.005; Figure 14.28 c). As with the TRACK task, increase in difficulty resulted in lower performances as seen by higher reaction times. Across sessions, reaction times decreased in all sub-tasks.

- **Cardiac activity**:HR was found to be significantly affected by difficulty and TOT (Difficulty: Easy - Medium t(6860)=-4.11 p<.001; Easy - Difficult t(6860)=7.44 p<.001; Medium - Difficult t(6860)=-11.51 p<.001; Figure 14.28 d). Contrast analysis showed that heart rate increased in the difficult condition, whereas the medium condition had the lowest heart rates. HR was shown to increase across sessions (Session 1 - Session 2 t(6860)=6.27 p<.001; Session 1 - Session 3 t(6860)=7.72 p<.001). However, the increase from session 2 to session 3 was not significant. Within sessions, an increase in heart rate was observed over time (TOT: t(6860)=8.47 p<.001).

  HRV, as measured by RMSSD, was significantly influenced by difficulty, across and within sessions. In the difficult condition, significantly lower RMSSD values were observed as compared to the medium and easy conditions. Session 1 differed significantly from both sessions 2 and 3 (Difficulty: Easy - Difficult t(6860)=-9.89 p<.001; Medium - Difficult t(6860)=9.75 p<.001; TOT: t(6860)=-4.72 p<.001; Session 1 - Session 2 t(6860)=-4.78 p<.001;

Session 1 - Session 3 t(6860)=-4.85 p<.001 ; Figure 14.28 e). The heart rate variability was significantly lower in the difficult condition. During the first session, the variability was higher than during the other two sessions and within sessions, the HRV decreased.

The results of the analysis of HRV as measured by SDNN showed similar results as the analysis using RMSSD. The difficult condition showed the lowest variability, while the easy condition had the highest SDNN values. SDNN values observed during session 1 were significantly higher as compared to the values observed during session 2(Difficulty: Easy - Difficult t(6860)=-16.29 p<.001; Easy - Medium t(6860)=-5.54 p<.001; Medium - Difficult t(6860)=10.74 p<.001; Session 1 - Session 2 t(6860)=-2.24 p<.001; Figure 14.28 e). Within sessions a decrease in HRV was observed (TOT: t(6860)=-3.32 p<.001).

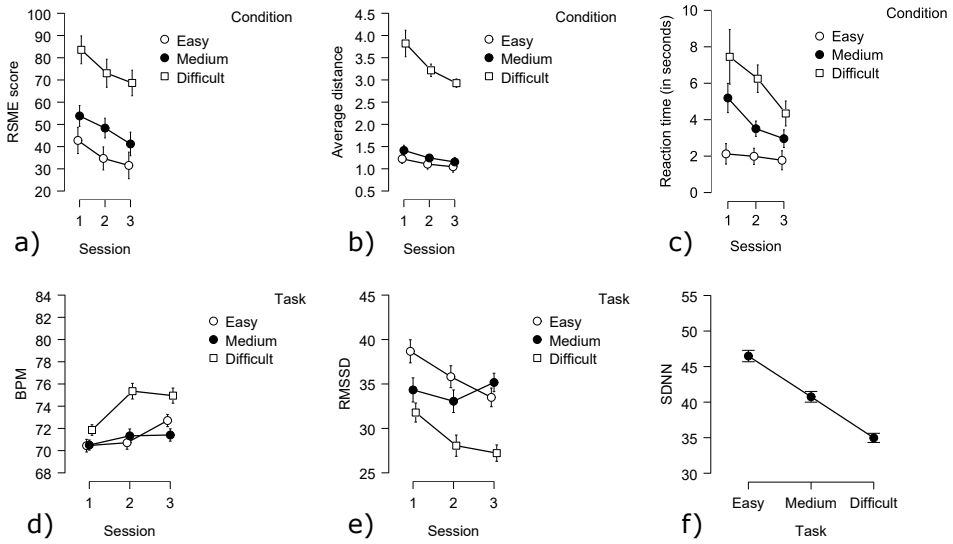

Fig. 14.28 Behavioral and cardiac MATB results: a) RMSE score divided by condition and by session; b) Average absolute TRACK distance divided by condition and by session; c) SYSMON reaction time divided by condition and by session d) BPM divided by condition and by session; e) RMSSD (measure of HRV) across conditions; f) SDNN (measure of HRV) across conditions.

- **Cerebral activity**: There was a significant effect of difficulty on the frontal and posterior theta power (Figure 14.29). A contrast analysis revealed that theta power during the easy condition was significantly lower compared to the medium and difficult conditions in both areas (Easy-Medium: frontal: t(256)=-4.35 p<.001; posterior: t(256)=-5.74 p<.001; Easy-Difficult: frontal: t(256)=-3.54 p<.01; posterior: t(256)=-6.18 p<.001; Figure 14.31). Concerning the alpha power, there was a significant effect of difficulty in the posterior area (Figure 14.30). A contrast analysis revealed that the alpha power was significantly lower in the easy than the difficult condition (Easy-Difficult: t(256)=-3.45 p<.01).

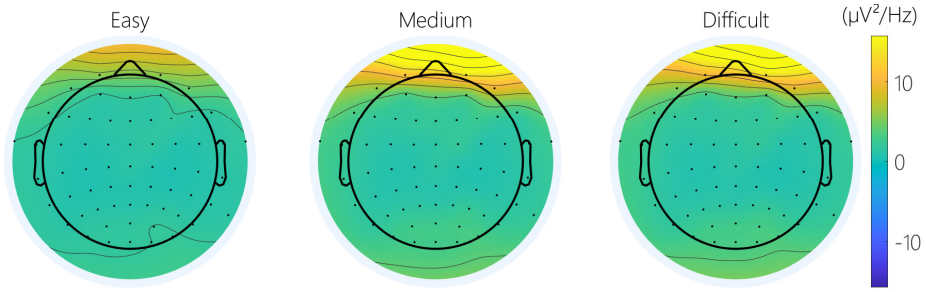

Fig. 14.29 Topography of the average theta power across difficulty on the MATB.

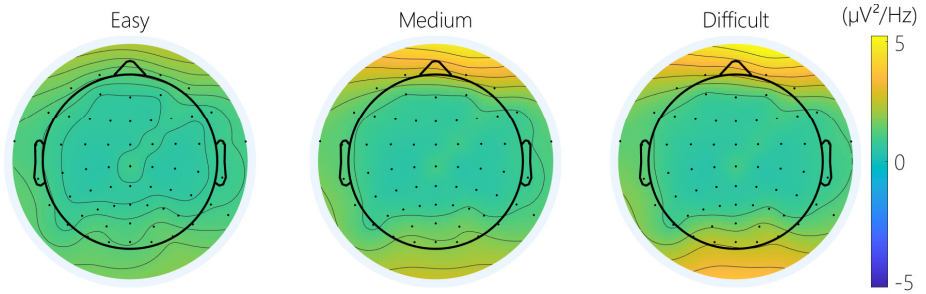

Fig. 14.30 Topography of the average alpha power across difficulty on the MATB.

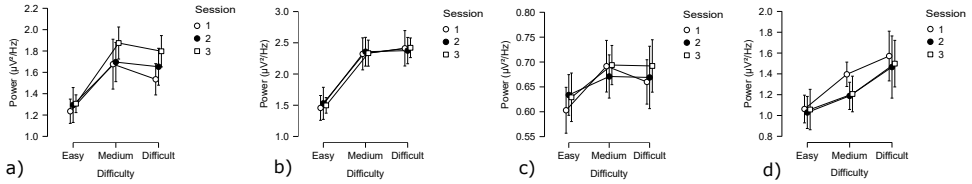

Fig. 14.31 Cerebral MATB results: a) Average theta power in the frontal area; b) Average theta power in the posterior area; c) Average alpha power in the frontal area d) Average alpha power on the posterior area.

N-Back task

- **RSME questionnaire**: RSME scores were significantly impacted by difficulty and session (Difficulty: F(2,104)=55.56 p<.005; Session F(2,104)=3.52 p<.05; Figure 14.32 a). An increase in difficulty was reflected by higher scores.

- **Behavior**: All GLM factors (session, condition, block and error) had significant main effects on reaction time (Session: t(28)=-8.89 p<.005; Condition: t(28)=57.53 p<.005; Block t(28)=6.29 p<.005; Error: t(28)=-10.29 p<.005; Figure 14.32 b). Reaction time decreased across sessions and increased within sessions. Errors were accompanied by longer reaction times. More difficult conditions (higher N) resulted in increases in reaction time. Regarding accuracy, a significant effect of condition on error rates was found (Condition: F(2,56)=52.25 p<.005; Figure 14.32 c). More difficult conditions resulted in lower accuracy.

- **Cardiac activity**: Heart rate was found to be significantly higher during the 2-back condition (Difficulty: 0-back - 2-back t(6953)=8.60 p<.001; 1-back - 2-back t(6953)=-7.52 p<.001; Figure 14.32 d). The lowest heart rate was observed during session 1, whereas the highest heart rate occurred during Session 2 (Session: Session 1 - Session 2: t(6953)=13.96 p<.001; Session 1 - Session 3: t(6953)=7.78 p<.001; Session 2 - Session 3: t(6953)=6.12 p<.001). Block did not have any significant effect on heart rate.

  HRV as measured by SDNN was influenced by difficulty , session and block (Difficulty: 0-back - 1-back t(6953)=-4.02 p<.001; 1-back - 2-back t(6953)=4.82 p<.001; 0-back - 2-back t(6953)=-8.75 p<.001; Session: Session 1 - Session 2: t(6953)=-9.45 p<.001; Session 2 - Session 3: t(6953)=-9.78 p<.001; Block: Block 1 - Block 2 t(6918)=-4.02 p<.001; Block 1 - Block 3 t(6918)=-6.67 p<.001; Figure 14.32 f). HRV was lower for more difficult tasks and decreased within each session. During session 2 a significantly lower HRV was observed. The first block had a significantly higher HRV as compared to both upcoming blocks.

  The results of the RMSSD metric largely confirmed these results. Difficulty, session and block showed significant effects on the RMSSD (Difficulty: 0-back - 2-back: t(6918)=-5.53 p<.001; 1-back - 2-back: t(6918)=4.18 p<.001; Session: Session 1 - Session 2: t(6953)=-11.07 p<.001; Session 1 - Session 3: t(6953)=-5.32 p<.001; Session 2 - Session 3: t(6953)=-5.71 p<.001; Block: Block 1 - Block 2 t(6918)=-4.02 p<.001; Block 1 - Block 3 t(6918)=-3.62 p<.001; 14.32 e). More difficult conditions resulted in a lower RMSSD value. As with the SDNN the lowest RMSSD value was observed during the second session. In addition session 1 had a significantly higher RMSSD as compared to session 3. The HRV (as observed by the RMSSD) was significantly higher during the first block as compared to the other two blocks.

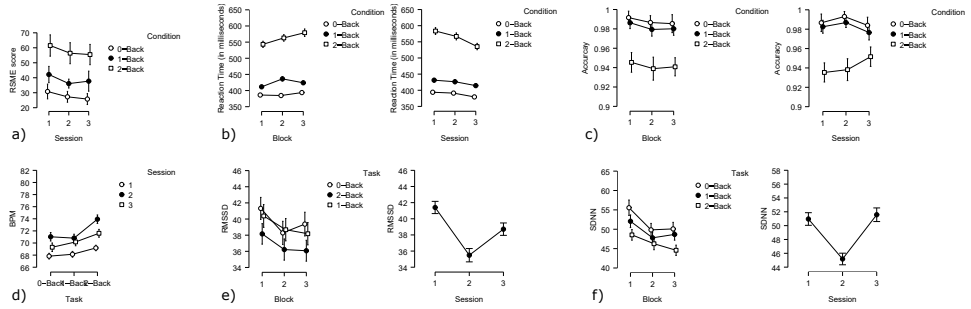

Fig. 14.32 Behavioral and cardiac N-Back results: a) RMSE score divided by condition and by session, showing an increase in subjective workload with increasing difficulty and a slight decrease in subjective workload over sessions b) Reaction time across sessions divided by condition and Reaction time within sessions divided by condition. Over sessions participants improved in performance (lower reaction time). Within sessions the reaction time of participants increased c) Accuracy (1=100 %) across conditions; d) BPM divided by condition and session; e) RMSSD (measure of HRV) across conditions and sessions; f) SDNN (measure of HRV) across conditions and sessions.

- **Cerebral activity**:

  Analysis of the EEG power showed that there was a significant effect of the session in the theta band on the frontal area (Figure 14.33). A contrast analysis revealed that theta

power during the third session was significantly higher compared to the first and second sessions in the frontal area (Session 1 - Session 3: t(256)=2.12 p<.05). No effect of difficulty on theta power was found. Also, no significant effects of the session nor difficulty were found on alpha power for this task.

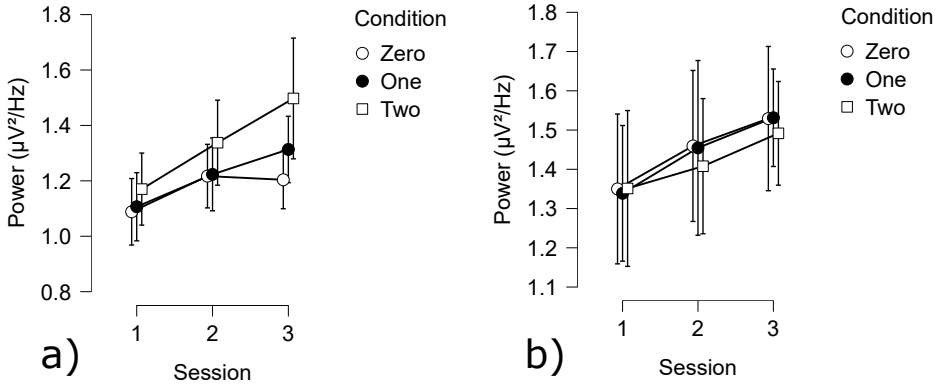

Fig. 14.33 Cerebral N-Back results: Average theta power at a) the frontal area; b) the posterior area.

Flanker task

- **RSME questionnaire**: RSME scores were significantly higher during session 1 as compared to the session 2 (t(54)=3.38 p<.005).

- **Behavior**: All main factors (response, congruency and session as well as the covariate TOT) showed significant effects on the reaction time to the Flanker trials (Response: t(10433)=46.84 p<.001; Congruency: t(10433)=-15.48 p<.001; Session: t(10433)=-14.83 p<.001; Figure 14.34 a,b). Contrast analysis revealed significant differences between: correct incongruent - correct congruent (t(10433)=23.03 p<.001); incorrect incongruent - correct congruent (t(9980)=11.29 p<.001); correct incongruent - incorrect congruent (t(9980)=15.65 p<.001). Reaction times did not change due to correct responses for incongruent trials. However Congruent correct trials were significantly faster than responses to all other trials. Congruent Incorrect responses were significantly slower than all other trials. The analysis with accuracy as a dependent variable showed significant main effects of congruency as well as session (Condition: F(2,28)=20.79 p<.005; Session: F(2,56)=10.33 p<.005; Figure 14.32 c). Participants' accuracy was significantly lower in incongruent trials. Their accuracy improved across sessions.

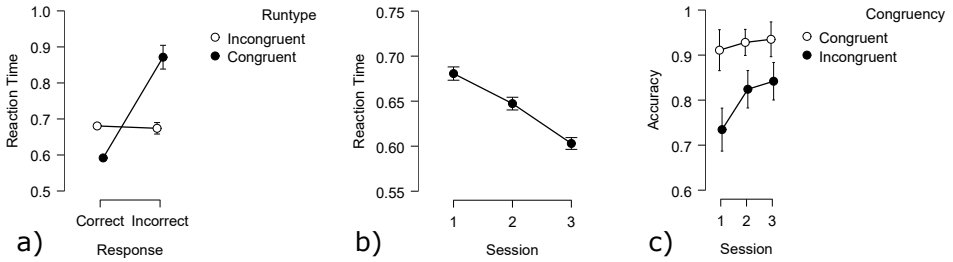

Fig. 14.34 Behavioral Flanker results: a) Reaction time for Correct and Incorrect trials divided by Congruent and Incongruent trials; b) Reaction time across sessions; c) Accuracy across conditions and sessions.

- **Cerebral activity**: Following the response of the participant, the event-related negativity ($ERN_{response}$) amplitude was sensitive to correctness, session and congruency (Correct: t(9941)=-3.51 p=.001; Session 1 - Session 2 t(9941)=-1.97 p<.05; Session 1 - Session 3 t(9941)=-5.24 p<.001; Session 2 - Session 3 t(9941)=3.29 p<.001; Congruency t(9941)=4.93 p<.001). Across sessions, the ERN value decreased. A contrast analysis of the interaction effects between Correct/Error and Incongruent/Congruent revealed several significant effects at the electrode site FCz: Incorrect incongruent - Correct congruent (t(9941)=3.51 p<.005); Incorrect incongruent - Incorrect congruent (t(9941)=-4.9 p<.001); incorrect congruent - correct congruent t(9941)=-5.02 p<.001). Incongruent trials have a significantly lower amplitude during correct trials as compared to congruent correct trials.(Figure 14.35 a & c). There was also a significant effect of congruency and session at the Fz electrode (Session 1 - Session 2 t(9941)=9.04 p<.001; Session 1 - Session 3 t(9941)=9.55 p<.001; Congruency t(9980)=3.11 p<.005). Contrast analysis showed the same effect as with electrode FCz: Incongruent trials had a significantly lower amplitude during correct trials as compared to congruent correct trials (correct congruent - correct incongruent t(9941)=-3.11 p<.01). The amplitude of session 1 was significantly lower as compared to the other sessions (Figures 14.35 b & d, Figure 14.39).

  When participants received feedback for the preceding trial, the feedback-related negativity (FRN) showed to be affected by multiple factors. At electrode FCz Correctness, congruency as well as sessions had an effect on the amplitude of the FRN (Correct: t(9941)=-13.24 p<.001; Congruency: t(9941)=2.95 p<.005; Session 1 - Session 2 t(9941)=-7.79 p<.001; Session 1 - Session 3 t(9941)=-12.42 p<.001; Session 2 - Session 3 t(9941)=4.65 p<.001) Correct responses resulted in a lower amplitude, while congruency increased the amplitude. Across sessions, the amplitude decreased. (Figures 14.36 a & c). Fz also showed more negative amplitudes following a correct response (Correct - Error: t(9941)=-8.07 p<.001; Figure 14.36 a and b), while congruency increased the amplitude (Congruent - Incongruent: t(9941)=2.46 p<.05). Across sessions the amplitude decreased (Session 1 - Session 2: t(9941)=-5.69 p<.001; Session 1 - Session 3 t(9941)=-10.29 p<.001; Session 2 - Session 3 t(9941)=4.61 p<.001; Figures 14.36 b & d).

  The P300 amplitude during feedback was also seen to be affected by several conditions. The P300 amplitude at electrode Fz was influenced by session, congruency and accuracy(Correct - Error: t(9941)=-34.4 p=.001; Session 1 - Session 2: t(9941)=-6.72 p<.001; Session 1 - Session 3: t(9941)=-8.97 p<.001; Session 2 - Session 3: t(9941)=2.26 p<.05; Congruent - Incongruent: t(9941)=3.07 p<.005). The P300 component was significantly larger

during incorrect trials and incongruent trials while decreasing across sessions (Figures 14.37 b & d, 14.38). The P300 component at electrode FCz was influenced by congruency, accuracy and session (Correct - Error: t(9941)=-12.0 p=.001; Session 2 - Session 3: t(9941)=-2.58 p=.01; Congruent - Incongruent: t(9989)=3.02 p<.005).The amplitude was larger for incorrect trials, as well as for congruent trials. The amplitude during session 3 was significantly lower as compared to session 2 (Figure 14.37 a & c).

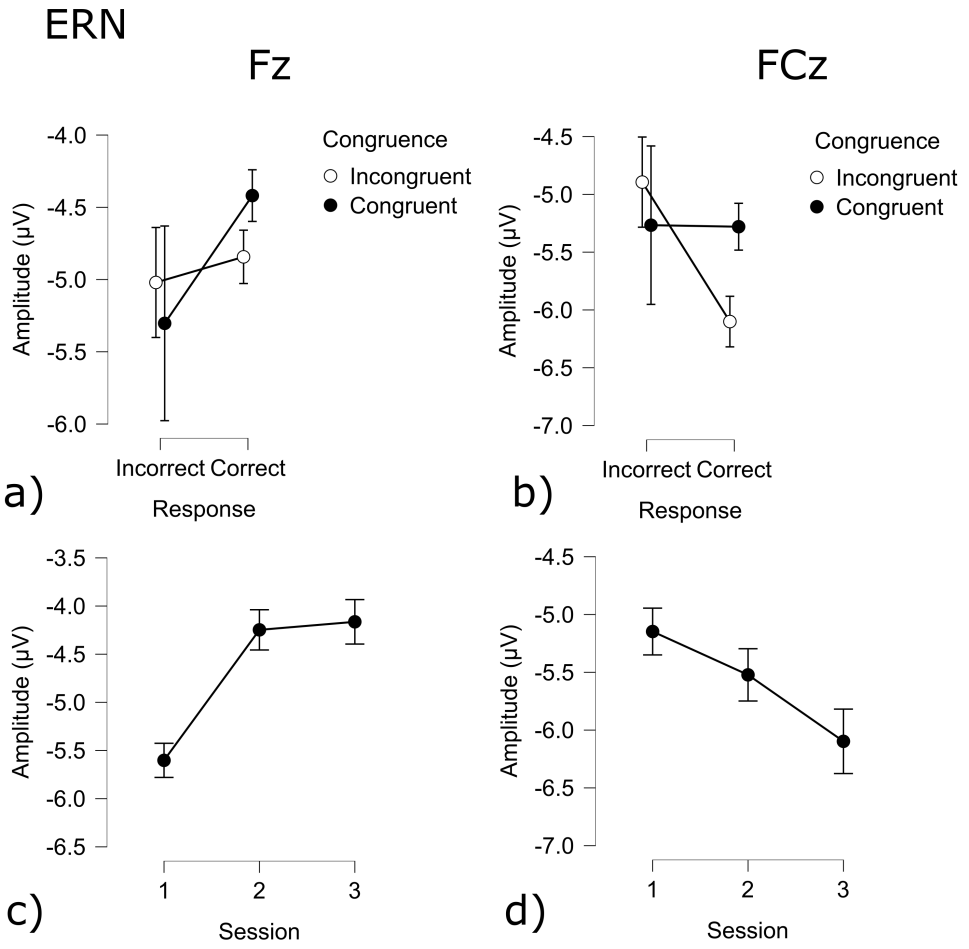

Fig. 14.35 ERN results for the Flanker task: a) Congruency and Response effect at the Fz electrode; b) Congruency and Response effect at the Cz electrode; c) Session effect at the Fz electrode; d) Session effect at the FCz electrode

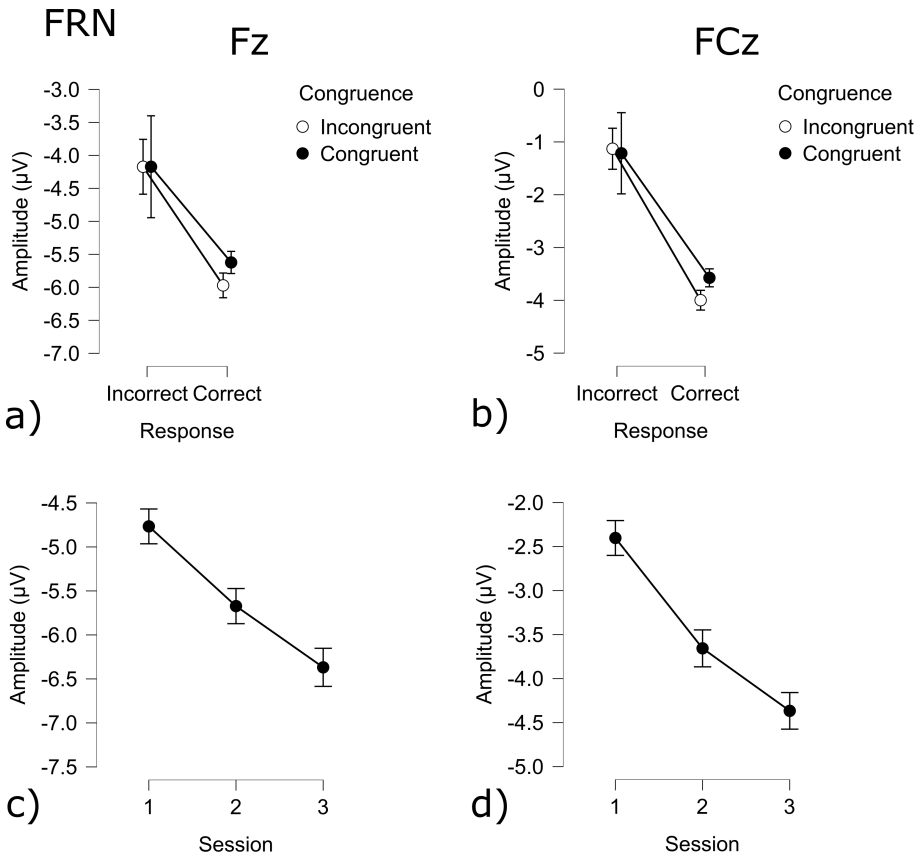

Fig. 14.36 FRN results for the Flanker task: a) Congruency and Response effect at the Fz electrode; b) Congruency and Response effect at the Cz electrode; c) Session effect at the Fz electrode; d) Session effect at the FCz electrode

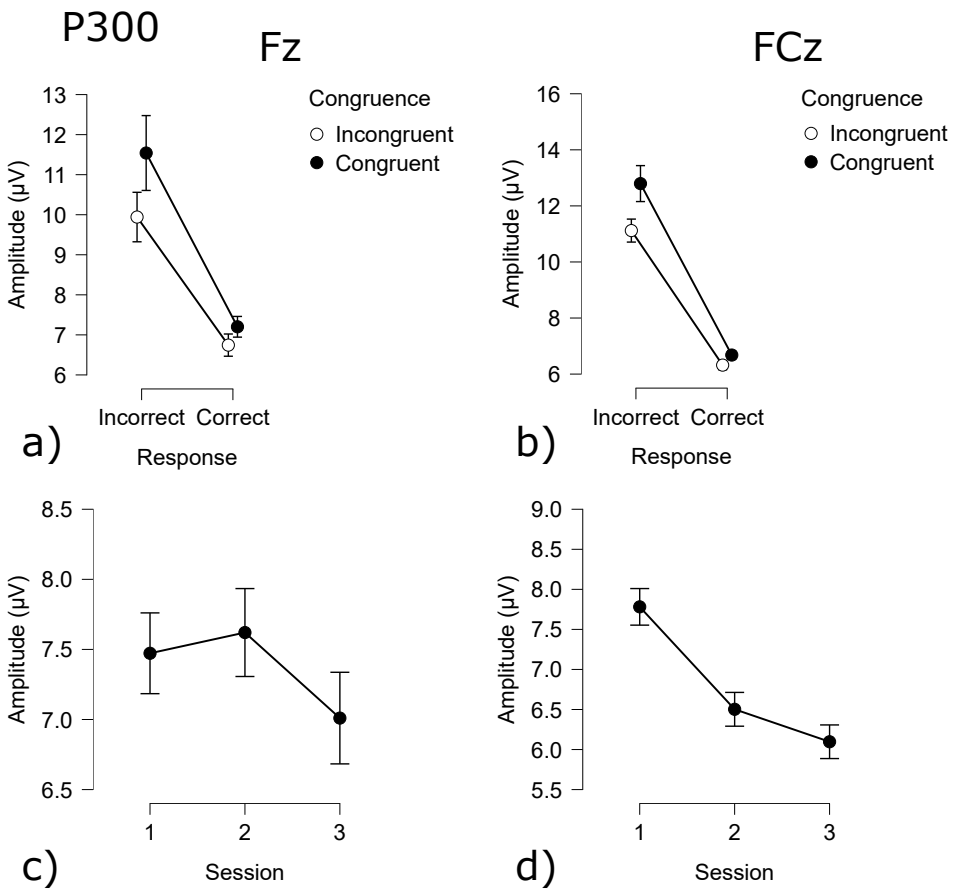

Fig. 14.37 P300 results for the Flanker task: a) Congruency and Response effect at the Fz electrode; b) Congruency and Response effect at the Cz electrode; c) Session effect at the Fz electrode; d) Session effect at the FCz electrode

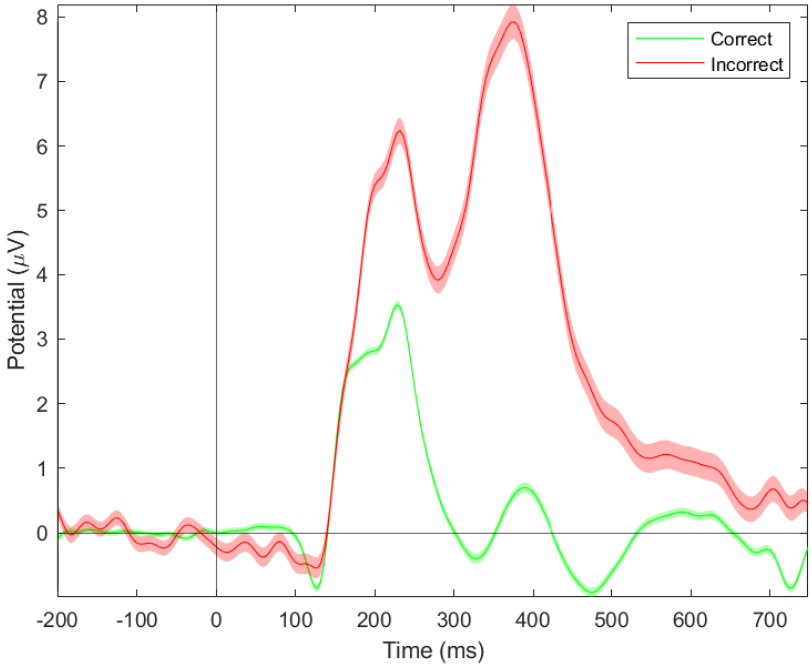

Fig. 14.38 Event-related potentials elicited by feedback (correct/positive and incorrect/negative)
at the FCz electrode (see ample P300 component in the incorrect condition).

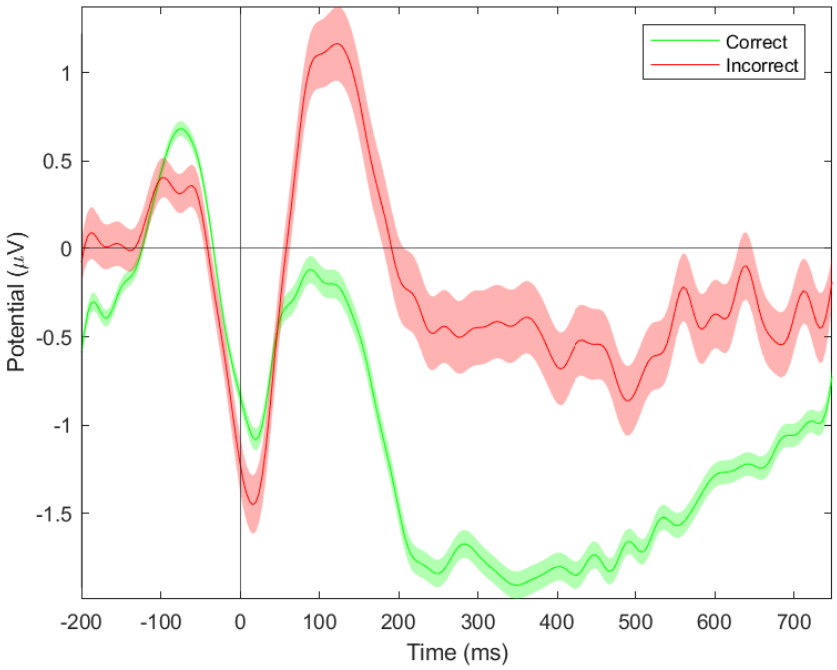

Fig. 14.39 Error-related potentials time-locked to the participants' response at the FCz electrode during an arrow-based flanker task. Grand averages across participants show the ERP in response to correct (green) and incorrect (red) trials with the Error-Related Negativity peaking 30ms post-response.

Mental Workload estimation

The estimation of mental workload using the Riemannian MDM classifier proved capable of detecting different levels of mental workload for both the MATB and the N-Back tasks. The RMANOVA with accuracy as a dependent variable and bandwidth (alpha or theta band), Session (1, 2, 3) and Task (MATB and N-BACK) showed several significant effects. With 69.40% (±12.50%), the accuracy for the MATB task was significantly higher than the accuracy of the N-Back (64.97 ± 12.99; F(1,28)=13.00 p<.001; Figure 14.40 a). Using the data filtered in the alpha band resulted in significantly higher accuracies (68.41 ± 12.37%) as compared to the theta band (65.97 ± 13.37%; F(1,28)=7.95 p<.05; Figure 14.40 a & c). The accuracies also changed across sessions (F(2,56)=4.96 p<.05; Figure 14.40 b). Accuracy was highest during the second session (70.30 ± 12.73%) followed by the first session (64.30 ± 12.52%) with the lowest accuracy during the third session (66.96 ± 12.87%). Multiple comparisons between the three tests showed a significant difference between session 1 and session 2 (t(28)=-3.14 p<.01). The interaction effect between Bandwidth and Session was also significant (F(2,56)=3.50 p<.05; Figure 14.40 b). The interaction shows a more pronounced difference between classification performances of the alpha and theta band during the first session (Alpha Session 1 - Theta Session 1: t(28)=3.67 p<.001). The strongest performance occurred with the alpha band during the second session (accuracy of 70.67%).

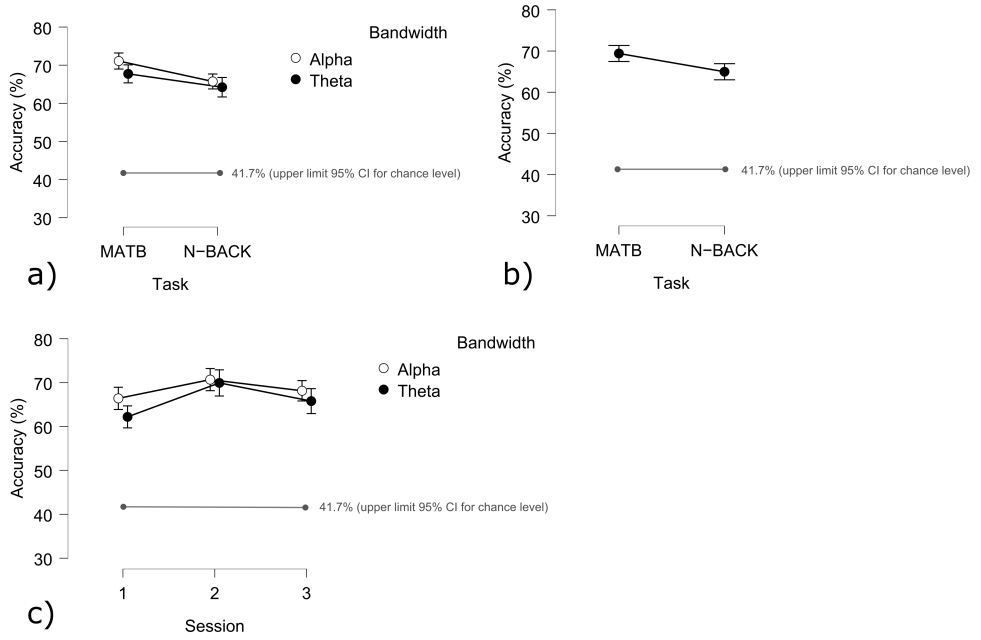

Fig. 14.40 Results of the Mental Workload estimation: a) Accuracy of estimation based on the alpha and theta band per task (i.e. MATB and N-back); b) Accuracy of estimation based on the alpha and theta band per session (i.e. session 1, 2 or 3).: c) Accuracy of the estimation based on the alpha and the theta band.

V.14.7.4 Conclusion of the dataset validation

To validate the acquired dataset, the tasks performed were assessed regarding their capability to elicit the well-known and expected effects on cognition. This validation occurred at three levels. Subjective levels of mental workload and vigilance decrement were assessed by the KSS and RSME subjective scales (Ghanbary Sartang et al., 2016; Kaida et al., 2006). Behavioral performance for each task was obtained in the form of accuracies and reaction times. Physiological changes were assessed using cardiac (ECG) and cerebral (EEG) activity measurements.

Regarding the Psychomotor Vigilance Task, it was expected that over its 10-minute duration the vigilance of participants would decrease (Lamond et al., 2005; Reifman et al., 2018). This was shown as reflected by the increase in reaction times as well as the increase in heart rate and decrease in HRV with increasing TOT. Concerning the cerebral activity, it was shown that slow EEG wave activity, including alpha, increases throughout the cortex as the person fatigues (Craig et al., 2012). The results obtained here confirm this effect; namely, the increase in alpha power with increasing TOT further confirms the vigilance decrement.

The N-Back task is a popular task for eliciting different levels of mental workload by taking working memory, while keeping visual stimuli as well as motor responses constant (Brouwer et al., 2012; Jaeggi et al., 2010). As shown here, differences in mental workload were observed at the subjective level with higher RSME scores (with increasing difficulty/mental workload), at the behavioral level with decreased performance with increasing difficulty, as well as at the physiological level with increased heart rates and decreased HRV during more difficult conditions. The current analyses revealed an effect of the session on cerebral activity. Theta power increased with sessions in the frontal area. This result suggests that task engagement increased over the sessions (Berka et al., 2007) – in accordance with the increase in performance.

The MATB task represents an approach towards a more ecologically valid task for eliciting different levels of mental workload (Santiago-Espada et al., 2011). As with the N-Back task, the analysis showed increasing mental workload at the subjective, behavioral and physiological levels as difficulty increases. As with the N-Back, more difficult conditions resulted in a higher heart rate and a lower heart rate variability. The expected results concerning the modulation of brain activity during MATB was a peak of activity in the frontal theta band as the difficulty increased (Smith et al., 2001). Regarding alpha power, a decrease in amplitude in the posterior sites was expected with increasing difficulty (Smith et al., 2001). The cerebral analysis found an increase in theta power during more difficult conditions. While the alpha power was sensitive to changes in mental workload the results diverge from the traditionally seen decrease in alpha power with increasing difficulty. The opposite effect was observed here. However, behavioral and subjective analysis and theta power confirm different levels of mental workload across conditions of varying difficulty. The Flanker task is a decision-making task that involves dealing with conflictual information (Eriksen & Eriksen, 1974). Commonly it is observed that incongruent trials result in less accurate and slower responses. On a cerebral level, it has been observed that errors result in stronger deflections of the ERN, FRN and P300 components. The behavioral results show a decrease in accuracy and an increase in reaction times for incongruent trials, compared to congruent trials (Chavarriaga & Millan, 2010; Chavarriaga et al., 2014). Furthermore, it was shown that errors resulted in significantly longer reaction times. The difference between erroneous and correct trials was also significant at the physiological level, as represented by the sensitivity of the ERN, FRN, and P300 components to error sessions and congruency. In addition to these validation measures, the collected data was also tested in its ability to be used for mental workload estimation. On both the MATB and the N-Back task, an MDM Riemannian classifier was used to classify epochs of EEG data between three levels of mental workload. The obtained average accuracy of over 65% in this 3-class problem (more than 20% above the upper limit of a 95% CI for chance level) confirms that the COG-BCI database (Hinss, Jahanpour, Somon, et al., 2022b) may be of use for developing and testing classification pipelines. A higher accuracy was reached on the MATB-II data, which may result from increased activity of the sensory-motor areas depending on task difficulty. In the N-back task, it is expected, that all conditions result in the same amount of motor activity. Overall, the results show that the COG-BCI database (Hinss, Jahanpour, Somon, et al., 2022b) is useful for testing novel methods within passive BCI research or related fields. It should be noted that the analysis presented here does not in any form present an exhaustive analysis of all effects that may be present within the data, as well as of all pipelines that may be designed to estimate users' cognitive state. The analysis aimed to validate the collected dataset's usability for working on passive BCIs and pave the way towards open science practices in pBCI research and development.