

Multimodal Biosignal Analysis of Attention and Stress in Task-Based Environments

An Experimental, Data-Driven Study Using EEG and Peripheral Biosignals

Bachelor Thesis

B.Sc. General Engineering, Technical University of Denmark

Authors:

Natalie Wilske
Kawsar Nabawi

Supervisors:

Sadasivan Puthusserypadky Kumaran
Jacopo Mattia de A Constantini



Multimodal Biosignal Analysis of Attention and Stress in Task-Based Environments
An Experimental, Data-Driven Study Using EEG and Peripheral Biosignals

Bachelor Thesis
January, 2026

By
Kawsar Nabawi
Natalie Wilske

Supervisors:
Sadasivan Puthusserypady Kumaran
Jacopo Mattia de A Constantini

Copyright: Reproduction of this publication in whole or in part must include the customary bibliographic citation, including author attribution, report title, etc.

Cover photo: Pawel Czerwinski, 2021

Published by: DTU, Department of Health Technology, Ørstedes Plads, Building 345B, 2800 Kgs. Lyngby Denmark
www.healthtech.dtu.dk/

ISSN: [0000-0000] (electronic version)

ISBN: [000-00-0000-000-0] (electronic version)

ISSN: [0000-0000] (printed version)

ISBN: [000-00-0000-000-0] (printed version)

Approval

This thesis has been conducted over a period of four months at the Research Group for Biomedical Signal Processing within the Section for Digital Health, Department of Health Technology, at the Technical University of Denmark (DTU), in partial fulfillment of the requirements for the Bachelor of Science degree in General Engineering, specializing in Cyber Systems.

It is assumed that the reader has basic knowledge in signal processing, physiological measurement, and basic statistical and machine learning methods.

Kawsar Nabawi

Natalie Wilske

.....
Signature

24.01.2026

.....
Date

.....
Signature

24.01.2026

.....
Date

Abstract

Attention and stress are two distinct yet related cognitive states that influence performance and well-being. This study explores how neural and peripheral biosignals reflect these states during controlled cognitive tasks. Data was collected during a designed experiment encompassing both stress-dominated and attention-dominated tasks. Collection tools included two EEG devices, namely the Emotiv MN8 headset and the OpenBCI Headband, and a physiological biosensor (EmotiBit). The signals collected included electroencephalography (EEG), heart activity (HR, HRV), electrodermal activity (EDA), and skin temperature (ST).

Alpha-band EEG activity was used to characterize attentional involvement, while peripheral signals captured autonomic stress response. Moreover, subjective ratings from the participating individuals regarding cognitive load, attention, and stress were collected after each task to support the interpretation of biosignals and to quantitatively and qualitatively understand perceived task experience and difficulty.

The results uncover a clear dissociation between the neural and peripheral measures. EEG was most sensitive to attentional demands, whereas peripheral signals reacted to physiological changes reflecting stress. High cognitive demand did not necessarily lead to high stress, as attention-focused tasks showed weaker autonomic responses than stress-inducing tasks such as breath-holding.

An exploratory machine learning analysis evaluated task-level classification of cognitive states. EEG-based features performed best for general attention classification, while stress was mainly classified using HRV and EDA. Specifically in machine learning, multimodal fusion did not consistently improve results, suggesting that modality choice matters more than modality count.

The work of this project thesis concludes with a comprehensive and reproducible framework for a task-based experiment, designed for attention and stress distinction. Furthermore, it clearly identified the key signals and metrics for cognitive state distinction and successfully classified the tasks based on attention and stress levels.

Acknowledgements

We would like to express our sincere and heartfelt thanks to everyone who contributed to the completion of this thesis.

First and foremost, we are deeply grateful to all the participants who took part in this study. Their time, patience, and willingness to participate made this work possible. We truly appreciate the effort they put into completing the experimental sessions and taking the time to participate, especially during exam season.

We would also like to extend our enormous gratitude to our supervisors for their guidance, support, and constructive feedback throughout this project. Their expertise and encouragement were invaluable at every stage of the work.

We would also like to express our sincere thanks to Helena for her design contributions, and to Anja and Jan for their careful reviewing and proofreading of the thesis. We are also grateful to Deniz for his help with executing the experimental sessions. In addition to their practical contributions, all of them provided invaluable moral support throughout the process, which we greatly appreciate.

We would also like to acknowledge our own efforts throughout this project, including the persistence, teamwork, and many hours of problem-solving it required. Completing this thesis was a challenging but rewarding journey, and we are proud of what we achieved together.

Finally, we would like to thank our family and friends for their constant support, understanding, and encouragement throughout this process. Their patience and belief in us, especially during demanding periods, meant more than we can express.

Table of Contents

Preface	ii
Abstract	iii
Acknowledgements	iv
List of Abbreviations	viii
1 Introduction	1
1.1 Introduction and Motivation	1
1.2 Research Questions and Objectives	2
1.3 Thesis Structure	3
Part I: Research	4
2 Literature Review	4
2.1 Literature Search Methodology	4
2.2 Background & Theory	6
2.3 Related Work Synthesis	9
Part II: Methodology	12
3 Experimental Design Overview	12
4 Hardware Setup	13
4.1 Emotiv MN8	13
4.2 OpenBCI EEG Headband	13
4.3 EmotiBit	14
4.4 Synchronization	15
5 Experimental Protocol	17
5.1 Pre-Recording Session	17
5.2 Task Execution	17
5.3 Post-Recording Session	18
6 Task Design	19
6.1 Overview of Task Design	19
6.2 Task Grouping and Targeted Cognitive States	19
6.3 Software Interface and Task Flow	19
6.4 Phase A: Attention-Dominant Tasks	21
6.5 Phase B: Stress-Dominant Tasks	23
6.6 Expected Physiological Responses	27
6.7 Synchronization	29
6.8 Participant Engagement and Motivation	30

7	Survey Design	32
7.1	Pre-Survey	32
7.2	Task-Survey	32
7.3	Post-Survey	32
8	Participant Information	34
9	Data Collection and Storage	35
9.1	Task Event Markers	35
9.2	Recording	35
9.3	File Format	36
9.4	Handling Missing Data	36
9.5	Data Storage	36
	Part III: Data Analysis	38
10	Overview of Analysis Pipeline	38
11	Data Preprocessing	40
11.1	Emotiv MN8 & OpenBCI Electroencephalography	40
11.2	Photoplethysmography Infrared	41
11.3	Electrodermal Activity	41
11.4	Skin Temperature	42
12	Feature Extraction	43
12.1	EEG Features	43
12.2	HRV Features	44
12.3	EDA Features	44
12.4	Skin Temperature Features	45
13	Statistical Analyses	46
13.1	Within-Trial Analysis	46
13.2	Between-Trial Analysis	49
14	Survey Analyses	58
14.1	Pre-experiment Survey	58
14.2	Task-level Subjective Ratings	59
14.3	Attention-Dominant and Stress-Dominant Task Profiles	62
14.4	Post-experiment Survey Results	64
14.5	Summary and Implications	66
15	Machine Learning Analyses	68
15.1	Problem Formulation	68
15.2	Ground Truth and Labeling	69
15.3	Feature Extraction and Representation	70
15.4	Model Selection	71
15.5	Training Procedure	71

15.6 Evaluation Metrics	73
15.7 Machine Learning Results	74
Part IV: Results	77
16 Discussion	77
16.1 Modality-Level Interpretation	78
16.2 Task-Centered Interpretation	79
16.3 Limitations	81
16.4 Implications and Suggestion for Future Work	83
17 Conclusion	84
Bibliography	87
Appendix	91
A.1 Synthesis Matrix	91
A.2 Consent Form Template	92
A.3 Signed Consent Forms	94
A.4 Surveys	94
A.5 xdf files	103
A.6 Available Data Streams	103
A.7 Data Completeness	104
A.8 Features Extracted Per Signal Modality	105
A.9 Within-Trial Plots Per Signal Modality	109
A.10 Between-Trial Plots Per Signal Modality	112
A.11 Statistical Results	118
A.12 Machine Learning Training Files and Outputs	118

List of Abbreviations

ANS autonomic nervous system

BH breath-holding task

CL cognitive load

CNNs convolutional neural networks

DL deep learning

ECG electrocardiogram

EDA electrodermal activity

EEG electroencephalography

FFT fast Fourier transform

GB gradient boosting

GSR galvanic skin response

GUI graphic user interface

HF high frequency

HR heart rate

HRV heart rate variability

LF low frequency

LF/HF low frequency/high frequency

LOSO Leave-One-Subject-Out

LSL Lab Streaming Layer

LSTM long short-term memory

MAD median absolute deviation

MAT mental arithmetic task

ML machine learning

NB naive Bayes

PermEn permutation entropy
PNS parasympathetic nervous system
PPG photoplethysmography
PPG IR PPG infrared
PRV pulse rate variability
PSD power spectral density
PSS Perceived Stress Scale

RBF radial basis function
RC reading comprehension
RFE recursive feature elimination
RMS root mean square
RMSSD Root Mean Square of Successive Differences

SCL skin conductance level
SCR skin conductance response
SCWT Stroop Color and Word Test
SD standard deviation
SEM standard error of the mean
SNS sympathetic nervous system
ST skin temperature
SVM support vector machine

NB: This thesis was written in American English.

1 Introduction

1.1 Introduction and Motivation

Modern life strains minds and emotions. Wars, economic uncertainty, rapid tech change, and constant connection create ongoing stress. News about conflict, higher prices, and job insecurity is hard to avoid. This news leaves people tense and anxious for long periods. Artificial intelligence and automation change how people work and learn, so people must adapt quickly. These pressures make it harder to focus and manage stress.

Humans are not designed to process large amounts of information or to stay mentally alert for long periods of time. Many people feel mentally tired, lose focus more easily, and struggle with decision-making. Fast and stimulating digital media often demands instant reactions, which further breaks concentration. When stress lasts too long, it can reduce attention even more and lead to burnout, anxiety, or depression. These effects can harm learning, work performance, and overall well-being, even without a formal medical diagnosis.

Stress is now widely recognized as a major public health issue. Global surveys show that more than one-third of adults experience high daily stress (Gallup, 2025). In Denmark alone, both young people and working adults report rising stress and more trouble paying attention (Psykiatrifonden, 2025; The Local, 2025). Many find it hard to stay focused for long. Distraction can increase safety risks like traffic accidents. These trends point to the need for reliable and objective ways to understand attention and stress in daily life.

Wearable biosensors offer a promising way forward. Consumer devices now measure brain activity, heart rate, skin conductance, and body temperature. These signals give objective data on cognitive demand and stress. As wearable technology becomes more accurate and accessible, it can help people detect problems early, manage cognitive load, and reduce stress. However, research is still limited. Few studies examine how attention and stress change together across different tasks with everyday devices. This thesis focuses on this gap. It studies whether wearable EEG and autonomic sensors can distinguish between attention-dominant and stress-dominant task states with the goal of supporting live monitoring in education, workplaces, and mental health settings.

1.2 Research Questions and Objectives

This thesis investigates how attention and stress can be distinguished using physiological signals recorded with wearable biosensors in controlled, task-based settings. To guide this investigation, one main research question was formulated, supported by a set of more specific sub-questions.

1.2.1 Research Questions

Main Research Question:

How can attention and stress be reliably distinguished in task-based environments using physiological signals measured with wearable biosensors?

With the following supporting research questions:

- What is the added value of combining neural and peripheral physiological signals when studying attention and stress?
- Which physiological modalities and extracted features are most sensitive to changes in attention and stress across diverse task conditions?
- How well do subjective self-reported measures of attention and stress align with objectively measured physiological responses?
- Can machine-learning models trained on multimodal biosignal features reliably classify attention-dominant and stress-dominant task states?

1.2.2 Project Objectives

To address the research questions outlined above, this project pursued the following objectives:

1. **Experimental design:** Develop a controlled, task-based experimental protocol capable of capturing different levels of attention and stress.
2. **Multimodal data collection:** Collect EEG and peripheral biosignals from participants using wearable devices during task execution.
3. **Physiological characterization:** Characterize EEG, HRV, EDA, and temperature responses associated with different levels of attention and stress across task conditions.
4. **Multimodal comparison:** Examine whether combining neural and peripheral biosignals provides more robust insights into attention and stress than single-modality analyses.
5. **Subjective-objective comparison:** Assess the correspondence between physiological measurements and participants' self-reported attention, stress, and mental demand.
6. **Exploratory machine learning:** Explore the feasibility of distinguishing attention-dominant and stress-dominant task states using machine-learning methods applied to multimodal physiological features.

7. **Reusable dataset creation:** Establish a structured and reusable dataset combining physiological signals and subjective reports to support future research in multimodal cognitive state assessment.

1.3 Thesis Structure

In order to address the project objectives introduced in Section 1.2.2, this thesis has been organized into four main parts, as illustrated in Figure 1:

Part I Research: introduces the research context and presents a structured literature review.

Part II Methodology: describes the methodological framework, detailing the experimental design, hardware setup and synchronization, participant recruitment, and the execution of the experiment and data collection procedures.

Part III Data Analyses: focuses on data analysis, covering signal preprocessing, feature extraction, statistical analyses, survey analysis, and exploratory machine learning classification.

Part IV Results: presents the results and discusses them in relation to both the physiological data and the survey responses, as well as the study's limitations and future work.

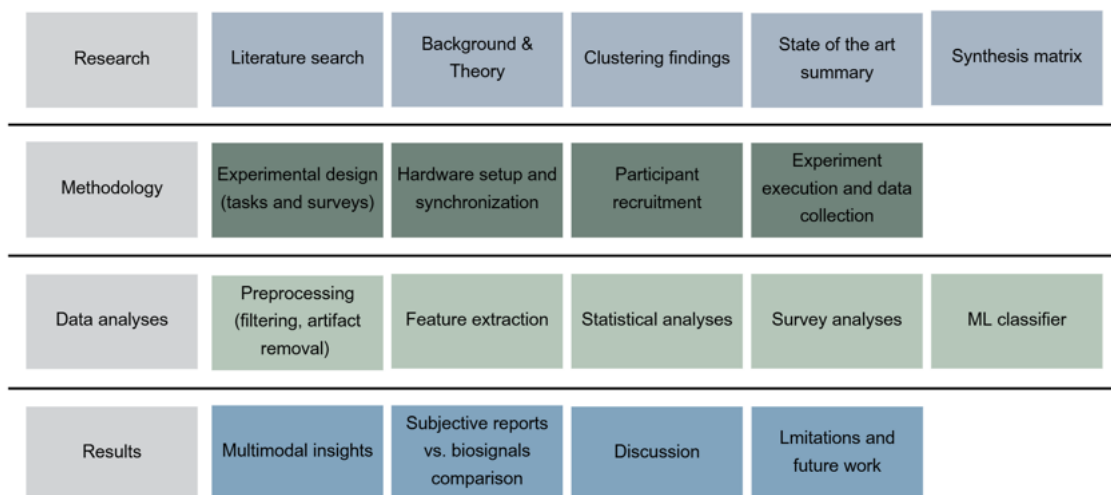


Figure 1: Overview of the thesis structure and methodological workflow.

Part I: Research

2 Literature Review

Research on the inference of attention states, cognitive load, and stress has been growing as wearable sensors become more widely available. By reviewing earlier work on attention and stress detection using physiological signals, this literature review aims to place this study within the context of existing research. To do so effectively, a systematic literature search methodology was implemented, which is described below to ensure transparency and reproducibility. The selected studies were then organized and analyzed using a synthesis matrix (Appendix A.1), which provided an overview across signal modalities, experimental designs, and analysis approaches. Based on this synthesis, related work is grouped by signal modality to highlight common trends, methodological differences, and remaining research gaps.

2.1 Literature Search Methodology

Before the systematic search, inclusion criteria (IC) and exclusion criteria (EC) were defined. Only papers that met all the IC listed below were selected:

Inclusion Criteria (IC):

- IC1:** The article must be written in English.
- IC2:** The publication date must be 2015 or later to ensure the inclusion of recent and relevant research.
- IC3:** The study must be based on the use of biosignals for inferring cognitive or emotional states.
- IC4:** The study must report on at least one of the following biosignals in relation to attention or stress detection:
 - electroencephalography (EEG)
 - heart rate variability (HRV)
 - electrodermal activity (EDA)
 - skin temperature (ST)

Studies that examine attention or stress levels across different tasks, or apply machine learning models for cognitive state prediction, are also eligible.

- IC5:** The study must be based on human data.

Exclusion Criteria (EC):

- EC1:** Papers that do not use biosignals for cognitive or emotional state inference (e.g., purely hardware validation).
- EC2:** Studies will be excluded if they do not involve any of the following:
 - electroencephalography (EEG)

- heart rate variability (HRV)
- electrodermal activity (EDA)
- skin temperature (ST)
- Multimodal biosignals
- Task-based stress or attention level assessment
- Machine learning applied to cognitive state prediction

EC3: Studies conducted exclusively on animal subjects.

EC4: Reviews, meta-analyses, or systematic reviews (only primary research articles were included).

To perform the systematic review, a search for articles in PubMed, ScienceDirect, IEEE, Google Scholar, and Scopus databases was performed using the keywords: EEG, HRV, temperature, attention detection/classification, stress detection/classification, mental workload, cognitive load, machine learning, deep learning, multimodal, and biosignals. Initially, 203,000 papers were found. After applying ICs and ECs, there were a total of 78,600 papers, and among them. Next, 10,600 duplicates were removed. These papers were ran through a Python script created to apply the ICs and ECs automatically to the title, which resulted in 1,950 papers. Then, a similar but more stricted automated script was implemented on the abstract sections of the collected studies resulting in in 84 papers. These papers were manually screened through and 61 papers were excluded based on title or abstracts, which left 23 papers. Figure 2, illustrates the screening procedure for research articles inclusion and exclusion.

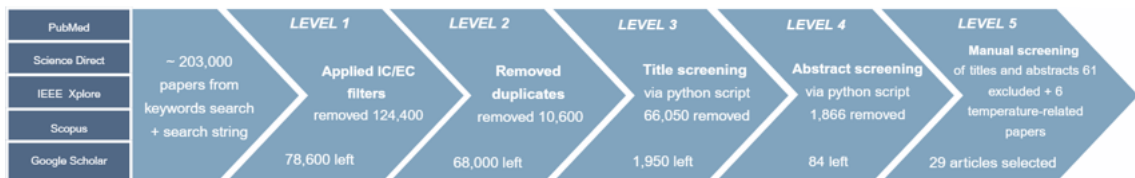


Figure 2: Overview of the literature search and screening process. The figure shows the databases queried, the initial number of retrieved records, and the filtering and screening steps applied.

Nevertheless, a gap in temperature-related studies was identified. Thus, a targeted search on DTU Findit was run with the search string "temperature and stress or attention". From the first 50 papers, sorted by relevance, a screening based on the title and skimming of the abstract was done. This resulted in 14 papers from which, after fully reading the abstracts, 6 were kept. Hence, a total of 29 core papers investigating how different biosignals (EEG, HRV, and temperature) relate to attention and/or stress were collected for this study's literature review.

2.2 Background & Theory

This study examines attention and stress through physiological signals. To facilitate comprehension, the following briefly introduces fundamental concepts related to stress, attention, the autonomic nervous system, and their associated physiological signals.

2.2.1 Attention and Stress

Attention is a multifaceted cognitive function that enables individuals to select, process, and respond to relevant stimuli in their immediate environment. Stress, in contrast, is generally conceived as a psycho-physiological phenomenon triggered when external or internal agents overload the individual's adaptive capacity (Arza et al., 2019; Eyad Talal Attar et al., 2021).

Neuropsychology classifies attention into different types:

- Selective attention: focusing on specific stimuli while disregarding others deemed irrelevant (Tiwari et al., 2019; Dzianok et al., 2021).
- Sustained attention: maintaining focus on a specific stimulus or task during an extended period of time (Luque-Casado et al., 2016; Kaushik et al., 2022).
- Divided attention: performing concurrent tasks or simultaneously processing multiple streams of information (Hernández-Sabaté et al., 2024).

Stress can be categorized into cognitive stress, referring to the mental strain caused by high cognitive demand, and physiological stress, which is the body's biological stress response. For instance, multitasking and solving arithmetic challenges can induce cognitive stress, while cold exposure and breath holding trigger physiological stress.

Attention and stress are closely linked through cognitive load (CL), the attention effort required to attempt a task (Arutyunova et al., 2024). If the CL meets or exceeds the individual's adaptive capacity, the resulting physiological response is stress. That is, high CL acts as a stress-inducing agent. Thus, there's a physiological and neural overlap between the states of intense attention and acute stress, both activating the autonomic nervous system (ANS).

2.2.2 Autonomic Nervous System (ANS)

The ANS is the part of the human nervous system that autonomously adapts the body to internal and external situational demands by regulating vital organs (Hanshans et al., 2024). The ANS is activated when a stressor is perceived, provoking a series of biological changes that prepare the individual for survival. The system is composed of two branches that work in a complementary and often opposing dynamic to regulate the body's state of arousal, namely, the sympathetic ("fight or flight") and the parasympathetic ("rest and digest") branches. The sympathetic nervous system (SNS) prepares the body for action, while the parasympathetic nervous system (PNS) functions as a recovery mechanism (Ahmadi et al., 2022; Hanshans et al., 2024).

The activation of the ANS produces measurable physiological changes. Notably, the heart shows altered heart rate (HR) and heart rate variability (HRV) (Tiwari et al., 2019; Eyad Talal Attar et al., 2021); the sweat glands, controlled by the SNS, increase sweat secretion, a change measured by electrodermal activity (EDA) (Ahmadi et al., 2022; Hanshans et al.,

2024); the redistribution of blood flow causes fluctuations in skin temperature (Ahmadi et al., 2022; Hanshans et al., 2024); and changes in the brain activity can be measured through an electroencephalography (EEG) by observing variations in Alpha and Beta waves.

2.2.3 Physiological Measures of Stress and Attention

Shifts in mental states, including stress and attention, manifest as measurable physiological changes in the SNS and PNS. As cognitive load increases, the sympathetic branch typically dominates (Arza et al., 2019; Ahmadi et al., 2022; Arutyunova et al., 2024).

To analyze signals that help identify changes in SNS, and therefore stress and attention, researchers turn to analysis in the time and frequency-domains. The first refers to the rate of change and statistical properties (mean, SDNN, and RMSSD) of a signal over a defined period (Tiwari et al., 2019). The latter assesses the distribution of a signal's energy and power across frequency bands, such as High Frequency (HF) and Low Frequency (LF), identifying periodic oscillations (Acı et al., 2019; Tiwari et al., 2019).

The signals covered in this study are:

- Cardiac indicators - HR, the speed at which the heart beats, and HRV, the beat-to-beat variation in the heart rhythm, measured by RMSSD or SDNN. Table 1 provides an overview of HRV metrics and their definitions.
- electrodermal activity (EDA) - also known as galvanic skin response (GSR), measures the skin's electrical conductance. It is composed of: (1) the background level or long-term trend of the electrodermal time series, also referred to as skin conductance level (SCL) or tonic level, and (2) the acute spikes or rapid fluctuations on top of the SCL, known as skin conductance response (SCR) or phasic responses. Both components serve to understand types of sympathetic activation, such as momentary or persistent stress.
- skin temperature (ST) - SNS activation redirects blood flow towards vital organs, leading to a rapid decrease in temperature in the individual's extremities.
- Neural indicators - EEG is a non-invasive technique that enables the recording of the brain's electrical activity through electrodes placed on the scalp's surface. As it can accurately track real-time fluctuations in cognitive attention, it's widely used to study attention.

Metric	Type	Definition
s _{dnn}	Time-domain	Standard deviation of all inter-beat intervals; reflects overall heart rate variability (HRV).
r _{mssd}	Time-domain	Root mean square of successive differences between inter-beat intervals; sensitive to short-term (parasympathetic) variability.
p _{nn50}	Time-domain	Percentage of successive inter-beat interval differences greater than 50 ms.
l _f	Frequency-domain	Power in the low-frequency band.
h _f	Frequency-domain	Power in the high-frequency band.
l _f /h _f	Frequency-domain	Ratio of low- to high-frequency power.

Table 1: Overview of HRV metrics and their definitions

The ranges of biological oscillations from the brain activity are commonly categorized into five frequency bands, as summarized in Table 2.

Band	Frequency Range	Associated Cognitive State
Delta	0.5–4 Hz	Deep sleep, sluggishness, or a resting phase of the body.
Theta	4–8 Hz	Drowsiness, memory recall, deep meditation, or light sleep.
Alpha	8–13 Hz	Relaxed, calm, conscious, and alert state.
Beta	13–30 Hz	Task-oriented, problem-solving, critical thinking, and rational logic.
Gamma	> 30 Hz	Very high mental activity, demanding cognitive or motor functions.

Table 2: Common EEG frequency bands and their associated mental states.

To quantify the energy within those frequency bands, mainly two methods are used: absolute and relative band power. Absolute band power measures the total power within a specific frequency range by integrating the power spectral density (PSD) over that specific frequency band, whereas relative band power is the ratio of power in a specific band to the total power across all frequency bands.

Given that mental states are multifaceted and no single marker can fully capture attention or stress, multimodal approaches are preferred when inferring mental states. A multimodal approach integrates measures and data sources, leading to higher accuracy.

2.3 Related Work Synthesis

The papers selected were categorized into studies relying on Electroencephalography (EEG) measures, cardiac dynamics such as heart rate (HR) and heart rate variability (HRV), and multimodal and peripheral fusion frameworks. In the following, the findings are presented by signal modality.

2.3.1 Electroencephalography (EEG)

EEG remains the most studied modality for detecting and measuring attention states, cognitive load, and stress. Most studies in this cluster target the frontal lobe, as this region provides significant insights into these states (AlShorman et al., 2022; Mukherjee et al., 2024; Nirabi et al., 2025). For instance, AlShorman et al. (2022) applied a fast Fourier transform (FFT) to measure all bands' power density for the frontal lobe and achieved an accuracy of 98.21% in detecting mental stress using support vector machine (SVM) and naive Bayes (NB) as machine learning (ML) classifiers.

Other studies utilized EEG brainwave frequencies to detect stress and attention. Kaushik et al. (2022) decoded attention states in naturalistic monastic debates, finding that attention is linked to increased left frontal alpha, increased left parietal theta, and decreased central delta. Their long short-term memory (LSTM) model achieved ~95% accuracy using delta and theta waves. Comparably, Acı et al. (2019) identified that changes in the frontal and parietal lobes occurring at 1–5 Hz and 10–15 Hz frequency bands are significant markers to distinguish between focused, unfocused, and drowsy states. In addition, Dzianok et al. (2021) successfully discriminated attentive states during reaction tasks by using Hilbert-transformed theta, alpha, and beta EEG bands and ML classifiers. Specifically, changes in alpha and beta activity have proven to be good indicators of mental states. Halim et al. (2020) investigated driving-induced stress and noted a significant drop in alpha power when transitioning from a relaxed state to stress, and Singh et al. (2022) utilized the ratio of absolute alpha and beta power alongside other power band features to classify three levels of stress with a mean accuracy of 76.04%. Regarding attention detection, both Gupta and Kumar (2021) and Lestari et al. (2020) monitored students' attention states using consumer-grade EEG devices and found that higher beta values correlate with higher attention.

It is important to note that while the majority of studies use ML, deep learning (DL) approaches have gained traction in recent studies. Qin et al. (2024), Mukherjee et al. (2024), and Hernández-Sabaté et al. (2024) used convolutional neural networks (CNNs) to extract features and successfully classify stress and cognitive load, with accuracies above 90%, across diverse tasks, including assembly tasks, games, and flight simulation.

2.3.2 Cardiac Dynamics

Cardiac signals, particularly heart rate (HR) and heart rate variability (HRV), are also widely used as indicators of psychological states due to their reflection of autonomic nervous system (ANS) activity. Since multiple studies have shown that high cognitive load (CL) or stress increase HR and reduce HRV, most studies in this cluster focus on identifying the metrics that best discriminate cognitive states. Luque-Casado et al. (2016) found that HRV metrics, specifically Root Mean Square of Successive Differences (RMSSD) and high frequency (HF) power, systematically decreased as task demands increased. While Arutyunova et al. (2024) also found that increased CL is accompanied

by higher HR and lower RMSSD, they dove into non-linear domain HRV measures and found that permutation entropy (PermEn) consistently increased as CL grew. Behradfar et al. (2025) extracted 93 HRV features, spanning time-domain, frequency-domain, and non-linear metrics, and applied recursive feature elimination (RFE) combined with gradient boosting (GB), achieving an accuracy of 76%, and identifying HRV-MCVNN (median absolute deviation of RR intervals normalized by the median) and normalized low-frequency power (LFn) as the most critical features for accurate stress detection.

Another prominent focus in cardiac research is the comparison between HRV derived from electrocardiogram (ECG) and from photoplethysmography (PPG). On one hand, Tiwari et al. (2019) used an ECG-based smart shirt and a PPG-based smart bracelet to record data from 196 hospital workers during their work shifts for a period of 10 weeks. They utilized an SVM classifier and found that the smart shirt (ECG) significantly outperformed the smart bracelet (PPG) for stress detection. On the other hand, Iovino et al. (2024) showed that pulse rate variability (PRV) derived from PPG can be as effective as ECG-based HRV to classify stress in a controlled environment.

However, in some studies, cardiac dynamics were not the best indicators for inferring cognitive states. Hanshans et al. (2024) explicitly state that in their study using a Virtual Reality (VR) horror stressor, HRV failed to reliably show a stress response. Similarly, Ding et al. (2020) found no significant differences in average heart rate or the LF/HF ratio across three different levels of mental arithmetic tasks.

2.3.3 Multimodal & Peripheral Fusion Approaches

While brain and heart signals are each informative on their own, recent work shows that the most reliable results are obtained when EEG and cardiac signals are combined, often together with peripheral measures such as electrodermal activity (EDA) and/or skin temperature (ST). For instance, Mukherjee et al. (2024) combined EEG with pulse rate data. By collecting data from participants solving a math question set of increasing complexity and using advanced deep learning (DL) architectures, they achieved an accuracy of 97.86% in classifying stress levels into low, medium, and high. Similarly, Eyad Talal Attar et al. (2021) simultaneously recorded EEG and HRV during Stroop Color and Word Test (SCWT) and meditation tasks, finding significant correlations between frontal EEG alpha/beta activity and HRV metrics (LF, HF, RMSSD), suggesting coupled cortical–cardiac dynamics.

The best results have been achieved by fusing cardiovascular signals, skin temperature, and skin conductance. Ding et al. (2020) demonstrated that a multimodal model incorporating ECG, EDA, and respiration obtained 96.4% accuracy in determining mental workload levels, whereas individual physiological signals had poor classification ability (below 48%). Umer (2022) reinforced this by showing that a fusion of HR, breathing rate, skin temperature, and skin conductance achieved 94.7% accuracy for the simultaneous monitoring of physical and mental stress. Martinez et al. (2017) and Rodríguez-Arce et al. (2020) got similar results: Martinez et al. (2017) leveraged both GSR and HRV within a Finite State Machine (FSM) architecture to achieve an F1 score of 0.984 for high-level stress detection, and Rodríguez-Arce et al. (2020) achieved an accuracy over 90% combining HR, ST, and oximetry.

Several studies highlight HR and ST as the most reliable stress predictors across diverse scenarios. Albaladejo-González et al. (2024) identified HR and ST as the best predictors of stress during cognitive tasks, outperforming executive function metrics. Arza et al. (2019) utilized the same signals to measure stress on a 0–100 au scale, and identified Pulse Transit Time (PTT) and skin temperature ratios (face and finger) as highly sensitive markers for stress-level distinction. Ahmadi et al. (2022) provided critical real-world validation during 12-hour ICU nursing shifts, finding significant positive correlations between stress and both HR and ST.

The reviewed sources reveal a difference in the reliability of EDA when derived from laboratory or from naturalistic settings. Hanshans et al. (2024) concluded that skin conductance response (SCR) is the most promising marker for detecting acute stress in VR because it is low-cost and highly resistant to motion artifacts. Uniformly, Rodríguez-Arce et al. (2020), Umer (2022), and Ding et al. (2020) highlighted EDA as possessing the highest discriminating power in controlled simulated environments. However, Ahmadi et al. (2022) found that phasic EDA did not correlate significantly with stress during 12-hour ICU nursing shifts.

2.3.4 Contribution

In conclusion, previous studies show that HRV offers useful autonomic insights and that EEG is a reliable and well-researched modality for cognitive state inference. At the same time, the literature shows that the highest accuracy is achieved through multimodal approaches. However, these studies typically combine only a subset of signals, such as EEG with EDA or skin temperature, or HRV with peripheral measures, rather than integrating all modalities simultaneously. Moreover, much of the existing evidence is derived from controlled laboratory studies that focus on a single cognitive state and task type and often rely on participant samples dominated by male subjects. Building on these findings, the present thesis extends prior work by adopting a multimodal framework that integrates EEG, HRV, EDA, skin temperature, and subjective feedback from participants to infer attention states across diverse task-based environments.

Part II: Methodology

3 Experimental Design Overview

This project's experiment was designed to study attention states and stress using a multimodal biosignal approach. A set of cognitive tasks, carefully designed to capture different levels of attentional demand and stress, was implemented in a controlled setting. Physiological responses were recorded with two EEG sensors and one wearable sensor capturing the peripheral biosignals. The emphasis of the study was primarily on the design of a robust experimental protocol and the collection of high-quality data from a participant sample (N=23) performing a diverse set of tasks, with the goal of enabling meaningful analysis and reuse of the dataset in future work. Furthermore, statistical analysis was conducted to evaluate task-related differences in physiological responses and to assess the suitability of the proposed experimental approach.

Prior to execution, a detailed experimental protocol was developed, and all tasks were designed within a dedicated task application. Each participant completed all tasks in a randomized order to mitigate potential order effects. The study followed a within-participant design, allowing comparisons to be made across tasks within the same individual and thereby reducing inter-subject variability.

In total, the experiment consisted of nine tasks: baseline with eyes open, baseline with eyes closed, a neutral fixation task, reading comprehension, oddball, a 2-Back task, Stroop Color and Word Test (SCWT), mental arithmetic task (MAT), and a breath-holding task. The task selection was supported by the reviewed literature. Reading tasks have previously been used to assess attention levels (Lestari et al., 2020; Gupta & Kumar, 2021), while the n-back paradigm is a well-established measure of working memory and mental workload, with the 2-Back variant commonly used to elicit higher cognitive load (Luque-Casado et al., 2016; Salankar et al., 2023; Hernández-Sabaté et al., 2024). The Stroop and mental arithmetic tasks are widely employed to induce cognitive stress (Ding et al., 2020; Eyad Talal Attar et al., 2021; Rai et al., 2022; Singh et al., 2022; Nirabi et al., 2025). In addition to their relevance for eliciting targeted cognitive states, all tasks were chosen for their suitability for wearable sensor recording, as they can be performed while seated and with minimal movement, thereby reducing motion-related artifacts.

4 Hardware Setup

The hardware configuration for the experiment’s multimodal approach consisted of three sensing systems: the Emotiv MN8 EEG headset (Emotiv, n.d.), the OpenBCI EEG Headband Kit paired with a Cyton Board (OpenBCI, n.d.), and the EmotiBit for peripheral signals (EmotiBit, n.d.). All devices were worn simultaneously by the participants, enabling synchronized acquisition of brain and peripheral responses during task execution. An example of how the sensors were worn by participants can be seen in Figure 7.

4.1 Emotiv MN8

The Emotiv MN8 is a consumer-grade, two-channel EEG headset designed for mobile and unobtrusive brain monitoring. Unlike traditional EEG systems, the Emotiv MN8 comes in form of earbuds, with electrodes positioned behind the ears to capture temporal brain activity (see Figure 3). This design allows for rapid setup and relatively high participant comfort.



Figure 3: Emotiv MN8 headset (Emotiv, n.d.)

The Emotiv MN8 samples data via Bluetooth at 128 Hz with a 14-bit resolution and operates within a frequency bandwidth of 0.5–45 Hz, with an integrated hardware low-pass filter at 45 Hz to reduce high-frequency noise (Emotiv MN8 Technical Specifications, n.d.).

Emotiv has its own software application, enabling users to manage the device and providing real-time feedback on electrode contact quality and EEG signal quality. This feature made the fitting process more efficient and ensured correct sensor placement across participants. In addition, multiple earbud sizes were available, allowing each participant to select the most comfortable and secure fit.

4.2 OpenBCI EEG Headband

To complement the Emotiv MN8 recordings and provide additional EEG coverage, the OpenBCI EEG Headband Kit was used in combination with the OpenBCI Cyton board. The Cyton board supports up to eight EEG channels and streams data wirelessly via Bluetooth at a sampling rate of 256 Hz (OpenBCI Docs, n.d.).

In this study, two EEG electrodes were mounted in the headband and positioned over the frontal pole regions (Fp1 and Fp2 according to the international 10–20 system; see Figure 4), targeting frontal activity, which has been associated with attentional and

executive processes. A third electrode positioned centrally between Fp1 and Fp2 served as the reference (ground) electrode, while an ear-clip electrode was used as the bias electrode to improve signal quality (see Figure 5). Although the hardware supports up to eight channels, only two channels were used to maintain consistency with Emotiv MN8's low-channel configuration and to reduce setup complexity.

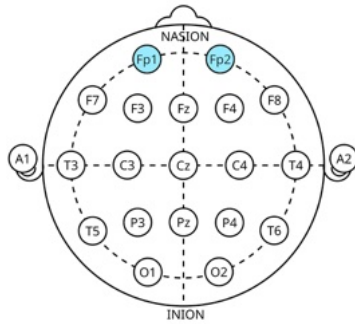


Figure 4: Fp1 and Fp2 electrode locations in the International 10-20 system for EEG recording (Wikipedia, n.d.)



Figure 5: OpenBCI headband setup (OpenBCI Docs, n.d.)

4.3 EmotiBit

Peripheral physiological signals were recorded using the EmotiBit, a wearable multimodal sensor designed for high-quality physiological and motion data acquisition. The EmotiBit integrates multiple sensing modalities within a single compact device, enabling simultaneous recording of cardiovascular, electrodermal, thermal, and motion-related signals.

The EmotiBit provides the following data streams and sampling rates:

Function	Data Type	Sampling Rate (Hz)
Motion	AX, AY, AZ, GX, GY, GZ, MX, MY, MZ	25
PPG	PI, PG, PR	25
Temperature	T0, TH	7.5
EDA	EDA	15

Table 3: EmotiBit data types and sampling rates (EmotiBit Docs, n.d.)

In the present study, only the PPG infrared (PPG IR) signal (used for heart rate and heart rate variability extraction), electrodermal activity (EDA), and skin temperature measured by the thermistor (TH) were analysed. These signals were selected due to their established relevance to stress and arousal.

Data was streamed wirelessly via Bluetooth, and its quality can be monitored in real time using EmotiBit's Oscillator software, which provides live visualization of all data streams.

The EmotiBit was secured to the participant's left upper arm using an adjustable strap (see Figure 6 for an illustrative example). This placement was chosen to balance comfort, signal stability, and reduced interference with task performance. After placement, signal quality was verified via EmotiBit's Oscillator before proceeding with the experiment.

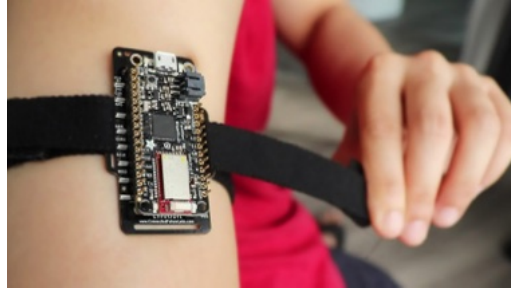


Figure 6: Example of EmotiBit placement on the upper arm. The image is provided for illustrative purposes only and does not depict the exact placement used in this study. (EmotiBit, n.d.)

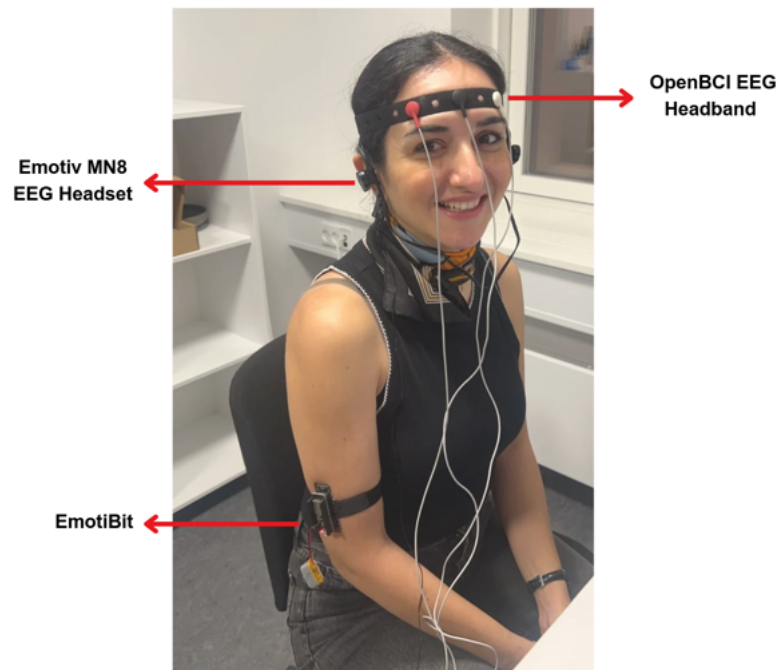


Figure 7: Example of the experimental setup showing a participant wearing the EMOTIV MN8 ear-EEG headset, the OpenBCI EEG headband, and the EmotiBit sensor for peripheral physiological recordings during the experiment.

4.4 Synchronization

All sensor data streams were synchronized and transmitted in real time using Lab Streaming Layer (LSL), an open-source, network-based middleware framework for the unified collection of time-series data from heterogeneous sources. LSL enables precise temporal alignment of neural, physiological, and behavioral data streams by mapping all samples to a shared global clock, achieving sub-millisecond synchronization accuracy despite differences in device clocks and sampling rates (Lab Streaming Layer, n.d.).

Both the OpenBCI system and the EmotiBit natively support LSL output. Consequently, data from these devices was streamed directly to LSL by configuring the respective

software interfaces (OpenBCI GUI and EmotiBit Oscillator) to use LSL as the output protocol.

The Emotiv MN8 does not natively support LSL streaming without the acquisition of the Emotiv Pro license. Instead, it communicates through the Emotiv SDK (Software Development Kit), which provides access to the raw EEG signals, device metadata, and internal timing information. To integrate Emotiv MN8's data into LSL, a streaming bridge was developed. This bridge extracted EEG samples from the SDK in real time and repackaged them into an LSL outlet. Once all devices were configured to output their respective data streams to LSL, the streams were selected and recorded using the LSL recording application *LabRecorder*. During recording, all incoming data streams were timestamped using LSL's shared global clock. As a result, EEG data from the Emotiv MN8 and OpenBCI systems, peripheral physiological signals from the EmotiBit, and task-related event markers were temporally aligned in the recorded data.

Figure 8 illustrates the streaming process of each sensor to LabRecorder.

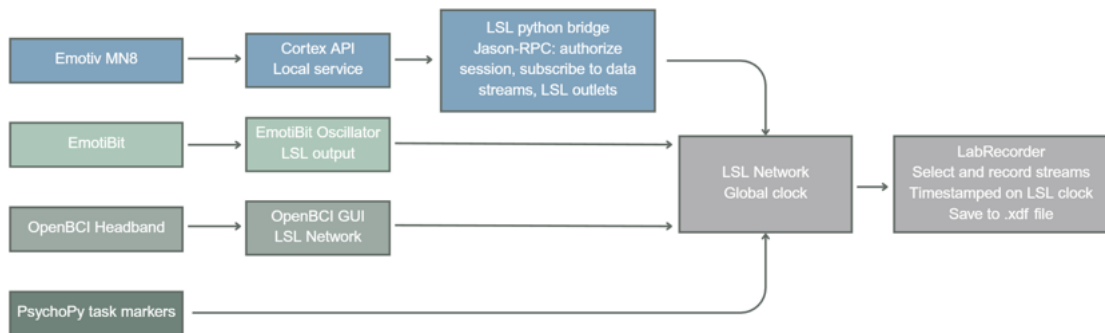


Figure 8: Path from each sensor to LabRecorder

5 Experimental Protocol

The experimental session consisted of three phases: a pre-recording session, a task execution session, and a post-recording session. All sessions were conducted in a quiet indoor room with normal room lighting and an average temperature of 20.61°C. Participants were seated in an upright position in front of a table, with their arms resting on the laptop keyboard. The experimenter remained in the same room throughout the session, often behind the participant to minimize distraction while remaining close by to provide instructions.

5.1 Pre-Recording Session

During the pre-recording session, participants were welcomed and given a brief explanation of the study. Written informed consent was obtained from each participant (Appendix A.3), after which a unique participant ID was assigned. Participants were then asked to complete the pre-survey (Appendix A.4.1) and their blood pressure was taken.

Following the survey, the sensors were placed and adjusted until satisfactory signal quality was obtained, as indicated by Emotiv MN8's contact and EEG quality and by stable signals in the OpenBCI and EmotiBit live monitoring software. In addition, participants were instructed to minimize movement, refrain from speaking during task execution, and ask any questions before the tasks began.

The pre-recording session lasted approximately 10 minutes per participant.

5.2 Task Execution

Participants completed all tasks in a single session while seated in front of a laptop displaying the tasks. All participants first performed two baseline recordings consisting of 1 minute with eyes open, followed by 1 minute with eyes closed.

After the baseline, participants proceeded to the task execution phase. Tasks were divided into two blocks, referred to as "Phase A" and "Phase B". The order of tasks within each block was randomized to reduce potential task order effects.

Phase A included tasks designed to elicit different levels of attention, namely the reading comprehension task, oddball task, and n-back task. Phase B consisted of stress-inducing tasks, including the Stroop Color and Word Test (SCWT), mental arithmetic task (MAT), and the breath-holding task. Before each phase, participants completed a neutral task. These are referred to as "Neutral 1" (neutral task before Phase A) and "Neutral 2" (neutral task before Phase B), and served as a baseline condition for subsequent analyses.

Before the start of each task, participants received task instructions from the experimenter. After completing each task, participants filled out a task-level questionnaire (Appendix A.4.2). Detailed descriptions of all tasks are provided in Chapter 6.

In cases of technical malfunctions, such as application crashes, interrupted tasks were repeated. The implications of these are discussed in Section 9.4.

The task execution phase lasted approximately 50 minutes per participant.

5.3 Post-Recording Session

After completion of all tasks, the recording was stopped in LabRecorder and the session data was saved in *.xdf* format. Participants were then asked to complete the post-survey (Appendix A.4.3). While participants filled out the survey, the experimenter removed the sensors. Finally, participants were thanked for their participation and provided with a small token of appreciation.

The post-recording session lasted approximately 5 minutes per participant.

All participants followed the same experimental protocol. The total session duration per participant was approximately 60-75 minutes. Figure 9 illustrates the overall structure of the experimental procedure.

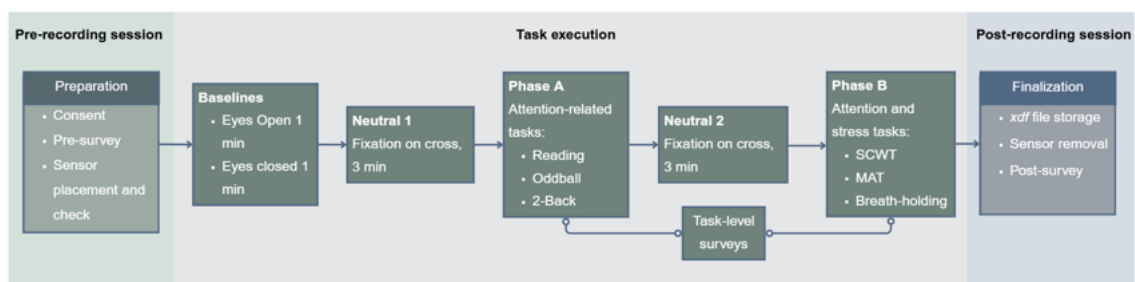


Figure 9: Experimental Protocol

6 Task Design

This chapter presents the design and purpose of the experimental tasks used in this study, detailing their intended cognitive effects and the measures collected during task performance.

The tasks were designed to elicit different cognitive and physiological states, with an emphasis on attention and stress.

6.1 Overview of Task Design

The tasks were designed to cover a range of cognitive demands, including sustained and selective attention, and to capture elevated stress. The tasks varied in difficulty, time-pressure, and performance conditions, targeting multiple cognitive states.

Tasks were grouped into two categories based on their main intended effect. Phase A consisted of attention-dominant tasks with minimal stress induction, while phase B included stress-inducing and stress-monotonous tasks. Participants also performed eyes-open and eyes-closed baseline tasks throughout the experiment.

Physiological signals were recorded simultaneously throughout task execution, and event markers were used to flag key events and task boundaries in the recording. The markers served as reference points for later analysis and were not used to synchronize tasks or participants' responses.

6.2 Task Grouping and Targeted Cognitive States

The tasks were grouped based on the main cognitive state they were designed to target. This grouping was used to describe the task design and to guide later analyses, in line with (Hernández-Sabaté et al., 2024).

Two task groups were defined. The first group included attention-dominant tasks. These tasks required sustained or selective focus and were designed to avoid strong stressors such as strict time limits or performance penalties.

The second group consisted of stress-dominant tasks designed to increase mental or physiological strain through time pressure, interference, difficulty, or physical discomfort. Additionally, in some cases, factors like background audio and the experimenter's presence may have contributed to performance-related pressure. While attention was necessary throughout, the primary objective of these tasks was to raise perceived stress and workload.

The overlap of attention and stress may lead to participants responding differently to the same task. The task grouping, therefore, reflects the proposed task design rather than a fixed cognitive state for all participants.

6.3 Software Interface and Task Flow

Python powered the experimental tasks, which used a custom graphical user interface. Scientific and experimental libraries supported the implementation. The software presented stimuli, collected participant responses, and sent event markers to indicate key events during task execution for subsequent synchronization and analysis.

Visual stimuli and interaction logic were handled using PsychoPy, which provided full-screen stimulus presentation, keyboard and mouse input control, and control over stimulus timing. Python libraries were used for functions, such as screen setup, timing, and marker handling, which were shared among tasks.

Each task script defined the task flow, stimulus sequence, response handling, and behavioral measures collected during execution. A shared configuration and utility module, `experimentGui.py`, was used to ensure consistent behavior across tasks.

A central experiment launcher controlled task execution. This launcher allowed the experimenter to run the full sequence, individual phases, or single tasks. In the main experiment, tasks were executed in a set order, with randomization within each phase as described in the protocol. Single-task execution was mainly used for debugging, and when the experiment crashed, it followed the same internal logic.

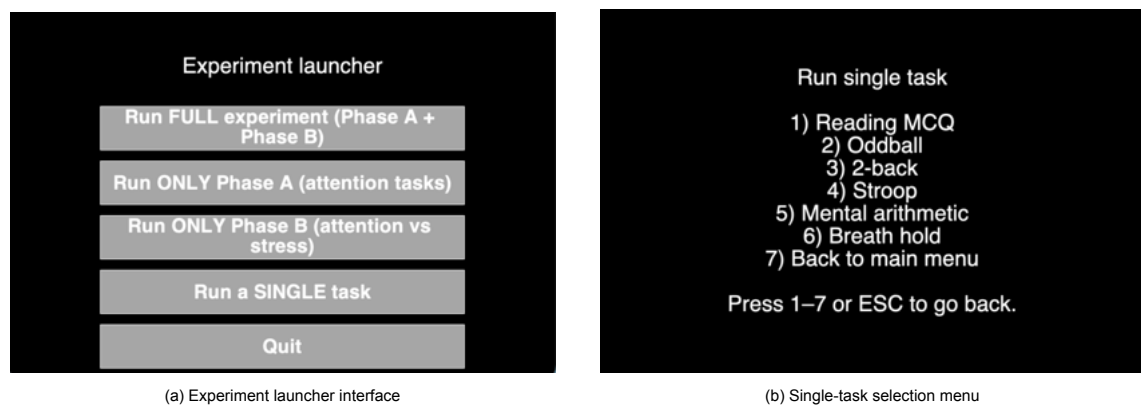


Figure 10: Custom software interface used to launch and control experimental tasks.

Behavioral measures, e.g., accuracy, reaction time, and task duration, were recorded during task execution and saved for later analysis.

Managing the experiment was facilitated by this modular structure, which also ensured all tasks followed a consistent execution pattern.

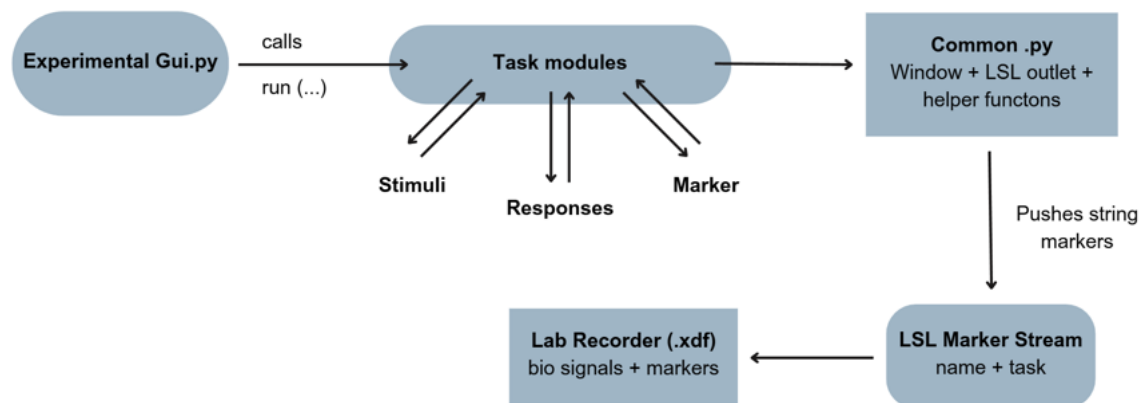


Figure 11: Software architecture of the experimental task framework

6.4 Phase A: Attention-Dominant Tasks

Tasks included in Phase A ranged from engaging different forms of attention while keeping stress at a lower level. The reading comprehension task was included to assess sustained attention at a moderate level. The oddball task was used to target selective attention and target detection, and lastly, the 2-back task was chosen to engage working memory and continuous attention. Together, these tasks provided attentional demands under neutral conditions, allowing attentional effects to be studied without the stress-related confounds.

6.4.1 Reading Comprehension Task

The reading comprehension task was designed to obtain sustained attention under low-stress conditions and to reflect natural, day-to-day attention Dzianok et al., 2021.

During the task, participants were shown a set of passages Britannica, n.d. on the screen and instructed to read them carefully at their own pace. Participants were given a maximum time limit of 3 minutes to avoid excessive reading; however, 3 minutes was more than enough, as the passage was short (4 paragraphs). This allowed participants to focus on comprehension rather than speed. The participants were also instructed that they could not go back and forth between passages or questions. Thus ensuring they paid maximum attention before moving on to the next task.

After reading the task, participants were presented with a set of multiple-choice questions about the passage's content. These questions require remembering and understanding the text.

Behavioral measures during this task included response accuracy and reaction time for the multiple-choice questions. To avoid stress, no explicit feedback was provided.

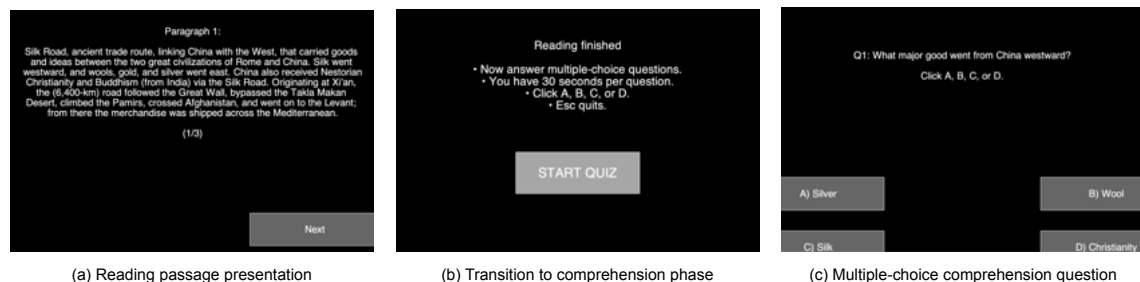


Figure 12: Screenshots from the reading comprehension task, illustrating the reading phase, the transition screen, and an example multiple-choice question used to assess understanding of the text.

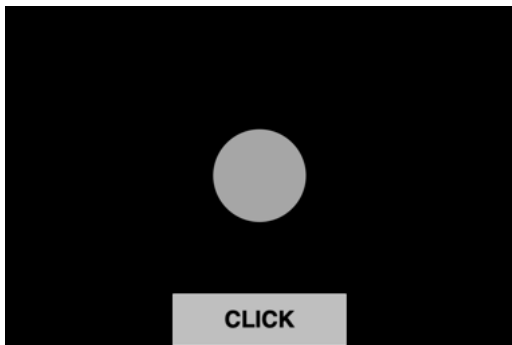
6.4.2 Oddball Task

The oddball task was used to evoke selective attention and target detection under low-stress conditions. In this task, participants continuously monitor incoming stimuli and respond to rare target events. Target stimuli occurred in approximately 20% of trials and participants were asked to respond quickly and accurately to targets and ignore non-targets. No additional time pressure was induced within the 0.4-second response window, and performance feedback was not given. This task was inspired by even or odd task conducted by (Iovino et al., 2024).

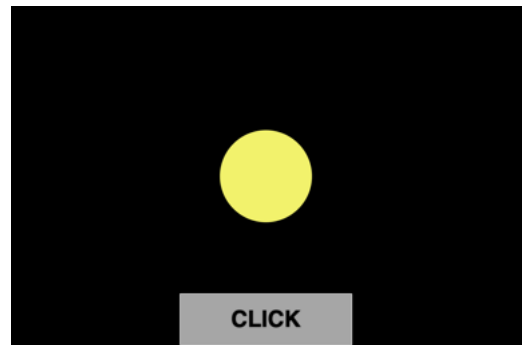
Behavioral measures collected during the oddball task included accuracy of response and reaction time for target stimuli. The task was designed to maintain attentional engagement while decreasing stress, rendering it suitable for assessing selective attention.



(a) Oddball task instructions presented before task onset



(b) Standard (non-target) stimulus



(c) Target stimulus requiring a response

Figure 13: Screenshots from the oddball task, showing the instruction screen and examples of standard and target stimuli used to elicit selective attention and target detection.

6.4.3 2-Back Task

The 2-back task was created to target a high level of attentional demand (Hernández-Sabaté et al., 2024). Compared to the other phase A tasks, this task required continuous information updates and sustained monitoring. This task required participants to compare the current stimulus presented with the stimulus presented two trials ago.

The relevant stimuli remained visible on the screen, meaning that the task relied more on recognition and comparison than on pure memory; however, some participants attempted to memorize the word instead.

Participants were instructed to compare each stimulus to the stimulus from two trials earlier. If the current and previous stimuli match, press the designated "match" button; if not, do nothing. The task did not include explicit time pressure or performance feedback.

Participants were instructed to indicate whether the current stimulus matched the one shown two trials before. The task did not include explicit time pressure.

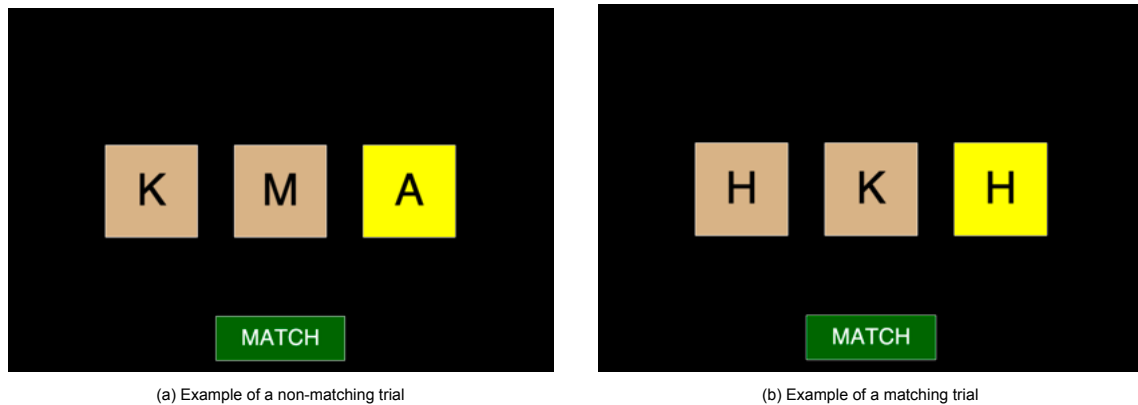


Figure 14: Screenshots from the 2-back task showing examples of non-matching and matching trials. Participants indicated whether the current stimulus matched the stimulus presented two trials earlier.

6.5 Phase B: Stress-Dominant Tasks

Phase B included tasks designed to induce perceived stress. These tasks introduced stress through time pressure, increasing difficulty, or physical discomfort.

Participants still needed to focus to complete each task. However, the primary goal of this phase was to increase stress levels, moving beyond tasks that only required attention.

In addition to task-specific stress demands, contextual factors were also introduced to satisfy this condition. A whiteboard in the experiment room displayed participants' phase B task performance, and participants were encouraged to beat the records set by previous participants. This introduced a competitive and evaluative element.

Phase B consisted of Stroop Color and Word Test (SCWT), mental arithmetic task (MAT) and, breath-holding task (BH).

6.5.1 Stroop Color-Word Task

The Stroop Color and Word Test (SCWT) is a task designed to evoke selective attention and stress (Nirabi et al., 2025). In this task, participants were asked to ignore the written word, which is the name of a color, and instead respond only to the ink the word is written in. This created a conflict on incongruent and sometimes congruent trials in which the word's meaning and ink did not match, or when they did, left participants in doubt.

At the start of the task, participants were shown a screen with instructions and a *Start* button. During each trial, a single color word (RED, BLUE, GREEN, or YELLOW) was displayed in the center of the screen in a large, bold font. To respond, participants used the keyboard with keys *R*, *G*, *B*, and *Y* matching the corresponding color.

Each trial had a maximum response time of 2.5 seconds, which was displayed in the top-right corner of the screen. As soon as the participants recorded a response or used up the 2.5-second window for each trial, a fixation cross was shown for 0.5 seconds.

The task included a mixture of congruent and incongruent trials, with an equal number of each for each color (64 in total). These were then randomly shuffled and distributed in a mixed, unpredictable order.

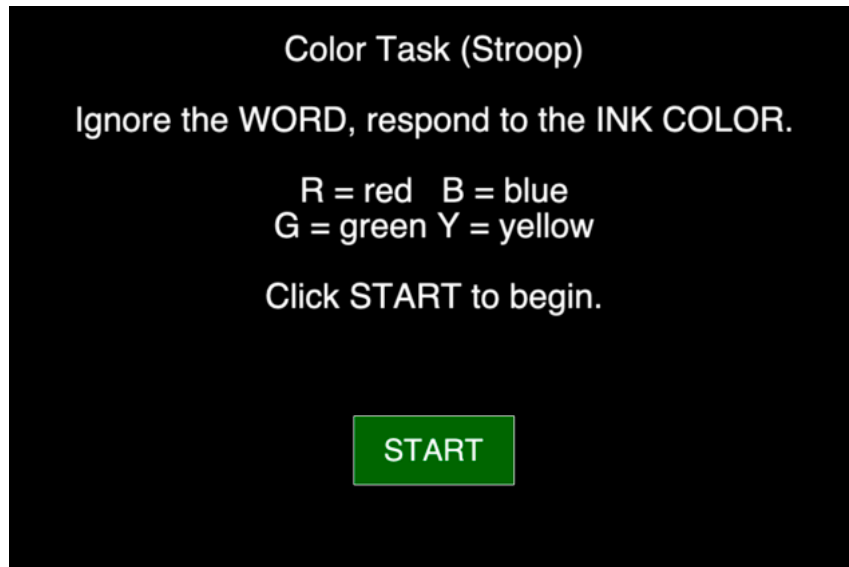
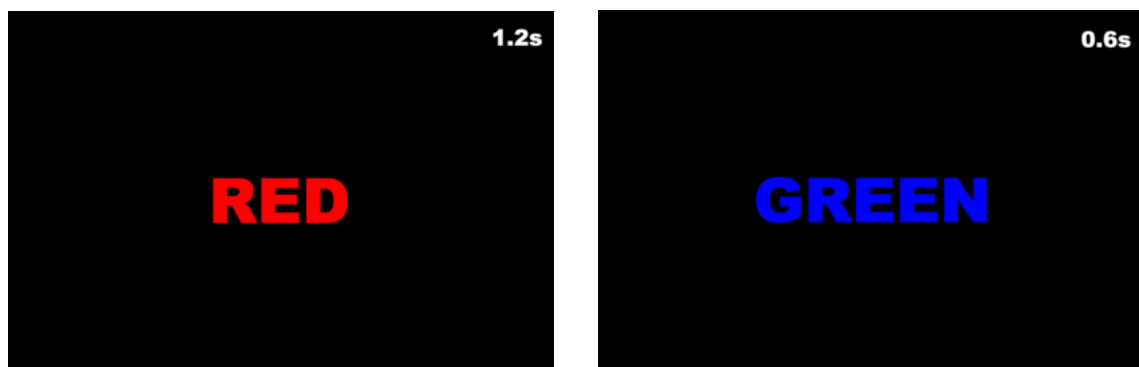


Figure 15: Instruction screen for the Stroop color task. Participants were instructed to ignore the word meaning and respond only to the ink color using predefined keyboard keys before starting the task.



(a) Congruent Stroop trial: the word meaning and ink color match.

(b) Incongruent Stroop trial: the word meaning and ink color do not match.

Figure 16: Example stimuli from the Stroop task. Participants were instructed to ignore the written word and respond only to the ink color. Congruent and incongruent trials were presented in randomized order.

To increase engagement and pressure on performance, the task was designed to show feedback every 15 trials. This feedback was a comparison message containing a motivational/demotivational message for 5 seconds, comparing the participant's task accuracy so far with the human accuracy rate for middle schoolers performing this task (94%; SD = 5%) (Gómez et al., 2015).



Figure 17: Examples of performance feedback screens presented during the Stroop task. Feedback was used to introduce performance pressure and maintain task engagement.

Behavioral measurements collected during this task were accuracy, reaction time, missed responses, and trial congruency.

6.5.2 Mental Arithmetic Task (MAT)

The mental arithmetic task (MAT) required participants to solve arithmetic problems as quickly and accurately as possible. They saw a problem on the screen, performed the calculation mentally, and entered their answer using the keyboard within a set time. The design aimed to induce stress through continuous effort and time constraints.

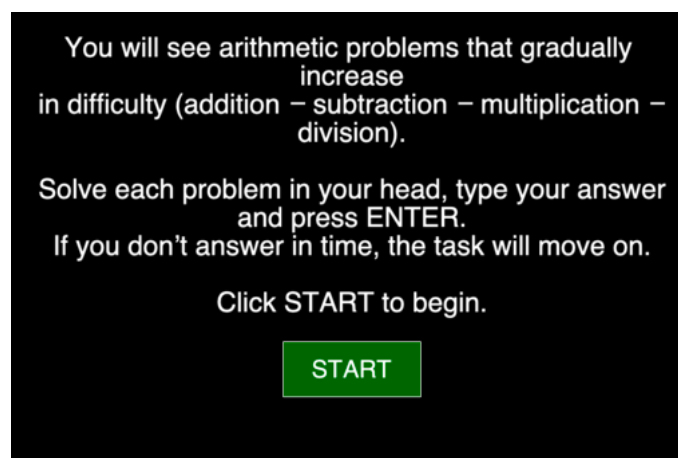


Figure 18: Instruction screen of the Mental Arithmetic Task (MAT). Participants are informed that arithmetic problems increase gradually in difficulty (addition to division) and that each problem must be solved within a limited time.

In each of the 60 trials, an arithmetic problem was shown in the format $a \text{ op } b = ?$, where the operands were single-digit numbers, except for division trials, which included single- or double-digit numbers. Participants responded by typing their answer.

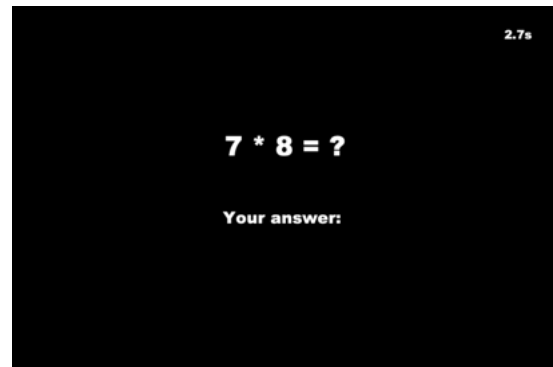
Participants entered their answers using the computer keyboard and saved their responses by pressing the ENTER key. If no answer was recorded within the time limit, the task proceeded to the next problem.

For each problem, a 4-second timer was provided, with a countdown visible in the top-right corner of the screen that turned red at the final second to emphasize the time pressure.

No planned rest was provided to participants, as problems were displayed continuously on the screen.



(a) Example MAT trial with less than 1 second remaining. The countdown timer turns red to emphasize time pressure.



(b) Example MAT trial with more than 1 second remaining, showing a neutral countdown timer.

Figure 19: Example trials from the Mental Arithmetic Task (MAT), illustrating the visual presentation of arithmetic problems and the on-screen countdown timer under different remaining time conditions.

The difficulty increased gradually over time. The task began with addition problems, subtraction with non-negative results, multiplication, and finally division problems with integer solutions.

Several stressors were used during this task. Background music Music, Freeport 2021 was played throughout the task. Based on informal verbal feedback collected after the task, approximately 80% of participants informally reported that the music increased their stress during the task. Participants described the music as pressuring. To add to the stress, the experiment facilitator stayed inside the experiment room and next to the participant for the first time to test an exploratory hypothesis that someone else's presence would increase stress; only 20% of the participants informally admitted that this increased their stress.

As with all tasks in Phase B, participants were encouraged to beat other participants' scores on the blackboard to create a competitive, evaluative element that increased pressure for some individuals.



Figure 20: Participant performing the Mental Arithmetic Task (MAT) during the experiment.

mental arithmetic task (MAT) was treated as a stress-dominant task. Attention was required to perform the calculations; the primary goal of the task design was to increase perceived stress rather than to measure attentional performance alone.

Behavioral measures collected during the MAT included the number of correct, incorrect, and missed responses, as well as reaction time.

6.5.3 Breath Hold (BH)

The breath-holding task (BH) task was used to induce short-term physical discomfort and physiological strain, as holding your breath can trigger the body's threat and arousal systems, even if there is no external danger (Berenz et al., 2012). Participants started the task by clicking a *START* button and were then asked to hold their breath for as long as they could. A large on-screen timer counted upward in seconds (updated continuously), and participants ended the task by clicking *STOP* when they were no longer able to continue.

To provide a simple performance incentive, the interface showed a short reward message every 10 seconds (e.g., "Nice work. You passed 70 seconds."), displayed for 1.5 seconds. Task start, reward events, stop time, and end were sent as event markers for later segmentation and analysis.

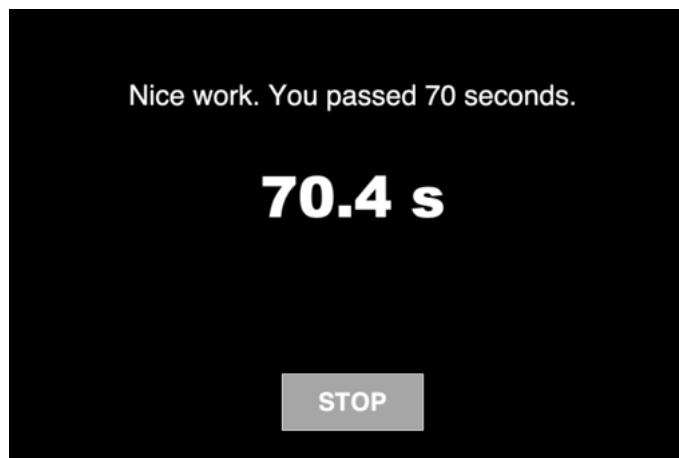


Figure 21: Breath Hold task feedback screen. The timer shows the current breath-hold duration, and a short reward message is displayed at each full 10-second milestone.

6.6 Expected Physiological Responses

Different physiological responses were expected across the tasks due to differences in attention demands, stress, and physical strain. These expectations are based on the literature and are meant to support the results of the conducted study.

Attention-Dominant Tasks

Tasks in Phase A were designed mainly to engage attention, with limited stress.

2-back task

This task involved continuous monitoring and comparison of stimuli. Earlier studies indicate increased heart rate, decreased heart rate variability, and changes in brain activity, such as increased theta activity in frontal and parietal regions, as well as changes in alpha and beta activity compared to low-demand conditions (Arutyunova et al., 2024; Hernández-Sabaté et al., 2024).

Reading comprehension

This task involved sustained attention and active processing associated with reduced alpha activity, particularly in frontal and occipital regions. Increased beta activity has also been observed during reading tasks and is commonly connected to focused attention and active thinking (Lestari et al., 2020).

Oddball

This task was included to elicit selective attention and target detection. However, the reviewed sources did not directly provide task-specific physiological expectations for this task. However, based on findings on **even-or-odd** task (Iovino et al., 2024) and on general attention research, mild changes in brain activity and autonomic measures may be expected, but no strong task-specific assumptions were made.

Stress-Dominant Tasks

The tasks implemented in Phase B were structured to introduce higher levels of stress through mechanisms such as time constraints, cognitive interference, and performance evaluation.

Stroop color-word task

Prior research links cognitive interference and acute stress to increased sympathetic nervous system activation, reflected in higher HR, reduced HRV, increased EDA, and reduced peripheral skin temperature (Martinez et al., 2017; Hanshans et al., 2024)

Mental arithmetic tasks

Generally, the tasks performed under time pressure are commonly used to trigger mental stress (Iovino et al., 2024). (Albaladejo-González et al., 2024; Mukherjee et al., 2024) report increased HR, reduced HRV, and changes in brain activity in frontal regions involved during problem-solving and executive control.

Breath-hold task

This task was included as a physical stressor. Holding the breath can lead to increased physical discomfort due to rising carbon dioxide levels, which may trigger a stress response. In previous research, shorter breath-hold duration has been linked to higher sensitivity to internal stress and lower tolerance to physical discomfort (Berenz et al., 2012)

Overall, tasks were expected to differ in the type and strength of responses: stress-dominant tasks were expected to trigger stronger autonomic and physiological responses than attention-dominant tasks, which were expected to primarily influence brain activity associated with focused cognitive processing.

Summary

Legend: ↑ increase, ↓ decrease, ~ no clear change expected, Δ change without consistent direction, – not specified in reviewed literature

Task	EEG	HR / HRV	EDA	ST
Reading (A)	↓ α , ↑ β	~ / ↓ HRV	~	~
Oddball (A)	–	–	–	–
2-back (A)	↑ θ , Δ α/β	↑ HR, ↓ HRV	~	~
Stroop (B)	Δ frontal activity	↑ HR, ↓ HRV	↑	↓
MAT (B)	Δ frontal activity	↑ HR, ↓ HRV	↑	↓ / ~
Breath-hold (B)	–	↑ autonomic strain	↑	Δ

Table 4: Overview of expected physiological responses for each task based on prior literature. Symbols indicate the expected direction of change relative to baseline conditions.

6.7 Synchronization

All physiological signals and task-related events were synchronized and recorded using the Lab Streaming Layer (LSL) framework with LabRecorder, which provided time-stamped streams in a shared timing framework to allow offline alignment. LabRecorder was set to save data in *xdf* format, following a BIDS-like folder and filename template (participant, session, task, run, modality).

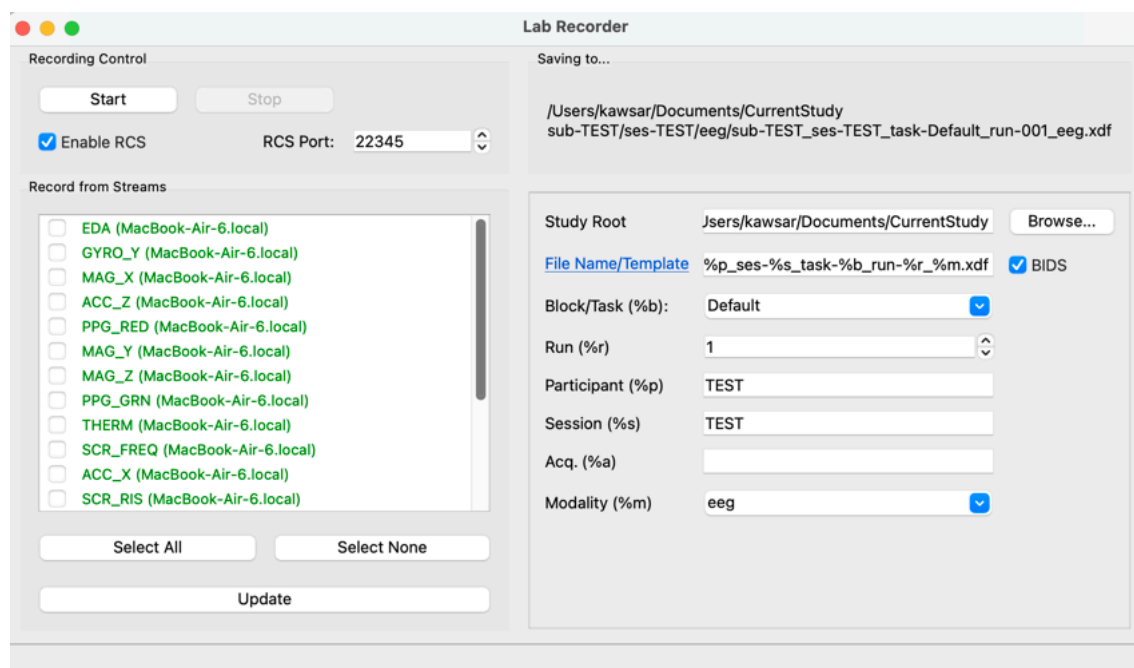


Figure 22: LabRecorder interface showing synchronized physiological data streams and the TaskMarkers event stream recorded via LSL.

Physiological data was recorded from three systems: the Emotiv MN8 and OpenBCI devices for EEG, and the EmotiBit device for peripheral signals. These continuous sensor streams, recorded alongside an LSL marker stream, captured task phases and participant behavior.

6.7.1 Marker Stream and Event Logging

Task events were transmitted via an LSL outlet implemented in PsychoPy using `pyls1`. A dedicated marker stream was used to send string-based event labels corresponding to task onsets, offsets, and key task events, which were recorded alongside physiological data for offline alignment.

Markers indicated the start and end of the experiment, baseline periods (eyes open and eyes closed), and neutral blocks. They also marked the start and end of Phases A and B, including the task order for each phase.

In each task, markers captured key events, including task start and end, stimulus presentation, trial start and end, and response outcomes, including reaction times when applicable.

Time Stamp	Relative Time (s)	Marker Value
251967.0551	0.00	EXP:START
251980.5744	13.52	BASELINE:EO:START
252040.5821	73.53	BASELINE:EO:END
252146.2033	179.15	NEUTRAL1:START
252326.2134	359.16	PHASE_A:START
252326.2135	359.16	PHASE_A_TASK_ORDER:[READING, ODDBALL, 2BACK]

Table 5: Example task event markers recorded via LabRecorder. Each marker encodes task structure and timing information and was used for offline alignment of physiological signals.

6.8 Participant Engagement and Motivation

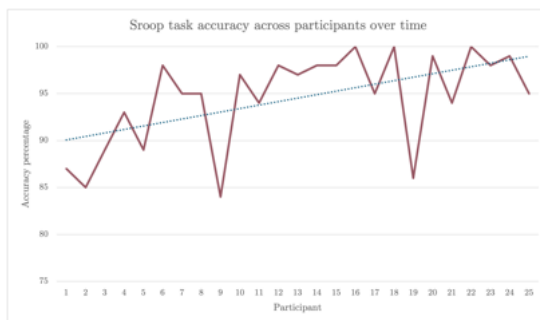
Participants were kept engaged to ensure they complied with tasks and maintained consistent effort.

In Phase A, engagement was supported by clear instructions, intuitive task interfaces, and short task durations, maintaining focus while minimizing fatigue or frustration. Subsequently, Phase B introduced additional motivational and evaluative elements to further increase engagement and, in some cases, perceived performance pressure.

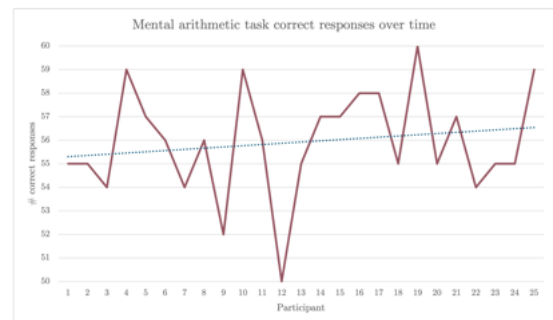
A whiteboard displayed previous participants' Phase B performances. Participants were told they could try to beat these records, creating a competitive, social-comparison element. Results improved over time. This feature was not intended to manipulate motivation quantitatively; rather, it served as an environmental factor.



Figure 23: Whiteboard used during Phase B to display task performance and encourage participants to beat previous records, introducing performance-related pressure. Participant identifiers have been blurred for privacy and GDPR compliance.



(a) Stroop task accuracy (%) across participants.



(b) Mental arithmetic task: number of correct responses.

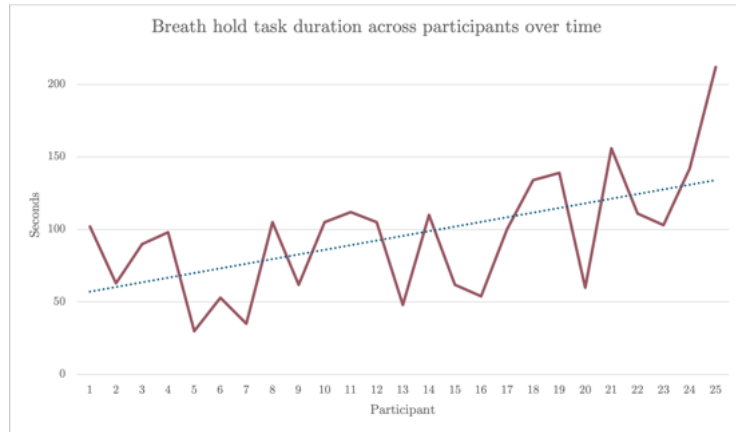


Figure 25: Breath-hold duration (s) across participants.

In addition, after completing the Phase B tasks, participants were given a small treat (e.g., doughnuts) as informal positive reinforcement and a gesture of appreciation. This was not tied to task performance and was provided equally to all participants.

Individual engagement and motivation varied, but these strategies broadly supported task completion.

7 Survey Design

In addition to physiological recordings, subjective data was collected to complement the statistical analyses and provide contextual information about participants' perceived attention, stress, and task experience. In total, each participant completed eight surveys: one pre-survey administered before task execution, six task-level surveys completed after each task in Phases A and B, and one post-survey at the end of the session. All surveys were administered digitally using SurveyXact (SurveyXact, n.d.). A description of each survey is provided below.

7.1 Pre-Survey

The pre-survey consisted of three sections: demographic information, baseline state assessment, and the Perceived Stress Scale (PSS).

The demographic section collected basic background information, including age, highest completed education level, hours of sleep prior to the session, recent caffeine intake, and the use of medication related to attention or stress.

The baseline state assessment section consisted of seven questions rated on a 5-point Likert scale ranging from 1 (Not at all) to 5 (Extremely). These questions were used to assess the participants' mental state at the start of the session, including perceived stress, fatigue, nervousness, and difficulties with concentration.

Finally, the participants completed the PSS, a 10-item self-report questionnaire that measures the extent to which individuals perceive situations in their lives as stressful (Townsend & Medvedev, 2025).

Overall, the purpose of the pre-survey was to collect background information and establish a subjective baseline of participants' mental state prior to task execution.

7.2 Task-Survey

The task-level survey was designed to obtain immediate subjective feedback after the completion of each task. It consisted of eight questions addressing perceived mental demand, effort, focus, stress, and frustration. Each question was rated on a 5-point Likert scale from 1 (Not at all) to 5 (Extremely).

Participants completed one task-level survey after each task in Phases A and B, resulting in a total of six task-level surveys per participant. Using the same set of questions for all tasks made it possible to compare participants' perceived task difficulty and stress across tasks.

7.3 Post-Survey

After completing all tasks, participants filled out a post-survey, consisting of two parts.

The first part included six questions rated on a 5-point Likert scale (1 = Not at all, 5 = Extremely), aiming to assess participants' perceived stress, fatigue, and mental exhaustion at the end of the experimental session.

The second part gathered overall feedback regarding device comfort, as well as participants' subjective assessment of which tasks they found most difficult, easiest, and most stressful.

The complete surveys can be found in Appendix A.4.

8 Participant Information

A total of 24 participants took part in the study. However, data from one participant were excluded from the analysis due to a corrupted *xdf* recording file. Of the remaining 23 participants, 34.8% were female. The mean age was 26 years ($SD = 6.9$), and 56.5% had completed a bachelor's degree. Participation was voluntary, and participants were recruited through personal contact. Those who agreed to participate were able to book a suitable time slot using an online scheduling tool (Calendly, n.d.).

Participants did not receive monetary compensation for their participation. After completing the session, they were offered a small sweet treat as a token of appreciation, but this was not mentioned in advance and was not intended as an incentive for participation.

Individuals with a history of neurological disorders, cardiovascular diseases, or other conditions that could affect EEG, heart rate, or electrodermal measurements were excluded from the study. Blood pressure was measured prior to the experiment to screen for unusually high or low values. All measurements were within acceptable ranges, and no exclusions were made based on blood pressure.

Prior to the experimental session, participants received an email reminder containing practical information about the study. They were asked to avoid caffeine, energy drinks, and heavy physical activity for at least three hours before the session, and to wear short sleeves or clothing that could be rolled up to facilitate sensor placement. The email also stated that participation in the study was completely voluntary and that participants could withdraw from the study at any time without providing a reason.

Before the experiment began, participants were handed a written consent form and given time to read and sign it. The consent form described the purpose of the study, what participation involved, potential risks or discomforts, data protection and confidentiality measures, and the voluntary nature of participation, as well as the research team's contact details.

The study was conducted in accordance with institutional ethical guidelines and the EU General Data Protection Regulation (GDPR). The full consent form is included in the Appendix A.2.

9 Data Collection and Storage

During each experimental session, continuous physiological data were recorded together with task-related event markers. In addition, survey responses collected before, during, and after the experiment were stored separately and linked to the physiological recordings using participant identifiers.

The following sections describe how task markers were generated, how the physiological data streams were recorded, and how the resulting data were stored and managed.

9.1 Task Event Markers

Task-related event markers were defined during task implementation in PsychoPy. These markers indicated key events throughout the experimental session, including the start and end of each task and phase. Additional event markers were also recorded during task execution.

All task markers were generated automatically by the PsychoPy task application and streamed to LabRecorder. The markers were recorded alongside the physiological data and shared the same timestamps.

9.2 Recording

As described in Section 4.4, all devices were configured to output their data streams to LSL, such that all recorded signals shared a common time base. A new recording was started and stopped in LabRecorder for each experimental session.

All available data streams were selected and recorded in LabRecorder to allow flexibility for later analyses and potential reuse of the dataset. However, only a subset of these streams was used in the analyses presented in this thesis. An overview of the data streams used for analysis is provided below, while a complete list of recorded streams is included in Appendix A.6.

Sensor	Data Stream	Description	Purpose
EmotiBit	EDA	Electrodermal Activity	Autonomic arousal and stress estimation.
EmotiBit	PPG_IR	PPG infrared (PPG IR)	HR & HRV extraction.
EmotiBit	HR	Heart rate	Instantaneous HR monitoring.
EmotiBit	THERM	Thermopile temperature	Stress-related changes.
Emotiv MN8	MN8_EEG	Raw EEG	EEG feature extraction.
System	TaskMarkers	Task and event markers from PsychoPy	Data segmentation.
OpenBCI	obci_eeg	OpenBCI raw EEG	Reference EEG for comparison.

Table 6: Subset of recorded streams used in the analyses

9.3 File Format

Each recording generated by LabRecorder was stored in the `xdf` file format. For analysis, the `xdf` files were converted into CSV files, with one file per data stream. Each CSV file contained a `timestamp` column based on LSL's global clock, together with the corresponding signal data. All `xdf` files collected during the experiment execution can be found in Appendix A.5.

9.4 Handling Missing Data

As mentioned in Section 5.2, there were instances in which PsychoPy crashed during the experiment. When this occurred, the recording was stopped, the task application was restarted from the beginning of the interrupted task, and a new recording was started in LabRecorder. As a result, some participants had multiple `xdf` files corresponding to different parts of the same session.

For these participants, the recordings were combined into a single dataset using a custom stitching script. The script operated on the exported CSV files for each recorded data stream. First, the `TaskMarkers` stream was used to identify task boundaries and to determine which tasks were completed successfully within each recording, based on a predefined task registry. Each session was cut at the first interrupted task, after which the script moved on to the next session and appended the completed tasks. This process was repeated until all sessions had been processed and all tasks in the protocol had been completed.

Moreover, a clock-alignment step was required as, in some cases, the physiological data streams were recorded using the global LSL clock while the `TaskMarkers` stream either used a local clock or switched between the two. When such mismatches were detected, the task marker timestamps were shifted to the global LSL time base using a physiological data stream as a reference.

After task selection and timestamp correction, all data streams were sliced using the corresponding `TaskMarkers` time windows and merged into a single continuous recording per participant.

There were also cases where certain data streams were missing or empty after conversion from `xdf` to CSV. In these cases, participants were not excluded from the study entirely. Instead, they were excluded only from analyses involving the missing data modality (for example, EEG-based analyses when EEG data were unavailable).

In a small number of cases, tasks were missing from some participants due to technical malfunctions. Specifically, one participant was missing one task and another two tasks. These tasks were therefore missing from the corresponding participant datasets and were not included in analyses involving those tasks. An overview of what streams and tasks were present for each participant can be found in Appendix A.7.

9.5 Data Storage

All collected data were stored in a SharePoint repository created for this project. Within the repository, a directory named `participant_data` was created, containing one subfolder per participant. Each participant folder was named using the participant ID and contained subfolders with the original `xdf` files as well as the corresponding CSV files.

For participants whose data required stitching, an additional subfolder named `stitched` was used to store the stitched CSV files for all recorded data streams.

In addition to the raw and stitched data, derived feature files generated during later analysis stages were stored in separate subfolders within each participant's directory. The extraction and use of these features are described in Chapter 12.

An overview of the data organization is shown in Figure 26.

```
participant_data/  
|-- P01/  
|   |-- xdf/  
|   |-- csv/  
|   |-- stitched/  
|   |-- features/  
|-- P02/  
|   |-- ...
```

Figure 26: Overview of the directory structure used to store participant data and derived outputs.

Part III: Data Analysis

10 Overview of Analysis Pipeline

The pipeline describes how the synchronized recordings collected during the experimental sessions were transformed into features and analyzed to investigate attention- and stress-related effects during the tasks. The purpose of this chapter is to outline the main stages of the pipeline, while detailed descriptions are provided in the following chapters.

Figure 27 illustrates the complete workflow, from raw data acquisition to statistical analysis and machine learning.

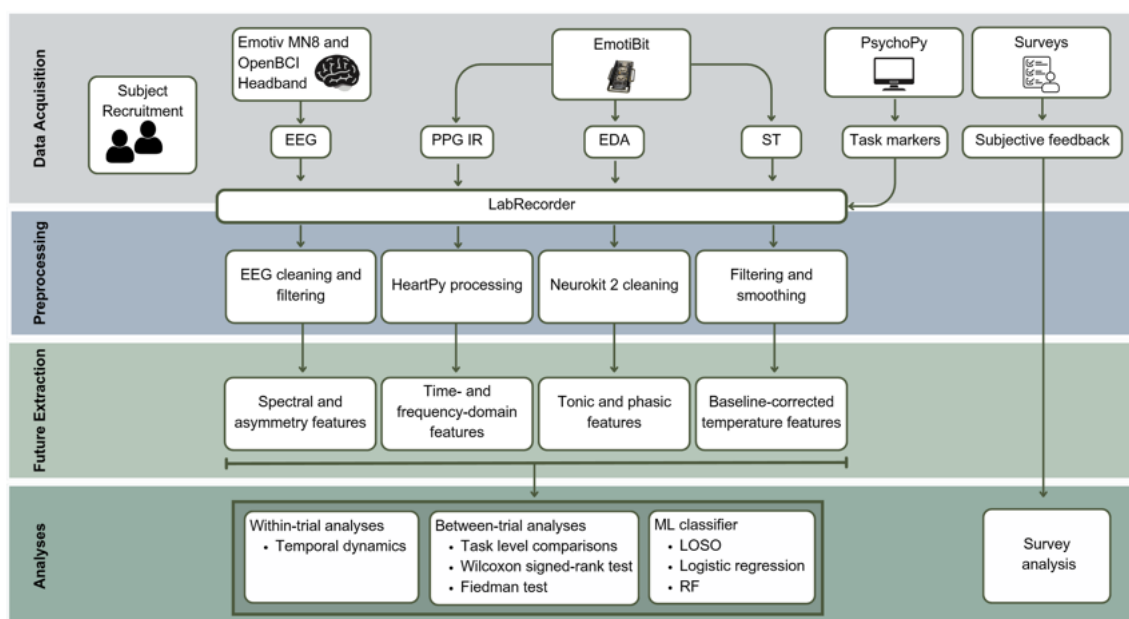


Figure 27: The framework is composed of four main steps: data acquisition, data preprocessing, feature extraction, and analyses (statistical and ML classifier).

Inputs

The inputs to the pipeline consist of the synchronized physiological data streams and task event markers recorded during the experiment (Table 6), as well as the subjective survey responses collected before, during, and after the experiment. All physiological data streams and task markers shared a common time base, which allowed task-specific segments to be identified within the continuous recordings.

Preprocessing

Before analysis, all physiological signals underwent preprocessing to prepare the data for further analysis. This stage involved filtering the raw signals and handling artifacts, all tailored to each signal modality.

Feature Extraction

Following preprocessing, features were extracted from the data. This stage transformed the continuous signals into a set of descriptive metrics. Feature extraction was performed

at two levels: over fixed time windows and over task segments. For task-level features, task event markers were used to segment the data. Features were extracted separately for each signal modality, resulting in feature sets that were later used for statistical analysis and machine learning.

At this stage, basic data validation and quality control were also performed by identifying and excluding windows or task segments that did not meet predefined quality criteria.

Analyses

The extracted features were used in a series of analyses. First, within-trial analyses were conducted to investigate temporal changes in the signals during task execution. Second, between-trial analyses were performed to compare physiological responses across different tasks and baseline conditions (neutral tasks). In addition, survey data were analyzed to identify which tasks were subjectively associated with higher cognitive demand, and to compare these ratings with the physiological findings to provide additional context. Finally, machine learning analyses were carried out to assess whether combinations of features could be used to infer cognitive states.

Outputs

The outputs of the analysis pipeline include statistical comparisons across tasks, temporal patterns observed within tasks, performance metrics from machine learning models, and summaries of subjective survey responses. Together, these outputs provide multiple perspectives on how attention- and stress-related processes manifest across different task conditions.

11 Data Preprocessing

As data was acquired raw from the sensors, it was necessary to clean the signals by applying filters and removing artifacts. Since each signal has different characteristics and the outputs desired were different, the preprocessing was tailored to each signal modality.

Figure 28 provides an overview of the preprocessing pipeline for all signal modalities.

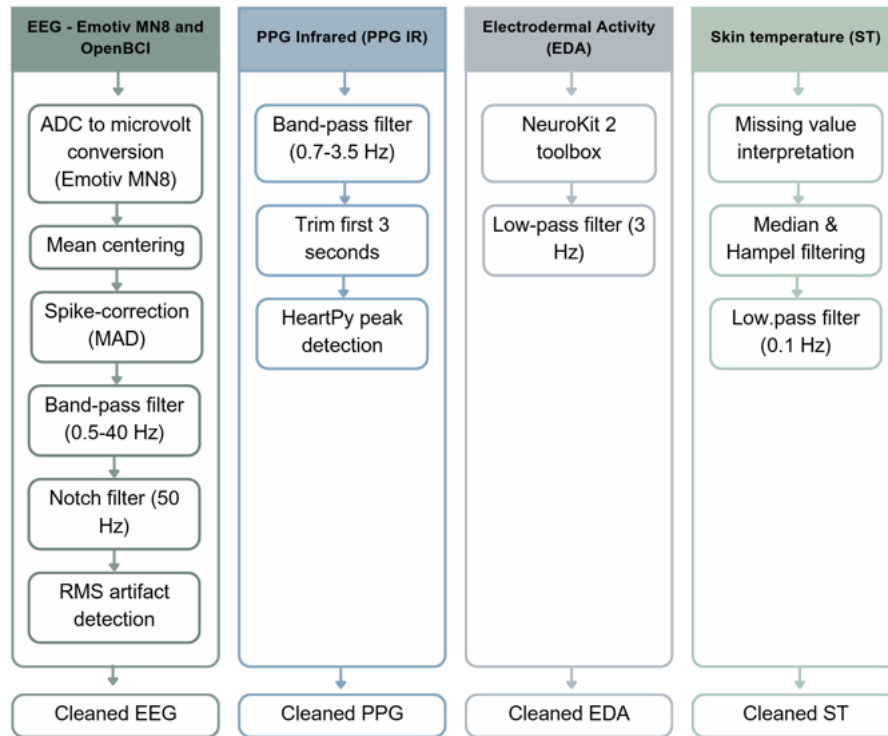


Figure 28: Overview of the preprocessing steps for each signal modality.

11.1 Emotiv MN8 & OpenBCI Electroencephalography

To ensure consistency across EEG devices and allow comparability between recordings, Emotiv MN8 and OpenBCI EEG data were preprocessed using the same pipeline, with only minor device-specific differences. One such difference concerns the signal units: while OpenBCI EEG data were already expressed in microvolts (μV), Emotiv MN8 EEG data were recorded in ADC counts. Consequently, the first preprocessing step for the Emotiv MN8 EEG data was the conversion of ADC counts to centered microvolts. For each channel, the mean of the ADC counts was computed and subtracted to center the signal around zero. The data were then scaled to microvolts using the device-specific dynamic range and bit resolution (Emotiv MN8 Technical Specifications, n.d.), according to the relation $\text{LSB} = \text{dynamic_range} / 2^{\text{bits}}$. Although such a conversion was not necessary for the OpenBCI EEG data, mean-centering was still applied in the same manner. All subsequent preprocessing steps were applied identically to both devices.

To reduce short-duration artifacts such as abrupt motion artifacts, cable movement, and headset movement, a spike correction procedure was applied. Impulsive artifacts were identified using an amplitude- and derivative-based outlier detection approach based

on the median absolute deviation (MAD). Samples were flagged as artifacts when their deviation from the median exceeded a predefined threshold, either in signal amplitude or in the first temporal derivative, allowing the detection of both large-amplitude spikes and sudden signal jumps. Detected artifacts were corrected via linear interpolation. To prevent overcorrection, channels in which more than 20% of samples were flagged were left unchanged and marked as low-quality for subsequent analyses.

After spike correction, the EEG signals were filtered using a zero-phase 0.5–40 Hz band-pass filter. The high-pass cutoff reduced slow baseline drift, while the low-pass cutoff attenuated high-frequency noise and muscle-related artifacts. In addition, a 50 Hz notch filter was applied to suppress power-line interference.

Finally, an additional artifact detection procedure was used to identify longer-lasting, high-energy signal contaminations, likely caused by motion. This method computed the root mean square (RMS) energy over short, overlapping time windows and applied z-score thresholding to detect segments with abnormally high energy. RMS energy was computed across all available EEG channels. Furthermore, as the OpenBCI EEG data contained many large-amplitude outliers, an additional amplitude-based mask was used to identify these segments. All detected artifacts, including impulsive spikes, high-energy bursts (RMS-based), and large-amplitude outliers (OpenBCI only), were flagged for exclusion during feature computation.

11.2 Photoplethysmography Infrared

The purpose of preprocessing the EmotiBit PPG IR channel was to prepare the data for HRV estimation. The first preprocessing step consisted of applying a band-pass filter in the range 0.7–3.5 Hz, corresponding to approximately 42–210 bpm. To remove transient effects introduced by the filtering, the first 3 seconds of the recording were trimmed.

The filtered PPG data were then processed using HeartPy, a heart rate analysis toolkit designed to handle noisy PPG recordings (HeartPy, n.d.). In brief, HeartPy detects pulse peaks in the filtered signal and computes inter-beat intervals (RR/IBI) from the detected peaks. It further applies segment-wise rejection to exclude periods of unreliable signal quality and filters implausible RR outliers. Finally, it computes heart rate and HRV metrics, including time-domain, frequency-domain (e.g. LF, HF, LF/HF), and additional descriptive statistics of the inter-beat interval series (quartiles, min/max, iqr).

11.3 Electrodermal Activity

The EDA stream was preprocessed using NeuroKit, a Python toolbox for neurophysiological signal processing (NeuroKit, n.d.). NeuroKit applies a 4th-order low-pass Butterworth filter with a 3 Hz cutoff frequency to reduce high-frequency noise. The cleaned signal was decomposed into tonic and phasic components, and skin conductance response (SCR)-related events, such as onsets, peaks, amplitudes, rise times, and recovery times, were detected.

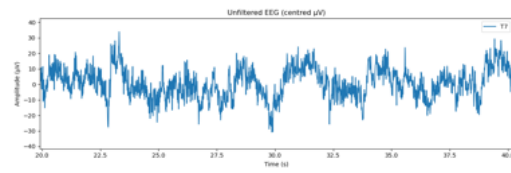
As with PPG, the first 3 seconds of the EDA recording were trimmed to remove initial transient effects.

11.4 Skin Temperature

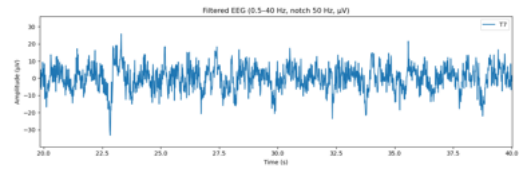
The EmotiBit skin temperature signal was preprocessed to remove missing values, suppress outliers, and smooth the signal. First, missing values were interpolated to obtain a continuous time series. Next, a median filter was applied to reduce isolated spikes while preserving the overall temperature trend. Subsequently, a Hampel filter based on a rolling median and the MAD was used to identify samples that deviated strongly from their local neighborhood. These outliers were replaced with the rolling median to remove abrupt jumps in the signal. Finally, a fourth-order low-pass Butterworth filter with a cutoff frequency of 0.1 Hz was applied to further smooth the data.

Representative examples of raw and preprocessed signals for each modality are shown in Figure 29.

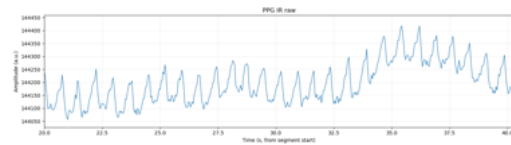
Raw vs. Preprocessed Data



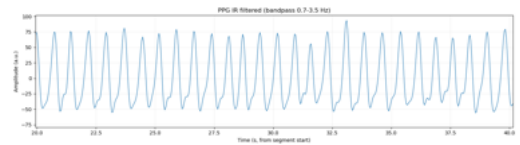
(a) EEG from Emotiv MN8 centered and in microvolts. (Note that Emotiv MN8 EEG already has an integrated hardware low-pass filter at 45 Hz).



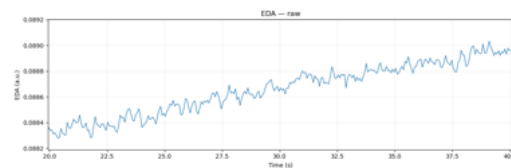
(b) Emotiv MN8 EEG after preprocessing.



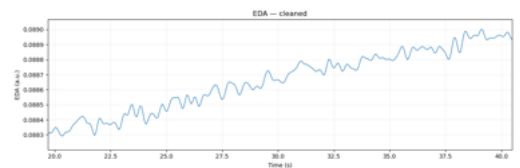
(c) Raw PPG Infrared (IR) from EmotiBit.



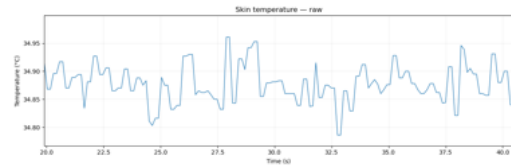
(d) PPG IR after preprocessing.



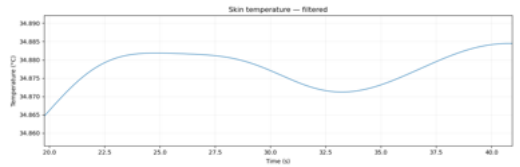
(e) Raw EDA from EmotiBit.



(f) EDA after preprocessing.



(g) Raw skin temperature from EmotiBit.



(h) Skin temperature after preprocessing.

Figure 29: Representative raw vs. preprocessed signals plots from a single participant over the same 20 s segment are shown for each modality.

12 Feature Extraction

For subsequent statistical analysis and machine learning, features were extracted from each signal modality on a per-participant basis. Feature extraction was performed at the window level to capture short-term dynamics within tasks and at the task level by aggregating window-level features using the task registry and task boundaries defined in the task markers CSV. Quality control was conducted at this stage to ensure that only high-quality windows contributed to task-level summaries.

A broad set of features was initially extracted for each modality to allow flexibility during analysis. However, only a subset of these features was used in the analyses presented in this thesis. Complete lists of all extracted features are provided in the appendix, while this chapter focuses on the features selected for analysis.

Figure 30 provides an overview of the features extracted from each signal modality.

12.1 EEG Features

EEG feature extraction was performed in the same way for the Emotiv MN8 and the OpenBCI headband to ensure consistency and comparability across devices.

EEG signals were segmented into sliding windows of 2 seconds with 50% overlap. This window length was selected based on prior work, which reported optimal classification performance using 2-second EEG windows (Hyun-Soo Choi et al., 2019). Feature extraction was performed independently for each window. Extensive quality control was applied at this stage: using the artifact detection procedures described in Section 11.1, windows were rejected if more than 20% of their samples were flagged as contaminated. Quality control metrics were computed for each task segment, including the number of samples and windows extracted, the number and percentage of retained windows, and the percentage of flagged samples. These QC summaries were generated separately for each participant.

For feature extraction, frequency bands were defined as follows: delta (0.5–4 Hz), theta (4–8 Hz), alpha (8–12 Hz), beta (12–30 Hz), and low gamma (30–40 Hz). A large set of time- and frequency-domain features was initially extracted from both EEG channels from both devices, including spectral band power, hemispheric band power asymmetries, time-domain statistics, and Hjorth parameters. This broader feature set was extracted to allow flexibility in subsequent analyses. An overview of all extracted EEG features is provided in Appendix A.8.1.

For the analyses presented in this thesis, only a subset of the extracted features was selected. Relative band power was used to facilitate comparisons across participants by normalizing for inter-individual differences in absolute EEG power. In addition, for the Emotiv MN8, hemispheric asymmetry was quantified using the log-transformed difference in band power between the left (T7) and right (T8) channels, a measure that has been associated with stress-related and affective responses (Halim et al., 2020; Eyad Talal Attar et al., 2021).

The EEG features selected for statistical analysis are summarized in Table 7.

Feature family	Channels	Frequency bands	Statistic
Log band asymmetry	Emotiv MN8: T7, T8 OpenBCI: ch1, ch2	Alpha, Beta, Gamma	Median (Q50)
Relative band power	Emotiv MN8: T7, T8 OpenBCI: ch1, ch2	Theta, Alpha, Beta, Gamma	Median (Q50)
Relative band power variability	Emotiv MN8: T7, T8	Theta, Alpha, Beta, Gamma	Interquartile range (IQR)

Table 7: EEG features used for statistical analysis.

Additionally, for the machine learning classifier, the theta-alpha and beta-alpha ratios were also used.

12.2 HRV Features

HRV features were computed over sliding windows within each task segment. For time-domain features, the window length was set to 100 s with a step size of 30 s. For frequency-domain features, a longer window length of 120 s was used, as frequency-domain HRV indices require longer segments for reliable estimation (Arutyunova et al., 2024). The complete list of extracted HRV features is provided in Appendix A.8.2.

The analyses presented in this thesis focused on a subset of HRV features. Mean heart rate (bpm) was included as a direct and intuitive measure of overall cardiovascular arousal. SDNN was used as an indicator of global HRV, while RMSSD was selected due to its sensitivity to parasympathetic activity. In the frequency domain, LF power is considered to reflect slower fluctuations in heart rhythm, HF power to reflect parasympathetic activity, and the LF/HF ratio as a rough indicator of the balance between sympathetic and parasympathetic responses.

Quality control and rejection decisions were performed at both the window and task-segment level and were based on signal quality and RR interval validity. Segments were rejected if HeartPy peak detection resulted in too few valid beats, if a large proportion of RR intervals were physiologically implausible, or if overall signal quality was insufficient. These criteria ensured that HRV features were computed only from reliable PPG segments. Window- and task-level quality control summaries were generated for each participant.

12.3 EDA Features

EDA feature extraction included descriptive statistics (e.g., mean, standard deviation, minimum and maximum values, and percentiles) computed from the cleaned EDA signal, as well as from its tonic and phasic components. In addition, SCR features were extracted from the phasic component, including the number of SCR peaks, SCR rate per minute, and SCR amplitude-related measures. The complete list of extracted EDA features is provided in Appendix A.8.3.

Features were computed at both the task and window level. For the windowed analysis, a window length of 60 s with a step size of 30 s was used. Task segments or windows

were rejected if they were empty or too short to allow reliable decomposition into tonic and phasic components.

For the statistical analyses presented in this thesis, a subset of EDA features was selected. Specifically, the mean tonic EDA level, SCR rate per minute, and the number of SCR peaks were used to represent both tonic and phasic aspects of EDA.

12.4 Skin Temperature Features

Similar to EDA, skin temperature (ST) features were computed at both the window and task level. Windows were defined with a length of 60 s and a step size of 30 s. Feature extraction was performed on the filtered temperature signal as well as on a baseline-corrected temperature signal. Baseline correction used the “Neutral 1” task as baseline for phase A tasks and “Neutral 2” for phase B tasks, as described in Chapter 13.

An overview of all extracted ST features is provided in Appendix A.8.4. For the analyses presented in this thesis, only baseline-corrected features were used. Specifically, the selected metrics were the change in mean temperature relative to baseline (delta mean), the slope of the temperature signal per minute (delta slope), and the difference between the end and start of the task segment (delta end-start).

Windows and task segments were rejected if they were empty or if an insufficient number of valid samples was available for feature computation.

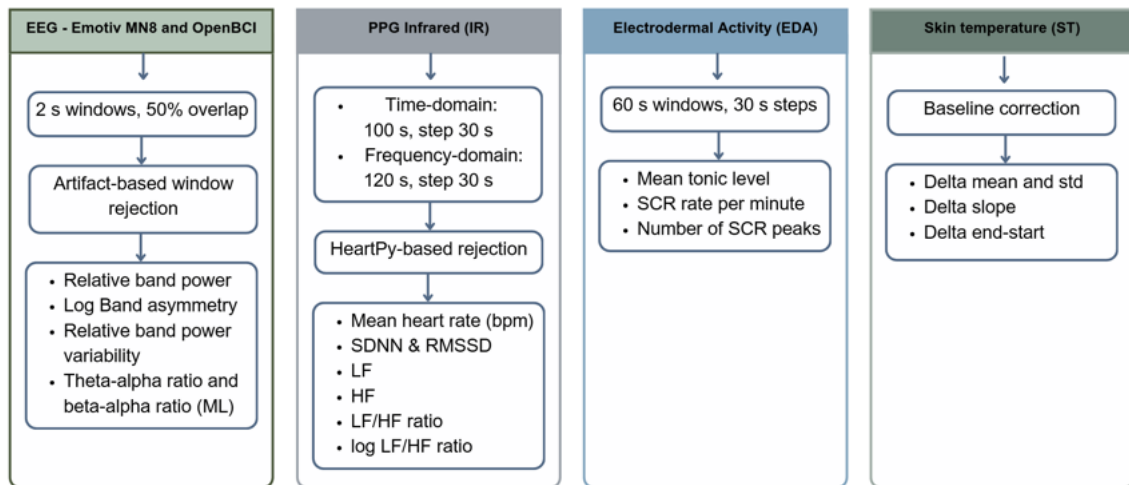


Figure 30: Overview of features extracted from each signal modality.

13 Statistical Analyses

To reduce inter-subject variability and enable comparability across participants, all features were baseline-corrected. With this approach, a value of zero indicates no change relative to baseline, positive values indicate an increase, and negative values indicate a decrease. By expressing features as changes relative to a neutral baseline, the analyses focus on within-subject responses induced by the active task conditions, which is the primary interest of this study. In the following, “active tasks” refers to all non-neutral task segments, namely reading, 2-Back, oddball, MAT, SCWT (Stroop), and breath-holding.

For baseline correction, “Neutral 1” was used as the reference for tasks in Phase A, as it represents a state minimally influenced by cognitive load, temporally close to the subsequent tasks, and unaffected by task-induced fatigue. Phase B tasks were performed after participants had already completed several tasks. Thus, their physiological state could not be assumed to match the initial baseline due to accumulated effects of sustained arousal, habituation, or fatigue. Therefore, a second neutral condition was recorded mid-experiment and used as the baseline for Phase B tasks. This approach allowed baseline correction to account for slow drifts in the biosignals across the session, while still isolating task-specific effects. Throughout this chapter, baseline-corrected features are referred to as “delta” features.

Two types of statistical analyses were performed: within-trial and between-trial analyses. Within-trial analyses focused on temporal dynamics during task execution, using windowed features to examine how physiological signals evolved over time and whether consistent patterns emerged across participants. Between-trial analyses, in contrast, operated on task-level aggregated features and were used to compare physiological responses between tasks and their baseline references, as well as between different tasks.

It is important to note that windows or task segments flagged during the quality control step at feature extraction (Chapter 12) were excluded from all analyses. As a result, the final sample size varied across tasks and modalities.

13.1 Within-Trial Analysis

The within-trial analysis aimed to address the question: how does a physiological signal change over the course of a task?

Within-trial analyses were performed primarily at the window level to capture temporal changes during task execution, allowing potential habituation, escalation, or recovery patterns to be observed. For peripheral signals, the windowed feature files generated during feature extraction were used. The selected features were baseline-corrected according to the baseline assigned to each task. Participant-level trajectories were then aligned in time, aggregated across participants, and visualized as mean time courses with standard error of the mean (SEM) shading.

In addition, summary metrics describing within-task dynamics were computed at the participant level from the windowed data (e.g. average change over time). These metrics were later aggregated at the task level and served as inputs to the between-trial analyses.

For EEG, window-level data were already summarized for each participant; therefore, within-trial analyses for EEG were restricted to the generation of task-level summaries.

13.1.1 HRV Within-Trial Results

Within-trial HRV analysis examined how heart rate (bpm) and time-domain HRV metrics (RMSSD and SDNN) evolved over the course of each task. Due to strict signal quality control for HRV estimation, several tasks resulted in a limited number of valid windows. Since the aim of the within-trial analysis was to examine overall trends in signal evolution across a task, only tasks with a sufficient sample size ($N > 10$) were included in the HRV within-trial analysis. Accordingly, temporal patterns are described for reading, 2-Back, and Stroop, while the remaining tasks (oddball, breath-holding, and MAT) are not discussed further.

In the 2-Back task, bpm remained relatively stable during the first two-thirds of the task and increased toward the end. RMSSD also showed stable values initially, followed by a marked decrease in the later windows. SDNN declined gradually over time. Together, these patterns indicate a reduction in heart rate variability as the task progressed, consistent with increasing cognitive demand.

During the reading task, bpm increased during the initial phase, decreased thereafter, and then rose sharply toward the end of the task. RMSSD and SDNN displayed fluctuating temporal patterns, characterized by an initial decrease followed by a pronounced increase and a subsequent decline. These non-monotonic patterns suggest heterogeneous cardiac responses across participants during sustained reading.

For the Stroop task, bpm decreased slightly at the beginning and then increased steadily over time, indicating sustained task engagement. Both RMSSD and SDNN decreased across task windows, with SDNN showing a pronounced and progressive decline. This pattern is consistent with increased arousal and cognitive load during Stroop performance.

Compared to reading and 2-Back, Stroop elicited more consistent and monotonic changes in both bpm and HRV metrics. Overall, within-trial HRV results suggest that HRV metrics were sensitive to sustained cognitive engagement, but also exhibited considerable inter-individual variability at the within-trial level.

Representative HRV time-course plots for the 2-Back task are shown in Figure 31. The full set of HRV within-trial plots is provided in Appendix A.9.1.

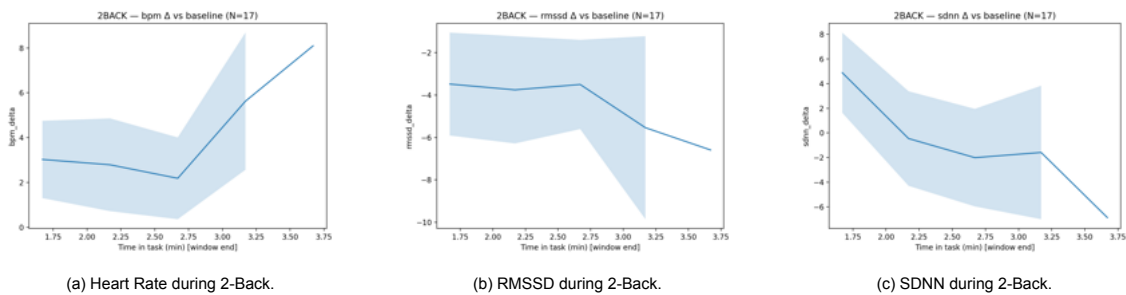


Figure 31: Temporal dynamics of HR, RMSSD, and SDNN during the 2-Back task.

13.1.2 EDA Within-Trial Results

Within-trial EDA analyses focused on the temporal dynamics of phasic activity, represented by the SCR rate, and tonic activity, represented by baseline-corrected tonic EDA.

Across tasks, SCR rate showed transient temporal dynamics rather than sustained increases, with patterns varying across task types. In general, SCR activity tended to be elevated early or mid-task and decreased toward the end, although the exact shape differed between tasks. For example, during reading, SCR showed an overall decreasing trend, with a brief increase around the middle of the task. The 2-Back task showed relatively stable SCR rates early on, followed by a late increase and subsequent decline. The MAT exhibited a gradual increase in SCR rate that plateaued toward the end, while oddball showed a slight gradual increase over time. Breath-holding elicited an early increase in SCR rate that remained relatively constant thereafter. Stroop showed an initial increase, followed by a relatively stable period and a decrease toward task completion. Inter-individual variability increased toward the later parts of the task, as reflected by a larger SEM. This effect is mainly explained by a reduced number of participants contributing data toward the end of the task.

In contrast, tonic activity displayed more consistent temporal patterns across tasks. For most tasks, tonic EDA increased during the early phase of the task and then gradually plateaued or decreased. Breath-holding elicited a rapid tonic increase followed by a sharp decline. Reading and 2-Back showed a more monotonic increase in tonic EDA over time, indicating sustained arousal.

Overall, tonic EDA captured slower, sustained changes in arousal, whereas SCR rate reflected more transient, event-related sympathetic responses.

Representative temporal plots for the 2-Back task are shown in Figure 32. Complete within-trial EDA time-course plots for all tasks are provided in Appendix A.9.2.

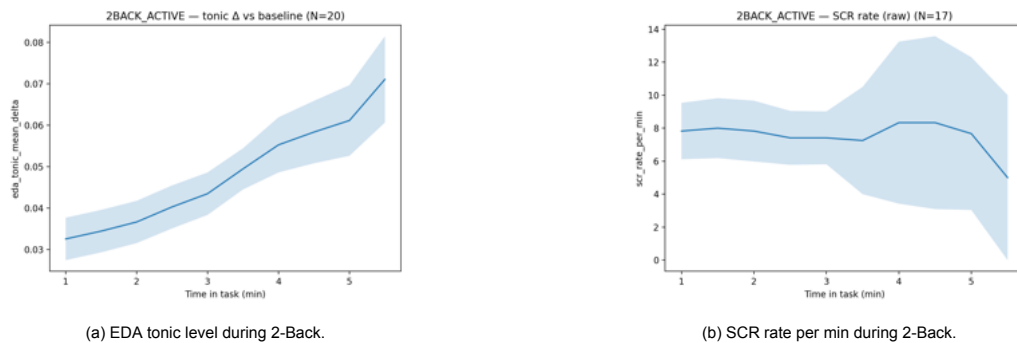


Figure 32: Temporal dynamics of EDA tonic and SCR rate per min during the 2-Back task.

13.1.3 ST Within-Trial Results

For skin temperature, within-trial analyses examined how baseline-corrected temperature evolved over the course of each task. Overall, ST showed slower temporal changes than EDA, with task-specific patterns and substantial inter-individual variability.

During reading, temperature increased gradually during the early phase of the task, followed by a pronounced drop around the middle and a gradual recovery thereafter.

Stroop and MAT both showed an early increase in temperature, with Stroop reaching an early peak and MAT continuing to increase for a longer duration, after which both tasks exhibited a marked decline.

In contrast, the 2-Back task showed relatively stable temperature values during the first half of the task, followed by a sharp increase in the second half and a subsequent decrease. Breath-holding showed relatively stable temperature values at the beginning of the task, followed by a brief increase of smaller magnitude. Oddball showed little systematic change over time, with a slight decrease toward the end of the task.

Taken together, skin temperature changes unfolded slowly over the course of task execution, indicating that skin temperature captured gradual, task-dependent shifts rather than rapid transient responses.

Representative time-course plots for the 2-Back and MAT task are shown in fig. 33. The full set of ST within-trial plots is provided in Appendix A.9.3.

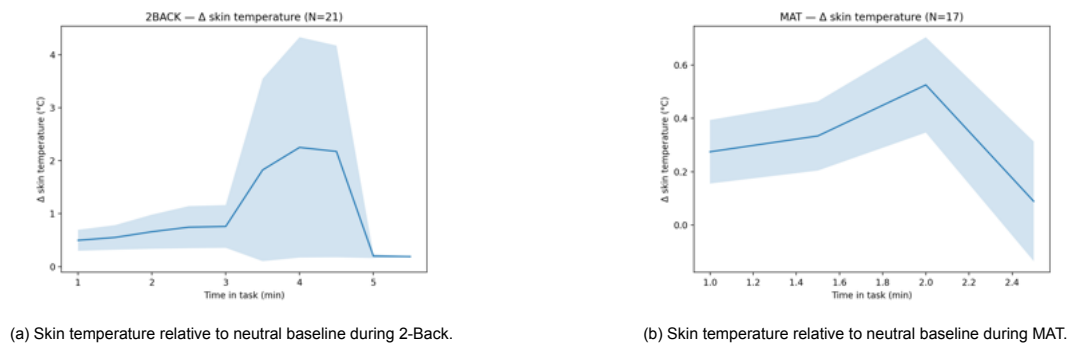


Figure 33: Temporal dynamics of skin temperature change (ΔST) during the 2-Back and MAT tasks.

13.2 Between-Trial Analysis

The between-trial analysis aimed to answer two questions: (1) do physiological responses differ between a task and its baseline condition, and (2) do tasks differ from each other in terms of their physiological impact?

Using task-level aggregated metrics from within-trial analysis, the first step was to generate task-level plots to visualize features' response magnitudes across tasks and modalities. These plots provide descriptive context and visualize group-level trends across tasks.

To address the first question, Wilcoxon signed-rank tests were conducted on baseline-corrected task-level metrics. For each task and feature, descriptive statistics (N, median, q25, q75) were computed. The Wilcoxon signed-rank test assessed whether the median delta differed from zero; zeros were excluded from hypothesis testing, and tests were only performed when at least five non-zero deltas were available. A significant result indicates a systematic task-related change relative to baseline. Effect sizes were quantified using the rank-biserial correlation, which ranges from -1 to $+1$, with values close to zero indicating little or no systematic difference. To control for multiple comparisons, Benjamini–Hochberg false discovery rate (BH-FDR) correction was applied per feature

across tasks. Nonparametric tests were chosen due to the repeated-measures design, small sample sizes, and non-normal feature distributions.

To assess whether certain tasks elicited stronger responses than others, Friedman tests were performed on task-level delta features. These analyses were restricted to participants for whom all tasks were available, which resulted in a reduced sample size due to quality control exclusions. When a Friedman test indicated a significant effect, pairwise Wilcoxon signed-rank tests were performed between task pairs, with BH-FDR correction applied across all comparisons for the given feature.

13.2.1 EEG Between-Trial Results

Task-Level Response Patterns

Emotiv MN8 EEG

Theta power showed a consistent decrease across all tasks, with the strongest decreases observed for MAT, Stroop, and 2-Back. This pattern was evident for both the median and the IQR, with the exception of the breath-holding task, where theta variability (IQR) increased. The overall suppression of relative theta suggests that theta did not reliably increase with cognitive demand in this dataset.

Across all cognitive tasks (reading, oddball, 2-Back, Stroop, and MAT), the task-level bar plots showed a suppression of relative alpha power (see Figure 34). The largest negative delta values were observed for Stroop, 2-Back, and MAT. This pattern was highly consistent across both temporal electrodes (T7 and T8) and across summary metrics (median and IQR). In contrast, breath-holding showed values close to zero with comparatively high variability, clearly separating it from the cognitively demanding tasks.

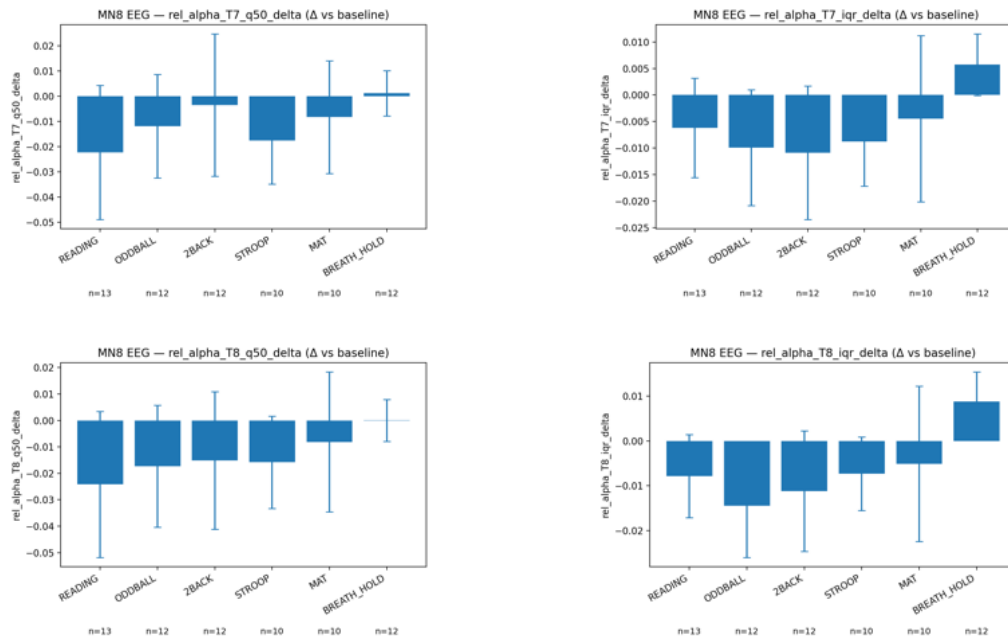


Figure 34: Task-level changes in relative alpha power across tasks for Emotiv MN8 EEG. Relative alpha power is shown for electrodes T7 (top) and T8 (bottom), summarized using the median and interquartile range (IQR). All values are expressed as change relative to baseline.

Relative beta power predominantly showed negative delta values across cognitive tasks, particularly for Stroop, 2-Back, and MAT, for both the median and IQR. Oddball and reading exhibited moderate positive changes for the median, while breath-holding showed negative values closer to zero. The tasks with the highest beta variability were Stroop and reading. These patterns were consistent across both T7 and T8 electrodes.

Low gamma effects were weaker and more variable. Stroop, MAT, and 2-Back generally showed small negative changes, whereas reading and oddball exhibited values close to zero or small positive changes. Breath-holding showed minimal changes but relatively high variability, indicating heterogeneous responses.

The full set of plots can be found in Appendix A.10.1.

OpenBCI EEG

For OpenBCI EEG, relative theta power (median) showed positive changes for reading, oddball, and 2-Back, with Stroop showing near-zero changes. In contrast, breath-holding exhibited a clear decrease in theta power. Theta variability (IQR) was generally small across tasks.

Alpha power showed consistent suppression across both EEG channels (ch1 and ch2) and both summary metrics (median and IQR) for all tasks, including breath-holding. The strongest decreases were observed for MAT, followed by Stroop and 2-Back, while oddball showed the weakest suppression (see Figure 35).

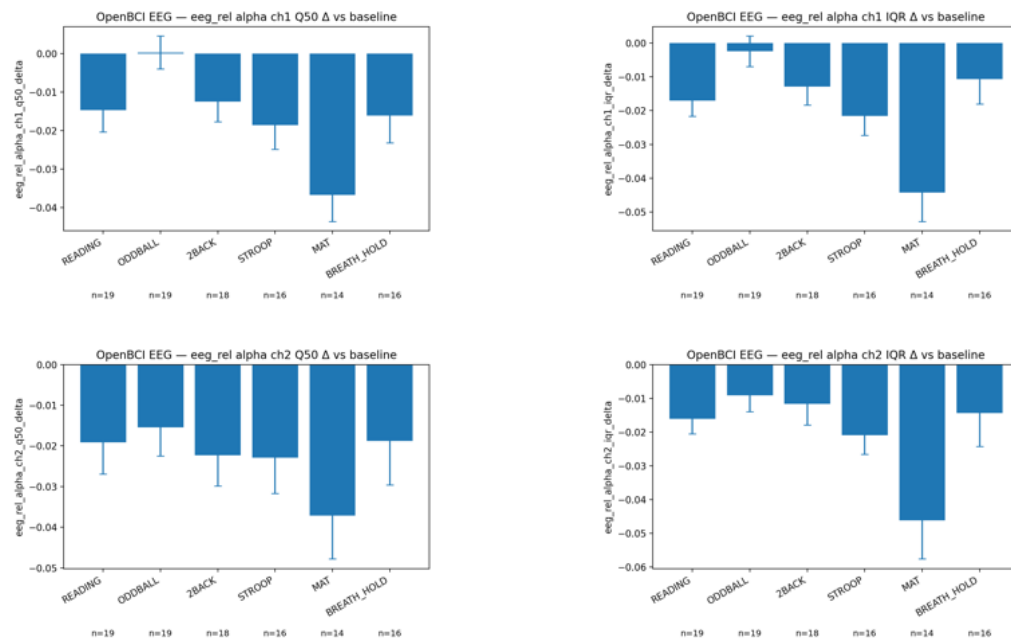


Figure 35: Task-level changes in relative alpha power across tasks for OpenBCI EEG. Relative alpha power is shown for electrodes Fp1 (ch1 - top) and Fp2 (ch2 - bottom), summarized using the median and interquartile range (IQR). All values are expressed as change relative to baseline.

Beta power exhibited a similar pattern to alpha, with consistent suppression across tasks. MAT showed the largest decrease in beta power and the lowest variability, suggesting a more uniform response across participants.

Gamma power showed only small changes across tasks. Although median values generally decreased, variability was high, making it difficult to distinguish task-related effects in the gamma band.

The full set of plots can be found in Appendix A.10.2.

Cross-Comparison Between OpenBCI and Emotiv MN8

Both Emotiv MN8 and OpenBCI EEG showed alpha suppression during cognitively demanding tasks, reflecting increased attentional engagement and cognitive effort. Stroop and MAT elicited the strongest alpha suppression, suggesting that alpha power reliably differentiates higher versus lower cognitive demand tasks.

MAT additionally showed reduced variability in both alpha and beta bands, suggesting more consistent responses across participants. In contrast, breath-holding consistently differed from the cognitive tasks across all frequency bands, with minimal power changes and higher variability, which is coherent given its non-cognitive and physiologically driven nature.

Theta power results differed between devices. OpenBCI EEG showed increased theta power for reading, oddball, and 2-Back, which aligns with prior findings linking frontal theta activity to sustained attention and working memory (Hernández-Sabaté et al., 2024). In contrast, MN8 EEG showed theta suppression across tasks, likely reflecting differences in sensor placement, sensitivity to frontal versus temporal theta activity, and the use of relative band power.

Wilcoxon and Friedman Results

Emotiv MN8 EEG

Across Emotiv MN8 EEG features, Wilcoxon signed-rank tests revealed only two features with statistically significant deviations from baseline after BH-FDR correction: beta-band variability (IQR) at T7 and theta median power at T8.

During the Stroop task, beta-band variability (IQR) at T7 remained significant after BH-FDR correction ($p_{\text{fdr}} = 0.035$), with a median delta of 0.025 and a large effect size ($r = 0.8$). This indicates that beta-band activity at T7 showed greater variability relative to baseline. This pattern is consistent with the more dynamic nature of the Stroop task.

Several tasks, particularly oddball and 2-Back, showed a reliable decrease in theta median power at T8. Both oddball and 2-Back reached statistical significance after BH-FDR correction ($p_{\text{fdr}} = 0.048$, $r = -0.50$ for both). Reading and MAT also showed trends toward theta reduction (Reading: $p_{\text{fdr}} = 0.053$; MAT: $p_{\text{fdr}} = 0.056$). This pattern is consistent with a reduction in slower background activity during cognitively demanding tasks.

Friedman tests conducted on complete-case participants ($N = 7$) identified a significant task effect only for beta-band variability (IQR) at T7. This aligns with the Wilcoxon results, indicating that Stroop stands out relative to other tasks for this feature. Post-hoc Wilcoxon comparisons suggested that beta variability during Stroop tended to be higher than during oddball and 2-Back. However, none of the pairwise contrasts survived BH-FDR correction, likely due to the reduced sample size.

Complete Wilcoxon and Friedman test outputs, including post-hoc comparisons, are reported in Appendix A.11.

OpenBCI EEG

Overall, OpenBCI EEG showed more widespread statistically significant effects than Emotiv MN8. Across both OpenBCI channels, relative alpha (IQR and median) exhibited negative median deltas during reading, Stroop, and MAT, with multiple effects remaining significant after BH-FDR correction. For example, alpha variability (IQR) at channel 1 was significantly reduced during reading ($p_{\text{fdr}} = 0.0035$, $r = -0.58$), Stroop ($p_{\text{fdr}} = 0.0054$, $r = -0.75$), and MAT ($p_{\text{fdr}} = 0.0007$, $r = -1.0$), with comparable effects observed at channel 2. This pattern is consistent with both alpha suppression during cognitively demanding tasks and reduced inter-individual variability during demanding tasks.

Relative beta median (q50) also showed significant reductions during 2-Back, Stroop, and MAT. At channel 1, beta median decreased during 2-Back ($p_{\text{fdr}} = 0.0080$, $r = -0.56$), Stroop ($p_{\text{fdr}} = 0.0039$, $r = -0.50$), and MAT ($p_{\text{fdr}} = 0.0007$, $r = -1.0$), with channel 2 showing even stronger and more consistent reductions across tasks. Relative gamma median exhibited a similar pattern, with significant decreases during 2-Back, Stroop, and MAT (e.g., channel 1 during MAT: $p_{\text{fdr}} = 0.0015$, $r = -0.86$). Together, these findings indicate task-related shifts in spectral composition during increased cognitive demand, rather than isolated changes within a single frequency band.

Friedman tests conducted on complete-case participants ($N = 9$) revealed multiple significant task effects across alpha, beta, and gamma features (e.g., alpha IQR at channel 1: $\chi^2 = 14.27$, $p = 0.014$; beta median at channel 1: $\chi^2 = 16.49$, $p = 0.0056$). Post-hoc Wilcoxon comparisons indicated that the oddball task differed from the more cognitively demanding tasks (reading, Stroop, and MAT), particularly for alpha variability and beta median features (e.g., oddball vs Stroop in alpha IQR at channel 1: $p_{\text{fdr}} = 0.039$). Additionally, theta median power at the second channel differed primarily for breath-holding when compared to reading, Stroop, and MAT (e.g., breath-holding vs Stroop: $p_{\text{fdr}} = 0.0195$), suggesting that theta variations in the OpenBCI recordings were driven mainly by changes in physiological state, such as breath-holding, rather than by differences in cognitive task demands.

Complete Wilcoxon and Friedman test outputs, including post-hoc comparisons, are reported in Appendix A.11.

13.2.2 HRV Between-Trial Results

Task-Level Response Patterns

Task-level plots of frequency- and time-domain HRV features were examined to identify overall response patterns across tasks. Heart rate (bpm) showed an increase relative to baseline for all cognitive tasks, with MAT showing the largest increase (see Figure 36a). In contrast, breath-holding exhibited a slight decrease. These findings align with elevated heart rate during increased mental effort, and with breath-holding breaking the pattern since respiratory manipulation alters cardiac dynamics.

RMSSD showed near-baseline values for reading and Stroop, positive shifts for oddball and MAT, and negative shifts for 2-Back and breath-holding. A decrease for 2-Back is consistent with reduced parasympathetic influence under working memory load. Vari-

ability was very high for Stroop and MAT, indicating heterogeneous responses across participants.

Across all cognitive tasks, SDNN shows small positive increases relative to baseline, whereas breath-holding produced a distinctly larger increase. For the LF/HF ratio and its log-transformed, reading, MAT, and Stroop exhibited positive shifts relative to baseline, while oddball and 2-Back showed negative or near-zero shifts. Breath-holding demonstrated a large positive increase, which is consistent with a strong sympathetic dominance (see Figure 36b).

Overall, heart rate turned out to be the most robust indicator of task-related arousal. Frequency-domain metrics primarily reflected physiological state changes rather than changes in cognitive load, and breath-holding differed consistently from all cognitive tasks across HRV features.

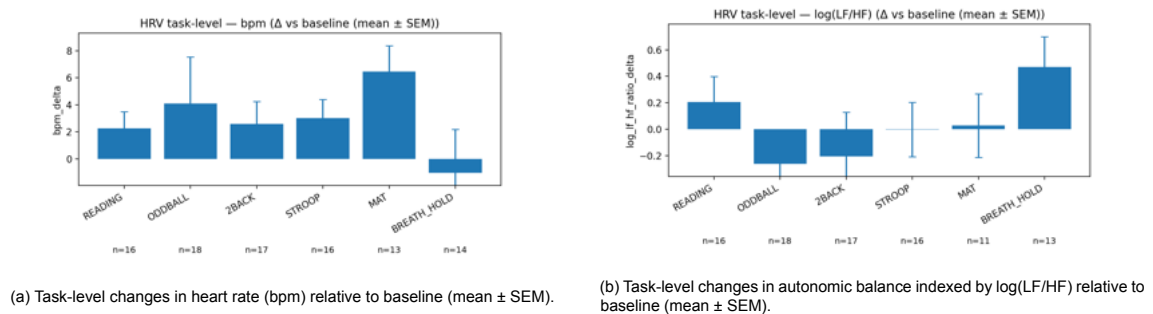


Figure 36: Task-level changes in heart rate and log(LF/HF) across tasks.

The full set of plots can be found in Appendix A.10.3.

Wilcoxon and Friedman Results

Wilcoxon signed-rank tests revealed that only MAT survived BH-FDR correction for heart rate, with a median increase of +7.9 bpm ($p_{\text{fdr}} = 0.048$) and a large effect size ($r = 0.90$). This indicates that MAT reliably induced an increase in heart rate relative to baseline across participants. Reading, 2-Back, and Stroop also showed positive median heart-rate changes and large effect sizes, but these effects did not survive BH-FDR correction, indicating that their effects are consistent but not statistically significant. This pattern aligns with the task-level bpm plots.

Breath-holding showed a significant increase in SDNN, with a median change of +28.8 ms ($p_{\text{fdr}} = 0.018$, $r = 0.92$), indicating that SDNN was primarily sensitive to physiological state changes rather than cognitive demand. RMSSD and frequency-domain HRV measures did not show systematic task-related effects.

Friedman tests indicated that tasks did not produce equivalent heart-rate responses ($\chi^2 = 12.46$, $p = 0.029$). However, post-hoc Wilcoxon tests for bpm did not yield any pairwise comparisons that survived BH-FDR correction. The largest effect sizes were observed for contrasts involving MAT, but these comparisons did not reach statistical significance, likely due to the reduced complete-case sample size ($N = 10$).

Overall, these findings indicate that heart rate is the most reliable HRV marker of cognitive demand in this dataset.

Complete Wilcoxon and Friedman test outputs, including post-hoc comparisons, are reported in Appendix A.11.

13.2.3 EDA Between-Trial Results

Task-Level Response Patterns

Task-level plots revealed positive tonic shifts relative to baseline across all tasks, with the largest increases observed for MAT and breath-holding (see Figure 37). The pronounced tonic increase during MAT indicates sustained sympathetic arousal under continuous mental effort, while the increase during breath-holding reflects physiological stress. Lower tonic shifts for reading, oddball, 2-Back, and Stroop suggest task engagement without prolonged stress.

Phasic EDA responses showed a different pattern. The number of SCR peaks increased most strongly during 2-Back and reading, while oddball and Stroop demonstrated positive but smaller shifts. In contrast, MAT and breath-holding showed reductions in SCR count. A similar pattern was observed for SCR rate, with reading, oddball, 2-Back, and Stroop showing positive increases relative to baseline, while MAT showed a slight decrease.

Taken together, phasic SCR features primarily reflect stimulus-driven and dynamically structured tasks, whereas sustained cognitive effort and physiological stress are characterized by elevated tonic EDA and suppressed phasic responding.

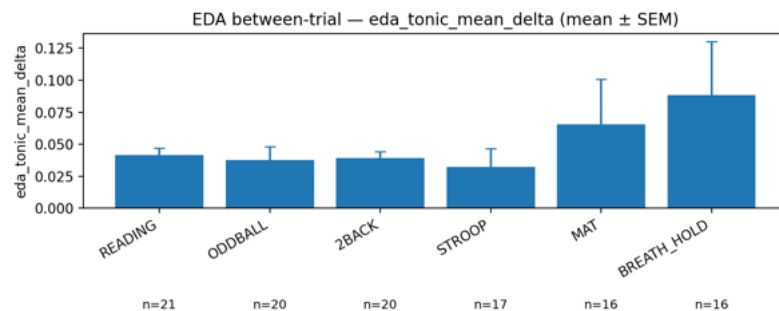


Figure 37: Task-level changes in tonic electrodermal activity relative to baseline (mean $\pm$ SEM).

The full set of plots can be found in Appendix A.10.4.

Wilcoxon and Friedman Results

Wilcoxon signed-rank tests showed consistent task-related increases in tonic EDA. All tasks, including breath-holding, exhibited statistically significant increases relative to baseline after BH-FDR correction, with large effect sizes (typically $r > 0.8$).

Phasic SCR rate increased significantly for the reading, oddball, and 2-Back tasks, with the largest effect observed for 2-Back ($r = 0.96$). This pattern indicates that phasic SCR rate is sensitive to stimulus-driven and dynamically structured tasks. Consistent with this, the number of SCR peaks also increased for reading and 2-Back. In contrast, the breath-holding task was associated with a reduction in phasic SCR activity, likely reflecting autonomic regulation.

Despite the clear task-versus-baseline effects, Friedman tests did not reveal significant differences between tasks for any EDA feature. This suggests that while EDA distinguishes task engagement from baseline, it does not support a consistent ordering of

cognitive tasks, likely due to inter-individual variability in phasic responses and reduced complete-case sample size ($N = 12$).

Complete Wilcoxon and Friedman test outputs are reported in Appendix A.11.

13.2.4 ST Between-Trial Results

Task-Level Response Patterns

In the task-level plots, the temperature slope showed near-zero to slightly positive values for reading, oddball, 2-Back, and breath-holding; slightly negative values for Stroop; and a pronounced negative slope for MAT. This indicates that temperature tended to remain stable for most tasks, whereas MAT was associated with a sustained decrease over time. This pattern suggests a gradual peripheral cooling response during MAT, consistent with sympathetic vasoconstriction during cognitively demanding or stressful tasks.

Temperature variability further differentiated the tasks. MAT showed low variability, indicating a consistent pattern across participants, while Stroop exhibited increased variability, reflecting more heterogeneous temperature responses. The remaining tasks showed very low variability.

Overall, temperature changed most during 2-Back and Stroop, while MAT was characterized by a gradual decrease across the task.

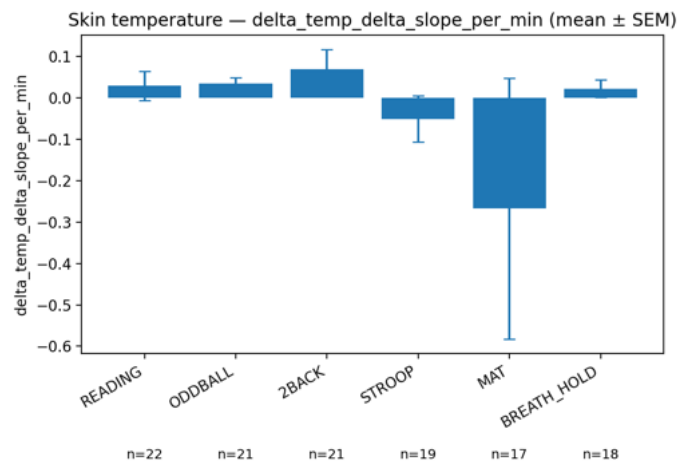


Figure 38: Task-level skin temperature slope ($^{\circ}\text{C}/\text{min}$, Δ vs. baseline; mean $\pm$ SEM).

The full set of plots can be found in Appendix A.10.5.

Wilcoxon and Friedman Results

Wilcoxon signed-rank tests showed that mean skin temperature increased relative to baseline for all tasks, with all tasks remaining significant after BH-FDR correction. This reflects engagement and physiological activation. In contrast, the temperature slope showed significant task-related trends only for the oddball and 2-Back tasks. Furthermore, 2-Back was the only task for which the end–start temperature change was both consistent across participants and statistically significant.

Friedman tests were non-significant, indicating that skin temperature features did not reliably differentiate between tasks. This is likely due to the slow and cumulative nature of

skin temperature dynamics, as well as substantial inter-individual variability and reduced complete-case sample size ($N = 12$).

Complete Wilcoxon and Friedman test outputs are reported in Appendix A.11.

14 Survey Analyses

This chapter presents the analysis of subjective survey data collected throughout the experiment in three main stages: before task execution, after each task, and after completing all tasks. The goal of these surveys was to understand each participant's subjective experience of attention, stress, cognitive workload, and perceived performance during task performance. Furthermore, the collected measures were selected to provide additional insight into the physiological recordings that may not be readily apparent from biosignals alone. The responses were analyzed at both participant and task levels to reveal subjective differences across tasks, thereby helping contextualize statistical and machine-learning analyses.

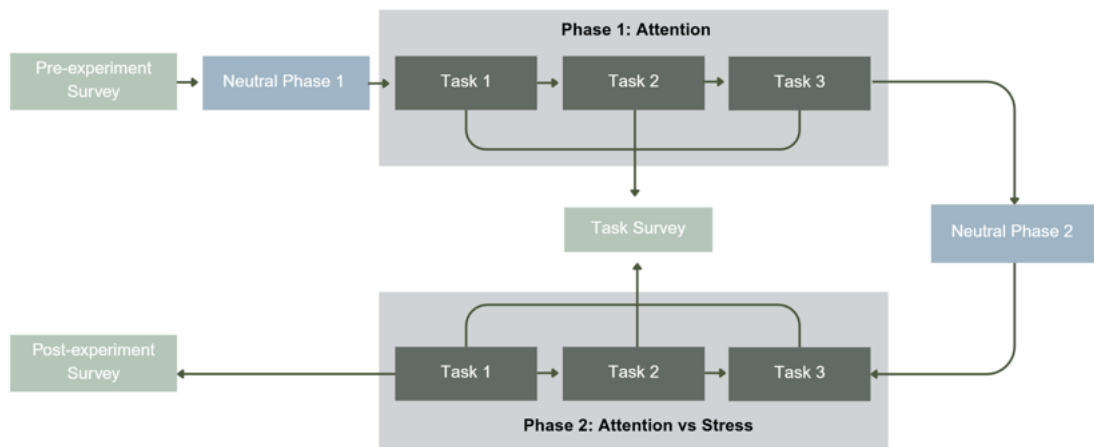


Figure 39: Timeline of survey administration and task execution during the experimental protocol.

14.1 Pre-experiment Survey

Before the experiment, participants completed a pre-experiment survey using the Perceived Stress Scale (PSS-10) (Cohen et al., 1983) to assess baseline perceived stress. The PSS-10 consists of 10 questions addressing experiences of unpredictability, lack of control, and perceived coping ability. Negatively worded items were reverse-scored accordingly, and a total PSS score was calculated for each participant.



Figure 40: Perceived Stress Scale (PSS-10) stress level categories.

Participants showed high baseline stress (mean 31.5 ± 2.96 , median 32, range 23–38), with no extreme outliers, indicating comparable starting conditions and inter-individual variability.

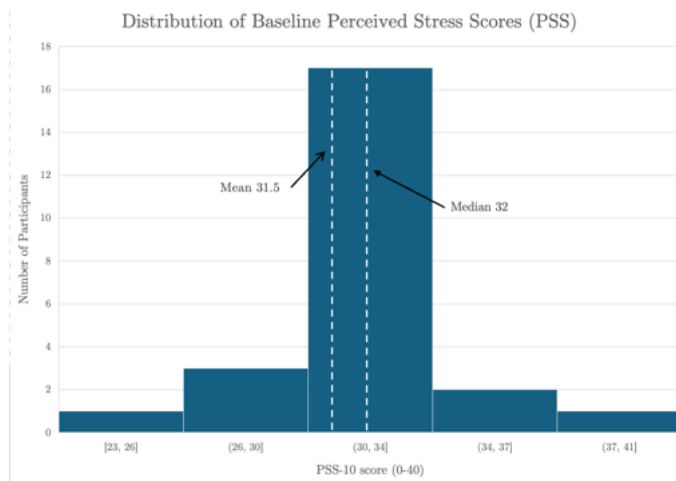


Figure 41: Distribution of baseline Perceived Stress Scale (PSS-10) scores across participants.

Given that PSS scores validly indicate participants' stress levels, these were compared with combined task-level ratings of mental effort, stress, and attention to assess how baseline stress influenced subsequent task experiences. Correlation analyses showed only weak associations ($|r| < 0.25$), showing that perceived stress was minimally related to participants' experiences of task difficulty, stress, or attention.

Thus, the pre-experiment survey described the baseline context rather than predicting outcomes. With no strong link between baseline stress and task ratings, differences in task experience appeared mostly due to the tasks' demands rather than prior stress levels.

14.2 Task-level Subjective Ratings

This section explains how participants experienced the experimental tasks and aims to identify systematic differences across tasks. These descriptive results provide a basis for building and comparing more complex subjective measures in the subsequent sections.

After each task, participants were asked to provide subjective ratings of their perceived experience during task execution. These included a questionnaire about cognitive demand, stress, and attention, with specific questions on mental demand, effort, focus, time pressure, stress, frustration, performance, and engagement. The ratings were provided on a Likert-type scale and averaged across participants, delivering task-level mean values.

14.2.1 Internal Consistency of Subjective Measures (Cronbach's α)

To determine if the questionnaire items could be grouped into a higher-level subjective construct related to the main research objectives, internal consistency was analyzed using Cronbach's alpha.

The Subjective Cognitive Load Index was assembled from items measuring mental health, effort, time-pressure, and frustration. This group showed good internal consistency ($\alpha = 0.82$). Likewise, the subjective stress index was derived from stress and frustration items, which showed excellent internal consistency ($\alpha = 0.92$). The Subjective Attention Index, based on focus and engagement items, demonstrated acceptable internal consistency ($\alpha = 0.66$), largely due to the conceptual distinction between focus and subjective engagement.

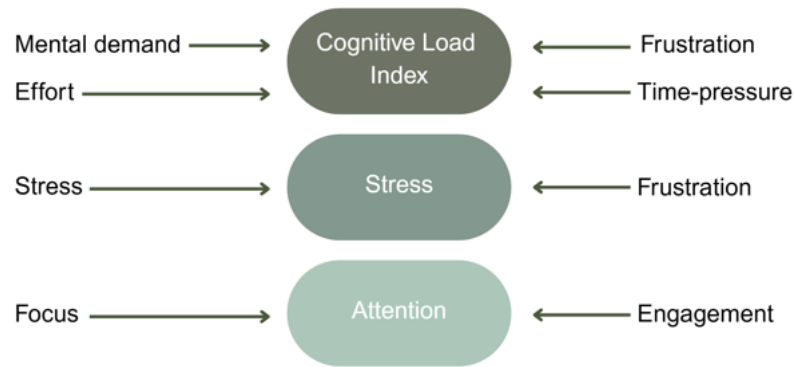


Figure 42: Mapping of task-level questionnaire items to composite subjective indices.

14.2.2 Subjective Mental Demand and Effort

Clear differences were observed across tasks in perceived mental demand and effort. MAT had the highest average mental demand score. This was followed by Stroop. These tasks also required high effort. Participants found them cognitively challenging and demanding continued mental effort.

In contrast, oddball and 2-Back tasks were rated as the least mentally demanding. Breath-holding task (BH) and reading comprehension (RC) tasks showed moderate mental demand. However, BH was linked to high perceived effort despite moderate cognitive demand. This was mainly due to its physiological nature and the competitive experimental environment. Similarly, RC required sustained effort. This may be due to participants' lack of familiarity with the text's subject and uncertainty about the upcoming question's difficulty or form.

Overall, the experimental task successfully varied in perceived cognitive challenge, supporting their application for examining variations in task-related cognitive load.

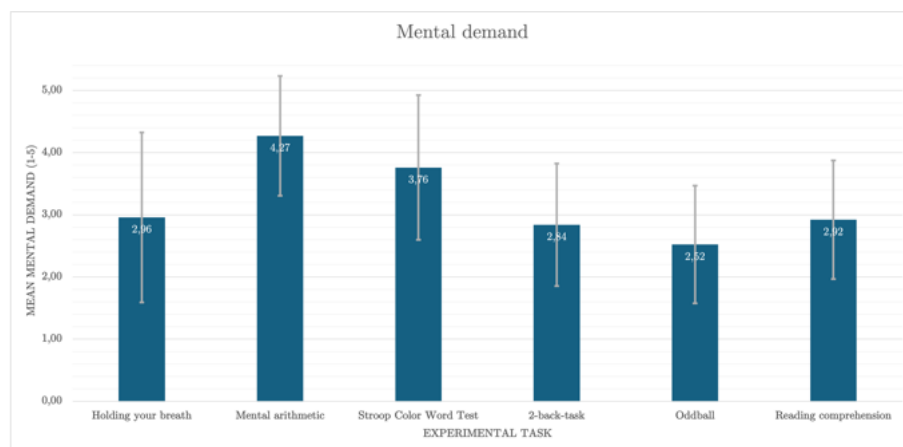


Figure 43: Mean self-reported Mental demand levels (1–5) across experimental tasks. Error bars denote standard deviation.

14.2.3 Perceived Stress and Frustration

Stress levels were reported differently across tasks. As expected, MAT had the highest stress scores for time pressure, auditory load, self-evaluative pressure, peer pressure,

and error sensitivity. This was followed by BH, which was associated with increased time, physiological stress, and a sense of competition. The 2-Back task showed moderate stress levels, while the Stroop task elicited comparatively lower stress despite its high cognitive demand.

The oddball and reading comprehension tasks were associated with the lowest stress ratings, as expected. Frustration ratings follow a similar pattern but at a lower overall intensity across all tasks, suggesting participants did not experience strong negative affect during task execution.

Overall, the variability observed in stress ratings across tasks was greater than that observed for focus-related measures, indicating that perceived stress was a more discriminative dimension of task experience.

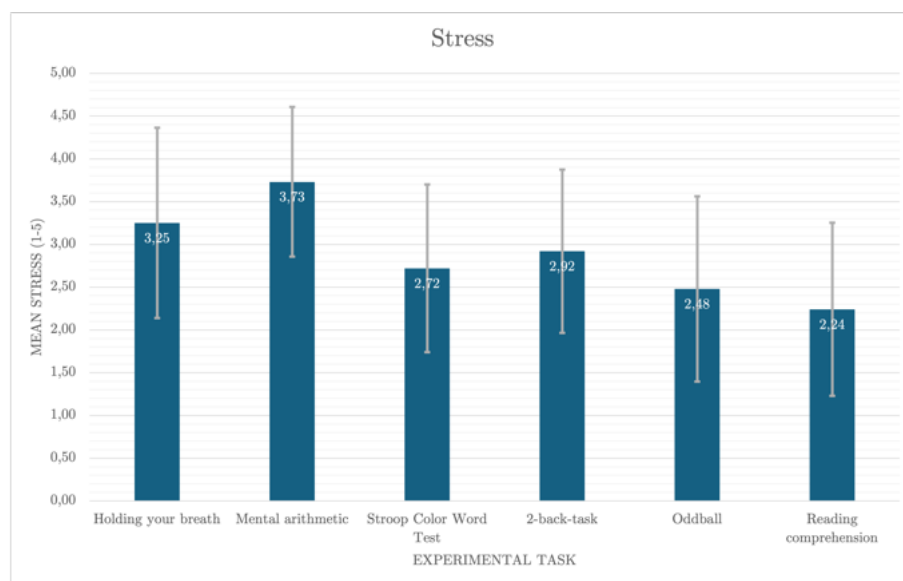


Figure 44: Mean self-reported stress levels (1–5) across experimental tasks. Error bars denote standard deviation.

14.2.4 Focus and Engagement

Focus ratings remained high and stable throughout the experiment. The MAT, SCWT, and 2-Back tasks resulted in the highest focus scores due to their sustained attentional demands. This trend continued even in tasks with lower focus requirements.

Engagement showed a similar trend: high-demand tasks such as MAT and SCWT elicited high engagement. Importantly, task difficulty did not reduce engagement; instead, participants found challenging tasks more engaging than unpleasant.

Overall, while cognitive demand increased, it did not reduce focus or engagement; rather, it led to sustained focus.

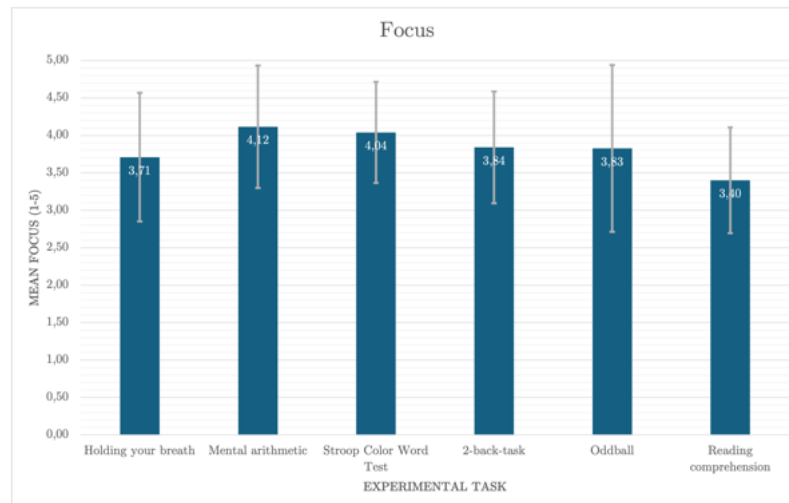


Figure 45: Mean self-reported focus levels (1–5) across experimental tasks. Error bars denote standard deviation.

14.2.5 Perceived Performance

The performance ratings received showed an interesting pattern throughout the experiment. Tasks with a higher cognitive load, such as MAT, SCWT, and the 2-Back task, were also associated with relatively high perceived performance. In contrast, RC was associated with lower perceived performance ratings, despite its lower stress and cognitive demand scores being on the higher end.

This pattern suggests that the participants felt more successful and confident in performing cognitively engaging tasks. These ratings reflect subjective perceptions of performance rather than objective task accuracy, and should therefore be interpreted as indicators of self-assessed task success.

14.2.6 Summary of Task-level Subjective Differences

In summary, subjective task-level ratings indicate that the experimental tasks were experienced differently across multiple dimensions. The tasks varied clearly in perceived cognitive load, stress, and effort, while focus and engagement remained consistently high between tasks. Stress-related ratings showed greater variability than focus-related ratings, highlighting stress as a key state differentiating the task experience.

These task-level differences provide a basis for constructing and comparing composite subjective indices, which are examined in the following section to further characterize attention- and stress-related task profiles.

14.3 Attention-Dominant and Stress-Dominant Task Profiles

Section 14.2 described subjective ratings across individual dimensions. This section integrates those results by comparing composite indices of cognitive load, stress, and attention across tasks. The goal is to provide an understanding of how each participant experienced the experimental task. This section focuses on finding which tasks were most strongly associated with attentional engagement, stress responses, or both. By exploring this, the study aims to identify tasks that primarily elicit attentional engagement, stress responses, or both.

Task-level indices were constructed as described in the Methods chapter. These were averaged across participants. The indices provide a higher-level representation of task experience and enable direct comparisons of tasks along the dimensions most relevant to the study objectives.

14.3.1 Comparison of Composite Subjective Indices Among Tasks

Comparison of the Cognitive Load Index, Stress Index, and Attention Index revealed clear differences in how tasks were subjectively experienced. Mental arithmetic showed the highest cognitive load and stress indices, along with a close attention index, indicating that this task was perceived as both cognitively demanding and stressful, while still requiring sustained attention. This pattern suggests that MAT represents a task with combined high cognitive and emotional demands.

The SCWT and the 2-Back task showed higher attention indices with lower stress levels than MAT. These tasks were perceived as demanding in terms of focus and cognitive control but did not produce equally strong stress responses.

Compared to other tasks, Oddball showed lower cognitive load and stress indices but a higher attention index. This indicates participants experienced Oddball as requiring sustained attention with minimal mental or emotional load, making it an attention-focused, low-stress task.

Following Oddball, the BH task showed moderate cognitive load, elevated stress, and moderate attention. This pattern, as observed previously, reflects its physiological nature: increased perceived difficulty and stress stem from physical constraints rather than cognitive complexity, resulting in a stress-dominant state with limited cognitive demand.

In contrast, the reading comprehension task involved relatively low stress and moderate cognitive load and attention. Unlike the other tasks, it was not characterized by stress or attention emphasis and was categorized as a low-arousal cognitive task in the experimental protocol.

All-in-all, identifying attention-dominant and stress-dominant task profiles provides contextual grounding for physiological and machine learning analyses. Using diverse tasks supports a multi-modal approach to studying mental states, rather than relying on a single task or signal type.

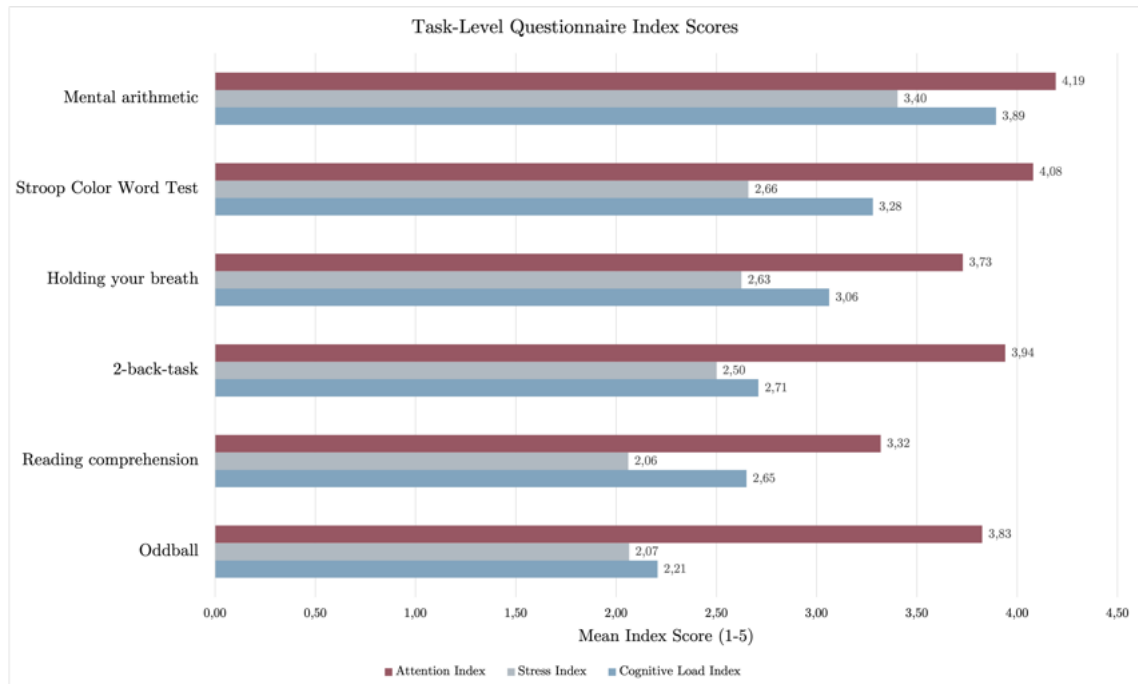


Figure 46: Mean task-level questionnaire index scores for attention, stress, and cognitive load across experimental tasks.

14.4 Post-experiment Survey Results

Upon completion of all experimental tasks, participants completed a post-experiment questionnaire designed to assess overall stress, cognitive workload, sensor comfort, and perceived distraction during the study. The objectives of this survey were threefold:

- Capture participants' cumulative subjective experiences after tasks
- Understand relationships between post-task cognitive load and stress
- Evaluate whether sensor discomfort or distraction may have influenced task experience or confounded physiological measurements.

Unlike the task-level surveys, which focused on immediate subjective responses to individual tasks, the post-experiment survey provides an evaluation of the experiment session as a whole.

14.4.1 Overall Post-task Stress and Cognitive Load

Mean post-task stress scores were relatively low. Mean cognitive load scores were moderately higher. This shows that participants experienced mental fatigue without correspondingly high stress.

The data distribution shows no extreme outliers. This suggests the experiment was demanding yet well-tolerated. The clear gap between stress and cognitive load indicates that sustained mental engagement did not result in increased emotional strain or distress.

Overall, the results show that the experiment successfully extracted cognitive effort while maintaining stress within acceptable limits, supporting its appropriateness for investigating attention and stress without overburdening participants.

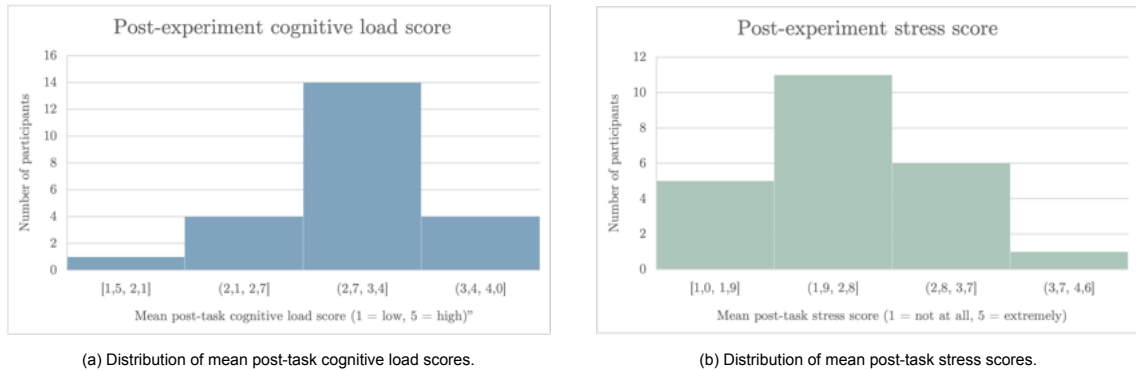


Figure 47: Distributions of post-task questionnaire scores across participants. Cognitive load (left) and stress (right) are shown as histograms of mean post-task ratings.

14.4.2 Relationship Between Post-task Cognitive Load and Stress

To better understand what participants went through during the experiment, post-task mental effort scores were compared with post-task stress scores. The results displayed a small to moderate association, indicating that people who exerted more mental effort usually felt a bit more stressed. However, this link is not very strong, as some people exerted a lot of mental effort without feeling more stressed.

This pattern supports the idea that thinking and feeling are distinct experiences. Feeling mentally tired and working hard did not always cause more stress.

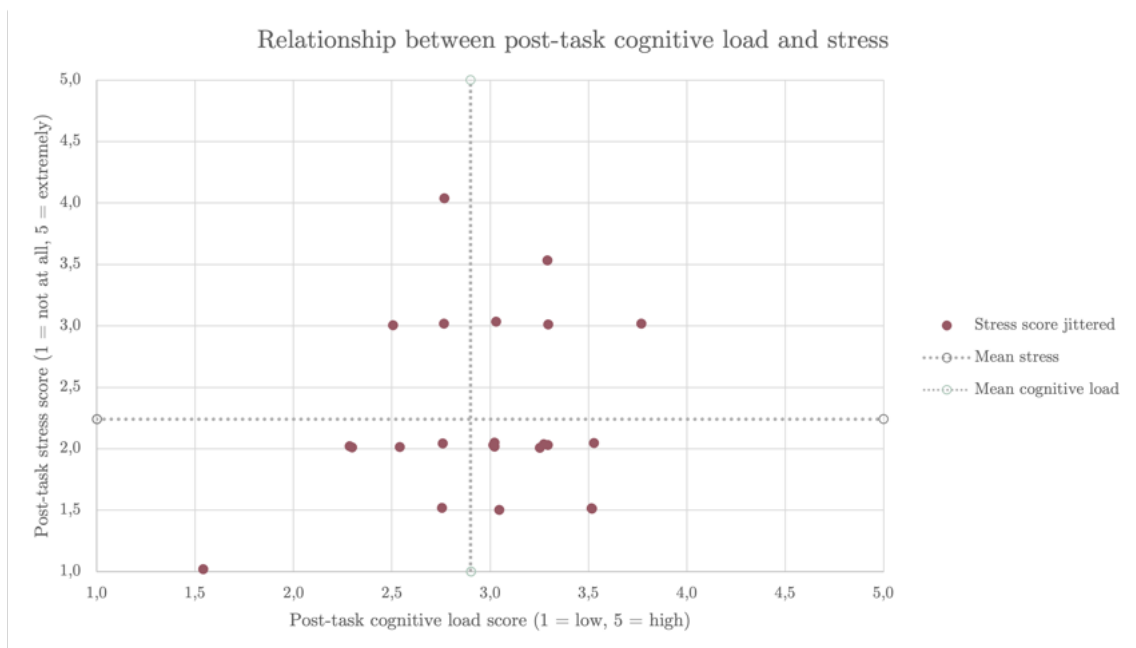


Figure 48: Relationship between post-task cognitive load and stress scores. Each point represents a task-level observation (jittered for visibility). Dotted lines indicate mean cognitive load and mean stress scores.

14.4.3 Sensor Comfort and Distraction Effects

Participants were asked to evaluate sensor discomfort and device-related distraction during the experiment. To determine whether these factors affected attention, Spearman's rank correlation was calculated between distraction and difficulty focusing. A moderate

positive relationship was observed ($\rho = 0.62$), indicating that participants who found sensors distracting were also more likely to find it harder to stay focused. However, the results suggest that although some individuals experience device distraction, the device's overall impact on attention is limited.

The sensors showed a weak relationship with focus difficulty ($\rho = 0.17$) and no clear relationship with post-task stress. Some participants reported mild to moderate discomfort, but this did not correlate with higher stress or attention issues across devices, further indicating a limited impact of the device on stress and focus.

Overall, while sensory distraction may have affected attention for some, it did not influence stress responses. This distinction supports the idea that the observed stress effects are mainly due to task demands rather than sensor artifacts or device effects.

14.4.4 Retrospective Task Difficulty and Stress Ranking

In this section, participants were asked to select the tasks they found most challenging and most stressful, respectively. As shown in Figure 49, MAT was the most challenging and stressful task. The SCWT was next most frequently selected as stress-inducing, and RC task was second most challenging, while the oddball task was consistently regarded as the easiest and least stressful. Furthermore, these judgments align closely with the task-level subjective ratings. Tasks rated as cognitively demanding and stressful in real time were also recalled as such afterward. This alignment between immediate ratings and post-experiment reflection supports confidence in the subjective survey data and increases the reliability of the task-level profiling used in the study.

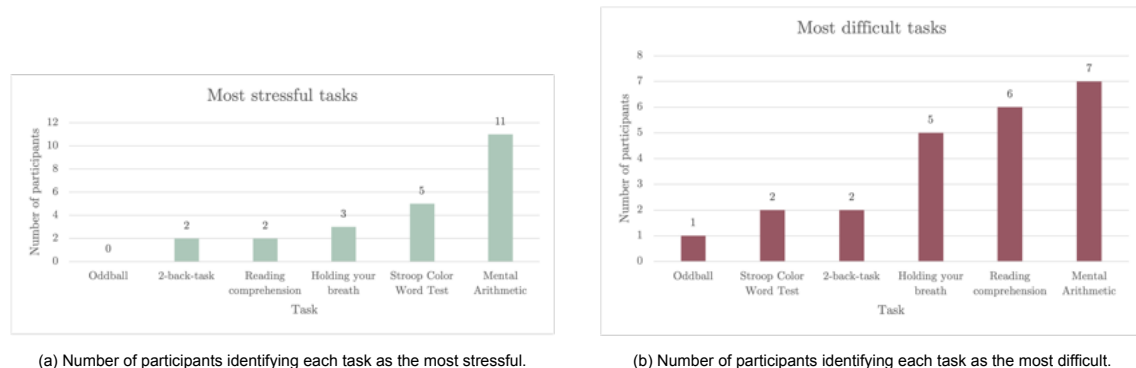


Figure 49: Participant-reported task rankings based on post-experiment questionnaires. Tasks most frequently rated as the most stressful (left) and most difficult (right) are shown.

14.5 Summary and Implications

To summarize, the post-experiment survey showed that the experimental protocol created sustained cognitive engagement. It did not induce excessive stress. Cognitive load and stress were only partially correlated. This finding supports treating them as separate constructs in statistical and machine learning analyses.

Sensor-related discomfort did not affect stress responses. While sensor distraction was associated with reduced focus, its influence was limited. Retrospective task evaluations validated the task-level findings. They confirmed the intended differentiation between attention-dominant and stress-dominant tasks.

Together, these results support the internal validity of the experimental design. They also provide important context for interpreting physiological measurements and for model classification in the next chapters.

15 Machine Learning Analyses

This chapter presents an exploratory machine learning (ML) analysis to evaluate to what extent task-level cognitive states can be distinguished using multimodal biosignal features under subject-independent validation. The ML aspect serves as a supplementary study rather than the main contribution.

Supervised classification methods are applied to task-level features extracted from EEG and peripheral physiological signals. The objective is to identify general patterns, evaluate different feature groups, and examine whether subject-independent classification is better than focusing on predictive accuracy.

Several model configurations were tested, including multimodal, single-modality, task-based, and ablation analyses. Each used Leave-One-Subject-Out (LOSO) validation to address participant differences.

The next sections describe the problem, labeling, model choice, training, assessment, and validation.

15.1 Problem Formulation

The machine learning (ML) task is formulated as a binary classification problem with the goal of distinguishing between task-level cognitive states. Each sample is a participant performing a task and is described by features from biosignals recorded during the experiment.

The model's output is a binary label. It indicates the cognitive state associated with the task, depending on the labeling strategy used. In this chapter, both attention-related classification based on EEG-derived labels and stress-related classification based on task-level stress proxies are explored. Formally:

Let $x_i \in \mathbb{R}^d$ denote the task-level feature vector for participant i , and let $y_i \in \{0, 1\}$ denote the corresponding class label. The objective is to learn a function

$$f(x_i) = y_i,$$

which maps the extracted biosignal features to the associated class label.

As a result of substantial variability in physiological responses, the problem is treated as subject-independent. Models are trained on data from multiple participants. They are evaluated on data from participants not seen during training in order to assess generalization across individuals.

was formed as a subject-independent classification, and models were trained on data from multiple participants and evaluated on data from participants not seen during training. This creates a realistic use case as it can be applied to new participants without having to calibrate them each time.

Additionally, the analysis includes a task-restricted and feature-balanced setting. These settings are used to explore how classification performance is affected by task type and by inclusion or exclusion of specific signals.

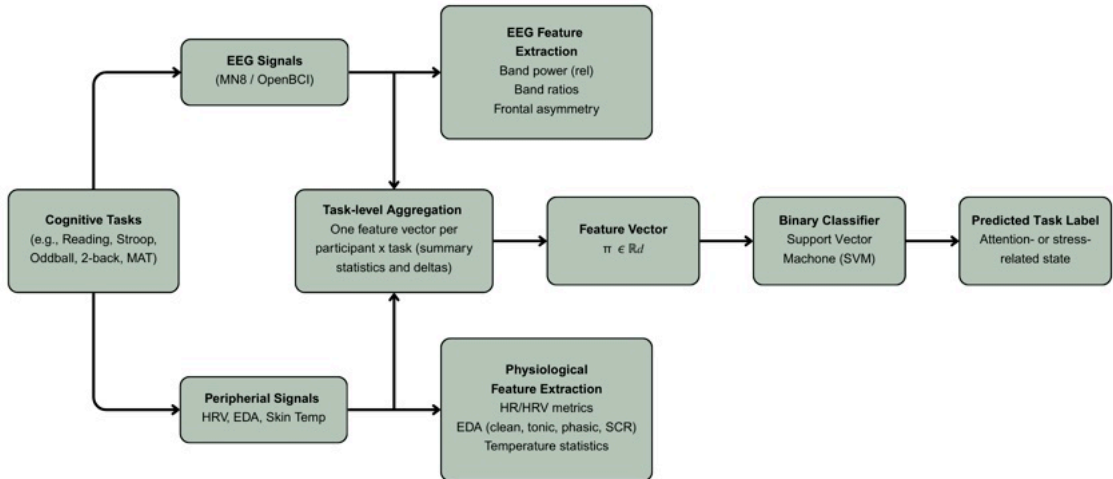


Figure 50: Detailed overview of the machine learning problem formulation. Multimodal biosignals recorded during cognitive tasks are transformed into modality-specific features and aggregated at the task level to form a single feature vector per participant and task. The resulting feature vectors are used as input to a binary classifier that predicts the task-level cognitive state under the chosen labeling formulation.

15.2 Ground Truth and Labeling

The ML analysis uses task-level labels to represent cognitive states. Since this analysis is exploratory, multiple labeling approaches are used, in which all labels are defined independently of the physiological features used by the model.

15.2.1 Attention Labeling

For attention classification, labels are derived from EEG alpha activity. Alpha power is known to be inversely related to attention (Acı et al., 2019). To account for differences between participants, a subject-relative approach is used.

For each participant, the median alpha value across all tasks is calculated. Task instances with alpha values below this median are labeled as closer attention, while instances above the median are labeled as lower attention. This produces a binary attention label for each participant-task instance pair.

These labels reflect each participant’s attention levels and are not intended to represent absolute or perceived attention.

15.2.2 Stress Labeling

For stress classification, labels are derived based on task type rather than physiological signals. Tasks designed to be cognitively demanding are treated as higher-stress, while those with lower demand are treated as lower-stress.

In the stress analysis, MAT and STROOP are labeled as higher-stress tasks, while READING and ODDBALL are labeled as lower-stress tasks. Participants received a label based on the task they performed.

These labels represent differences between task conditions and do not capture individual stress responses.

15.2.3 Labeling Overview

Table 8 summarizes the two labeling strategies used in this chapter.

The labels used in this analysis are proxies designed to support exploratory, subject-independent classification rather than to define absolute cognitive or emotional states.

Label type	How the label is defined	What it represents
Attention	Binary label based on whether task-level EEG alpha is below or above a participant-specific median	Relative attention within each participant
Stress	Binary label based on task category (e.g., MAT, STROOP vs. READING, ODDBALL)	Differences in stress demand between task types

Table 8: Overview of labeling strategies used in the machine learning analysis.

15.3 Feature Extraction and Representation

Machine learning models do not operate directly on raw physiological signals. Instead, the recorded biosignals are transformed into a set of numerical features that summarize relevant characteristics of each signal during a task.

ML models cannot be operated on raw biosignals; the recorded signals are transformed into a set of characteristics of each signal during a task.

For each participant and each task, signal segments corresponding to the task duration are extracted and summarized using simple statistical measures, which results in one feature vector per participant-task instant.

15.3.1 EEG Features

Relative band power is computed across selected frequency bands, with focus on theta and beta activity (AlShorman et al., 2022; Qin et al., 2024). In addition, ratios between frequency bands, such as theta-to-alpha and beta-to-alpha ratios, are calculated. These ratio-based features are used because they are less sensitive to absolute power differences between participants (Dzianok et al., 2021; Iovino et al., 2024).

In some analyses, frontal asymmetry features are also included (Halim et al., 2020). All EEG features are aggregated at the task level using summary statistics.

15.3.2 Peripheral Physiological Features

Peripheral physiological features are extracted from HRV, EDA, and skin temperature signals.

HRV features include standard time-domain metrics that reflect variability in heart rate over the task duration. EDA features summarize tonic and phasic components of skin conductance, representing overall arousal during the task. Skin temperature features consist of simple summary statistics computed over the task window.

As with EEG, all peripheral features are computed independently for each task and participant and then aggregated into task-level summaries.

15.3.3 Feature Representation

The final representation for each sample is a feature vector.

$$x \in \mathbb{R}^d$$

where d depends on the selected feature set. Different feature configurations are chosen by selectively including or excluding specific modalities. These configurations are used in later analyses to examine the contribution of each signal modality to classification performance.

The feature extraction process is identical across all experiments to ensure consistency and comparability between models.

15.4 Model Selection

Support vector machine (SVM) is used as the classification model in all machine learning experiments. The choice of model is guided by the exploratory nature of the analysis, the size of the available dataset, and its popularity in the related literature on machine learning models for classifying biosignals.

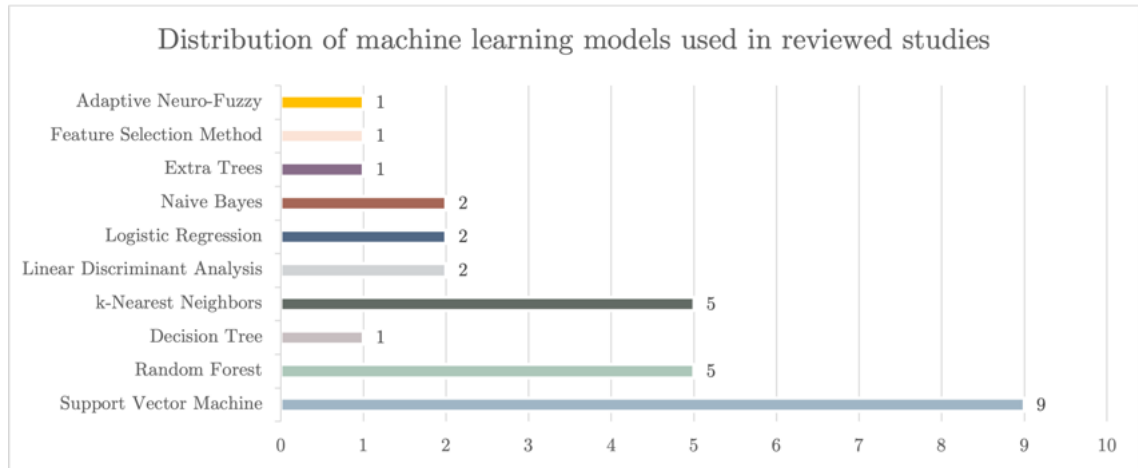


Figure 51: An overview of the machine learning models most frequently used in the reviewed studies.

SVM is ideal for small to medium-sized datasets and can work with feature domains with multiple signals. They provide steady performance without requiring large amounts of training data, which makes them appropriate for subject-independent classification of physiological data.

An radial basis function (RBF) kernel is used to model non-linear relationships (Kaushik et al., 2022). The model configuration is kept fixed across all experiments to allow fair comparison between feature sets and tasks.

More complex models, such as deep learning approaches, are not considered. Given the limited sample size, simpler, more controlled models are preferred.

15.5 Training Procedure

All machine learning experiments follow the same general pipeline: task-level feature vectors are used as input, preprocessing is applied only within the training folds, and

models are evaluated using leave-one-subject-out (LOSO) validation. Differences between experiments arise from the labeling strategy (attention vs. stress), the task subset used, and the selected feature set (single-modality vs. multimodal ablation).

15.5.1 Dataset Construction

Task-level features were compiled into a master dataset. This dataset contains 23 participants and 132 participant-task samples, with 83 columns. The dataset contains all tasks conducted throughout the experiment. In the dataset, each row corresponds to one participant performing one task and is represented by a feature vector derived from EEG and peripheral physiological signals.

15.5.2 Training and Validation

Models are trained and evaluated using a leave-one-subject-out (LOSO) scheme. In each iteration, data from one participant are held out for testing and the model is trained on data from all remaining participants. This process is repeated until each participant has served as the test subject once.

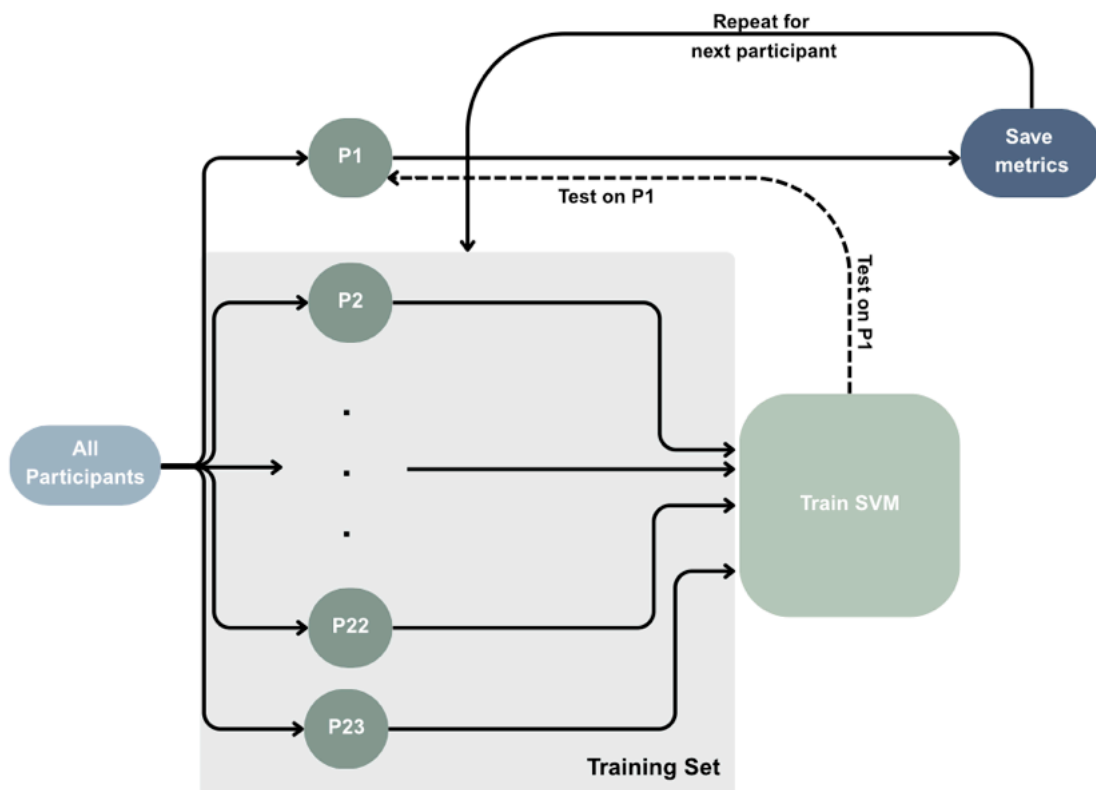


Figure 52: An overview cross-validation procedure for training and evaluating the SVM classifier

15.5.3 Attention Classification

For attention classification, subject-relative EEG alpha-based labels are used in two settings: first, a full attention analysis including all available labeled samples. Second, multimodal attention experiments use a reduced subset of participants for whom complete multimodal data are available. In addition, task-restricted attention experiments are performed to examine how modality contributions change under specific task conditions.

15.5.4 Stress Classification

Stress classification uses task-based labels, which were defined by expected task demands. Stress analyses are conducted in task-restricted settings to determine which tasks are higher- or lower-demanding. The same LOSO training and preprocessing pipeline is used as in the attention analyses.

15.5.5 Feature Ablation

To evaluate the contribution of individual biosignals, feature ablation experiments are performed. Models are trained using different feature sets, ranging from EEG-only to combinations including HRV, EDA, and skin temperature. The training procedure is kept the same across all configurations so that performance differences can be attributed solely to the included feature modalities rather than to task composition or training changes.

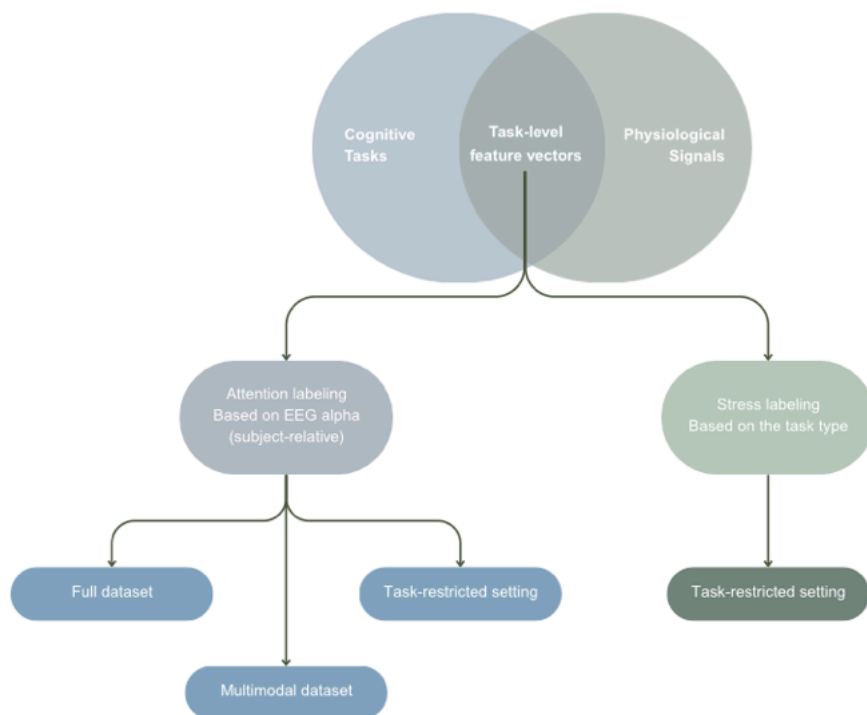


Figure 53: Overview of the machine learning analysis. Physiological signals recorded during cognitive tasks are converted into task-level features. Two labeling approaches are explored: attention labeling based on subject-relative EEG alpha activity and stress labeling based on task-level proxies. All experiments share the same preprocessing, training, and validation pipeline, allowing direct comparison of feature modalities and task settings.

15.6 Evaluation Metrics

Model performance is evaluated applying metrics suitable for binary classification and unequal class sizes (Kaushik et al., 2022). The same metrics are used for both attention and stress classification.

The main metric is **balanced accuracy**. Balanced accuracy is the average of the recall for the two classes. It is used because class distributions are not always balanced, especially in task-restricted analyses (Tiwari et al., 2019).

The **F1-score** is also reported for the positive class. The F1-score combines precision and recall into a single value. For attention classification, the positive class represents higher

attention. For stress classification, the positive class represents higher stress (Tiwari et al., 2019).

Standard accuracy is reported for reference but is not used as the primary metric. Confusion matrices are included to help interpret model behavior.

All metrics are computed for each leave-one-subject-out fold and then summarized using the mean and standard deviation (Hernández-Sabaté et al., 2024).

15.7 Machine Learning Results

This section provides a detailed overview of findings from the exploratory machine learning analysis. Results are reported separately for attention classification and stress classification. All results are obtained using LOSO validation and are summarized using balanced accuracy and F1-score, as defined in 15.6.

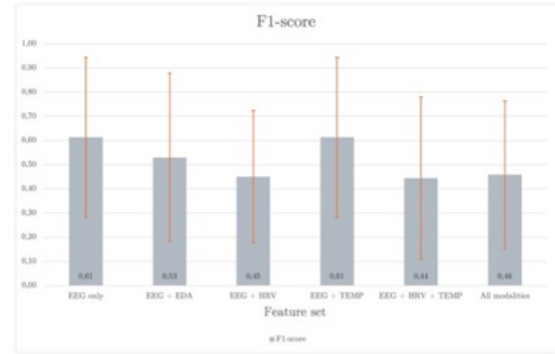
15.7.1 Attention Classification

Attention classification was evaluated in three settings to understand which biosignals are most informative and how this depends on task type.

Full Attention Dataset Using subject-relative EEG alpha-based labels across all tasks, EEG-only models achieved the best performance, with a balanced accuracy of approximately and an F1-score of approximately 0.61 for the attention class. Feature ablation showed that EEG features based on relative band power and frequency ratios performed better than absolute power and frontal asymmetry features.

Multimodal Attention Ablation When restricting the analysis to participants with complete multimodal data (12 participants, 48 samples), EEG-only models again performed best. Adding skin temperature did not affect performance, whereas adding EDA slightly reduced it. The inclusion of HRV features clearly decreased both balanced accuracy and F1-score. This shows that adding peripheral signals does not improve attention classification at the global task level.

Task-Restricted Attention Analysis When the analysis was limited to the oddball and 2-back tasks, EEG-only models performed close to chance level (balanced accuracy ≈ 0.50). The best results were obtained when combining EEG with EDA features (balanced accuracy ≈ 0.56). Adding HRV or skin temperature lowered performance. Figure 54 summarizes these results. Overall, this indicates that for short, stimulus-driven attention tasks, EDA provides useful complementary information to EEG, while slower physiological signals are less relevant.



(b) F1-score for the attention class.

Figure 54: Performance of attention classification across feature sets using leave-one-subject-out validation. Balanced accuracy reflects overall class balance, while the F1-score highlights performance on the attention class.

Overall, this suggests that for short, stimulus-driven attention tasks, autonomic arousal captured by EDA provides complementary information to EEG, whereas slower physiological signals do not align well with rapid attentional dynamics.

15.7.2 Stress Classification

Stress classification was evaluated to examine which physiological signals are most informative for detecting task-related stress under subject-independent validation.

Full stress dataset Using task-based stress labels in all tasks, models based on autonomic physiological signals performed best. The combination of HRV and EDA features achieved the highest balanced accuracy (≈ 0.60), which shows the most stable discrimination between stress and non-stress states. EEG-only models showed moderate balanced accuracy (≈ 0.53) but achieved a relatively higher F1-score for stress, showing increased sensitivity at the cost of more false positives. Skin temperature alone showed limited performance, while HRV alone performed poorly.

Multimodal stress ablation Feature-ablation experiments showed that combining HRV and EDA outperformed single-modality models in balanced accuracy. Including all modalities resulted in performance comparable to HRV+EDA but did not provide more improvement. Autonomic signals generalize better in terms of balanced accuracy for stress prediction and EEG features primarily increase sensitivity at the cost of specificity.

Task-restricted stress analysis For stress-relevant tasks, the relative performance order stayed the same. Models based on HRV and EDA provided the most balanced performance, and EEG features stayed moderately sensitive but less specific. Overall, these results show that stress is more strongly reflected in autonomic nervous system activity than in EEG features under subject-independent evaluation.

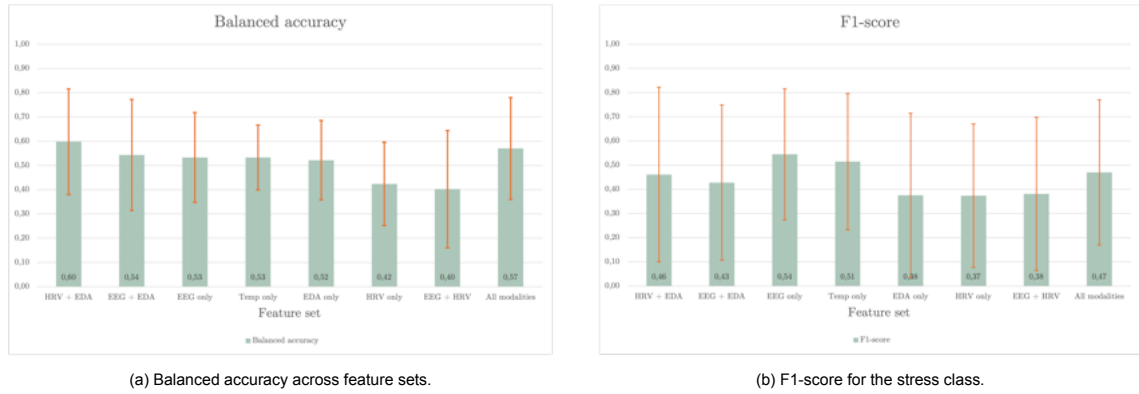


Figure 55: Balanced accuracy of stress classification across feature sets using LOSO validation. Error bars indicate the standard deviation across folds. Models based on autonomic signals (HRV + EDA) achieve the highest balanced accuracy, while EEG-based models show lower overall balance.

Figure 55 shows the balanced accuracy across feature sets for stress classification.

Results show that some physiological modalities align with different cognitive states in both attention and stress classification. EEG-based features show the highest generalizability for attention at the global task level, whereas combining EEG and EDA generates excellent performance for short, stimulus-driven attention tasks. Contrarily, stress classification relies predominantly on autonomic signals, with HRV and EDA providing the most consistent performance across participants. Simple multimodal fusion does not consistently improve classification effectiveness and may decrease generalizability under subject-independent validation. These findings underscore the significance of aligning physiological modality with cognitive state rather than merely increasing the number of sensors.

All training files and output results from the ML model can be found in Appendix A.12.

Part IV: Results

16 Discussion

This study investigated multimodal physiological responses to cognitive and physiological challenges by combining electroencephalography (EEG) with peripheral signals (HRV, EDA, and ST) across tasks varying in attentional demand, cognitive load, and stress. Specifically, participants completed a reading comprehension task, oddball, 2-Back, Stroop Color and Word Test (SCWT), mental arithmetic task (MAT), and a physiologically stressful breath-holding task. Physiological responses were examined using both within-trial temporal dynamics and between-trial task-level comparisons, complemented by nonparametric statistical tests and subjective self-report measures. The discussion of results throughout this chapter is based upon the generated learnings from Part III and highlights the most significant findings across the investigated multimodal responses.

Across modalities, EEG showed the clearest and most consistent differentiation between cognitive tasks. In particular, alpha-band suppression emerged as one of the most robust markers of increased cognitive demand and attentional engagement, showing consistent patterns across both EEG devices (Emotiv MN8 and OpenBCI Headband) and tasks. Heart rate (HR) also showed systematic changes with task demands, particularly for tasks involving high cognitive effort. Other HRV features, however, appeared more strongly influenced by physiological state than by task-specific cognitive demands.

In contrast, Electrodermal activity (EDA) and skin temperature reflected general arousal and physiological activation rather than differences between cognitive tasks. While both modalities reliably distinguished task engagement from baseline, they did not support a separation of tasks by attentional demand. Hence, these signals appeared sensitive to sustained engagement, autonomic activation, and stress.

The breath-holding task consistently differed from the cognitive tasks across modalities, particularly in peripheral signals (HRV, EDA, and ST). This separation supports its interpretation as a physiological stressor rather than a cognitively demanding task, and it serves as a useful reference for contrasting stress-related autonomic responses with neural measures of cognitive load.

Overall, the findings point to a functional dissociation between neural and peripheral measures. EEG is most informative for characterizing cognitive demand and attention, whereas peripheral signals capture arousal and stress-related processes. Subjective ratings generally mirrored the physiological findings, indicating that tasks perceived as cognitively demanding were not necessarily experienced as stressful.

Together, the multimodal physiological measures and continuous subjective reports enabled a separation between cognitive engagement and physiological stress. This, in turn, facilitated the interpretation and categorization of the experimental tasks.

16.1 Modality-Level Interpretation

EEG

Relative alpha suppression emerged as the most consistent neural marker of increased attentional demand across tasks and devices. Both Emotiv MN8 (temporal electrodes) and OpenBCI Headband (frontal electrodes) showed clear reductions in relative alpha power during cognitively demanding tasks such as 2-Back, Stroop, and MAT. For OpenBCI EEG in particular, several of these effects remained significant after BH-FDR correction.

Theta-band activity differed across devices: Emotiv MN8 recordings over temporal regions mainly showed relative theta decreases during cognitive tasks, whereas OpenBCI recordings over frontal regions showed small theta increases. This divergence is consistent with prior literature linking frontal theta to working memory and cognitive control, while temporal theta may reflect different neural processes or background activity.

Similarly, beta-band patterns were also inconsistent across devices. While Emotiv MN8 EEG showed beta increase for some tasks and a significant increase in beta variability during Stroop, OpenBCI EEG showed reductions in beta median power during cognitively demanding tasks. These differences likely relate to electrode placement and the use of relative band power.

Despite these differences, EEG reliably differentiated cognitive tasks from baseline and supported a general ordering of tasks by cognitive demand, particularly through alpha suppression.

HRV

Within HRV measures, heart rate showed the clearest and most consistent task-related effects. Heart rate increased relative to baseline across all tasks, with MAT showing the largest and only statistically significant increase after BH-FDR correction. Friedman tests also suggested task-related differences in heart rate, although post-hoc comparisons did not survive correction, likely due to the reduced complete-case sample size.

Other HRV features showed more variable patterns across tasks. RMSSD both increased and decreased across tasks and did not consistently track cognitive demand. SDNN increased for several tasks, but only breath-holding showed a statistically significant increase, suggesting that SDNN was more sensitive to physiological state changes than to cognitive load. Frequency-domain measures showed inconsistent task-level patterns and did not support task differentiation.

Overall, cardiac indicators, particularly heart rate, reflected general arousal and physiological activation and clearly distinguished the breath-holding task from the cognitive tasks. However, other HRV measures were less effective in differentiating levels of cognitive demand between tasks.

EDA

EDA primarily reflected general task engagement and sympathetic activation rather than differences between cognitive tasks. Tonic EDA significantly increased relative to baseline for all tasks, including breath-holding, with large effect sizes. This indicates that tonic EDA reliably distinguishes task engagement from resting state but does not support ranking tasks by cognitive demand.

Phasic EDA measures showed more task-specific effects. SCR rate increased significantly for reading, oddball, and 2-Back, consistent with stimulus-driven and dynamic tasks. In contrast, MAT and breath-holding were associated with reduced SCR counts, suggesting suppression of phasic responses during sustained cognitive strain and physiological stress. Despite these differences at the task-versus-baseline level, Friedman tests did not reveal significant differences between tasks.

Taken together, EDA appears effective for detecting engagement and autonomic activation but less suitable for distinguishing between different levels of cognitive load or stress across tasks.

Skin Temperature

Skin temperature reflected slow and cumulative physiological changes rather than rapid task-specific effects. Mean skin temperature increased significantly relative to baseline for all tasks, indicating general physiological activation.

Task-level plots provided additional contextual information about temperature dynamics over time. For most tasks, temperature slopes were close to zero or slightly positive, whereas Stroop and MAT showed negative slopes, reflecting a gradual cooling trend. End–start temperature changes were positive for all tasks, with larger accumulated changes for 2-Back and mental arithmetic. Despite these patterns, Friedman tests did not identify significant differences between tasks for any skin temperature feature.

Overall, skin temperature was sensitive to prolonged physiological states and sustained stress, but its slow dynamics and inter-individual variability limited its usefulness for task differentiation. As such, skin temperature complemented faster peripheral measures by capturing arousal rather than task-specific effects.

16.2 Task-Centered Interpretation

This section interprets each task in terms of the overall cognitive and physiological state it triggered, based on converging patterns across EEG, peripheral signals, and subjective reports. The goal is to characterize tasks by attentional demand, cognitive load, and physiological stress rather than by individual features.

Low-Demand and Sustained-Attention Tasks

Oddball and reading represented states of stable cognitive engagement without pronounced physiological stress.

Oddball induced high attentional vigilance, reflected in increased tonic EDA and a significant increase in SCR rate, indicating frequent orienting responses. At the same time, autonomic measures suggested a relaxed physiological state, with increased RMSSD and SDNN and a significant positive skin temperature slope, consistent with parasympathetic dominance and the absence of stress-related vasoconstriction. Subjective reports strongly aligned with this profile: participants rated oddball as the least mentally demanding and least stressful task, despite reporting high levels of attention. Together, this pattern characterizes oddball as a vigilance-driven attentional state with minimal cognitive load and stress.

Reading showed sustained engagement with low pressure. EEG measures indicated reliable alpha suppression and reduced alpha variability, reflecting continuous attentional

allocation. Peripheral signals showed moderate sympathetic activation, most clearly expressed through a significant increase in tonic EDA and significant increases in SCR count and rate, while heart rate remained relatively stable. Subjective reports closely matched this profile, with participants describing reading as requiring sustained effort and attention but little stress. This convergence supports interpreting reading as a stable, low-demand attentional task.

High Cognitive Load with Limited Stress

The 2-Back task represented a state of high cognitive demand without pronounced physiological stress. EEG patterns reflected strong cognitive load, while peripheral measures indicated controlled arousal rather than stress. EDA measures showed significant increases in tonic activity, SCR count, and SCR rate, reflecting strong phasic and tonic arousal. Skin temperature dynamics further supported this interpretation: the temperature slope was significantly positive, and 2-Back was the only task showing a significant positive end–start temperature change, indicating sustained physiological activation without cooling.

Subjective ratings closely aligned with these physiological patterns. Participants reported high focus and mental demand but only moderate stress levels. Together, these findings suggest that 2-Back induced a performance-oriented state characterized by intensive cognitive effort under relatively stable autonomic regulation.

Combined Cognitive Load and Physiological Stress

MAT and Stroop both induced states of high cognitive demand; however, their stress profiles differed considerably.

MAT showed the clearest and most consistent markers of physiological stress. It was the only task associated with a statistically significant increase in heart rate and exhibited the strongest negative ST slope, indicating sustained sympathetic activation and vasoconstriction. EDA measures further reflected elevated tonic arousal alongside reduced phasic responding, consistent with sustained strain rather than stimulus-driven engagement. Subjective reports strongly aligned with these findings, with participants rating MAT as the most mentally demanding and most stressful task. This convergence supports interpreting MAT as a task combining high cognitive load with sustained physiological stress.

Stroop, in contrast, revealed a notable divergence between subjective experience and physiological responses. EEG measures showed strong cortical engagement, reflected by pronounced alpha suppression, alongside heightened interference processing, indicated by elevated beta-band dynamics. Peripheral signals suggested acute physiological stress: skin temperature showed a clear cooling trend following an initial increase, and tonic EDA was elevated. Despite this, participants generally rated Stroop as demanding but not highly stressful. This mismatch suggests that Stroop induces acute stress driven by cognitive conflict and interference, even when this stress is not fully captured by subjective reports. Stroop therefore highlights the value of physiological measures in revealing stress responses that may remain outside conscious awareness.

Physiological Stress without Cognitive Demand

Breath-holding served as a non-cognitive stress reference and demonstrated a clear dissociation between physiological stress and cognitive engagement. Peripheral signals

showed clear autonomic disruption, including decreased HR and RMSSD, a significant increase in SDNN, elevated tonic EDA, a significant reduction in SCR count, and consistent changes in ST. Together, these patterns indicate a strong shift in the SNS and physiological strain.

In contrast, EEG measures did not reflect increased cognitive engagement, consistent with the absence of task-related cognitive processing. Subjective reports confirmed this dissociation, with participants reporting high effort and discomfort but low cognitive demand. Breath-holding thus validated the interpretation of peripheral signals as markers of physiological stress rather than cognitive load.

Overall Interpretation and Task Ordering

Across modalities, the results support a clear dissociation between cognitive demand and physiological stress. EEG measures, particularly relative alpha suppression, provided the strongest and most consistent basis for ordering tasks by cognitive demand, with oddball and reading at the low end, followed by 2-Back, Stroop, and MAT. This ordering was evident in task-level EEG patterns and supported by statistically significant effects in the OpenBCI data.

Peripheral signals refined this picture by differentiating stress-related activation from cognitive effort. MAT showed the clearest markers of physiological stress, including the only significant heart-rate increase and the strongest negative skin temperature slope. Stroop also exhibited stress-related physiological responses, despite lower perceived stress, suggesting stress driven by cognitive interference rather than subjective experience. In contrast, 2-Back provoked strong engagement and arousal without comparable stress signatures.

Oddball and reading consistently reflected sustained attention with low stress, while breath-holding formed a distinct category characterized by pronounced autonomic stress without cognitive demand. Based on the convergence of physiological signals and subjective reports, tasks could be ordered differently along the two dimensions: cognitive demand (oddball < reading < breath-holding < 2-Back < Stroop < MAT) and physiological stress (oddball < reading < 2-Back < Stroop < breath-holding < MAT). This dissociation highlights the value of a multimodal approach for separating cognitive engagement from physiological stress.

16.3 Limitations

This study has several limitations, including sample size and statistical power, inter-individual variability, and constraints related to wearable sensors and device differences.

First, the **sample size** was relatively small (N=23), and due to hardware issues and quality-control exclusions, the complete-case sample size was further reduced (N=12). As a result, some task-specific effects visible within-trial temporal analyses and task-level plots did not reach statistical significance. This particularly affected the Friedman tests and post-hoc Wilcoxon comparisons used to assess differences between tasks. For this reason, sample sizes were reported for all plots, and effect sizes were included alongside statistical significance. When comparing results across modalities, these varying sample sizes should be taken into account.

Physiological responses also varied substantially across participants, reflecting individual differences in stress sensitivity, baseline physiology, and task strategies. This **inter-individual variability** further reduced statistical power and limited the consistency of task-level effects across modalities.

Device-related differences also represent a limitation. EEG results differed between the Emotiv MN8 and OpenBCI systems, likely due to differences in sensor placement and sensitivity to distinct neural processes. While these differences limit direct quantitative comparisons across devices, they also provide complementary insights into task-related neural dynamics across temporal and frontal regions.

Wearable constraints played a significant role in data quality. While EDA and ST signals were generally stable, EEG and PPG were highly **sensitive to motion**. For the Emotiv MN8, electrode contact quality was typically stable, but EEG signals were strongly affected by movement. OpenBCI EEG was similarly motion-sensitive, resulting in noisy recordings and frequent artifacts. PPG recordings also triggered multiple low-quality warnings in HeartPy, which needed strict quality control procedures. Although devices such as EmotiBit and Emotiv MN8 are comfortable and suitable for wearable use, these findings indicate that substantial signal processing and quality control are required for reliable inference of cognitive states, particularly in real-time applications.

Device data latency and offsets. Although all data streams were recorded via LabRecorder and shared a common LSL time base, small temporal offsets between devices cannot be fully ruled out. This is particularly relevant for the Emotiv MN8 EEG data, which followed a longer data path from the device through the Cortex API to LSL. Exploratory tests were conducted to assess potential delays, but no consistent or fixed latency was identified. Nevertheless, the absence of formal latency validation means that minor misalignments may still be present and should be considered when interpreting cross-modal temporal dynamics.

Task design also imposed limitations: some tasks were relatively short, reducing the number of analysis windows available, particularly for signals that require longer periods to stabilize, such as HRV and ST. In addition, although the task order was randomized within blocks, later tasks may have been influenced by cumulative cognitive load or fatigue.

Stress was inferred indirectly through autonomic and peripheral physiological measures. While these measures are widely used, they do not uniquely distinguish stress from other forms of arousal or effort. The **absence of hormonal measures**, such as cortisol, limits the specificity of stress-related conclusions.

Finally, the machine learning analysis was exploratory and aimed primarily at assessing feasibility and comparing feature sets. The small sample size **limited model complexity and feature selection**, and the resulting ML performance should be interpreted as complementary to the statistical and survey analyses rather than as standalone evidence.

Despite these limitations, the convergence of findings across EEG, HRV, EDA, skin temperature, and subjective reports supports the robustness of the main conclusions.

16.4 Implications and Suggestion for Future Work

This study highlights the value of a multimodal physiological approach for distinguishing cognitive load, attentional engagement, and physiological stress. The results show that EEG is particularly well-suited for capturing differences in cognitive demand, whereas peripheral signals such as HRV and EDA are more informative for identifying stress-related autonomic responses. Including a non-cognitive physiological stressor (breath-holding) was particularly useful in validating this dissociation. The machine learning results reflected the same pattern: EEG-based models performed best for attention-related classification, while models relying on HRV and EDA features were more informative for stress-related distinctions.

An important implication of these findings is that high attentional or executive demands do not necessarily translate into physiological stress. For example, the oddball task required sustained attention but did not trigger strong stress responses. Reading involved prolonged engagement with relatively stable physiological patterns, while the 2-back task showed increased arousal without clear stress-related signatures. Opposite, mental arithmetic combined cognitive demand with time pressure and consistently produced stress-related autonomic responses. These distinctions underline the importance of separating cognitive engagement from stress rather than treating them as interchangeable states.

From an applied perspective, the results suggest that multimodal physiological monitoring could be useful in real-world settings where distinguishing productive engagement from harmful stress is important. In adaptive learning environments, such approaches could help differentiate focused effort from overload, while in workplace or human–computer interaction contexts, they could support the identification of sustained cognitive effort versus early signs of stress or burnout. A further advantage of multimodal systems is their robustness: in real-world conditions, signal quality is often compromised by motion or environmental factors, and combining multiple modalities increases the likelihood that meaningful information can still be extracted even if one signal degrades.

Several **directions for future work** emerge from this study. **Larger and more diverse participant samples** are needed to improve statistical power and generalizability, particularly with respect to gender balance and age range. **Improved signal cleaning and artifact handling** could allow a greater proportion of segmented windows to pass quality control, increasing data availability for analysis. In addition, although many features were extracted in this study, only a subset was used in the final analyses; future work could **explore a broader range of features**, especially for modalities such as EDA and skin temperature that were sensitive to engagement but less effective at differentiating tasks.

Future studies should also include **dedicated validation of synchronization across devices**. Controlled timing events or hardware-based triggers could be used to quantify end-to-end latency and ensure precise temporal alignment between modalities. In terms of machine learning, future work should **build on larger datasets, explore a wider range of models**, and **systematically evaluate additional feature combinations**.

17 Conclusion

The purpose of this thesis project was to investigate and identify multimodal physiological responses to both cognitive and physiological challenges by combining neural signals with peripheral physiological signals. This was achieved via a specifically designed experimental study in which participants engaged in a set of cognitive tasks.

The tasks in the experiment were conducted in a randomized task order, and consisted of variable cognitive effort, attentional demand, and stress levels. Selected signals were recorded by utilizing a set of specialized hardware equipment: (i) Emotiv MN8 Headset, (ii) Open BCI EEG Headband, and (iii) EmotiBit.

All biosensor data were synchronized across devices for each participant, stored, and subsequently processed. The recorded signals were segmented into analysis windows, from which features were extracted. Both within-trial temporal dynamics and task-level feature summaries were examined across participants. Task-level features were statistically evaluated using non-parametric methods, specifically Wilcoxon signed-rank tests and Friedman tests, to assess differences relative to baseline and differences between tasks.

The results of the statistical analyses, together with observed patterns in the physiological signals, were compared with subjective participant reports collected before, during, and after task execution. In addition, an exploratory machine learning approach was implemented to assess the feasibility of classifying tasks into attention-dominant and stress-dominant categories.

The combined findings from these analyses were presented and discussed in Part IV of this thesis. Building on these results, the research questions and project objectives are addressed below.

Main Research Question: *How can attention and stress be reliably distinguished in task-based environments using physiological signals measured with wearable biosensors?* Based on the work generated in this thesis project, by combining literature research with the signal responses collected during the experiments conducted, and after synthesizing key features, statistical analyses and machine learning classifier, EEG clearly distinguishes itself as the most reliable indicator for cognitive demand, while peripheral signals, primarily heart rate and EDA, are the best indicators to track arousal.

With the following supporting research questions:

- *What is the added value of combining brain-based and peripheral physiological signals when studying attention and stress?*
EEG is the most reliable modality for identifying cognitive demand and attention. Peripheral signals (heart rate, EDA, ST) capture physiological arousal and stress. Their combination improves robustness and classification performance by capturing complementary neural and physiological responses.
- *Which physiological modalities and extracted features are most sensitive to changes in attention and stress across diverse task conditions?*

Relative alpha-band power is the most robust indicator for attention, whereas heart rate and tonic EDA are the most sensitive for stress across the tasks' conditions.

- *How well do subjective self-reported measures of attention and stress align with objectively measured physiological responses?*

Subjective ratings generally mirror physiological findings. Some differences were observed, for example in Stroop: low-stress perceived but signals reflect physiological stress.

- *Can machine-learning models trained on multimodal biosignal features reliably classify attention-dominant and stress-dominant task states?*

Yes, but their reliability depends on the modality used. EEG-based models achieved the highest accuracy for attention classification (balanced accuracy $\approx$). For stress classification, models relying on autonomic signals (HRV and EDA) provided the most consistent performance (balanced accuracy $\approx$ 0.60).

Project Objectives:

1. **Experimental design:** Develop a controlled, task-based experimental protocol capable of eliciting varying levels of attention and stress.
A controlled experimental protocol was developed to systematically induce attention and stress through a set of validated tasks. The design of the tasks and experiment is described in Chapter 3, Chapter 5 and Chapter 6.
2. **Multimodal and subjective data collection:** Collect EEG and peripheral biosignals from participants using wearable devices during task execution.
Multimodal data, including EEG, HRV, EDA, and skin temperature, were collected from 23 participants using synchronized wearable devices. This work was extensively discussed in Chapter 4, Chapter 5, and Chapter 9. Pre-, post- experiment surveys, as well as in-between task-surveys were collected. A description of the surveys and their purposed is described in Chapter 7.
3. **Physiological characterization:** Characterize EEG, HRV, EDA, and temperature responses associated with different levels of attention and stress across task conditions.
Task-dependent changes in EEG, HRV, EDA, and skin temperature were quantified across attention and stress conditions, tested for significance as elaborated under Chapter 11, Chapter 12, Chapter 13, Chapter 16.
4. **Multimodal comparison:** Examine whether combining brain-based and peripheral biosignals provides more robust insights into attention and stress than single-modality analyses.
Combining findings from EEG with peripheral signals yielded clearer task differentiation. The implication and details of this results, as well as the increased robustness of these insights, were crystallized from the work performed in sections Chapter 13, and Section 16.1.
5. **Subjective–objective comparison:** Assess the correspondence between physiological measurements and participants' self-reported attention, stress, and mental demand.

Physiological responses were compared with self-reported attention, stress, and mental demand ratings and were found to be generally consistent. A detailed description is provided via the triangulation and cross-referencing of the Chapter 7, the Chapter 14, and the Section 16.2.

6. **Exploratory machine learning:** Explore the feasibility of distinguishing attention-dominant and stress-dominant task states using machine-learning methods applied to multimodal physiological features.

The study evaluated SVM classifiers using Leave-One-Subject-Out validation, finding that EEG-only models best generalize for global attention classification (Chapter 15).

7. **Reusable dataset creation:** Establish a structured and reusable dataset combining physiological signals and subjective reports to support future research in multimodal cognitive state assessment.

A structured dataset integrating synchronized biosignals, task markers, and survey data was established. The overview and structure of the frameworks applied is further detailed in Part II, particularly in Chapter 3, and Chapter 9.

Overall, this study provides a systematic multimodal characterization of cognitive and physiological responses across tasks with varying demands. By combining neural and peripheral signals with continuous subjective reports, it offers a foundation for future research on monitoring cognitive load and stress in both experimental and applied settings. While also delivering a set of frameworks that are reusable for future research in multimodal cognitive assessment.

Bibliography

- Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A global measure of perceived stress. *Journal of Health and Social Behavior*, 24(4), 385–396. <https://doi.org/10.2307/2136404>
- Berenz, E. C., Vujanovic, A. A., Coffey, S. F., & Zvolensky, M. J. (2012). Anxiety sensitivity and breath-holding duration in relation to PTSD symptom severity among trauma exposed adults. *Journal of Anxiety Disorders*, 26(1), 134–139. <https://doi.org/10.1016/j.janxdis.2011.10.004>
- Gómez, D. M., Jiménez, A., Bobadilla, R., Reyes, C., & Dartnell, P. (2015). The effect of inhibitory control on general mathematics achievement and fraction comparison in middle school children. *ZDM*, 47(5), 801–811. <https://doi.org/10.1007/s11858-015-0685-4>
- Luque-Casado, A., Perales, J. C., Cárdenas, D., & Sanabria, D. (2016). Heart rate variability and cognitive processing: The autonomic response to task demands. *Biological Psychology*, 113, 83–90. <https://doi.org/10.1016/j.biopsycho.2015.11.013>
- Martinez, R., Martinez, R., Irigoyen, E., Arruti, A., Martin, J. I., & Muguerza, J. (2017). A real-time stress classification system based on arousal analysis of the nervous system by an f-state machine. *Computer Methods and Programs in Biomedicine*, 148, 81–90. <https://doi.org/10.1016/j.cmpb.2017.06.010>
- Aci, Ç. İ., Kaya, M., & Mishchenko, Y. (2019). Distinguishing mental attention states of humans via an EEG-based passive BCI using machine learning methods. *Expert Systems with Applications*, 134, 153–166. <https://doi.org/10.1016/j.eswa.2019.05.057>
- Arza, A., Garzón-Rey, J. M., Lázaro, J., Gil, E., Lopez-Anton, R., de la Camara, C., Laguna, P., Bailon, R., & Aguiló, J. (2019). Measuring acute stress response through physiological signals: Towards a quantitative assessment of stress. *Medical & Biological Engineering & Computing*, 57(1), 271–287. <https://doi.org/10.1007/s11517-018-1879-z>
- Hyun-Soo Choi, Siwon Kim, Ho Bae, Jee-Eun Yoon, Inha Hwang, Dana Oh, Chang-Ho Yun, & Sungroh Yoon. (2019). Learning-based instantaneous drowsiness detection using wired and wireless electroencephalography. *IEEE Access*, 7, 146390–146402. <https://doi.org/10.1109/access.2019.2946053>
- Tiwari, A., Cassani, R., Narayanan, S., & Falk, T. H. (2019). A comparative study of stress and anxiety estimation in ecological settings using a smart-shirt and a smart-bracelet. *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2213–2216. <https://doi.org/10.1109/EMBC.2019.8857890>
- Ding, Y., Ding, Y., Cao, Y., Duffy, V. G., Wang, Y., & Zhang, X. (2020). Measurement and identification of mental workload during simulated computer tasks with multimodal methods and machine learning. *Ergonomics*, 63(7), 896–908. <https://doi.org/10.1080/00140139.2020.1759699>

- Halim, Z., Halim, Z., Halim, Z., & Rehan, M. (2020). On identification of driving-induced stress using electroencephalogram signals: A framework based on wearable safety-critical scheme and machine learning. *Information Fusion*, 53, 66–79. <https://doi.org/10.1016/j.inffus.2019.06.006>
- Lestari, D., Muhammad, P. S., & Zaini, I. A. (2020). Implementation artificial neural network on microcontroller for student attention level monitoring device using EEG. *2020 6th International Conference on Science in Information Technology (ICSITech)*, 57–61. <https://doi.org/10.1109/ICSITech49800.2020.9392043>
- Rodríguez-Arce, J., Lara-Flores, L., Portillo-Rodríguez, O., & Martínez-Méndez, R. (2020). Towards an anxiety and stress recognition system for academic environments based on physiological features. *Computer Methods and Programs in Biomedicine*, 190, 105408. <https://doi.org/10.1016/j.cmpb.2020.105408>
- Dzianok, P., Dzianok, P., Kołodziej, M., & Kublik, E. (2021). Detecting attention in hilbert-transformed EEG brain signals from simple-reaction and choice-reaction cognitive tasks. *2021 IEEE 21st International Conference on Bioinformatics and Bioengineering (BIBE)*, 1–4. <https://doi.org/10.1109/bibe52308.2021.9635187>
- Eyad Talal Attar, Vignesh Balasubramanian, Ersoy Subasi, & Mehmet Kaya. (2021). Stress analysis based on simultaneous heart rate variability and EEG monitoring. *IEEE Journal of Translational Engineering in Health and Medicine*, 9, 2700607. <https://doi.org/10.1109/JTEHM.2021.3106803>
- Gupta, S., & Kumar, P. (2021). Attention recognition system in online learning platform using EEG signals [Series Title: Lecture Notes in Electrical Engineering]. In P. K. Bora, S. Nandi, & S. Laskar (Eds.), *Emerging technologies for smart cities* (pp. 139–152). Springer Singapore. https://doi.org/10.1007/978-981-16-1550-4_15
- Ahmadi, N., Sasangohar, F., Nisar, T., Danesh, V., Larsen, E., Sultana, I., & Bosetti, R. (2022). Quantifying occupational stress in intensive care unit nurses: An applied naturalistic study of correlations among stress, heart rate, electrodermal activity, and skin temperature [Publisher: SAGE Publications Inc]. *Human Factors*, 64(1), 159–172. <https://doi.org/10.1177/00187208211040889>
- AlShorman, O., AlShorman, O., Masadeh, M., Heyat, M. B. B., Akhtar, F., Almahasneh, H., Ashraf, G. M., & Alexiou, A. (2022). Frontal lobe real-time EEG analysis using machine learning techniques for mental stress detection. *Journal of Integrative Neuroscience*, 21(1). <https://doi.org/10.31083/j.jin2101020>
- Kaushik, P., Moye, A., Vugt, M. V., & Roy, P. P. (2022). Decoding the cognitive states of attention and distraction in a real-life setting using EEG. *Scientific Reports*, 12(1), 20649. <https://doi.org/10.1038/s41598-022-24417-w>
- Rai, A. A., Rai, A. A., & Ahirwal, M. K. (2022). Electroencephalogram-based cognitive load classification during mental arithmetic task [Journal Abbreviation: Lecture Notes in Electrical Engineering]. *Lecture notes in electrical engineering* (pp. 479–487). https://doi.org/10.1007/978-981-19-0019-8_36
- Singh, P., Singla, R., & Kesari, A. (2022). An EEG based approach for the detection of mental stress level: An application of BCI [Series Title: Lecture Notes in Mechanical Engineering]. In M. Vashista, G. Manik, O. P. Verma, & B. Bhardwaj (Eds.), *Recent innovations in mechanical engineering* (pp. 49–57). Springer Singapore. https://doi.org/10.1007/978-981-16-9236-9_5

- Umer, W. (2022). Simultaneous monitoring of physical and mental stress for construction tasks using physiological measures. *Journal of Building Engineering*, 46, 103777. <https://doi.org/10.1016/j.jobbe.2021.103777>
- Salankar, N., Salankar, N., & Qaisar, S. M. (2023). EEG based stress classification by using difference plots of variational modes and machine learning. *PLOS ONE*, 18(11), e0293610. <https://doi.org/10.1371/journal.pone.0293610>
- Albaladejo-González, M., Gaspar-Marco, R., Tsai, N., Gómez Mármol, F., & Ruipérez-Valiente, J. A. (2024). A multimodal study of the interplay between stress, executive function, and biometrics in game-based assessment. *Expert Systems with Applications*, 242, 122864. <https://doi.org/10.1016/j.eswa.2023.122864>
- Arutyunova, K. R., Arutyunova, K. R., Bakhchina, A. V., Konovalov, D. I., Margaryan, M., Filimonov, A. V., & Shishalov, I. S. (2024). Heart rate dynamics for cognitive load estimation in a driving simulation task. *Scientific Reports*, 14(1). <https://doi.org/10.1038/s41598-024-79728-x>
- Hanshans, C., Amler, T., Zauner, J., & Bröll, L. (2024). Inducing and measuring acute stress in virtual reality: Evaluation of canonical physiological stress markers and measuring methods. *Journal of Environmental Psychology*, 94, 102107. <https://doi.org/10.1016/j.jenvp.2023.102107>
- Hernández-Sabaté, A., Yauri, J., Folch, P., Álvarez, D., & Gil, D. (2024). EEG dataset collection for mental workload predictions in flight-deck environment [Publisher: Multidisciplinary Digital Publishing Institute (MDPI)]. *Sensors*, 24(4). <https://doi.org/10.3390/s24041174>
- Iovino, M., Iovino, M., Lazic, I., Loncar-Turukalo, T., Javorka, M., Pernice, R., & Faes, L. (2024). Comparison of automatic and physiologically-based feature selection methods for classifying physiological stress using heart rate and pulse rate variability indices. *Physiological Measurement*, 45(11), 115004. <https://doi.org/10.1088/1361-6579/ad9234>
- Mukherjee, P., Mukherjee, P., & Halder Roy, A. (2024). A deep learning-based approach for distinguishing different stress levels of human brain using EEG and pulse rate. *2024 International Conference on Intelligent Algorithms for Computational Intelligence Systems (IACIS)*, 1–6. <https://doi.org/10.1109/iacis61494.2024.10721733>
- Qin, Y., Qin, Y., Bulbul, T., & Withers, J. (2024). EEG-based classification of cognitive load and task conditions for AR supported construction assembly: A deep learning approach. *Computing in Civil Engineering 2023*, 248–256. <https://doi.org/10.1061/9780784485231.030>
- Behradfar, M., Behradfar, M., Roy, S., & Nuamah, J. (2025). Optimizing mental stress detection via heart rate variability feature selection. *Sensors*, 25(13), 4154. <https://doi.org/10.3390/s25134154>
- Gallup. (2025). *State of the world's emotional health: Connecting global peace, wellbeing and health*. Retrieved January 6, 2026, from <https://www.gallup.com/analytics/349280/state-of-worlds-emotional-health.aspx>
- Nirabi, A., Nirabi, A., Rahman, F. A., Habaebi, M. H., Sidek, K. A., & Yusoff, S. (2025). Cognitive load assessment through EEG: A dataset from arithmetic and stroop tasks. *Data in Brief*, 60, 111477. <https://doi.org/10.1016/j.dib.2025.111477>

- Psykiatrifonden. (2025). *Rapport viser høje tal af stressramte unge*. Retrieved January 6, 2026, from <https://psykiatrifonden.dk/viden/fakta-forskning/egne-rapporter/rapport-viser-hoeje-tal-stressramte-unge?utm>
- The Local. (2025). *'One in five' workers in Denmark say they feel stressed*. Retrieved January 6, 2026, from <https://www.thelocal.dk/20250618/one-in-five-workers-in-denmark-say-they-feel-stressed?utm>
- Townsend, S., & Medvedev, O. N. (2025). Perceived Stress Scale (PSS). In O. N. Medvedev, C. U. Krägeloh, R. J. Siegert, & N. N. Singh (Eds.), *Handbook of Assessment in Mindfulness Research* (pp. 1849–1861). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-47219-0_91
- Britannica. (n.d.). *Silk road*. Retrieved January 6, 2026, from <https://www.britannica.com/topic/Silk-Road-trade-route>
- Calendly. (n.d.). *Calendly*. Retrieved January 13, 2026, from <https://calendly.com/>
- EmotiBit. (n.d.). *EmotiBit Shop - Wearable biometric sensing for any project*. Retrieved January 12, 2026, from <https://www.emotibit.com/>
- EmotiBit Docs. (n.d.). *EmotiBit_docs/Working_with_emotibit_data.md at master · EmotiBit/EmotiBit_docs*. Retrieved January 12, 2026, from https://github.com/EmotiBit/EmotiBit_Docs/blob/master/Working_with_emotibit_data.md
- Emotiv. (n.d.). *MN8 – 2 Channel EEG Headphones – EMOTIV*. Retrieved January 12, 2026, from <https://www.emotiv.com/products/mn8>
- Emotiv MN8 Technical Specifications. (n.d.). *Emotiv mn8 technical specifications*. Retrieved January 13, 2026, from <https://emotiv.gitbook.io/mn8-user-manual/introduction/technical-specifications>
- HeartPy. (n.d.). *Heartpy documentation*. Retrieved January 13, 2026, from <https://python-heart-rate-analysis-toolkit.readthedocs.io/en/latest/>
- Lab Streaming Layer. (n.d.). *Lsl-website*. Retrieved January 12, 2026, from <https://labstreaminglayer.org/#/>
- MN8 Data Streams. (n.d.). *Emotiv mn8 data streams*. Retrieved January 13, 2026, from <https://emotiv.gitbook.io/cortex-api/data-subscription>
- NeuroKit. (n.d.). *Neurokit github*. Retrieved January 13, 2026, from <https://github.com/neuropsychology/NeuroKit>
- OpenBCI. (n.d.). *OpenBCI EEG Headband Kit*. Retrieved January 12, 2026, from <https://shop.openbci.com/products/openbci-eeg-headband-kit>
- OpenBCI Docs. (n.d.). *Openbci documentation*. Retrieved January 12, 2026, from <https://docs.openbci.com/AddOns/Headwear/HeadBand/>
- SurveyXact. (n.d.). *Surveyxact*. Retrieved January 13, 2026, from <https://rambollxact.com/>
- Wikipedia. (n.d.). *International 10-20 system*. Retrieved January 17, 2026, from https://commons.wikimedia.org/wiki/File:21_electrodes_of_International_10-20_system_for_EEG.svg
- Music, W. (Freeport 2021). *Tenet official soundtrack | freeport - ludwig göransson | watertower* [YouTube video, accessed: 24 Jan 2026]. <https://www.youtube.com/watch?v=yywzQbZcDL0>

Appendix

This appendix contains supplementary tables, figures, and documents referenced throughout the thesis. Some materials, including signed consent forms and complete statistical test outputs, are not included here due to their size and are instead provided via an external repository at the following link: [Additional material](#).

A.1 Synthesis Matrix

[Link to Synthesis Matrix](#)

A.2 Consent Form Template

Informed Consent Form

Project Title: Inference of Attention States From Biosignals in Task-Based Environments

Investigators:

- Kawsar Nabawi
- Natalie Wilske

Supervisor:

- Sadasivan Puthusserypady Kumaran
- Jacopo Mattia de A Constantini

Institution: Technical University of Denmark (DTU)

Purpose of the Study

You are invited to take part in a study investigating how wearable sensors (EEG around the ear, heart rate variability, and temperature sensors) reflect attention and stress during simple tasks such as reading, games, and mental challenges.

The goal is both to improve our understanding of how these physiological signals relate to cognitive states and to build a high-quality anonymized dataset that can be used for further scientific analysis in this area.

What Participation Involves

If you choose to participate:

- You will complete a series of short tasks on a laptop (8 tasks in total)
- You will wear **three non-invasive sensors**:
 - an EEG earbud device (MN8)
 - a headband EEG system (OpenBCI)
 - an arm-based sensor (EmotiBit) measuring heart rate, PPG, and temperature
- You will answer short questionnaires before, between, and after the tasks (a total of 8 questionnaires)
- **Total duration: approximately 60 minutes.**

Risks and Discomfort

There are **no known risks** associated with this study.

Possible minor discomforts include:

- pressure from wearing the earbuds or strap
- mild fatigue from completing tasks

You may take breaks or stop the session at any time.

Benefits

There is no direct personal benefit to you from participating.

However, by taking part you are helping us complete our Bachelor's project and contributing

to the creation of a high-quality scientific dataset that can support future research on attention and stress.

Your participation genuinely makes a difference, and we really appreciate your help.

Data Protection and Confidentiality

- All data will be **pseudonymized**. Your name or contact information will not be linked to the data.
- Data will be stored securely and only accessible to the research team.
- Data will be used solely for academic purposes, including publications, and may be shared in anonymized form.
- This study complies with the EU General Data Protection Regulation (GDPR).

The collected data will form a research dataset that may be used for our Bachelor thesis and for future non-commercial scientific studies, always in anonymized form.

Voluntary Participation

Participation is entirely voluntary. You may decline to answer any question or withdraw at any time.

Contact Information

If you have questions about the study or want to withdraw your data, contact:

- Kawsar Nabawi
- Natalie Wilske

Consent Declaration

☐ I have read and understood the information above.

☐ I have had the opportunity to ask questions.

☐ I agree to participate in this study and consent to the collection of my data as described.

☐ I understand that I can withdraw at any time without giving a reason.

Name: _____

Signature: _____

Date: _____

A.3 Signed Consent Forms

A.4 Surveys

In the following, the pre-, task-, and post-surveys administered during the experiment can be found.

A.4.1 Pre-Survey

Participant Number

Pre-survey

Section 1: Demographics

What is your age?

What is your gender?

- ☐ Female
- ☐ Male
- ☐ Non-binary
- ☐ Prefer not to say

Are you right- or left-handed?

- ☐ Right-handed
- ☐ Left-handed
- ☐ Both/Ambidextrous

What is your highest completed education level?

- ☐ High school / Gymnasium
- ☐ Bachelor's degree
- ☐ Master's degree
- ☐ PhD or higher
- ☐ Other (please specify) _____

How many hours did you sleep last night?

Did you consume caffeine in the last 3 hours (coffee, energy drinks, etc.)?

- ☐ Yes
- ☐ No

Do you currently take medications affecting attention/stress?

- ☐ Yes
- ☐ No

Section 2: Baseline State

Please rate the following questions from 1 (Not at all) to 5 (Extremely).

How stressed do you feel right now?

- ☐ 1 - Not at all
- ☐ 2 - Slightly
- ☐ 3 - Moderately
- ☐ 4 - Very
- ☐ 5 - Extremely

How anxious or tense do you feel right now?

- ☐ 1 - Not at all
- ☐ 2 - Slightly
- ☐ 3 - Moderately
- ☐ 4 - Very
- ☐ 5 - Extremely

How calm do you feel right now?

- ☐ 1 - Not at all calm
- ☐ 2 - Slightly calm
- ☐ 3 - Moderately calm
- ☐ 4 - Very calm
- ☐ 5 - Extremely calm

How mentally tired do you feel?

- ☐ 1 - Not at all
- ☐ 2 - Slightly
- ☐ 3 - Moderately
- ☐ 4 - Very
- ☐ 5 - Extremely

How physically tired do you feel?

- ☐ 1 - Not at all
- ☐ 2 - Slightly
- ☐ 3 - Moderately
- ☐ 4 - Very
- ☐ 5 - Extremely

How easy is it for you to focus right now?

- ☐ 1 - Not at all
- ☐ 2 - Slightly
- ☐ 3 - Moderately
- ☐ 4 - Very
- ☐ 5 - Extremely

How much is your mind wandering today?

- ☐ 1 - Not at all
- ☐ 2 - Slightly
- ☐ 3 - Moderately
- ☐ 4 - Very
- ☐ 5 - Extremely

Section 3: Perceived Stress Scale

The questions in this scale ask you about your feelings and thoughts during the last month. In each case, you will be asked to indicate how often you felt or thought a certain way.

In the last month, how often have you been upset because of something that happened unexpectedly?

- ☐ 1 - Never
- ☐ 2 - Almost Never

- ☐ 3 - Sometimes
- ☐ 4 - Fairly often
- ☐ 5 - Very often

In the last month, how often have you felt that you were unable to control the important things in your life?

- ☐ 1 - Never
- ☐ 2 - Almost Never
- ☐ 3 - Sometimes
- ☐ 4 - Fairly often
- ☐ 5 - Very often

In the last month, how often have you felt nervous and stressed?

- ☐ 1 - Never
- ☐ 2 - Almost Never
- ☐ 3 - Sometimes
- ☐ 4 - Fairly often
- ☐ 5 - Very often

In the last month, how often have you felt confident about your ability to handle your personal problems?

- ☐ 1 - Never
- ☐ 2 - Almost Never
- ☐ 3 - Sometimes
- ☐ 4 - Fairly often
- ☐ 5 - Very often

In the last month, how often have you felt that things were going your way?

- ☐ 1 - Never
- ☐ 2 - Almost Never
- ☐ 3 - Sometimes
- ☐ 4 - Fairly often
- ☐ 5 - Very often

In the last month, how often have you found that you could not cope with all the things that you had to do?

- ☐ 1 - Never
- ☐ 2 - Almost Never
- ☐ 3 - Sometimes
- ☐ 4 - Fairly often
- ☐ 5 - Very often

In the last month, how often have you been able to control irritations in your life?

- ☐ 1 - Never
- ☐ 2 - Almost Never
- ☐ 3 - Sometimes
- ☐ 4 - Fairly often
- ☐ 5 - Very often

In the last month, how often have you felt that you were on top of things?

- ☐ 1 - Never
- ☐ 2 - Almost Never
- ☐ 3 - Sometimes
- ☐ 4 - Fairly often
- ☐ 5 - Very often

In the last month, how often have you been angered because of things that happened that were outside of your control?

- ☐ 1 - Never
- ☐ 2 - Almost Never
- ☐ 3 - Sometimes
- ☐ 4 - Fairly often
- ☐ 5 - Very often

In the last month, how often have you felt difficulties were piling up so high that you could not overcome them?

- ☐ 1 - Never
- ☐ 2 - Almost Never
- ☐ 3 - Sometimes
- ☐ 4 - Fairly often
- ☐ 5 - Very often

A.4.2 Task-Survey

Participant number

What task did you just complete?

- ☐ Reading comprehension
- ☐ Oddball
- ☐ 2-back-task
- ☐ Stroop Color Word Test
- ☐ Mental arithmetic
- ☐ Holding your breath

Task survey

Please indicate how you experienced the task you just completed.
Rate each statement on a scale from 1 (Not at all) to 5 (Extremely).

How mentally demanding was this task?

- ☐ 1 - Not at all
- ☐ 2 - Slightly
- ☐ 3 - Moderately
- ☐ 4 - Very
- ☐ 5 - Extremely

How hard did you have to work to accomplish this task?

- ☐ 1 - Not at all
- ☐ 2 - Slightly
- ☐ 3 - Moderately
- ☐ 4 - Very
- ☐ 5 - Extremely

How focused were you during this task?

- ☐ 1 - Not at all
- ☐ 2 - Slightly
- ☐ 3 - Moderately
- ☐ 4 - Very
- ☐ 5 - Extremely

How hurried or time-pressured did you feel?

- ☐ 1 - Not at all
- ☐ 2 - Slightly
- ☐ 3 - Moderately
- ☐ 4 - Very
- ☐ 5 - Extremely

How stressed or tense did you feel during the task?

- ☐ 1 - Not at all
- ☐ 2 - Slightly
- ☐ 3 - Moderately
- ☐ 4 - Very
- ☐ 5 - Extremely

How frustrated or irritated did you feel?

- ☐ 1 - Not at all
- ☐ 2 - Slightly
- ☐ 3 - Moderately
- ☐ 4 - Very
- ☐ 5 - Extremely

How successful or confident were you in performing the task?

- ☐ 1 - Not at all
- ☐ 2 - Slightly
- ☐ 3 - Moderately
- ☐ 4 - Very
- ☐ 5 - Extremely

How engaged or immersed did you feel?

- ☐ 1 - Not at all
- ☐ 2 - Slightly
- ☐ 3 - Moderately
- ☐ 4 - Very
- ☐ 5 - Extremely

A.4.3 Post-Survey

Participant number

Post-survey

Section 1: State Stress & Fatigue

Please rate the following from 1 (Not at all) to 5 (Extremely).

How stressed do you feel right now?

- ☐ 1 - Not at all
- ☐ 2 - Slightly
- ☐ 3 - Moderately
- ☐ 4 - Very
- ☐ 5 - Extremely

How anxious or tense do you feel right now?

- ☐ 1 - Not at all
- ☐ 2 - Slightly
- ☐ 3 - Moderately
- ☐ 4 - Very
- ☐ 5 - Extremely

How mentally exhausted do you feel right now after completing all the tasks?

- ☐ 1 - Not at all
- ☐ 2 - Slightly
- ☐ 3 - Moderately
- ☐ 4 - Very
- ☐ 5 - Extremely

How mentally demanding did the overall experiment feel?

- ☐ 1 - Not at all
- ☐ 2 - Slightly
- ☐ 3 - Moderately
- ☐ 4 - Very
- ☐ 5 - Extremely

How much effort did the experiment require from you?

- ☐ 1 - Not at all
- ☐ 2 - Slightly
- ☐ 3 - Moderately
- ☐ 4 - Very
- ☐ 5 - Extremely

How difficult was it to stay focused throughout the study?

- ☐ 1 - Not at all
- ☐ 2 - Slightly
- ☐ 3 - Moderately
- ☐ 4 - Very
- ☐ 5 - Extremely

Section 2: Device & Experimental Experience

Please rate the following from 1 (Not at all) to 5 (Extremely).

How comfortable were the sensors?

- ☐ 1 - Not at all
- ☐ 2 - Slightly
- ☐ 3 - Moderately
- ☐ 4 - Very
- ☐ 5 - Extremely

To what extent did the sensors distract you during the tasks?

- ☐ 1 - Not at all
- ☐ 2 - Slightly
- ☐ 3 - Moderately
- ☐ 4 - Very
- ☐ 5 - Extremely

Did you experience discomfort with the devices? If yes, please specify with one (EmotiBit - upper arm, MN8 - earbuds, OpenBCI - headband)

- ☐ Yes _____
- ☐ No

Which task did you find most difficult?

- ☐ Reading comprehension
- ☐ Oddball
- ☐ 2-back-task
- ☐ Stroop Color Word Test
- ☐ Mental Arithmetic
- ☐ Holding your breath

Which task felt the easiest?

- ☐ Reading comprehension
- ☐ Oddball
- ☐ 2-back-task
- ☐ Stroop Color Word Test
- ☐ Mental Arithmetic
- ☐ Holding your breath

Which task felt the most stressful?

- ☐ Reading comprehension
- ☐ Oddball
- ☐ 2-back-task
- ☐ Stroop Color Word Test
- ☐ Mental Arithmetic
- ☐ Holding your breath

Any comments about your experience?

A.5 xdf files

Link to repository with xdf files from the experiment sessions.

A.6 Available Data Streams

Sensor	Data Stream	Description
EmotiBit	SCR_AMP	Skin Conductance Response (SCR) Amplitude
EmotiBit	SCR_FREQ	Skin Conductance Response (SCR) Frequency
EmotiBit	SCR_RIS	Skin Conductance Response (SCR) Rise Time
EmotiBit	EDA	Electrodermal Activity
EmotiBit	PPG_IR	Photoplethysmography (PPG) infrared
EmotiBit	PPG_RED	Photoplethysmography (PPG) red
EmotiBit	PPG_GRN	Photoplethysmography (PPG) green
EmotiBit	HR	Heart rate
EmotiBit	TEMP1	Skin temperature
EmotiBit	THERM	Temperature via medical-grade thermopile
EmotiBit	ACC_X	Acceleration (X-axis)
EmotiBit	ACC_Y	Acceleration (Y-axis)
EmotiBit	ACC_Z	Acceleration (Z-axis)
EmotiBit	GYRO_X	Gyroscope (X-axis)
EmotiBit	GYRO_Y	Gyroscope (Y-axis)
EmotiBit	GYRO_Z	Gyroscope (Z-axis)
EmotiBit	MAG_X	Magnetometer (X-axis)
EmotiBit	MAG_Y	Magnetometer (Y-axis)
EmotiBit	MAG_Z	Magnetometer (Z-axis)
Emotiv MN8	MN8_EEG	Raw EEG
Emotiv MN8	MN8_POW	Band power of each EEG sensor
Emotiv MN8	MN8_EQ	EEG quality of each EEG sensor
Emotiv MN8	MN8_MOT	Motion data
Emotiv MN8	MN8_MET	Performance metrics output
Emotiv MN8	MN8_DEV	Device data (battery, signal strength, contact quality)
System	TaskMarkers	Task and event markers from PsychoPy
OpenBCI	obci_eeg1	OpenBCI raw EEG
OpenBCI	obci_aux	OpenBCI motion data

Table 9: Overview of all recorded processing streams (EmotiBit Docs, n.d., MN8 Data Streams, n.d.)

A.7 Data Completeness

Participant ID	Streams complete	Streams missing	Tasks complete	Tasks missing
P02	×	MN8_EEG	✓	—
P03	×	MN8_EEG; obci_eeg	×	MAT
P04	×	obci_eeg	✓	—
P05	×	obci_eeg	✓	—
P06	✓	—	✓	—
P07	✓	—	✓	—
P08	✓	—	✓	—
P09	×	EDA; HR; PPG_IR; THERM	✓	—
P10	×	EDA; HR; PPG_IR; THERM (half session)	✓	—
P11	✓	—	✓	—
P12	✓	—	✓	—
P13	✓	—	✓	—
P14	✓	—	✓	—
P15	✓	—	✓	—
P16	✓	—	✓	—
P17	✓	—	✓	—
P18	✓	—	✓	—
P19	✓	—	✓	—
P20	✓	—	✓	—
P22	✓	—	✓	—
P23	✓	—	✓	—
P24	✓	—	✓	—
P25	✓	—	×	ODDBALL; NEUTRAL2

Table 10: Overview of data completeness across participants, indicating missing data streams and incomplete tasks.

A.8 Features Extracted Per Signal Modality

A.8.1 EEG Features

Feature family	Domain	Channels	Frequency bands	Aggregation
Absolute band power	Frequency	Emotiv MN8: T7, T8 OpenBCI: ch1, ch2	$\delta, \theta, \alpha, \beta, \gamma$	Q25, Q50, Q75
Relative band power	Frequency	Emotiv MN8: T7, T8 OpenBCI: ch1, ch2	$\delta, \theta, \alpha, \beta, \gamma$	Q25, Q50, Q75
Log band asymmetry	Frequency	Emotiv MN8: T7, T8	$\delta, \theta, \alpha, \beta, \gamma$	Q25, Q50, Q75
Time-domain statistics (mean, std, RMS)	Time	Emotiv MN8: T7, T8 OpenBCI: ch1, ch2	–	Q25, Q50, Q75
Hjorth parameters (activity, mobility, complexity)	Time	Emotiv MN8: T7, T8 OpenBCI: ch1, ch2	–	Q25, Q50, Q75

Table 11: Consolidated EEG feature extraction for MN8 and OpenBCI. Features were computed in 2 s windows with 50% overlap and aggregated across windows per segment using quantiles (Q25, Q50, Q75).

A.8.2 HRV Features

Feature family	Domain	Statistics included	Description
Heart rate	Cardiac	Mean HR	Average heart rate estimated from the PPG signal.
Inter-beat interval (IBI)	Cardiac	Mean, median, min, max	Descriptive statistics of inter-beat intervals.
Time-domain HRV	HRV (time)	SDNN, RMSSD, pNN50	Standard time-domain HRV metrics reflecting beat-to-beat variability.
Frequency-domain HRV	HRV (frequency)	LF, HF, LF/HF	Spectral HRV measures capturing autonomic balance.
RR interval variability	RR statistics	Std, variance, IQR	Variability measures of inter-beat intervals.
RR interval distribution	RR statistics	Q25, Q75	Distributional measures of inter-beat intervals.
Beat count	RR statistics	Number of RR intervals	Number of detected inter-beat intervals in the segment.

Table 12: Consolidated overview of heart rate variability (HRV) features extracted from the PPG signal. Features were computed from inter-beat intervals and summarized at the segment level.

A.8.3 EDA Features

Feature family	EDA component	Domain	Statistics	Aggregation
Cleaned EDA level	Cleaned signal	Time	Mean, median, std, min, max	Q25, Q50, Q75
Cleaned EDA variability	Cleaned signal	Time	IQR, CV	Q25, Q50, Q75
Cleaned EDA trend	Cleaned signal	Time	Slope, end–start difference	Segment-level
Tonic EDA level	Tonic component	Time	Mean, median, std, min, max	Q25, Q50, Q75
Tonic EDA variability	Tonic component	Time	IQR, CV	Q25, Q50, Q75
Tonic EDA trend	Tonic component	Time	Slope, end–start difference	Segment-level
Phasic EDA level	Phasic component	Time	Mean, median, std, min, max	Q25, Q50, Q75
Phasic EDA variability	Phasic component	Time	IQR, CV	Q25, Q50, Q75
Phasic EDA trend	Phasic component	Time	Slope, end–start difference	Segment-level
SCR activity	SCR events	Event-based	Count, rate	Segment-level
SCR amplitude	SCR events	Event-based	Mean, median, max, sum	Segment-level
SCR timing	SCR events	Event-based	ISI mean, ISI std	Segment-level

Table 13: Consolidated overview of EDA features extracted in this study. Features were computed from the cleaned EDA signal, as well as tonic and phasic components, and summarized at the window or segment level using descriptive statistics.

A.8.4 Skin Temperature Features

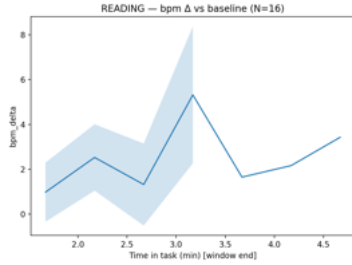
Feature family	Signal variant	Statistics	Description
Central tendency	Filtered, Delta	Mean, median	Average skin temperature level and baseline-relative change.
Variability	Filtered, Delta	Std, Q25, Q75	Variability and distribution of skin temperature values.
Range	Filtered, Delta	Min, max	Minimum and maximum skin temperature values.
Temporal trend	Filtered, Delta	Linear slope	Linear change in skin temperature over time (°C/min).
Segment change	Filtered, Delta	End–start difference	Difference between the final and initial temperature values of the segment.

Table 14: Consolidated overview of skin temperature features extracted in this study.

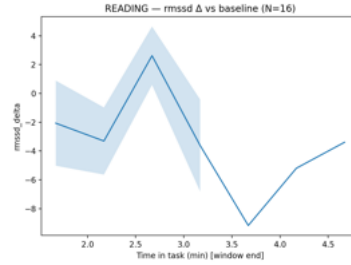
A.9 Within-Trial Plots Per Signal Modality

A.9.1 HRV within-trial plots

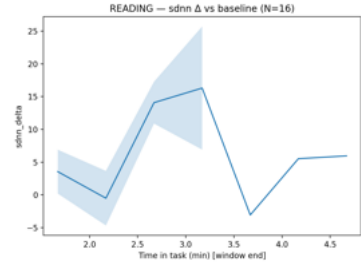
Reading task



(a) Heart Rate during reading.

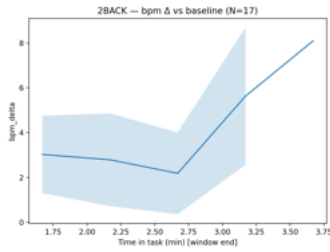


(b) RMSSD during reading.

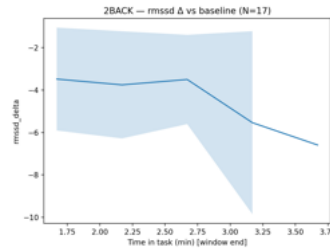


(c) SDNN during reading.

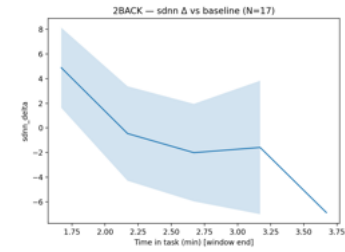
2-Back task



(d) Heart Rate during 2-Back.

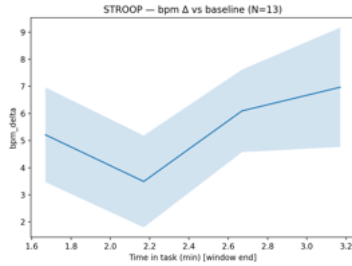


(e) RMSSD during 2-Back.

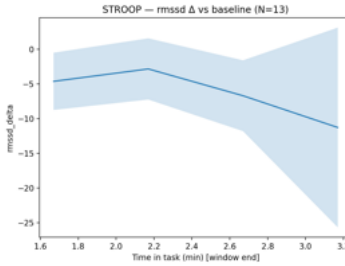


(f) SDNN during 2-Back.

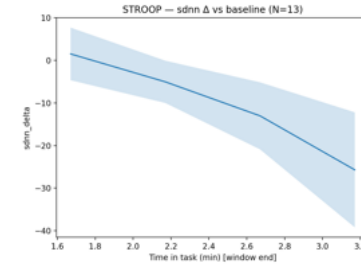
Stroop task



(g) Heart Rate during Stroop.



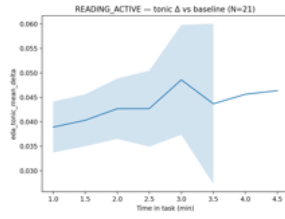
(h) RMSSD during Stroop.



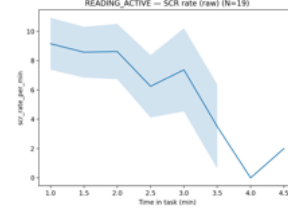
(i) SDNN during Stroop.

Figure 56: Temporal dynamics of HR, RMSSD, and SDNN during the reading, 2-Back, and Stroop task.

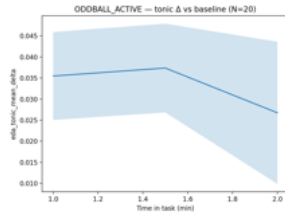
A.9.2 EDA Within-Trial Plots



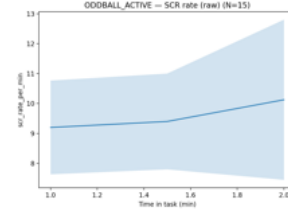
(a) EDA tonic level during Reading.



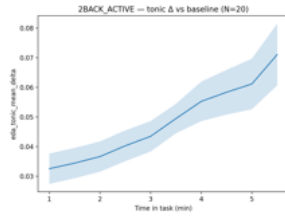
(b) SCR rate (per min) during Reading.



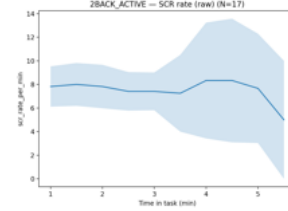
(c) EDA tonic level during Oddball.



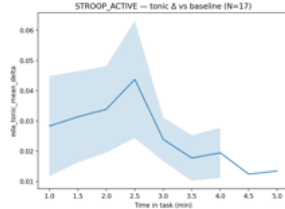
(d) SCR rate (per min) during Oddball.



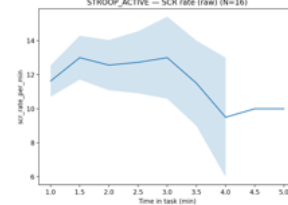
(e) EDA tonic level during 2-Back.



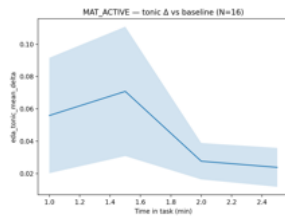
(f) SCR rate (per min) during 2-Back.



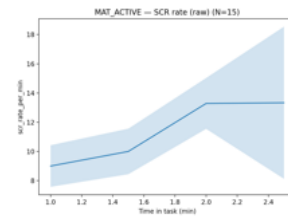
(g) EDA tonic level during Stroop.



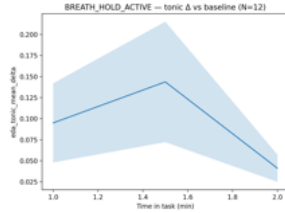
(h) SCR rate (per min) during Stroop.



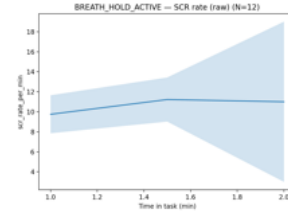
(i) EDA tonic level during MAT.



(j) SCR rate (per min) during MAT.



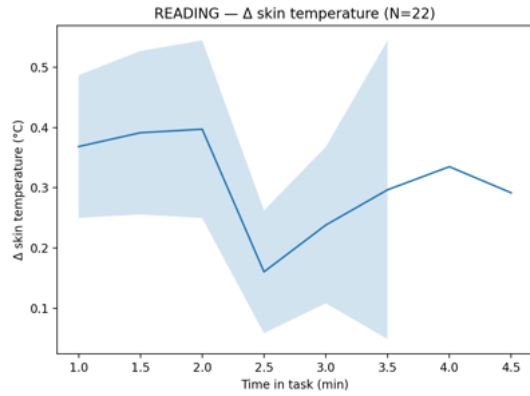
(k) EDA tonic level during Breath-holding.



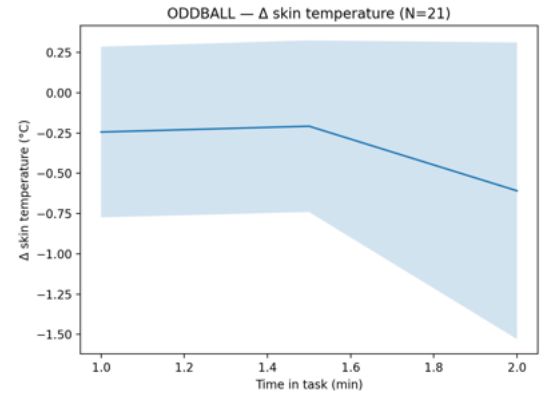
(l) SCR rate (per min) during Breath-holding.

Figure 57: Temporal dynamics of EDA tonic level and SCR rate (per min) during Reading, Oddball, 2-Back, Stroop, MAT and Breath-holding tasks.

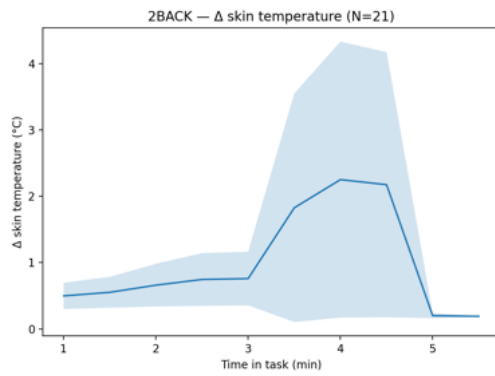
A.9.3 ST Within-Trial Plots



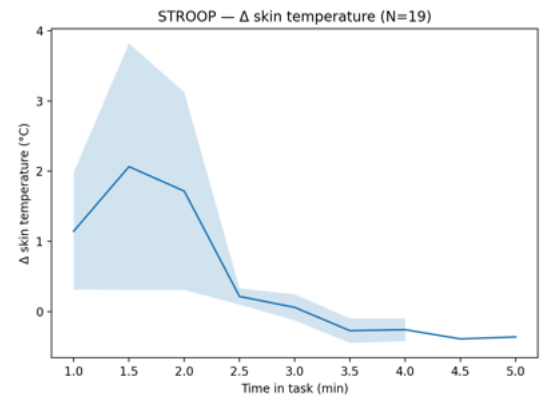
(a) Skin temperature relative to neutral baseline during Reading.



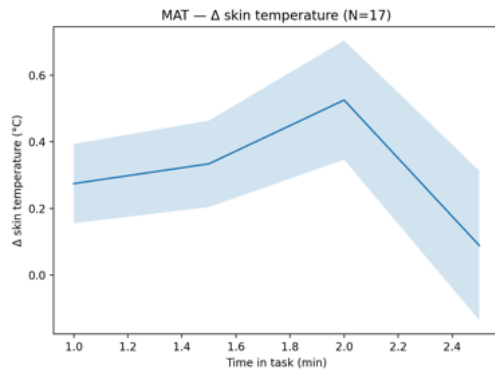
(b) Skin temperature relative to neutral baseline during Oddball.



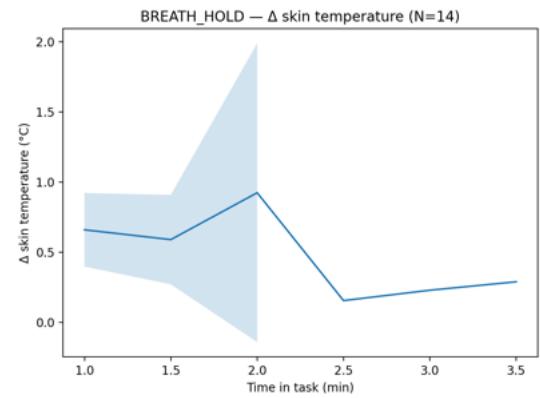
(c) Skin temperature relative to neutral baseline during 2-Back.



(d) Skin temperature relative to neutral baseline during Stroop.



(e) Skin temperature relative to neutral baseline during MAT.



(f) Skin temperature relative to neutral baseline during Breath-holding.

Figure 58: Temporal dynamics of skin temperature change (Δ ST) across tasks.

A.10 Between-Trial Plots Per Signal Modality

A.10.1 Emotiv MN8 EEG Between-Trial Plots

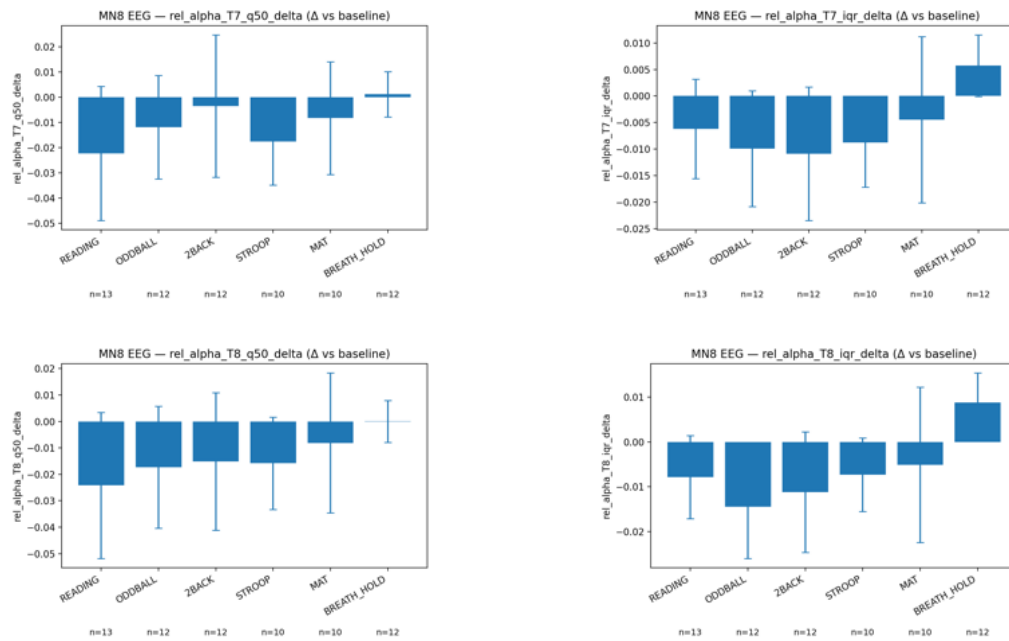


Figure 59: Task-level changes in relative alpha power across tasks for Emotiv MN8 EEG. Relative alpha power is shown for electrodes T7 (top) and T8 (bottom), summarized using the median and interquartile range (IQR). All values are expressed as change relative to baseline.

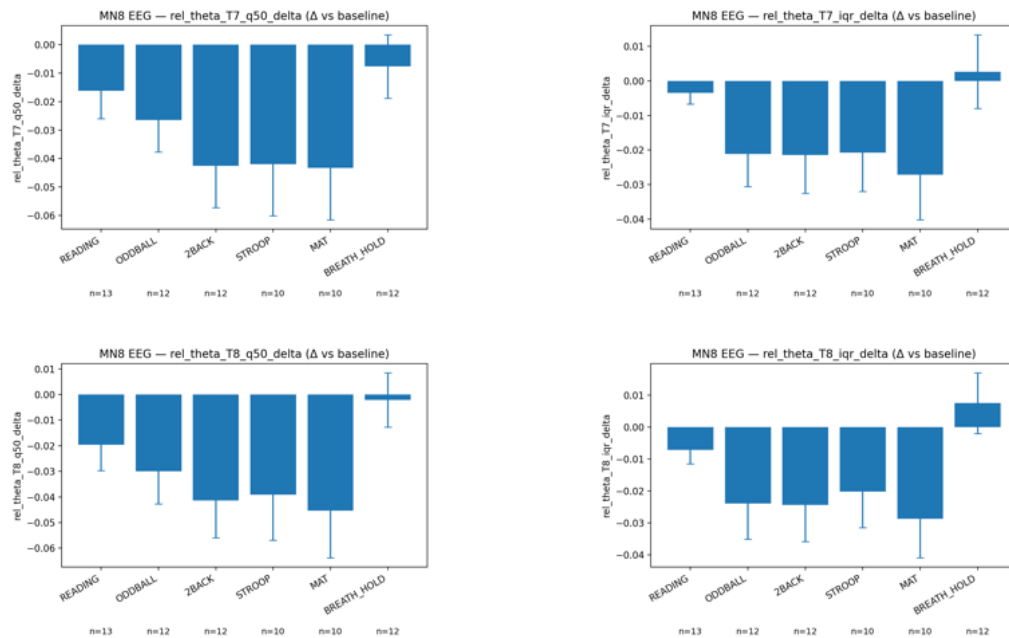


Figure 60: Task-level changes in relative theta power across tasks for Emotiv MN8 EEG. Relative alpha power is shown for electrodes T7 (top) and T8 (bottom), summarized using the median and interquartile range (IQR). All values are expressed as change relative to baseline.

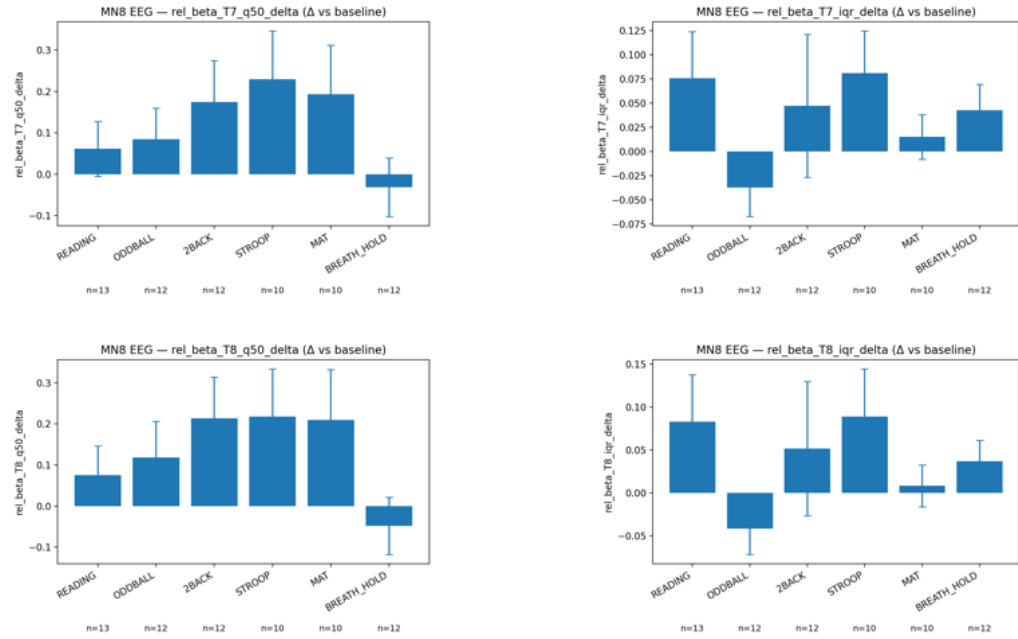


Figure 61: Task-level changes in relative beta power across tasks for Emotiv MN8 EEG. Relative alpha power is shown for electrodes T7 (top) and T8 (bottom), summarized using the median and interquartile range (IQR). All values are expressed as change relative to baseline.

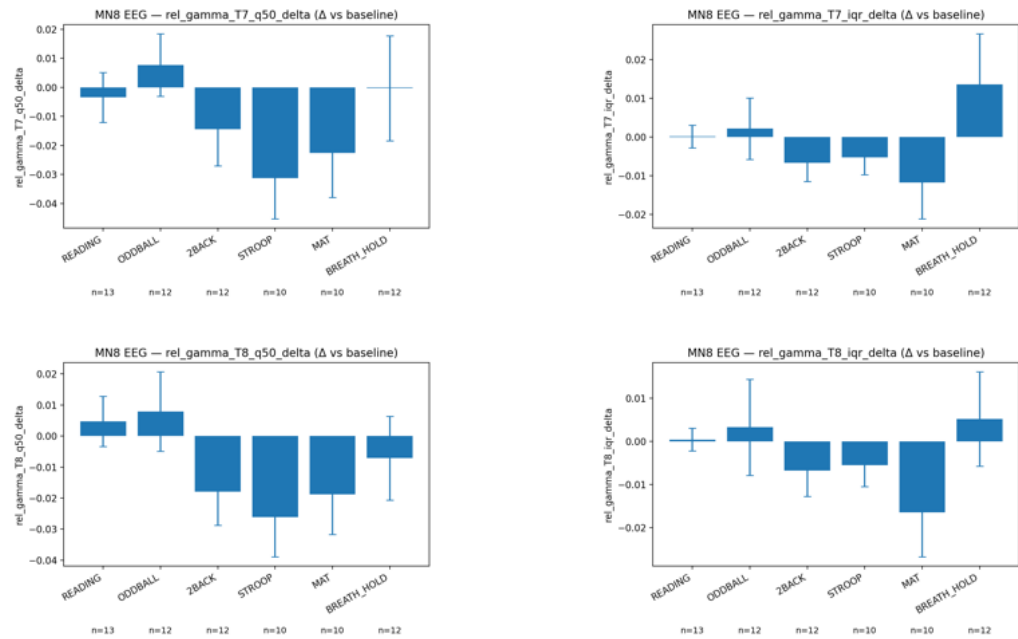


Figure 62: Task-level changes in relative gamma power across tasks for Emotiv MN8 EEG. Relative alpha power is shown for electrodes T7 (top) and T8 (bottom), summarized using the median and interquartile range (IQR). All values are expressed as change relative to baseline.

A.10.2 OpenBCI EEG Between-Trial Plots

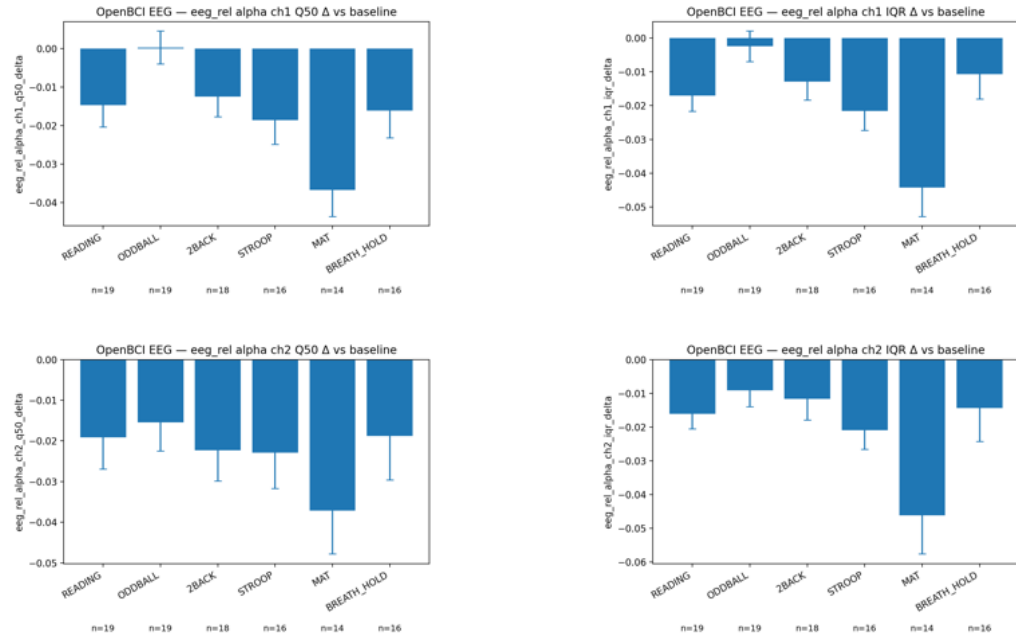


Figure 63: Task-level changes in relative alpha power across tasks for OpenBCI EEG. Relative alpha power is shown for electrodes Fp1 (ch1 - top) and Fp2 (ch2 - bottom), summarized using the median and interquartile range (IQR). All values are expressed as change relative to baseline.

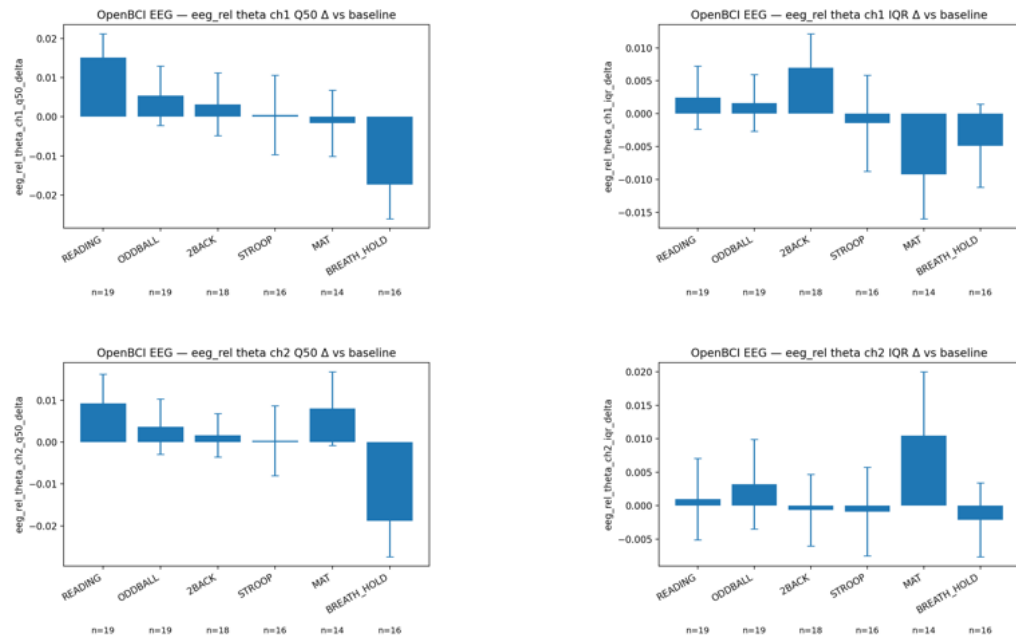


Figure 64: Task-level changes in relative theta power across tasks for OpenBCI EEG. Relative alpha power is shown for electrodes Fp1 (ch1 top) and Fp2 (ch2 bottom), summarized using the median and interquartile range (IQR). All values are expressed as change relative to baseline.

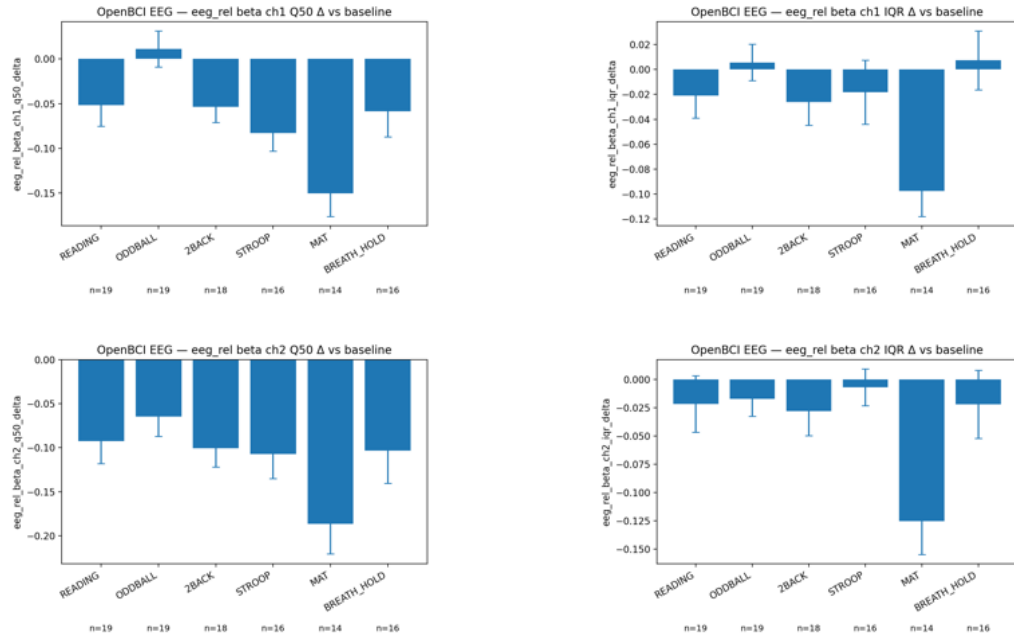


Figure 65: Task-level changes in relative beta power across tasks for OpenBCI EEG. Relative alpha power is shown for electrodes Fp1 (ch1 top) and Fp2 (ch2 bottom), summarized using the median and interquartile range (IQR). All values are expressed as change relative to baseline.

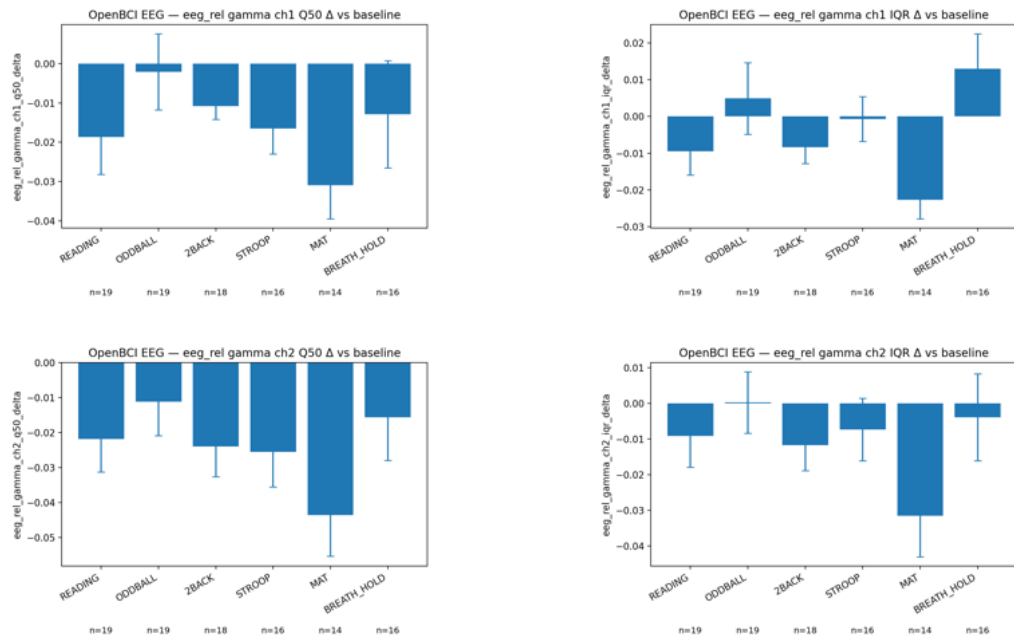
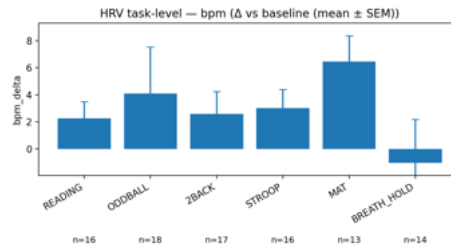
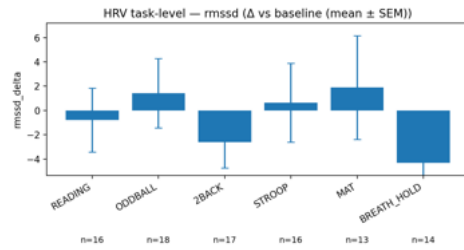


Figure 66: Task-level changes in relative gamma power across tasks for OpenBCI EEG. Relative alpha power is shown for electrodes Fp1 (ch1 top) and Fp2 (ch2 bottom), summarized using the median and interquartile range (IQR). All values are expressed as change relative to baseline.

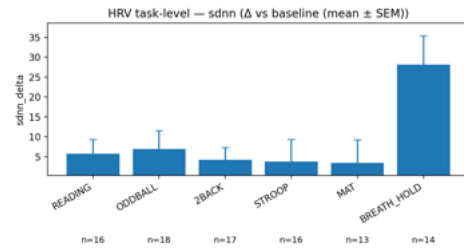
A.10.3 HRV Between-Trial Plots



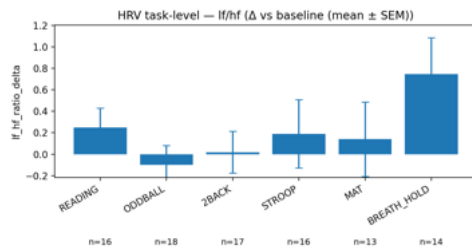
(a) Task-level changes in heart rate (bpm) relative to baseline (mean $\pm$ SEM).



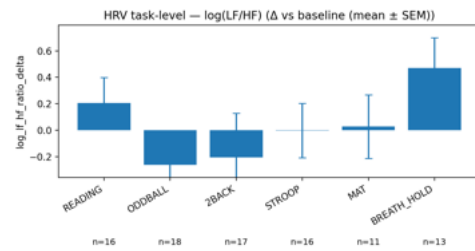
(b) Task-level changes in RMSSD relative to baseline (mean $\pm$ SEM).



(c) Task-level changes in SDNN relative to baseline (mean $\pm$ SEM).



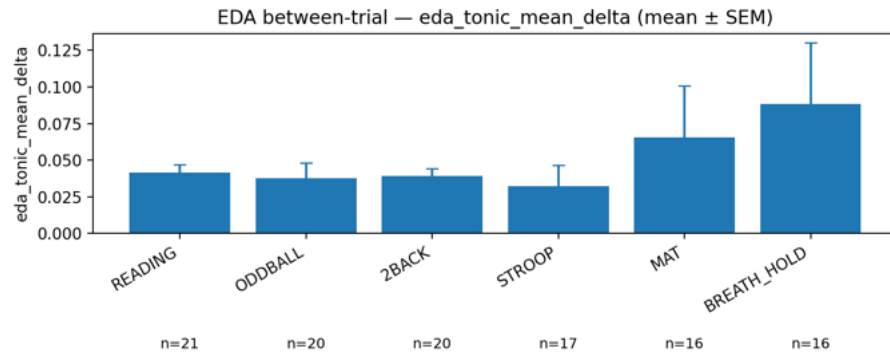
(d) Task-level changes in autonomic balance indexed by LF/HF ratio relative to baseline (mean $\pm$ SEM).



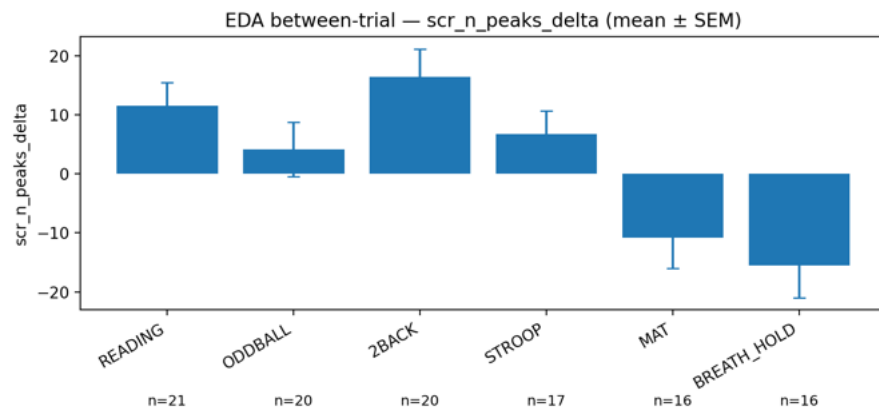
(e) Task-level changes in autonomic balance indexed by log(LF/HF) relative to baseline (mean $\pm$ SEM).

Figure 67: Task-level changes in HRV time- and frequency-domain relative to baseline (mean $\pm$ SEM).

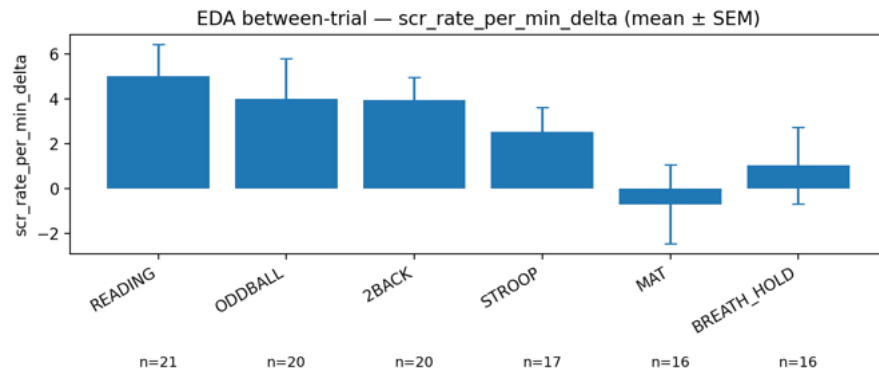
A.10.4 EDA Between-Trial Plots



(a) Task-level changes in tonic electrodermal activity relative to baseline (mean $\pm$ SEM).



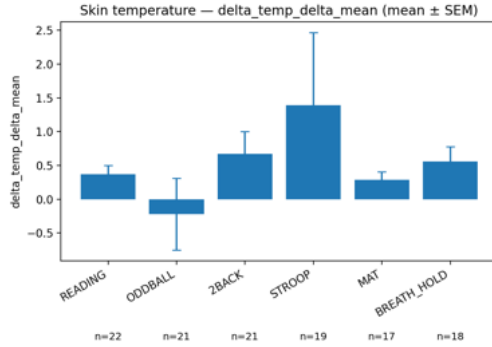
(b) Task-level changes in phasic electrodermal activity indexed by the number of skin conductance responses relative to baseline (mean $\pm$ SEM).



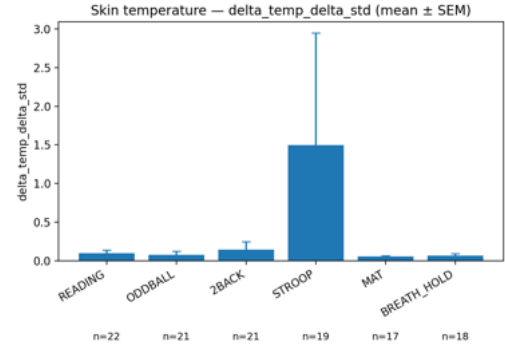
(c) Task-level changes in phasic electrodermal activity indexed by SCR rate per minute relative to baseline (mean $\pm$ SEM).

Figure 68: Task-level changes in tonic and phasic electrodermal activity relative to baseline (mean $\pm$ SEM).

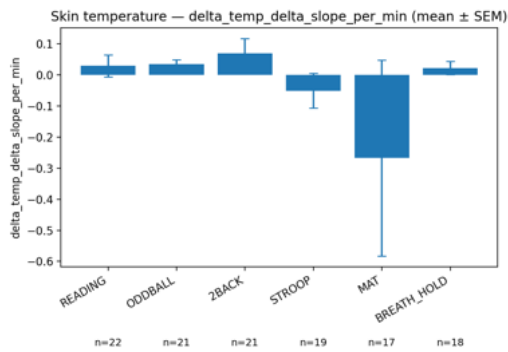
A.10.5 ST Between-Trial Plots



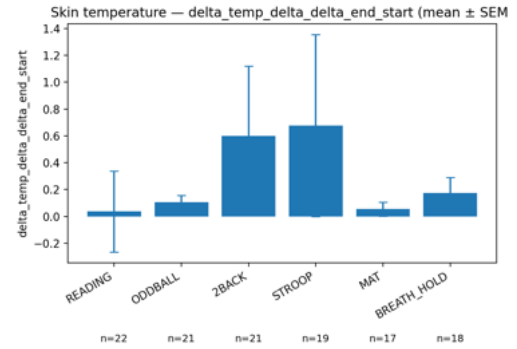
(a) Task-level changes in mean skin temperature relative to baseline (Δ mean; mean $\pm$ SEM).



(b) Task-level skin temperature variability (standard deviation, Δ vs. baseline; mean $\pm$ SEM).



(c) Task-level skin temperature slope per minute ($^{\circ}\text{C}/\text{min}$; mean $\pm$ SEM).



(d) Task-level net change in skin temperature (end-start, Δ vs. baseline; mean $\pm$ SEM).

Figure 69: Task-level changes in skin temperature relative to baseline (mean $\pm$ SEM).

A.11 Statistical Results

[Link to Emotiv MN8 EEG statistical results.](#)

[Link to OpenBCI EEG statistical results.](#)

[Link to HRV statistical results.](#)

[Link to EDA statistical results.](#)

[Link to ST statistical results.](#)

A.12 Machine Learning Training Files and Outputs

[Link to ML training files and outputs.](#)

Technical
University of
Denmark

Ørstedes Plads, Building 345B
2800 Kgs. Lyngby
Tlf. 4525 1700

www.healthtech.dtu.dk/