University of Pannonia

DOCTORAL (PhD) DISSERTATION

Abdulrahman Khalid Eesee Al-sabaawi

2026

University of Pannonia

Doctoral (PhD) Dissertation

Human Cognitive Load and Awareness Monitoring Using AI and Biosignals in Assembly Operations

DOI:10.18136/PE.2026.985

Author:
Abdulrahman Khalid Eesee Al-sabaawi

Supervisors:
Dr. Tamás Ruppert
Dr. habil. György Eigner

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

in the

Doctoral School in Chemical Engineering and Material Sciences,
Faculty of Engineering,
University of Pannonia
Department of System Engineering

University of Pannonia
Faculty of Engineering

2026

Human Cognitive Load and Awareness Monitoring Using AI and Biosignals in Assembly Operations

The thesis was prepared for the award of a doctoral degree (PhD) within the framework of the Chemical Engineering and Material Sciences Doctoral School at University of Pannonia

in the discipline of Bio-, environmental-, and chemical engineering

written by: Abdulrahman Khalid Eesee Al-sabaawi

Supervisors: Dr. Tamás Ruppert, Dr. habil. György Eigner

I recommend the dissertation for acceptance: yes / no.

…….
Supervisor (Dr. Tamás Ruppert)

I recommend the dissertation for acceptance: yes / no.

…….
Supervisor (Dr. habil. György Eigner)

I recommend the dissertation for peer review.

…….
chair of the DDHC

The PhD-candidate has achieved ……. % at the public debate.

The composition of the Final Examination Committee:

chair:…….

reviewers:.……

members:…….

Veszprém,

…….
chair of the committee

Qualification of degree: ….....

Veszprém,

…….
chair of the UDHC

UNIVERSITY OF PANNONIA

Abstract

Faculty of Engineering
Department of System Engineering

Doctor of Philosophy

Human Cognitive Load and Awareness Monitoring Using AI and Biosignals in Assembly Operations

by Abdulrahman Khalid Eesee Al-sabaawi

Human cognitive load limits the efficiency of manual and collaborative assembly. However, it is rarely measurable during ongoing work. This thesis develops an assistance architecture that connects task characteristics with physiological signals extracted from the human body in order to support timely adaptation. A multimodal analysis pipeline, Extended Human Asset Administration (Extended-HAAS), was developed, combining computational, artificial intelligence (AI), and statistical methods, utilizing heart rate variability (HRV), electrodermal activity (EDA), and wrist acceleration (ACC). These signals were processed in short, event-aligned windows with individualized baselines and feature normalization to capture fluctuations in cognitive load over time.

Four experimental studies were conducted focusing on different aspects of human performance: situational awareness during dual tasking, sequencing between mental and physical work blocks, the influence of instruction format (visual versus code-based), and multitasking with a robot partner at two levels of attentional difficulty. Across these studies, subjective workload was evaluated using the NASA Task Load Index (NASA-TLX), the Instantaneous Self-Assessment of workload (ISA), and the Short Dundee Stress State Questionnaire (Short DSSQ). Performance was assessed through task completion time (TCT), number of task repetitions (NTR), and assembly precision. Physiological features were analyzed using AI models, mixed models, and logistic models, with careful data processing to reduce the influence of motion-related noise.

The studies provided strong evidence for the reliability of the extracted features from the recorded biosignals in tracking dynamic human cognitive load. Human situational awareness in a dual-task scenario was found to be predictable using physiological signals, kinematic features, and reaction time. Alternating short physical and mental tasks showed a general trend toward lower perceived load at the session level. Heart rate variability shifted toward parasympathetic dominance, while electrodermal activity increased later in the session, indicating a balance between recovery and arousal over time. The instruction format also produced notable outcomes, showing lower cognitive load for the straightforward visual format but higher precision for the more cognitively demanding code-based format, reflecting a speed–accuracy trade-off. In human-robot collaboration multitasking, the more demanding attentional condition showed no significant differences but a consistent trend toward longer task time, more errors, and greater perceived effort.

These results support the use of a closed-loop assistance system through the Extended-HAAS, which integrates HRV, EDA, and ACC, along with information about the human environment and the current task, to manage cognitive load. Although the findings are based on controlled studies with relatively small samples, they highlight practical decision points and reporting standards for developing physiology-informed, human-centered production systems.

UNIVERSITY OF PANNONIA

Abstract

Faculty of Engineering

Department of System Engineering

Doctor of Philosophy

Human Cognitive Load and Awareness Monitoring Using AI and Biosignals in Assembly Operations

الخلاصة

يحدّ الحمل المعرفي لدى الإنسان من كفاءة عمليات التجميع اليدوية والتعاونية. ومع ذلك، نادرًا ما يمكن قياسه أثناء العمل الجاري. تطوّر هذه الأطروحة بنية مساعدة تربط خصائص المَهَمَّة بالإشارات الفسيولوجية الحيوية المستخلصة من جسم الإنسان لدعم التكيّف في الوقت المناسب. وقد طوّرت هذه الأطروحة آلية تحليل متعددة الوسائط، تُسمى إدارة الأصول البشرية المُوسّعة (Extended HAAS)، تجمع بين الأساليب الحاسوبية والذكاء الإصطناعي والأساليب الإحصائية، بإستخدام تقلب معدل ضربات القلب (HRV)، والنشاط الكهربي لتوصيلية الجلد (EDA)، وتسارع حركة اليد عند المعصم (ACC). وجرى تحليل هذه الإشارات ضمن نوافذ زمنية قصيرة متزامنة مع أحداث المَهمَّة، مع تخصيص خط أساس لكل فرد وتوحيد مقاييس الميزات، لرصد تقلبات الحمل المعرفي مع مرور الوقت.

أُجريت أربع دراسات تجريبية ركزت على جوانب مختلفة من الأداء البشري: الوعي الظرفي أثناء أداء مهمتين في آن واحد، والتسلسل بين مراحل العمل الذهني والبدني، وتأثير شكل التعليمات (المرئية مقابل البرمجية)، وتعدد المهام مع روبوت شريك على مستويين من صعوبة الانتباه. في هذه الدراسات، قُيِّم عبء العمل الذاتي باستخدام مؤشر ناسا لعبء المهام (NASA-TLX)، والتقييم الذاتي الفوري لعبء العمل (ISA)، واستبيان دندي المختصر لحالة الإجهاد (Short DSSQ). وقُيِّم الأداء من خلال زمن إنجاز المهمة (TCT)، وعدد مرات تكرار المهمة (NTR)، ودقة التجميع. كما جرى تحليل الإشارات الفيسيولوجية والحيوية باستخدام نماذج الذكاء الاصطناعي، والنماذج المختلطة (Mixed models)، والنماذج اللوجستية (Logistic models)، مع معالجة دقيقة للبيانات لتقليل تأثير التشويش الناتج عن الحركة.

قدمت الدراسات أدلة قوية على موثوقية الخصائص المستخرجة من الإشارات الحيوية المسجلة في تتبع الحمل المعرفي البشري الديناميكي. وُجد أن الوعي الظرفي البشري في سيناريو المهام المزدوجة قابل للتنبؤ باستخدام الإشارات الفيسيولوجية الحيوية والخصائص الحركية وزمن رد الفعل. كما أظهر التناوب بين المهام البدنية والذهنية القصيرة اتجاهًا

عامًا نحو انخفاض الحمل المُدرك على مستوى الفترة المحددة. واتجهت تقلبات معدل ضربات القلب نحو هيمنة الجهاز العصبي اللاودي، بينما ازداد النشاط الكهربائي الجلدي لاحقًا في الجلسة، مما يشير إلى توازن بين الاستعادة الفسيولوجية والاستثارة بمرور الوقت. وأظهرت صيغة التعليمات أيضًا نتائج لافتة، إذ ارتبطت الصيغة البصرية المباشرة بحمل معرفي أقل، بينما ارتبطت صيغة التعليمات القائمة على الرموز المشفرة بحمل معرفي عالٍ لكن بدقة أعلى بما يعكس مقايضة بين السرعة والدقة. وفي سياق تعدد المهام في التعاون بين الإنسان والروبوت، لم تُظهر الحالة الأكثر صعوبة من حيث الانتباه فروقًا ذات دلالة، لكنها أظهرت اتجاهًا ثابتًا نحو زمن مهمة أطول وأخطاء أكثر وجهد مُدرَك أعلى.

تدعم هذه النتائج استخدام نظام مساعدة مغلق (Closed-loop) عبر إدارة الأصول البشرية الموسّعة (Extended HAAS)، الذي يدمج بيانات تقلب معدل ضربات القلب (HRV) ووالنشاط الكهربائي لتوصيلية الجلد (EDA) وتسارع حركة معصم اليد (ACC) مع المعلومات المتعلقة بما يحيط الإنسان والمَهَمَّة الحالية التي يقوم بها المشارك لإدارة الحمل المعرفي. ورغم أن النتائج مبنية على دراسات ضمن إعداد/سياق تجريبي مُنضبط وبأحجام عينات صغيرة نسبيًا، فإنها تُحدِّد بوضوح مواضع القرارات العملية التي يجب حسمها، وتضع خطوطاً إرشادية لما ينبغي الإبلاغ عنه منهجياً عند تصميم أنظمة إنتاج تتمحور حول الإنسان وتستند إلى مؤشرات فسيولوجية حيوية.

Acknowledgements

I would like to express my deepest gratitude to my supervisor, Dr. Tamás Ruppert, for his continuous guidance, encouragement, and trust throughout my PhD journey. His kindness, flexibility, and optimism made every challenge easier to overcome. I am especially grateful for his ability to find solutions in difficult situations and for believing in my capabilities from the very beginning. Working under his supervision has been both a privilege and a valuable learning experience.

I would also like to thank my co-supervisor, Dr. György Eigner, for his kind support and valuable feedback whenever we met. His input contributed to shaping the direction of this work.

I am also grateful to David Kostolani from TU-Wien for his support during my PhD. His constructive comments, collaboration, and help with the experimental work and analyses were very valuable for the development of this thesis.

My sincere appreciation goes to the Industry 5.0 Laboratory at the University of Pannonia, where I carried out my experiments. The lab provided not only excellent facilities and resources but also an inspiring environment for research. I am especially thankful to my colleagues and friends in the lab for their collaboration, technical help, and many valuable discussions that enriched this work. Their support and teamwork made the experimental phases both productive and enjoyable.

Finally, I would like to thank my family for their endless love and support. My wife, whose patience and understanding gave me strength through the most challenging times; my daughters, who filled my days with joy; and my parents, brothers, and sisters, who have always stood by me. I also thank my friends for their encouragement and companionship during this long journey.

Contents

Introduction and motivation of the thesis

1.1 Motivation and Problem Space

As industries shift from mass production to more tailored products, human adaptability and problem-solving capacity remain difficult for automation to fully replace. With growing product variation, collaborative automation focuses on complementing human capabilities instead of replacing them [1]. The foundation of Industry 4.0 (I4.0) is the smart factory, in which products, machines, storage systems, and data are interconnected as cyber-physical production systems. While I4.0 has improved human–machine interaction on the technical side, its social sustainability requires thoughtful recognition of the human role [2].

The forthcoming 5th Industrial Revolution, Industry 5.0 (I5.0), aims to integrate human intellect into autonomous or semi-autonomous production processes, thereby mitigating the drawbacks of I4.0 by embracing human-centricity, sustainability, and resilience [3, 4, 5]. In contrast to Operator 4.0 in I4.0 [6], the human operator in I5.0, known as Operator 5.0, is central in the production system and uses technologies to enhance production quality [4, 7]. This strategic approach emphasizes human involvement in the production processes, prioritizing operator well-being, and promoting sustainability and resilience in manufacturing systems.

As production systems become more human-centric, simply acknowledging the human role is not enough; the system must also be able to estimate the operator's current functional state and adapt to it. Human factors such as fatigue, momentary overload, or poorly designed instructions can quickly erode the benefits of collaborative or assistive technologies, even in smart factories. In other words, a system that intends to keep the human "in the center" must know how much

mental demand it is imposing and whether that demand is sustainable for the worker over time. This will lead to enhanced system quality and efficiency, and improved working environments. Numerous studies have shown that integrating this strategy provides win-win outcomes [8, 9, 10, 11].

In this context, cognitive load becomes a key construct, because it captures the amount of mental effort required to perceive information, interpret instructions, and execute tasks in dynamic environments. Industrial assistance, digital work instructions, augmented reality/mixed reality (AR/MR) guidance, or task alternation will only be effective if they respect the limits of human working memory and do not impose unnecessary processing demands. This is precisely what Cognitive Load Theory (CLT) formalizes: while long-term memory is broad, working memory is narrow, and poor task or instruction design can overload it even when the task itself is not intrinsically difficult. Hence the need to distinguish between intrinsic load (what the task must require), extraneous load (what the system adds through its design), and germane load (what supports learning and engagement) [12, 13].

A central difficulty, however, is that cognitive load is not directly observable during work. Operators rarely report overload while they are assembling, monitoring, or switching between tasks, and many industrial tasks are too short or too dynamic for repeated questionnaires. For human-centric systems to adapt task difficulty, instruction modality, or environmental conditions, the worker's state must therefore be made measurable. This is where physiological signals become valuable: they offer an objective window into ongoing arousal, workload, and adaptation during task execution. Several studies have shown that specific physiological markers react significantly when task demands increase [14]. In human-in-the-loop production, a practical subset of signals is especially relevant: electrodermal activity (EDA, also reported as galvanic skin response, GSR) as a direct index of sympathetic activation; cardiac activity from an Electrocardiogram (ECG) or Photoplethysmogram (PPG), from which heart-rate variability (HRV) can be derived to reflect autonomic balance; and wrist/hand acceleration (ACC) as a proxy for movement demands and motor effort.

Because cognitive load is multifaceted, relying on a single signal is rarely sufficient. Prior work has combined EDA, PPG/ECG, electroencephalogram (EEG), eye tracking, skin temperature, or even functional near-infrared spectroscopy (fNIRS) to capture different facets of mental effort [15, 16, 17, 18]. Classic reviews on

physiological workload measurement already stressed that no single channel can cover all dimensions of cognitive load [19, 20]. This motivates the use of a multimodal approach, such as those adopted throughout this thesis.

Physiologically, rising task difficulty acts as a stressor and triggers the autonomic nervous system (ANS), engaging both sympathetic and parasympathetic branches; this can support problem solving but, if prolonged, may also relate to emotional instability or performance decline [21, 22]. HRV features are known to be sensitive to such workload-related shifts, but their interpretation must account for recording duration, context, and inter-individual differences; combining time-, frequency-, and nonlinear features improves robustness [23, 24]. A similar logic applies to EDA: fast, phasic skin conductance responses (SCRs) reflect immediate sympathetic reactions, while the slower tonic skin conductance level (SCL) reflects sustained activation; decomposing the signal into these two components makes it easier to link observed changes to task-induced load rather than to external factors [25, 26]. Taken together, these elements are summarized in the graphical abstract in Figure 1.1.

This graphical abstract presents the main conceptual model of this thesis. At its core is the Extended Human Asset Administration Shell (HAAS), which is used to manage worker cognitive load. The left side of the figure summarizes the inputs that describe the worker and the task/context descriptors. These include worker characteristics such as physical and mental strength, manual dexterity, training and education, experience, cognitive abilities, and emotions, together with task-related and environmental descriptors such as task type, work instructions, and ambient conditions. Some of these characteristics, such as emotions or detailed aspects of experience, are not analyzed in depth in this thesis, but they are included because they will be important for future extensions of the model. In this conceptual model, these variables are combined within the HAAS to estimate the required cognitive load for a given situation and to classify it into three levels: low, medium, or high.

The right side of the figure groups the measurable outputs that are used to evaluate how demanding the current situation actually is. These include physiological metrics (EDA, HRV, and related features), kinematic data from body motion, situational awareness indicators, and performance measures such as errors and task time. Together, these data provide the evaluated cognitive load, which is again categorized into low, medium, or high. By comparing the required and evaluated

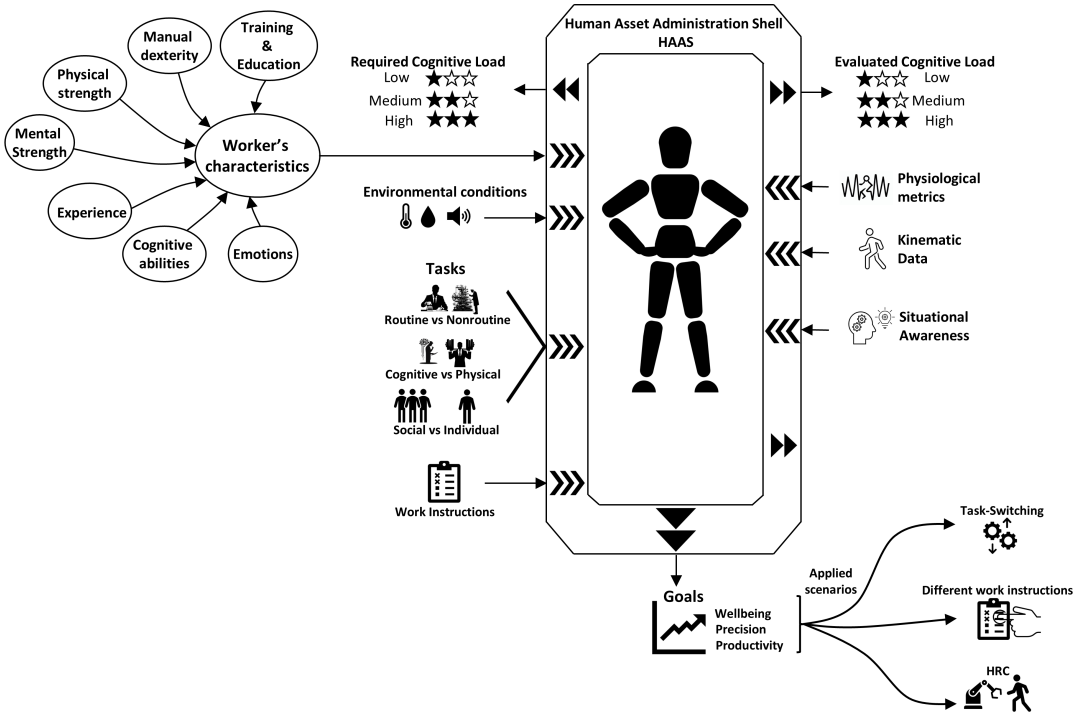


FIGURE 1.1: Overall concept of the dissertation. Worker characteristics and task/context descriptors are provided to the HAAS, which estimates the required cognitive load and compares it with the evaluated cognitive load derived from physiological and kinematic data, in order to support worker well-being, precision, and productivity.

cognitive load, the Extended HAAS is intended to support decisions that increase individual well-being, improve work precision, and maintain or enhance productivity. The subsequent parts of this thesis develop and test this conceptual model in applied scenarios, showing how these elements can be combined in practice to reach these goals.

1.2 Research Questions and Hypotheses

Human-centered industrial systems can only adapt to the operator if they (1) have a model that links tasks, worker characteristics, and context, and (2) can read, in real-time, how demanding the current situation is. The studies collected in this thesis therefore revolve around two recurring ideas: (a) physiology and performance can make cognitive load and attention observable, and (b) once the

state is observable, task design (instructions, sequencing, multitasking setup) can be optimized. This logic is reflected in the following research questions (RQs):

RQ1 (Framework / HAAS extension – Thesis I).

Can a human-centric assistance framework be extended with physiology-based modules so that estimated task demands (from task, worker, and environment) can be compared with the actual, measured load (from GSR/EDA and HRV), and used to trigger adaptations?

Hypothesis 1. An extended Human Asset Administration Shell (HAAS) that includes physiological metrics and task/context descriptors will be able to distinguish at least three meaningful load levels (low/medium/high) and indicate when the current task setting is not aligned with the worker's state.

RQ2 (Physiology for cognitive load tracking – Thesis II).

Do EDA-, HRV-, and ACC-derived features change systematically across repeated, instruction-guided assembly attempts so that they can be used to track within-session cognitive load and adaptation?

Hypothesis 2. Features derived from EDA and HRV (time- and frequency-domain features), supported by wrist acceleration, will show significant differences between baseline and task, and will be sensitive to repeated attempts within the same session.

RQ3 (Awareness & attention in dual-task situations – Thesis III.1).

Can event-based physiology (EDA/HRV, and acceleration) together with reaction-time measures explain fluctuations in situational awareness during dual-task or multitask assembly?

Hypothesis 3. Periods of slower reactions and lower awareness are expected to align with higher physiological signs of mental load.

RQ4 (Task sequencing and recovery–arousal balance – Thesis III.2).

Does interleaving brief physical work with mental work lead to a different arousal/recovery pattern–physiological and perceived–than performing the same work in blocks?

Hypothesis 4. Alternating (mental–physical–mental–physical) sequences will show HRV shifts toward parasympathetic dominance and equal or lower perceived load than continuous (mental–mental–physical–physical) sequences.

RQ5 (Instruction design and cognitive load – Thesis III.3).

How do code-based vs. visual-based work instructions affect subjective workload, performance, and physiology, and can a feature set combining biosignals and hand acceleration discriminate between these two instruction conditions?

Hypothesis 5a. Code-based instructions will produce higher subjective load (NASA-TLX / DSSQ) and fewer completed attempts than visual instructions.

Hypothesis 5b. A multimodal feature set (EDA + HRV + ACC) will classify instruction type above chance, indicating that the physiological response carries information about extraneous load.

RQ6 (Attentional multitasking in HRC – Thesis III.4).

In collaborative human–robot assembly, how does adding a parallel attention-demanding task (Go/No-Go) affect task time, assembly errors, misses in the secondary task, and perceived mental load?

Hypothesis 6. The higher-demand attentional condition will yield higher perceived mental load and a trend toward longer task times and more performance slips (assembly errors and/or Go/No-Go misses) compared with the simpler condition.

1.3 Aims and Objectives

The overall aim of this thesis is to make human-centered industrial systems capable of reading the operator's state from physiological signals and using that information to design or adapt tasks. To achieve this aim, the thesis pursues the following specific objectives:

1. **To extend the Human Asset Administration Shell (HAAS)** with modules that include task/context descriptors and physiology-based evaluation of cognitive load (EDA, HRV, ACC).

2. **To design and carry out controlled experiments** that induce different levels and types of cognitive demand in assembly-like industrial tasks (dual-tasking, task alternation, different work instructions, HRC multitasking).

3. **To compare alternative task and instruction designs** (blocked vs. alternating sequences, visual vs. code-based instructions, single- vs. dual-task HRC) in terms of subjective workload, performance, and physiological response.

4. **To extract and validate multimodal features** from the recorded biosignals and kinematic data that can later be used in adaptive or assistive human-in-the-loop systems.

1.4 Contributions of the Thesis

This thesis contributes to the development of human-centered and adaptive industrial systems by linking physiological sensing with task and instruction design. The main scientific and methodological contributions are summarized below, following the order of the research questions, parts, and chapters:

- **Thesis I:** Proposes an extended Human Asset Administration Shell (HAAS) that incorporates physiological and contextual data to estimate and compare required versus evaluated cognitive load, forming a conceptual basis for adaptive human-centric manufacturing.

- **Thesis II:** Designs a methodological framework for tracking cognitive load within industrial-like assembly sessions using multimodal physiological and kinematic signals (EDA, HRV, ACC).

- **Chapter III.1:** Demonstrates that event-based physiological and performance measures can reveal changes in attention and situational awareness during dual-task assembly scenarios.

- **Chapter III.2:** Provides experimental evidence that task sequencing (alternating vs. continuous mental–physical work) influences the balance between recovery and arousal, linking physiological and perceived load dynamics.

- **Chapter III.3:** Evaluates the impact of instruction design (code-based vs. visual-based) on workload, performance, and physiological response, showing how multimodal features can characterize extraneous cognitive load.
- **Chapter III.4:** Explores attentional multitasking in human–robot collaboration (HRC), identifying its effects on workload, performance, and attentional strategies in collaborative assembly contexts.

Collectively, these studies bridge industrial engineering and physiological sensing, providing conceptual, methodological, and empirical foundations for designing human-centric, cognitively adaptive production systems.

1.5 Dissertation Structure

The studies presented in this Dissertation are organized into three main theses: Thesis I, Thesis II, and Thesis III. The Dissertation begins with Chapter 1, which introduces the motivation, research questions, and overall concept of physiology-informed, human-centered production.

Thesis I presents the extended HAAS framework that links task, worker, and context descriptors with physiology-based load estimation. Thesis II describes the experimental setup and signal-processing pipeline used to track cognitive load during repeated, instruction-guided assembly attempts.

Thesis III includes four chapters: Chapter III.1 reports the dual-task and situational-awareness study based on event-related physiological measures; Chapter III.2 analyzes task sequencing and the balance between arousal and recovery when mental and physical work are alternated; Chapter III.3 examines how different work-instruction formats influence subjective load, performance, and physiological responses; and Chapter III.4 explores attentional multitasking in collaborative human–robot assembly.

Finally, the conclusion summarizes and concludes the main findings and provides directions for future work.

Thesis I

Modeling Worker, Task, and Context in an Extended Human AAS

I.1 Human-Asset Administration Shell for Cognitive Load Management

Implementing the recent approaches of Industries 4.0 and 5.0 (I4.0 and I5.0) that incorporate human-in-the-loop control systems showed that the cognitive load on operators in work environments has significantly increased, primarily due to the increased volume of data requiring advanced mental processing [27, 28]. Cognitive load is described as a multi-dimensional structure expressing the burden that a given task exerts on the worker. It also indicates the perceived effort required for learning, reasoning, and thinking as a measure of working memory pressure during the execution of the task [29]. The cognitive load brought on by an abundance of complicated knowledge has developed into a potential problem. Despite this, structured knowledge systems are still extensively employed, irrespective of the fact that individuals have varying rates of information intake [30].

For more efficient management of mental workload during crucial decision-making, there is an immediate need to design a smart data system that can adjust to the information-processing capacity of each individual [30]. The goal that researchers are looking for is to decrease the cognitive load, which will be reflected in production efficiency. A numerical simulation study suggests that the adoption of I4.0 technologies alleviates this load by decreasing the amount of information an operator needs to manage for a task, subsequently lowering cognitive effort. This increased processing capacity enables operators to handle more complex tasks involving multiple actions [28].

Technologies in I4.0 and I5.0 enable the creation of digital representations of industrial entities, supporting production systems with considerable advantages and capabilities [31]. To achieve resilient, sustainable manufacturing systems, researchers have started making digital twin models (DT) that represent the physical assets in the virtual world [32]. Asset Administration Shell (AAS) is the only DT definition that explicitly supports industry-standard protocols and data formats, according to Michael J. et al. [33]. AAS is an I4.0 architecture that specifies the technical characteristics of an asset. It was designed to convey information as well as data in an organized way, hence facilitating interoperability between DTs models [34].

Humans are increasingly being digitalized in the cyber field through the principle of human centricity. However, most studies in this field, according to Du et al. [30], focus on the system level in modeling information processing rather than modeling behaviors at the personal level. As a result, the Human Digital Twin (HDT) was proposed to integrate human workers in the I4.0 field, which supported data collection, scheduling, communication handling, and so on [35, 36]. HDT is the cyber-phase of the human entity, which is fed by dynamic real-time parameters to represent the human in the physical phase. These parameters include but are not limited to workers' characteristics, behaviors [37], geospatial and psychophysical conditions, contextual parameters, intentions, cognitions, emotional state, food intake [38], motion, and other physiological parameters such as electromyogram (EMG), heart rate, heart rhythm, respiration, blood pressure, Galvanic Skin Response (GSR) [35, 36, 39, 40]. Despite the rising number of papers that talk about HDTs and the possible influence they might have in the future, there is no existing agreement on exactly how to design these kinds of systems [41].

This thesis focuses on two key challenges in the transition to industrial digitization and proposes ways to address them. First, *what level of cognitive load does a given task place on an operator, and what is the cognitive limit beyond which performance declines, taking into account individual skills?* Second, based on this understanding, *is it necessary to regulate the task's cognitive load, and if so, how can this be achieved?* In this part, an Extended Human-AAS (HAAS) model has been proposed that incorporates the human worker as a central element of the assessment framework.

I.2 Human-in-the-Loop

Human-in-the-Loop places the human operator at the center of I4.0/5.0 systems. Rather than replacing individuals, digital tools and automation collaborate with them, so understanding how operators receive information and make decisions is essential. This section prepares the ground for the Extended HAAS by covering three elements: (1) the theory of cognitive load and its three dimensions, explaining how information presentation affects performance; (2) real-time assessment methods using physiological signals (GSR and HRV) within a Cognitive Work Analysis (CWA) perspective; and (3) a three-dimensional task classification (routine/nonroutine, cognitive/physical, individual/social) to map work demands

for digitization. Together, these elements enable human-aware monitoring and control, allowing HAAS to align required demands with the operator's state and adapt tasks and conditions accordingly.

I.2.1 Cognitive Load and Its Three Dimensions

It is highly personal how individuals receive the information presented to them; some may struggle with processing visual–spatial information, while others may resist verbal instructions [30]. Moreover, the same person's attitude toward information intake can change significantly depending on their cognitive state, for example, showing a stronger preference for visual–spatial content during periods of emotional disturbance [42, 30]. This implies that approaches for reducing the cognitive load caused by information intake should be tailored both to the individual and to the situation [30]. Based on this principle, three approaches under what is known as the "Cognitive Load Theory" (CLT), intrinsic cognitive load, extraneous cognitive load, and germane cognitive load, have been introduced to address two central questions: "What information should be presented to the worker?" and "How should this information be presented in order to reduce cognitive load?" The CLT further assumes that the capacity of working memory is limited, whereas the capacity of long-term memory is considered unlimited [13].

The first approach, intrinsic cognitive load, deals with the level of sophistication of new information being processed [12]. Strengthening a worker's prior knowledge, represented in long-term memory, can reduce this load. Alternatively, the task difficulty itself may be adjusted; for instance, sequential processing places less demand on working memory than simultaneous processing [13]. This dimension of CLT is therefore referred to as intrinsic cognitive load [13, 12].

The second approach, extraneous cognitive load, arises from the way instructions are presented and from the design of the system itself. Therefore, any factor that distracts workers from their primary goals should be avoided. Since this type of load can be directly influenced by trainers and system designers, it can often be minimized or even eliminated by improving the format in which strategies or instructions are delivered. Extraneous cognitive load should always be reduced, and there are no circumstances in which increasing it is beneficial [12, 13].

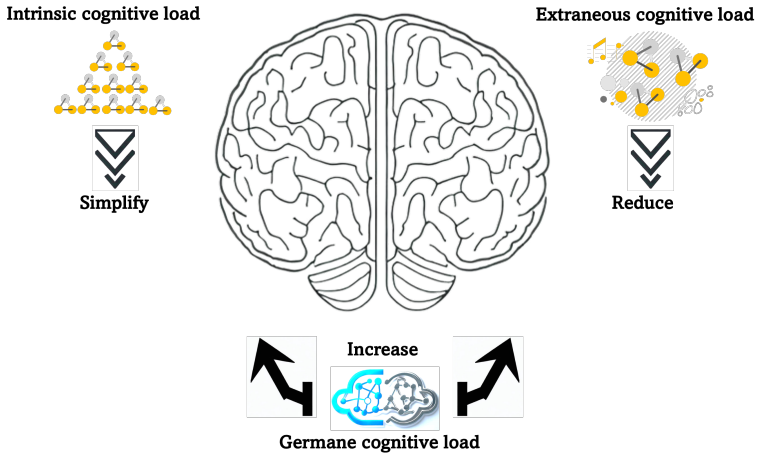


FIGURE I.1: Cognitive Load Theory (CLT) with its three components: intrinsic, extraneous, and germane cognitive load.

The final approach, germane cognitive load, differs from the previous approaches in its positive effects on workers by supporting schema processing and construction. It emphasizes the cognitive processes of learners or workers, encouraging them to invest effort in learning and thereby facilitating knowledge acquisition [12, 13]. Figure I.1 illustrates the CLT framework, showing the three components positioned around the brain to symbolize their interaction. On the left, intrinsic cognitive load is represented with a downward-pointing arrow, highlighting the need for task simplification to reduce this load. On the right, extraneous cognitive load is also represented with a downward arrow, stressing the importance of decreasing it at all times. At the bottom, germane cognitive load is shown with upward arrows, underscoring the value of increasing this element to counterbalance the other two.

I.2.2 Monitoring Cognitive Load in Real-Time

To monitor and measure cognitive load, it is crucial to carry out these processes in real time, as this allows feedback that can be used to reduce cognitive load, thereby improving workers' performance and supporting their decision-making. Cognitive load can be assessed subjectively using psychological self-assessments, such as the NASA Task Load Index and the Instantaneous Self-Assessment of Workload. These instruments are valuable because they capture perceived workload and related experiences, but they typically provide feedback after the task has been completed, which limits their use for continuous, in-task adaptation.

Cognitive load can also be assessed objectively and continuously through physiological measurements, including galvanic skin response (GSR), heart rate variability (HRV), electroencephalography (EEG), and blood pressure [30].

While these signals support real-time monitoring, cognitive load is a multifaceted construct, and physiological changes are not uniquely attributable to workload alone. For this reason, subjective self-assessments remain important for interpreting and validating physiological patterns, helping to determine whether observed signal changes correspond to perceived workload rather than other concurrent influences. Nevertheless, while real-time monitoring is vital for understanding and managing the demands placed on workers, it is equally important to consider how the work environment and system design can be structured to accommodate these demands. This is where Cognitive Work Analysis (CWA) becomes essential.

CWA is a framework for analyzing and designing work systems in which human operators interact with complex technological environments [43]. It is founded on the idea that effective system design requires an understanding of the cognitive processes humans use to perform their tasks [43, 44]. CWA provides a set of principles and methods to optimize the fit between system demands and human capabilities [43]. It focuses on how individuals perceive, think, and act in their work context, and how the design of the work environment can either support or hinder these processes [30]. In recent studies, wearable sensors have been applied within the CWA framework to capture data on both the physical and cognitive demands of tasks [45]

As discussed above, real-time monitoring and CWA are crucial for understanding and managing cognitive load in complex work systems. However, the increasing integration of advanced technologies into the workplace introduces new challenges. While these technologies aim to enhance workers' capabilities, they can also elevate stress and workload [46]. For instance, Dixon et al. (2003) observed that the use of diagnostic automation with an accuracy below 80 percent in unmanned aircraft operations actually increased stress and workload compared to performing the tasks without automation [47, 46]. With the rapid developments of I4.0 and Industry 5.0, and the growing presence of the human-in-the-loop, it has therefore become essential to explore technical methods that can estimate workers' workload in real time and support their capabilities by tracking their intentions.

This thesis proposes the use of two physiological signals, GSR and HRV, within a conceptual model (Extended HAAS) to monitor cognitive load in real time. Recent research has shown that GSR can be a reliable physiological indicator of cognitive load [15]. GSR, also referred to as electrodermal activity (EDA), denotes electrical activity of the skin [48]. It reflects changes in eccrine sweat-gland activity, which is directly controlled by the autonomic nervous system (ANS) [49]. The intensity of emotional arousal varies with the context, for instance, threatening, pleasant, or otherwise affective stimuli, and these events alter sweating, which in turn changes skin conductance. However, regardless of the stimulus type, skin conductance primarily indexes the intensity (not the kind) of emotion [50]. GSR is commonly characterized by two components: (1) the skin conductance level (SCL), representing the tonic baseline, and (2) the skin conductance response (SCR), representing phasic, stimulus-related changes [51, 48]. Complementing GSR, HRV quantifies the variability in the intervals between successive heartbeats. Mental effort is typically associated with its features like power spectrum, heart rate, etc. Thus, HRV provides a complementary, sensitive marker of cognitive load alongside GSR [52].

I.2.3 Task Classification Framework for HAAS Integration

Since the commencement of the I4.0 movement, the industrial shop floor has been undergoing a transformation driven by smart and digital technologies. However, to achieve an efficient smart and digital representation of the diverse assets in the proposed HAAS model, it is crucial to clearly understand the human operators' tasks and classify them according to specific characteristics that effectively support their digitization. Based on these criteria, a three-dimensional task classification was suggested by Cimini et al. [53] as follows:

- **Routine and Nonroutine tasks**: Routine tasks are activities that can be performed based on preprogrammed rules. In manufacturing, they are typically repetitive, such as loading, unloading, assembling, and packaging. These tasks are easily codable because their steps are already well defined. By contrast, nonroutine tasks are less clearly defined and therefore difficult to codify. They often involve abstract activities, such as management, technical, or creative roles, which require problem-solving skills, anticipation, and analytical abilities.

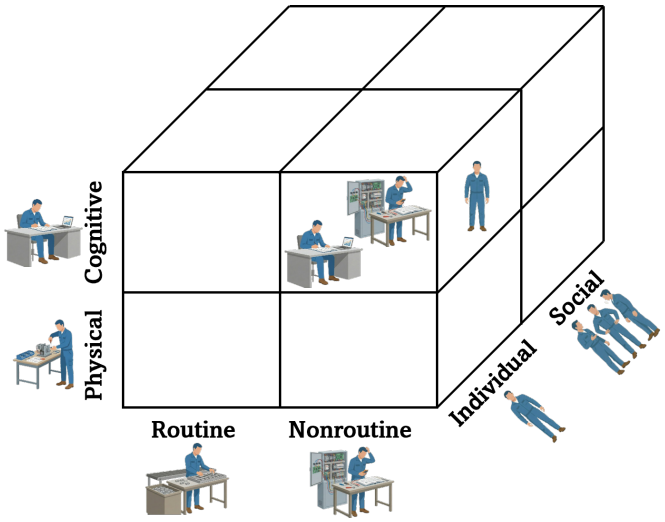


FIGURE I.2: Three-dimensional task classification model (adapted from Cimini et al., 2023) [53]

- **Cognitive and Physical tasks**: Cognitive tasks involve mental activities such as reasoning, decision-making, and problem-solving. In contrast, physical or manual tasks consist of a sequence of actions and rely on sensory and motor skills.

- **Individual and Social Tasks**: The level of interaction involved in an activity determines whether it is classified as an individual or a social task. Individual tasks can be performed independently by a single person and are characterized by little or no social interaction. In contrast, social tasks require a higher degree of interaction and coordination with others. According to Frey and Osborne (2017), tasks with lower levels of social interaction are generally easier to digitize [54].

Based on this three-dimensional classification, Figure I.2 illustrates the general model that can be used to map any type of task. A given task may represent a combination of these dimensions. For example, assembling tasks can be categorized as routine/physical, whereas maintenance tasks fall under nonroutine/physical. Likewise, cognitive tasks such as data collection may be either routine or nonroutine, depending on the characteristics of the duty. At the same time, the third dimension of the task may take an individualistic or sociocentric form, depending on the presence of collaborative elements such as operators, robots, or other assets.

I.3 Extended HAAS

Building an accurate HDT requires collecting data about diverse assets, such as machines, tasks, workers' data like acceleration, physiological and psychological data, workers' awareness and cognitive load, and the necessary skills. HAAS will collect these parameters and provide the interconnecting processes between these different assets to ensure efficient interoperability both within assets and with external systems. To develop an effective HAAS, it is necessary to identify a wide range of standards and classifications that describe assets and tasks.

I.3.1 Definition of HAAS

I4.0 uses information and communication technologies to connect actors in industrial processes within an intelligent network. The AAS plays a central role in this process, as it supports the implementation of I4.0 digital twins and enables communication interoperability between solutions from different vendors [55, 56]. In essence, the AAS is a digital representation of an asset [57]. In the context of I4.0, the AAS of machines and certain software components was well defined from the beginning, but the inclusion of humans as manufacturing entities was not nearly as straightforward. The worker can be represented in two major groups of AAS/digitized data: psychophysical data (like heart rate, galvanic skin response, electromyogram, etc.) and static data (such as height, weight, smoking habits). A key requirement of the AAS is to provide a minimal but sufficient description of the asset according to its use cases. At the same time, it is also expected that existing standards can be mapped to the definition of an AAS.

I.3.2 HAAS structure

Existing research shows a clear tendency to integrate humans into the production environment through the HDT approach, where the Human-AAS (HAAS) is considered an extension of the AAS concept. Current findings indicate, however, that this process is not yet well defined and that several questions remain open. For example: How should a human sub-module be designed? and, Should a generic AAS include a HAAS, or should these two entities be separated? Nonetheless, some works demonstrate that such integration is realistically achievable. Among

them, the thesis of Niko Bonomi [58] and Sparrow Dale Eric [59] highlight that humans can be integrated into AAS in a standardized way, provided additional standard components are involved.

The AAS consists of two main components: the header and the body, as shown in Figure I.3. The header contains basic information about the asset, such as identification, while the body manages the different submodels within the AAS. A single submodel is a hierarchical structure of properties, referring to the information and functions associated with an asset in a given domain. This approach enables the collection of standard and fixed information, such as a barcode or serial number. At the same time, it is possible to collect dynamic data, such as the temperature of a probe or the current value of a pressure sensor. In addition, the submodel includes functional properties that allow a program or routine to be started and stopped directly on the asset.

The requirements specify the structure, parameters, and properties of the AAS [60]. Meeting these requirements ensures the design of fully functional and I4.0-compatible AAS. In total, there are 22 requirements, which can be divided into three major groups [58]:

- General requirements.
- Requirements regarding the Administration Shell.
- Requirements regarding identifiers.

Before addressing the HAAS, it is important to establish the nature of the data produced in HDT and the specific standards that govern it. International standards are essential for achieving a high level of interoperability between different systems. The AAS standardizes the representation of data and the ways in which such data can be related to others. In addition, it allows each piece of data to be expressed within a wide variety of sub-structures. Importantly, the AAS does not prescribe the exact types of data that should be published or the manner of their publication; instead, this responsibility lies with the implementer. In practice, each submodel or property may refer to a specific standard that defines its details.

While relatively well-defined standards exist in this area for machinery (e.g., ecl@ss [61]), references to human-related standards remain limited in industrial production. Taxonomies of human capabilities, however, include for example:

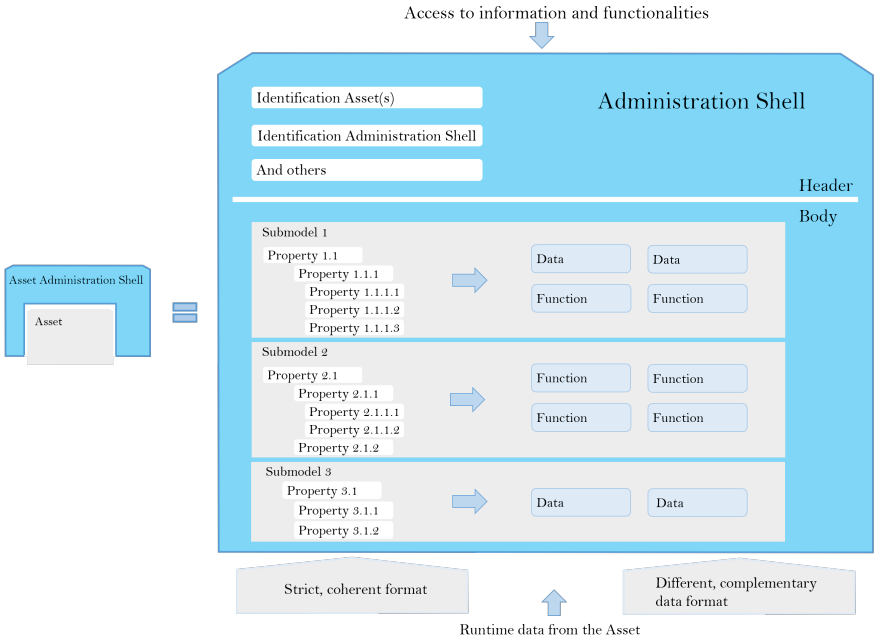


FIGURE I.3: General structure of an AAS

- O*NET: The Occupational Information Network [62], a database containing hundreds of job definitions, sponsored by the U.S. Department of Labor/Employment and Training Administration (USDOL/ETA).
- A.S. Chuilef et al. proposed a hierarchical taxonomy for human goals [63].
- ESCO: The European Skills, Competences, Qualifications, and Occupations classification system [64].
- Xiao et al. developed a taxonomy in the field of human public health [65].
- P.A. David proposed a taxonomy structure related to human classification [66].

For simplicity and clarity, P.A. David's grouping is used as an illustrative example (see Figure I.4) [66]. According to this classification, human capital can be divided into two main categories: tangible assets and intangible assets. Tangible assets include, for example, health or physiological conditions, while intangible assets refer to cognitive capacity or problem-solving ability. Building on the insights of Bettoni et al. [67], a new layer of abstraction is introduced in the human data, and each element can be categorized into either tangible or intangible groups. Based on this, the characteristics, parameters, and conditions of the worker can be described as follows:

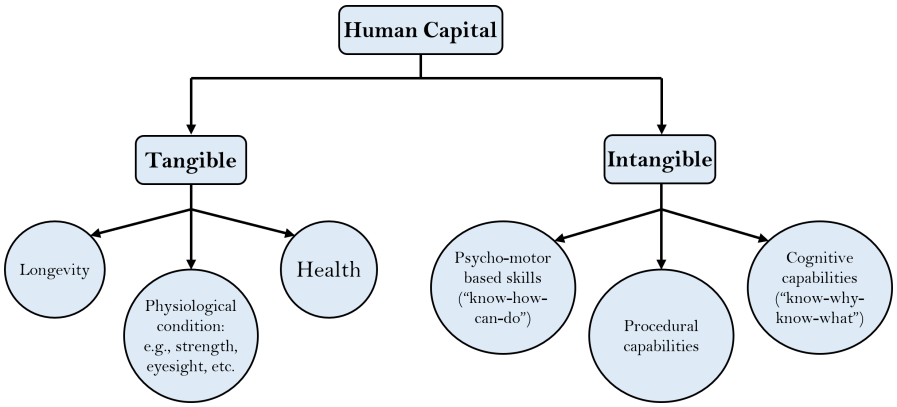


FIGURE I.4: Human capital classification

- **Characteristic:** An intrinsic or extrinsic quality of the worker, such as height or level of experience. In general, these characteristics are quasi-static, meaning they change very rarely or not at all.
- **Parameter:** A value that can change several times during the day. Such values are typically collected automatically through data acquisition techniques and continuously monitored. Examples include GSR, HRV, or a worker's position within a room.
- **Condition:** The worker's actual state, which may be intangible, such as current emotion, level of exertion, or a health-related issue (e.g., irritability or disability). In many cases, conditions can only be inferred indirectly.

Building on these categorizations, several research efforts have attempted to integrate human-related data into the AAS framework. As Marcon et al. [68] have shown, ideally, each production component in the I4.0 approach should have its own AAS, and the human is no exception. In their study, the operator wore a smart jacket equipped with sensors to measure parameters of both the wearer (like temperature) and the environment. The concept relied on multiple technologies, including human–machine interfaces (HMIs) and industry-standard communication protocols. While this represents a promising and forward-looking concept, it still lacks a clear formalization of the human role within the AAS.

Similarly, the research of Al Assadi et al. [69] introduced a Human Administration Shell (HAS) that uses smart devices (e.g., smartwatches, smartphones) to collect and provide information. Their approach distinguishes between two categories: condition monitoring (real-time data such as heart rate, location, and

access data) and service provision (e.g., personal skills and knowledge). This division aligns with the grouping proposed by A. David [66], in which human capital was separated into tangible (condition monitoring) and intangible (service provision). The concept proved valuable in several practical applications, such as automatic adjustment of ergonomic workstations, authentication, and adaptive HMIs. However, while this approach considers the Human AAS, it still lacks some necessary components to treat humans as a fully integrated entity in the I4.0 environment. In its current form, it can be regarded more as a separate entity than as a fully embedded element of the production system.

While previous studies have demonstrated different ways to integrate human data into the AAS framework, they also highlight limitations in fully embedding the human as an active component of I4.0 systems. To address this gap, Sparrow (2021) [59] outlined the key responsibilities that a HAAS should ideally fulfill, which can be summarized as follows:

- **Delegation representing the human:** Human operators can respond to commands or instructions from surrounding assets through various interfaces such as touchscreens or keyboards. However, this interaction often creates bottlenecks in the communication process. To avoid this, the HAAS can act as a representative of the human operator by monitoring and recording activities, behaviors, and working schedules. This allows humans to respond to more detailed inquiries about their tasks and capabilities without being hindered by interface limitations.
- **Facilitating human interfacing:** The HAAS requires bi-directional communication with human workers to ensure continuous data exchange. This enables the collection of essential information such as body posture, movements, eye-tracking, and physiological signals.
- **Enhancing digital processing and information management:** Human information processing relies on abstraction, pattern matching, heuristics, and creativity. These processes are relatively slow, on the order of 200 milliseconds, and become even slower during complex decision-making. By contrast, digital assets in the HAAS can process and transfer data with high precision, transmit events statistically, and operate within milliseconds.

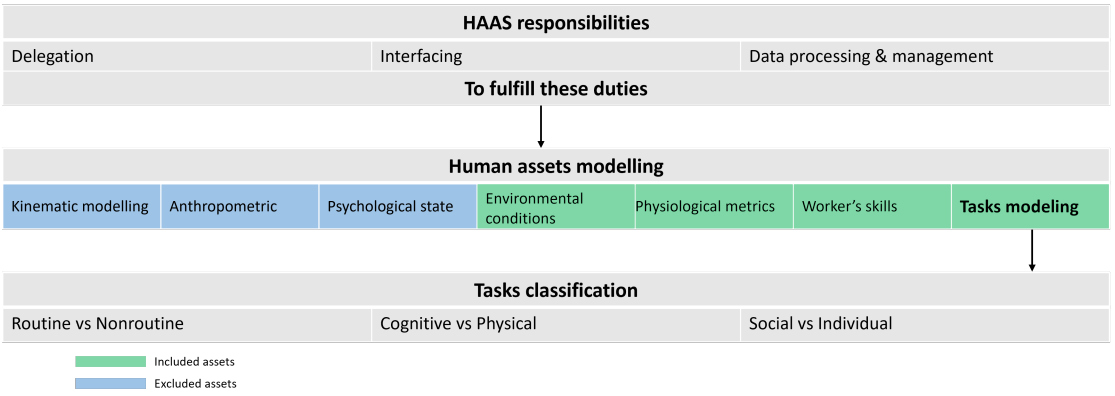


FIGURE I.5: Responsibilities and requirements of the extended HAAS

To provide a clear overview of the requirements and responsibilities of the HAAS, Figure I.5 presents them in a tabular format. The figure distinguishes between the modules that will be modeled for the construction of the extended HAAS and those that are excluded because their scope lies beyond the focus of this study.

I.4 Cognitive Load Management by the Extended HAAS

The extended HAAS model proposed in this thesis is built around four fundamental modules that form the basis of its structure and operation. These modules include physiological measures, workers' characteristics, task type and level, and environmental conditions. The physiological module in this thesis focuses on the GSR and HRV data, which are acquired using sensors placed on the individuals performing the tasks. This module records these metrics and then evaluates and categorizes the workers' cognitive load into three levels: low, medium, and high, referred to as the *"Evaluated Cognitive Load"*.

The next module is the workers' characteristics, which is designed to capture the distinct skills and proficiencies of each worker and acknowledges that individual capabilities can vary widely. This module is updated based on a preliminary assessment of each worker's abilities prior to performing specific tasks. The task type and level module considers not only the categorical nature of a task but also integrates the worker's inherent skills and capacities into its analysis. The final module is the environmental condition, which recognizes that external factors play

a crucial role in determining cognitive load. This module continuously monitors and adjusts for environmental changes, such as noise or temperature. Together, these three modules estimate the required cognitive load based on their inputs and classify it into three levels: low, medium, and high, referred to as the *"Required Cognitive Load"*.

With these four modules in place, the extended HAAS model establishes a dynamic interplay. By comparing the evaluated and required cognitive loads, it adjusts tasks and environmental conditions to maintain a balance between workers' wellbeing and their productivity.

Given the comprehensive nature of this study, which incorporates four modules from diverse disciplines, it was necessary to integrate these modules into a management system to enhance flexibility and improve decision-making. For this purpose, the proposed HAAS was integrated with the OODA Loop framework. The OODA Loop, introduced by military strategist John Boyd, is a decision-making process consisting of four sequential steps: Observe, Orient, Decide, and Act [70]. The aim of this integration is to maximize productivity, ensure safety, and promote overall well-being.

For simplification, the proposed model was divided into five methodical phases within the OODA loop, as shown in Figure I.6. The first step, Observe, includes Phase 1, which establishes the cognitive load thresholds, and Phase 2, which handles operational monitoring and cognitive load assessment. The second step, Orient, corresponds to Phase 3, which provides contextual information through the other modules. The third step, Decide, is represented by Phase 4, which evaluates and compares cognitive loads. Finally, the Act step is reflected in Phase 5, which performs adjustments and loops the process forward.

Each phase serves a distinct purpose and systematically builds on the knowledge and results of the preceding phase. At the core of this concept is a responsive feedback loop that continuously monitors, assesses, and adapts using real-time data, with the primary objective of regulating and managing cognitive load. This dynamic approach supports the creation of optimal working conditions, enhancing performance and reducing worker fatigue. The subsequent points present a detailed analysis of the steps:

- **OODA Step 1: Observe**

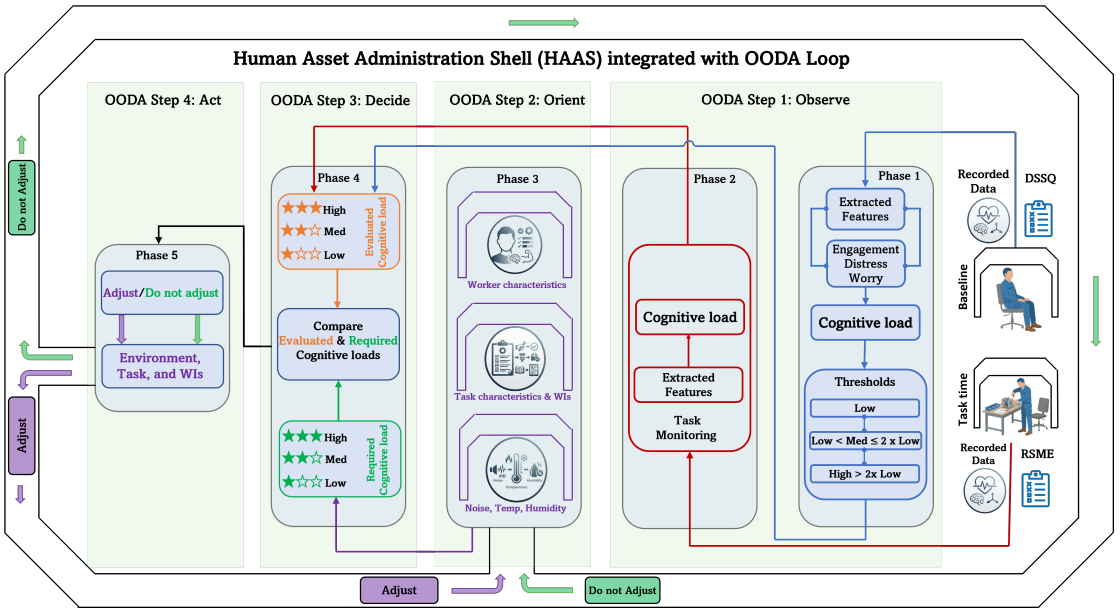


FIGURE I.6: Extended HAAS framework integrated with the OODA Loop for cognitive load management.

This step includes two phases for observing both baseline and operational conditions through physiological characteristics, as follows:

Phase 1 – Establishing the Cognitive Load Thresholds

1. Neutral Condition Monitoring: The objective of this phase is to consistently observe and analyze the worker's physiological characteristics (HRV and GSR) in a neutral situation free from tasks, noise, or distractions. In the same neutral condition, a brief self-assessment is collected using the short Dundee Stress State Questionnaire (Short DSSQ) to obtain the worker's initial profile of engagement, distress, and worry. These signals must be processed to derive cognitive load, which is considered an intangible factor.

2. Low Cognitive Load Threshold: The cognitive load calculated in this phase represents a low level. This value is taken as the threshold for low cognitive load. Any measured cognitive load that falls below or equals this threshold indicates that the worker is operating under low cognitive load. In addition, the short DSSQ profile serves as a subjective anchor for this calibration, so the low threshold is not based on physiology alone but is interpreted together with low distress and low worry and an appropriate level of engagement in the neutral condition.

3. Medium and High Cognitive Load Thresholds: The thresholds for medium and high cognitive load are determined relative to the low threshold established in **Phase 1**/2. The medium threshold is defined as twice the value of the low threshold, and any measurement above this medium threshold is classified as a high cognitive load. The short DSSQ is used as an anchor during this calibration to support the interpretation of borderline cases and reduce the risk of treating engagement-related arousal as overload, by considering whether increased physiological activation aligns with distress and worry rather than engagement alone.

Phase 2 – Operational Monitoring and Cognitive Load Assessment

1. Operational Condition Monitoring: The worker begins executing various tasks under potential auditory disturbances and interruptions. During this period, it is essential to consistently record the worker's physiological metrics (HRV and GSR).
2. Cognitive Load Calculation: The acquired physiological data are processed continuously to compute the cognitive load experienced during task execution under these conditions.

- **OODA Step 2: Orient**

This step of the loop is not only about collecting data, as in the observation step, but also about interpreting it in context with respect to workers' characteristics, task difficulty, and environmental conditions in order to understand the cognitive load. This step is represented by **Phase 3** as follows:

Phase 3 – Other Modules for Contextual Information

1. Worker's Characteristics Module: This module includes intrinsic qualities and talents originating from the worker, as well as extrinsic factors and influences coming from outside the worker.

   Some intrinsic qualities:

   - Physical Strength and Stamina: Many industrial tasks require significant physical effort, which demands sufficient strength and stamina.

- Manual Dexterity: Precise motor skills are essential for tasks such as manipulating tools or assembling small parts.
- Mental Strength: Includes resilience, determination, problem-solving skills, critical thinking, and the ability to learn and adapt to new circumstances.
- Cognitive Abilities: The ability to understand instructions, follow procedures, make quick decisions, and maintain attention to detail.

Some extrinsic qualities:

- Training and Education: Skills and expertise gained through systematic training and academic studies, covering both practical skills and theoretical understanding of equipment or processes.
- Work Experience: A worker's performance is strongly influenced by prior experience in similar roles or sectors [71].

2. Task Characteristics and Work Instructions (WIs) Module: This module is defined based on measurable parameters and characteristics of the task itself and its work-instruction design, which together contribute to estimating the Required Cognitive Load. Some examples include:

   - Complexity: Number of steps involved, degree of detail, or level of accuracy required.
   - Time Pressure: The time available to complete the task, including pacing control that is self-paced or system-paced and any enforced timing.
   - Familiarity: How common or unusual the task is for the worker.
   - Required Skill Level: Specific qualifications or skills needed to perform the task.
   - Physical Demands: The level of strength or stamina required.
   - Cognitive Demands: The level of concentration, problem-solving, or decision-making needed.
   - Step Granularity: Number of actions per step.
   - Information Density: Elements per screen or page and the number of symbols and constraints per step.
   - Visual Guidance and Signaling: Highlights, cues, warnings, and the clarity of the next action.

- Feedback and Verification: Confirmations, checklists, error messages, and recovery guidance.

3. Environment Module: Continuous observation and assessment of environmental factors, such as temperature, noise levels, humidity, and other relevant conditions.

- **OODA Step 3: Decide**

This step determines the next action based on **Phase 4**. It will decide whether to adjust or not adjust the modules in **Phase 3**.

Phase 4 – Evaluation and Comparison

1. Cognitive Load Assessment: Use the data from the cognitive load calculation (**Phase 2**/2) to evaluate the worker's cognitive load in the context of the environment, referencing the threshold values (**Phase 1**/2, 3), to classify it into one of three levels: low, medium, or high. In addition, after a preset task specific time window, the worker provides a brief self-perception rating using the RSME (Rating Scale Mental Effort). The RSME is a unidimensional subjective scale that captures the perceived mental effort required by the task at that moment [72]. It is not completed during the first round because it relies on the worker having experienced the task sufficiently to judge its effort demand.

   The RSME score is then used as an additional anchor for the evaluated cognitive load, complementing the DSSQ profile established in **Phase 1** by helping interpret whether the current state reflects task engagement with manageable effort, distress with high effort, or worry with elevated effort under uncertainty. This intermittent self-rating is inserted at predefined moments while continuous physiological monitoring continues, so it strengthens the cognitive load estimate without interrupting the principle of continuous monitoring.

2. Required Cognitive Load Assessment: Use the data from the Worker's Characteristics module, the task Characteristics and Work Instructions module, and the Environment module (**Phase 3**) to evaluate the expected cognitive load for a specific task, given the worker's skill level and environmental conditions.

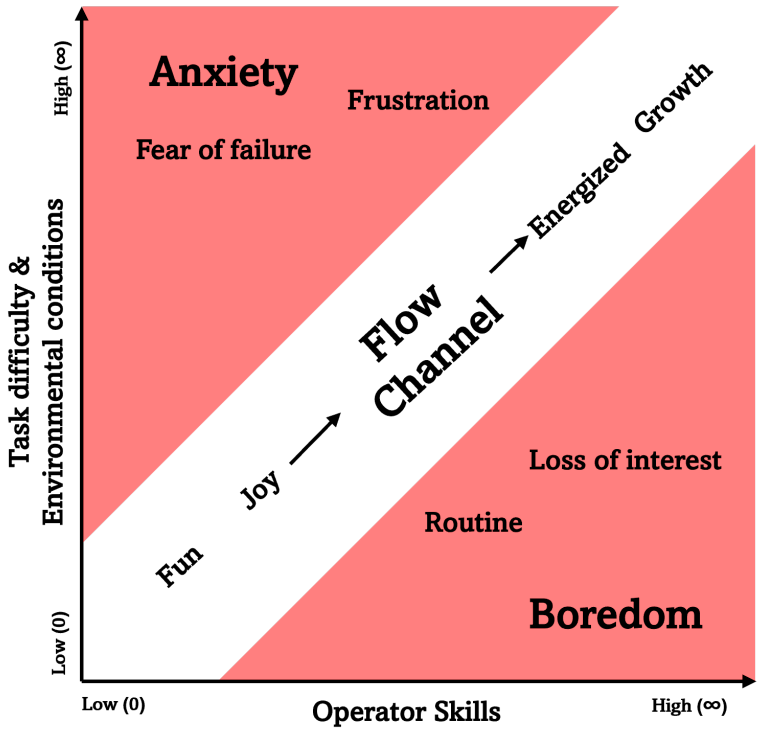


FIGURE I.7: Flow channel between the anxiety and boredom zones in the industrial fields

3. Comparison and Adjustment: Compare the cognitive load evaluated in (**Phase 4**/1) with the required cognitive load assessed in (**Phase 4**/2). The objective of this model is to optimize cognitive load for maximum productivity while maintaining a comfortable work environment, since low cognitive load is not always desirable if it indicates under-challenge and reduced engagement. To guide this process, Csikszentmihalyi's criteria were adopted, as illustrated in Figure I.7 [73], which define two zones, Anxiety and Boredom, with a "flow channel" between them. The aim is to keep the worker operating within this flow channel. When skill level is low and task or environmental demands are high, the worker experiences anxiety. Conversely, if the task is too simple and the skill level is high, boredom occurs. Based on this comparison, adjustments should be made to either the environment or the task characteristics to keep the worker within the flow channel, thereby maximizing both productivity and comfort.

- **OODA Step 4: Act**

This step involves implementing changes based on the decision made in the previous step of the OODA loop. The action may either involve monitoring the

system without alteration or adjusting the surroundings or task difficulty. **Phase 5** corresponds to this step, as follows:

Phase 5 – Looping for Continuous Monitoring and Adjustment

- Do Not Adjust: If the worker is within the flow channel, no changes to the environment or task are needed. The action is limited to continuously monitoring the cognitive load.
- Adjust the Surroundings: To optimize the working environment, environmental controls can be applied to reduce noise levels and remove potential distractions.
- Adjust the Task: If environmental modifications alone are not effective, the difficulty of the task may need to be adjusted. This could include altering the nature of the activity or providing supplementary aids, such as automated tools or additional instructions, to help the worker perform the task effectively under high cognitive load.
- Continuously Loop: Return to (**Phase 2**, Operational Monitoring Phase) and repeat the cycle. This ensures continuous monitoring and adjustments based on real-time data.

I.5 Extended HAAS Implementation

This section outlines the implementation procedures for the assets associated with each module in the extended HAAS. The diagram shown in Figure I.8 illustrates how the three assets are interconnected through I4.0-compliant communication to form the framework of the proposed HAAS. The primary objective of this conceptual framework is to manage workers' cognitive load effectively.

Given the clear relationship between physiological signals (GSR and HRV) and worker characteristics, an AAS was created that includes two submodels to capture and analyze these data. The detailed table describing the HAAS implementation is placed at the end of the thesis, in the Appendices. Table A.1 presents the header of the primary components of the first AAS. This table contains essential information about the asset, such as its title, unique identifier, and the submodels it contains.

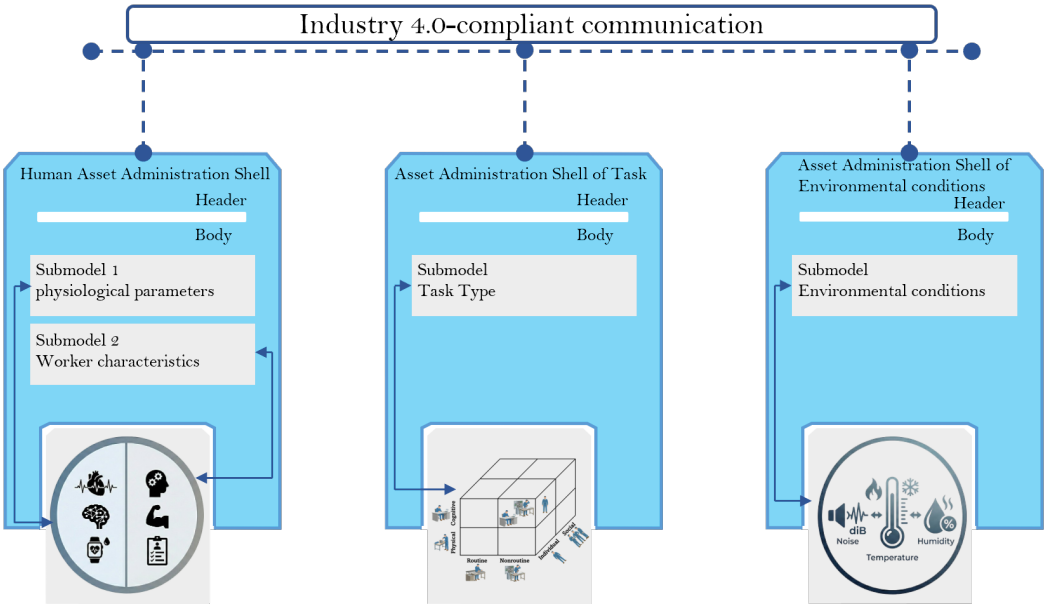


FIGURE I.8: Diagram of interconnecting the assets together through an I4.0-compliant connection to form the developed HAAS

The details of the two submodels, the worker physiological parameters submodel and the worker characteristics submodel, are provided in Tables A.2 and A.3, respectively. The first submodel covers physiological features derived from GSR and HRV signals. Each feature has a distinct guide code for later processing within the AAS. The second submodel focuses on worker characteristics, identifying key attributes that distinguish each individual, such as gender, age, and educational background.

Tables A.4 and A.5 present the header and body of the second AAS, which describes task types and characteristics in detail. This structure enables the definition of specific activities using the Task Classification Model shown in Figure I.2. In addition, a set of requirements was established for each task, including factors such as time constraints, physical and cognitive demands, required skills, and other relevant considerations.

The third asset covers the environmental conditions that affect human workers during shop-floor activities. As with the other assets, two Tables, A.6 and A.7, represent its header and body. This asset monitors three key aspects of the environment: noise, temperature, and humidity.

I.6 Conclusion

Significant changes have occurred in the design and functioning of industrial processes as a direct result of I4.0 and the associated digitalization. These changes have greatly increased the use of automation. With the advent of Industry 5.0, which emphasizes human centricity, manufacturing and logistics environments have become more detailed and complex. This complexity has, in turn, led to a higher mental workload for human operators. The study presented in this thesis introduces a new approach to managing cognitive load through an extended Human Asset Administration Shell (HAAS), which gathers real-time data from a variety of assets.

The proposed HAAS goes beyond the conventional role of the AAS in providing interoperability between assets. It extends its function to include the adaptation of tasks and environmental conditions based on multivariate parameters. The extended HAAS incorporates four modules: physiological data module, workers' characteristics module (to personalize the HAAS), task module, and finally, environmental conditions module that accounts for the operator's surroundings.

Thesis II

Event-Based Physiological Markers of Cognitive Load: Viability and Limits

II.1 Introduction

The frameworks of the 5th industrial revolution (I5.0) and Operator 5.0 currently recognize human-centricity as a fundamental principle [74, 75]. The concept of this term puts the workers' well-being, health, and safety at the center of production systems, alongside other metrics like production cost and quality [76, 4, 75]. Presently, traditional user interfaces (UIs) facilitate interactions between human operators and processes; however, these interfaces typically lack capabilities for adaptive learning from operators or recognition of their mental and physical states.

I5.0 enablers are trying to increase automation and efficiency in the production while ensuring human centricity and worker well-being through the use of various intelligent technologies [77, 78]. The demands placed on individuals are growing in pace with the automation of more and more tasks; as a result, humans will need to be prepared and monitored to deal with more complex and exceptional situations [79, 78]. Ensuring the effectiveness of adaptive technologies and varied instructional approaches is dependent on the ability to accurately tracking workers' cognitive load, as it directly impacts performance, safety, and overall well-being.

Physiological signals provide an indirect yet sensitive way to monitor individuals' cognitive load and emotional state, as they are tightly linked to the activation of the autonomic nervous system (ANS) [20, 21, 22]. Measures such as EDA and HRV can capture short-term fluctuations in sympathetic and parasympathetic activities.

High cognitive tasks have been found to predominantly activate sympathetic division of the ANS responses, as detected by the photoplethysmogram (PPG) and electrocardiogram (ECG) in individuals under stress. Participants who used more effort to complete activities were found to show a decrease in the amplitude of their pulse waves, which is an indication of vasoconstriction [80, 81, 22]. This highlights their significance in assessing the heightened sympathetic influence on cardiac rhythms [82, 22].

Furthermore, the increased demand for cognitive resources to complete challenging tasks has been found to be affecting electrodermal activity (EDA). Studies have reported that the mean amplitude of EDA rises as task difficulty increases [83], which supports its role as a key indicator of increased sympathetic activation [22].

TABLE II.1: Summary of multimodal datasets for cognitive load research and the unique contribution of this study in enabling attempt-level tracking within repeated tasks.

Dataset	Number of Participants	Task domain	Task repeating	Objective modalities	Subjective modalities	Attempt-level learning analysis	Time granularity
MOCAS [88]	21	CCTV monitoring task	(Limited) different workload settings	EEG, PPG, GSR, ACC, skin temperature, Webcam	Included	No	Task blocks/sessions
ADABase [89]	51	N-back, Autonomous Driving Simulation	(Limited) Multiple levels per session	ECG, EDA, EMG, PPG, respiration, skin temperature, eye tracking, facial video, salivary cortisol	Included	No	2 min rolling window with step size 5 sec
WheelSimPhysio -2023 [90]	58	Wheelchair training simulator	No	EDA, ACC, skin temp, EEG, eye-gaze, head pose, system logs	Included	No	Event markers
CLARE [91]	24	Multi-Attribute Task Battery-II	(Limited) 4 sessions 9 complexity levels	ECG, EDA, EEG, Gaze	Included	No	Labels every 10 sec
UNIVERSE [92]	24	Mental Arithmetic, Sudoku, N-back, Stroop, real home office tasks	(Limited) 2 sessions (Easy vs. Hard.)	ACC, EDA, Gyroscope, EEG, PPG, and Log Data	Included	No	60-sec window with 80% overlap
This work	**30**	**Assembly-like task**	**Yes**	**EDA, PPG(HRV), Wrist ACC**	**Recorded but not included**	**Yes**	**Attempt-level timestamps**

While many studies have examined how different task designs or instruction formats affect overall workload [84, 85, 86, 87], less attention has been given to how cognitive load evolves during repeated attempts of the same task. These measures can offer insight into how mental effort changes across successive repeated attempts of tasks. However, to the best of current knowledge, studies that track attempt-level dynamics of cognitive load within the same session remain limited, and it is still unclear how physiological signals reflect these changes [20], particularly during repeated instances of the same task.

In accordance with prior findings linking autonomic nervous system (ANS) activity to physiological responses, this thesis aims to identify the most reliable features extracted from PPG-derived HRV, EDA, and dominant-hand acceleration for tracking attempt-by-attempt changes in cognitive load and its effort-related physiological correlates during repeated task execution. This thesis analyzes the same experimental dataset as Chapter III.3. Subjective workload measures (NASA-TLX, short DSSQ) and behavioral/performance outcomes (such as task completion time, number of repetitions, and assembly precision) are reported and discussed in Chapter III.3 and are not repeated here to avoid duplication.

Building on this aim, this thesis focuses on within-session, attempt-by-attempt dynamics in an assembly-like task with two instruction formats. Table II.1 positions this study against recent datasets, which mostly provide session- or block-level labels with limited repetition of the same task, which makes attempt-level learning difficult to analyze. It is hypothesized that physiological and kinematic features can effectively capture these intra-session dynamics, offering new insights into how cognitive load fluctuates across repeated task attempts.

The next sections of this thesis are structured as follows: Section II.2 explains the experimental design, Section II.3 presents the data preprocessing methodology, Section II.4 shows the results, and Sections II.5 and II.6 discuss the findings and offer conclusions, respectively.

II.2 Experiment

II.2.1 Participants

A total of 30 participants were recruited from the University of Pannonia community (university students and researchers) to take part in a controlled assembly-like experiment comparing two instruction formats (visual-based vs. code-based). The primary outcomes of interest were attempt-level changes in cognitive load indicators derived from physiological signals (PPG-derived HRV and EDA) and wrist acceleration, complemented by task-related performance measures described in the corresponding study of Chapter III.3.

Participants were approached through university mailing lists and laboratory announcements and enrolled on a voluntary basis. Inclusion criteria were adult age and the ability to perform the manual assembly task; participants with self-reported color-vision deficiency were excluded because both instruction formats relied on color information. Vision aids were permitted (participants could use glasses or contact lenses). Handedness was recorded using the Edinburgh Handedness Inventory; three participants were left-handed, and no participant reported limitations in hand or finger movement that would affect task execution.

The sample included 12 males and 18 females, aged 19–39 years ($M = 24.7$, $SD = 5.2$). Ethical approval was obtained from the Institutional Review Board (Approval No. $KEB\_MK\_FIT\_2024\_01$). All procedures were explained orally and in writing, and all participants provided written informed consent prior to participation.

A priori sample size estimation was not conducted, as it would have required effect-size and variance assumptions based on closely comparable prior work or pilot data. This limitation is acknowledged; however, the cohort size is consistent with common practice in controlled, sensor-based cognitive-load studies.

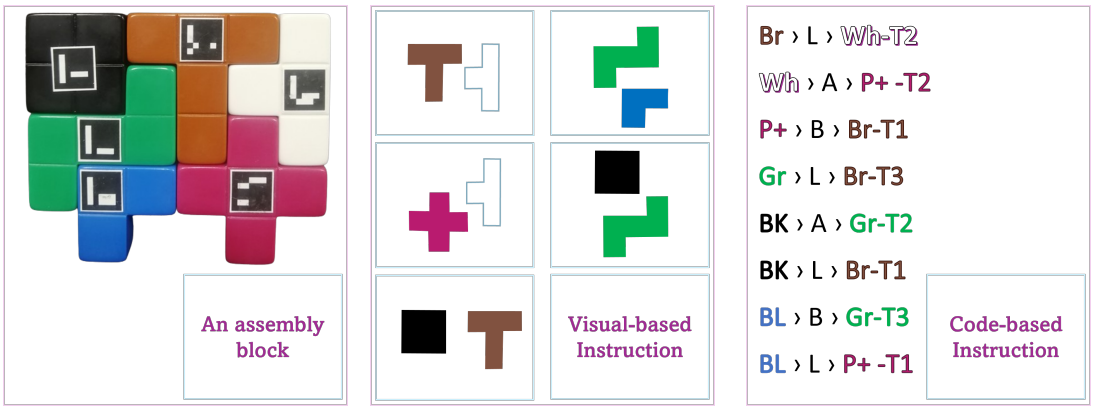
II.2.2 Experiment Design

In this study, simple pieces with unique shapes and colors from a game called "Make 'N' Break Extreme Game" have been used. To counterbalance task difficulty and avoid undesirable learning effects, four unique blocks of piece collection were created. Figure II.1(a) presents one of the assembled blocks of these pieces. Two kinds of work instructions have been used to assemble these blocks: *visual-based instruction* and *code-based instruction*. For each kind of these instructions, participants assembled two distinct blocks.

The visual-based instruction was based on presenting small images of the pieces that were used to assemble the pieces of the task. Each pair of images clearly shows which sides of the pieces should touch. This allows participants to visually align the pieces until they finish the task. This type of instruction is assumed to induce low cognitive load, as it should not deplete working memory resources. Figure II.1(b) shows a sample of this visual-based instruction.

The second instruction is code-based. It is assumed that this type of instruction induces a higher cognitive load than the visual-based one because it requires more interpretive effort. Since each assembly piece had a distinct color, they were labeled with the first two letters of their color names. These letter labels were displayed in the same color as the corresponding pieces. For example, the blue piece was represented by "BL" letters in blue text, while the green piece was represented by "Gr" symbols in green text. This instruction indicates the location of the piece and its contact regions in relation to neighboring pieces. To show spatial connections, "A" means "above," "B" means "below," "L" means "left of," and "R" means "right of". The degree of contact between neighboring pieces is described as "T1" for one area, with increments up to "T4" for four regions. Figure II.1(c) shows one sample of this instruction.

Both types of instructions were presented in printed format in front of the participants during the experiment. Since the experiment involved assembling blocks of exactly six pieces, the intrinsic cognitive load associated with task complexity remained constant across both instruction conditions. Therefore, the difference in cognitive load primarily arose from extraneous load due to variations in instruction format.



(a) Assembly block (b) Visual instruction (c) Code instruction

FIGURE II.1: Example of an assembly block, along with two instruction formats used to build it (visual and code-based).

Two conditions were set for each instruction session: First, each session had a five-minute time limit. During this time, participants had to assemble each of the two blocks at least three times (e.g., block 1, then block 2, repeated three times). Second, if participants did not complete the minimum number of repetitions (six assembly actions per session), they had to continue beyond the five-minute limit until they fulfilled the repetition requirement.

No structured behavioral observations, think-aloud protocols, or post-task interviews were included in the study design. Behavioral outcomes were therefore limited to task- and attempt-level performance logs (completion time and repetition counts), complemented by objective assembly-precision metrics described in Chapter III.3.

II.2.3 Sensors Utilized for Data Collection

For the physiological signals, EDA and PPG, the Shimmer3 sensor was used. Dried snap-style finger electrodes were wrapped around the proximal phalanges of the middle and index fingers of the non-dominant hand to capture the EDA. For the PPG signal, one of two electrodes was used, depending on the best quality offered: an optical pulse ear clip or an optical pulse probe on the distal phalanx of the thumb of the non-dominant hand. The sampling frequency for both signals was set at 250 Hz. The other sensor used to monitor the acceleration of the dominant hand was the Metamotion, a wristwatch-like wearable sensor placed on the wrist

of the dominant hand. The sampling frequency of the acceleration data was 12.5 Hz.

II.3 Methods

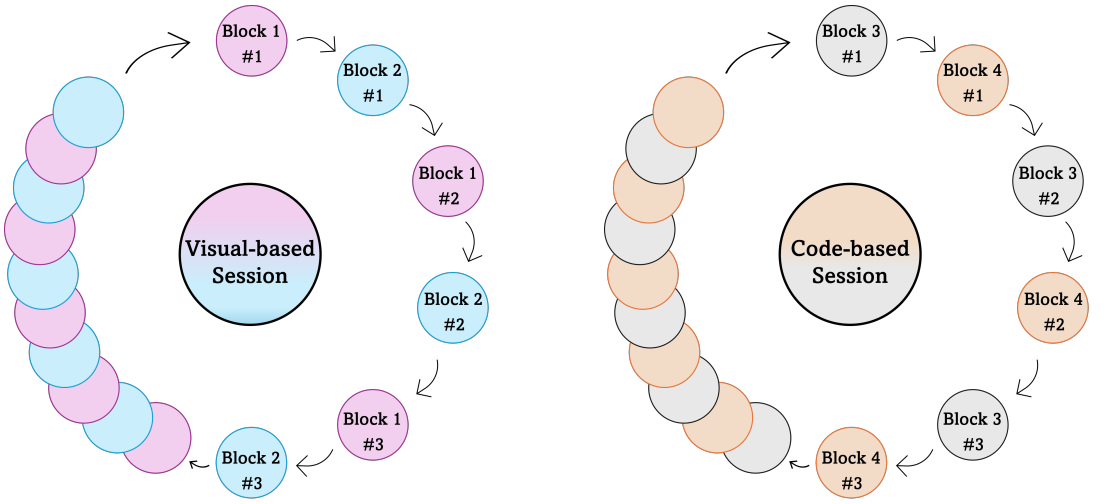
Repeating the same task with shorter durations indicates that individuals have gained knowledge [93], which can lead to a reduction in cognitive load. Therefore, the physiological features measured during each iteration are expected to reflect these changing cognitive demands. The first repetition, which is the most cognitively demanding, can differ significantly from the subsequent ones, as mastery of the assembly steps improves with practice. Capturing these changes enables the investigation of how the physiological features evolve over repeated assemblies.

The experiment used four unique assembly blocks with the same number of pieces: two associated with each instruction type. Figure II.2 shows how each block was assembled at least three times per session, with some participants completing as many as eight iterations. Although Figure II.2(a) labels blocks 1 and 2 for the visual instruction session and Figure II.2(b) labels blocks 3 and 4 for the code instruction session, these pairings were reversed for half of the participants to balance difficulty and avoid block-specific biases. After tracking the timestamps for each repetition (attempt), the data were segmented accordingly, and the physiological and acceleration features were extracted from these segments, as summarized in Tables A.8, A.9, and A.10, which are provided in the final part of the thesis, the Appendices.

This process resulted in unequal durations of data. Features that were affected by the duration of the data, cumulative features, were averaged according to their durations over these attempts to avoid inconsistencies caused by different durations.

II.3.1 Electrodermal Features (EDA)

At the beginning, a fourth-order low-pass filter with a cutoff frequency of 1 Hz was applied to the raw skin conductance to attenuate high-frequency noise. To separate the skin conductance signal into its two components, phasic (SCR) and tonic (SCL)



(a) Repetition Attempts – Visual Instruction

(b) Repetition Attempts – Code Instruction

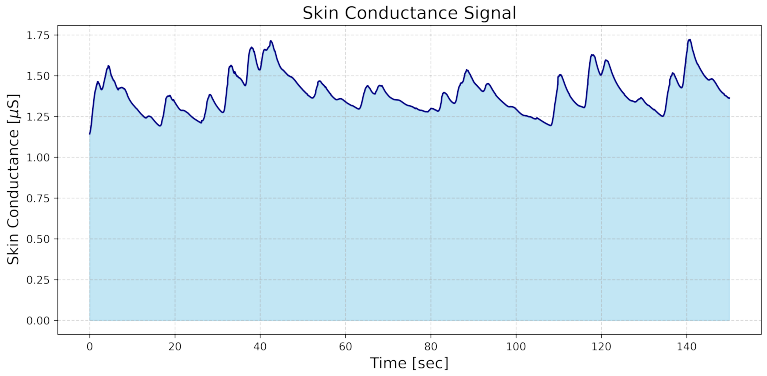
FIGURE II.2: Overview of attempt distribution across visual and code-based instruction sessions.

responses, the Matlab-based Ledalab software $V3.4.9$ [94] was used, which employs a standard deconvolution algorithm known as continuous decomposition analysis (CDA). Feature extraction followed the guidelines of Braithwaite et al. [95], which provide practical recommendations for EDA analysis. A threshold of $0.01\mu S$ was applied to identify valid SCRs; phasic deflections below this value were excluded. This ensured that only meaningful peaks were considered, which directly influenced three features: the number of peaks, rising time, and decay time.

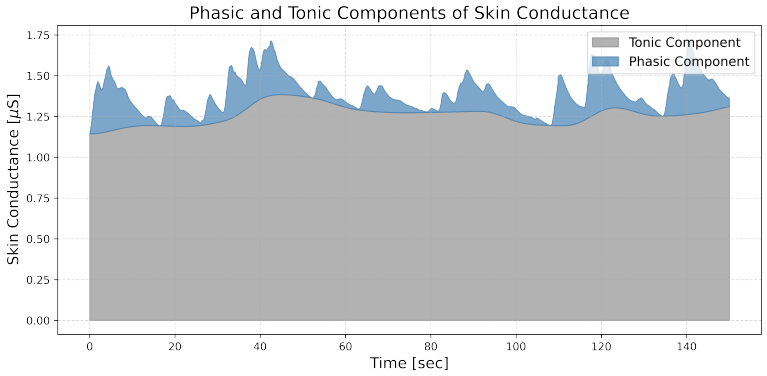
A sample of 150 seconds for both the skin conductance signal and its two components, the phasic and tonic components, are shown in Figures II.3(a) and II.3(b), respectively. Nine features in addition to the duration of each repetition (attempt) were extracted from the skin conductance signal and its components and presented in Table A.8.

II.3.2 Photoplethysmogram Features (HRV)

For the PPG signal, the same criterion applied to the EDA was followed, using the timestamps to segment the data corresponding to each repetition (attempt). For preprocessing, a Butterworth band-pass filter with cut-off frequencies set at 0.8 Hz and 3.5 Hz was applied to the raw PPG. This filtering step was necessary to retain the frequency components typically associated with heart rate activity while



(a) Raw skin conductance (SC) signal over time.



(b) Decomposed SC signal showing tonic (SCL) and phasic (SCR) components.

FIGURE II.3: Example of skin conductance signal and its decomposition into tonic and phasic components.

removing low-frequency baseline drift and high-frequency noise. After filtering, a common fluctuation trend in PPG signals that can complicate HRV feature extraction was addressed. First, the DC level offset was removed to ensure that the signal oscillated around zero. Then, a Savitzky–Golay filter was applied to reduce low-frequency trend noise [96]. This filter, configured with a polynomial order of three and a frame length of 501 samples, is designed to preserve the shape and height of waveform peaks while smoothing the data.

After cleaning the signal, a peak detection method was used to identify the heartbeat peaks, from which the interbeat intervals (RR) were calculated. To ensure physiological plausibility, detected peaks were constrained to a heart rate range of 50–120 beats per minute, with any peaks outside this range excluded. These intervals were then used to extract a variety of HRV features. Figure II.4 shows a sample of the processed PPG waveform along with the detected peaks. The

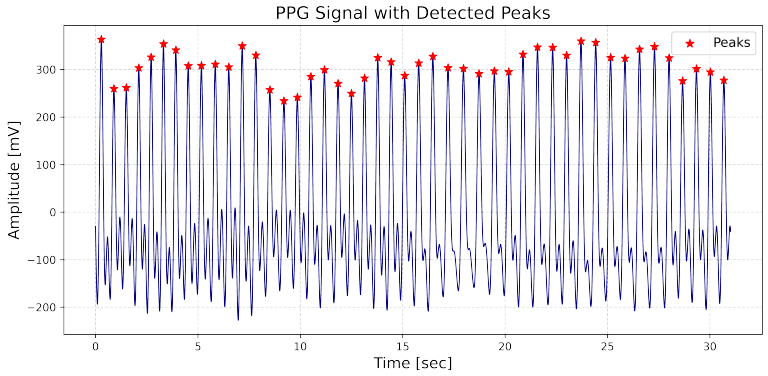


FIGURE II.4: Preprocessed PPG signal with identified heartbeat peaks used for HRV feature extraction.

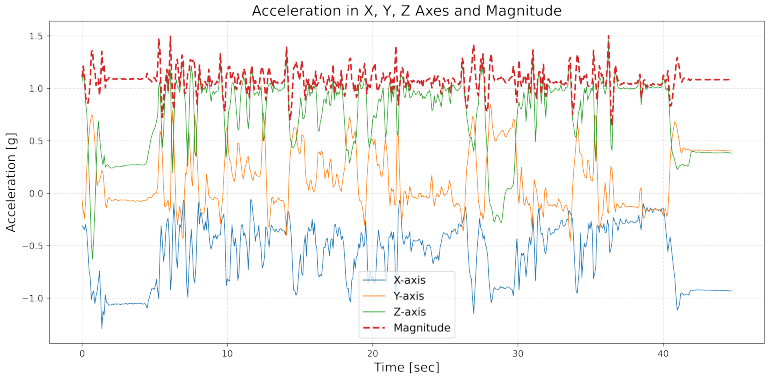


FIGURE II.5: Sample of raw acceleration data across X, Y, Z axes and their magnitude.

list and description of the HRV features extracted for this study are presented in Table A.9.

II.3.3 Kinematic Features (Acceleration)

The acceleration data were also segmented based on the timestamps of each repetition assembly process (attempt). Since the acceleration data were represented along the three axes—X, Y, and Z—and to avoid an excessive number of features, the *acceleration magnitude* of these three axes was calculated, and the analysis was conducted on the resulting signal: $Magnitude = \sqrt{X^2 + Y^2 + Z^2}$. Figure II.5 presents a 45-second sample of acceleration data from one participant, showing the three axes (X, Y, and Z), along with the calculated magnitude. The features extracted from the magnitude signal are presented in Table A.10.

II.3.4 Analyses Framework

The main goal of this study is to evaluate whether physiological features can track changes in cognitive load across repeated task attempts under different instruction formats. To address the variation in the number of repetitions between participants, linear mixed models (LMMs) were applied, as they effectively handle unbalanced data and capture attempt-level dynamics. The LMM provides flexibility by including both fixed and random effects. In the analysis, fixed effects were session type (in this case, baseline vs. instruction format) and attempt number, representing repeated assembly trials within a session. Random effects captured variability across participants, acknowledging that physiological responses to cognitive load differ between individuals.

The LMM Equation for the variables of this study is presented in Equation 1. This equation outlines the relationships among the key variables and serves as the foundation for the analyses conducted in this research.

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\,\mathbf{u} + \boldsymbol{\varepsilon} \tag{1}$$

$\mathbf{y}$ Vector of specific feature values.

$\mathbf{X}$ Fixed-effects design matrix (indicator variables for the Attempts)

$\boldsymbol{\beta}$ vector of fixed-effect coefficients $(\beta_0, \beta_2, \ldots, \beta_M)$.

$\mathbf{Z}$ Random-effects design matrix contains indicator variables for the random intercept of Participant.

$\mathbf{u}$ Vector of random intercepts for each participant $u \sim \mathcal{N}(0, \sigma_u^2)$.

$\boldsymbol{\varepsilon}$ Vector of residual errors $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2)$.

To isolate repetition effects on a single, identical task, each instruction session (which alternated two assembly blocks, see Figure II.2) was analyzed using only one block per session. For the retained block in each session, the attempt timestamps were used to segment the continuous signals and extract features within those attempt intervals. For the visual session, block 1 (Figure II.2(a)) was used and block

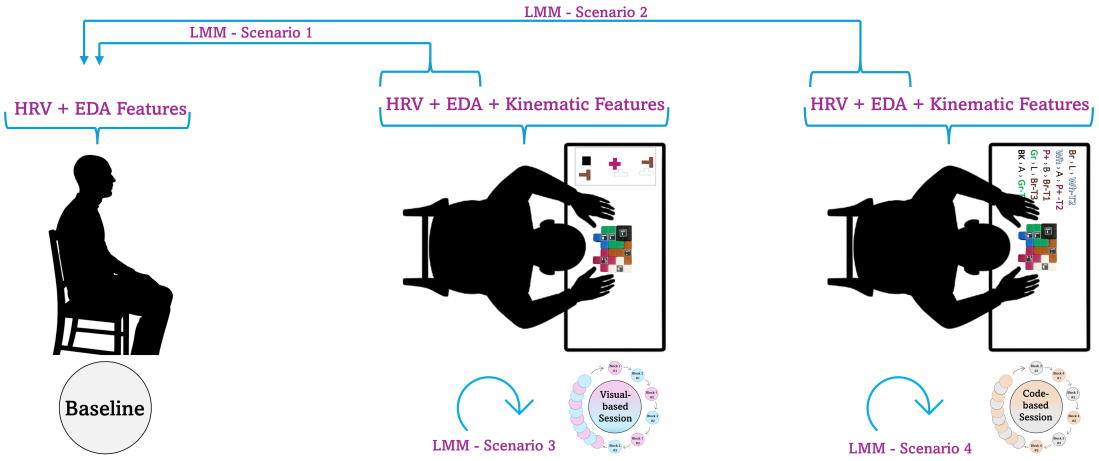
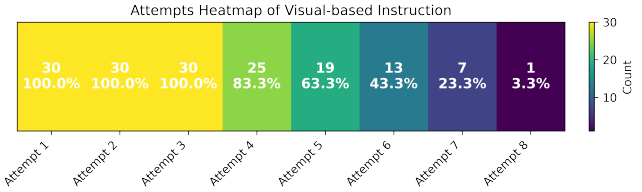


FIGURE II.6: Four LMM scenarios: (1) Comparing all task attempts of visual instruction using HRV and EDA features to the baseline; (2) Comparing all task attempts of code-based instruction using HRV and EDA features to the baseline; (3) Analyzing HRV, EDA, and kinematic features changes across attempts within the visual-based session; (4) Analyzing HRV, EDA, and kinematic features changes across attempts within the code-based session.

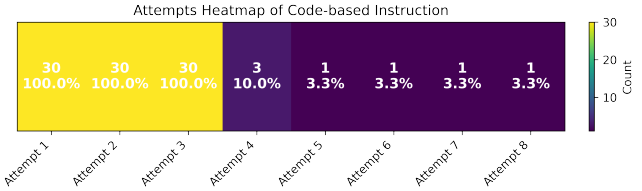
2 was excluded; for the code session, block 3 (Figure II.2(b)) was used and block 4 was excluded. Restricting analyses to repeated attempts of the same block eliminates cross-block variability, avoids contamination between task configurations, and ensures that observed changes reflect within-session learning/adaptation and the dynamic evolution of cognitive load across repeated attempts.

To address the research goal, four LMM scenarios were applied to capture how physiological and kinematic features change across baseline and instruction conditions, as well as across repeated task attempts (Figure II.6). In the first scenario, all visual-instruction attempts were compared to the baseline using HRV and EDA features. The second scenario applied the same comparison to the code instruction attempts. The third and fourth scenarios shifted the focus to within-session dynamics by analyzing how HRV, EDA, and wrist kinematic features evolved across successive attempts during both instruction sessions. Together, these scenarios allowed the evaluation of both instruction-related effects and attempt-level adaptations.

Before applying the LMM to the features of the attempts, heatmaps were created to visualize data distribution across attempts, as shown in Figure II.7. A gradual decrease in the number of participants occurs over attempts as the experiment proceeds in the visual instruction session (Figure II.7(a)), whereas a sharp decrease in



(a) Participant distribution across visual instruction attempts.



(b) Participant distribution across code-based instruction attempts.

FIGURE II.7: Heatmaps of participants distribution during the attempts of visual and code-based instruction sessions.

the number of participants has been observed in the code-based instruction after attempt 3 (Figure II.7(b)). Lower data distribution in an attempt will lead to an increase in the standard error (SE) of that attempt, which decreases the reliability of the statistical parameters extracted from the LMM. Hence, all attempts that were limited to a single participant were excluded: attempt 8 in the visual instruction and attempts 5 to 8 in the code-based instruction.

II.4 Results

II.4.1 Scenarios 1 and 2

In the first scenario, changes in physiological features were examined as participants transitioned from the baseline (rest) to the attempts of the visual instruction session. This scenario captures the effect of task execution under visual guidance and shows how attempt-wise responses diverge from baseline. The LMM used a treatment coding scheme, with the baseline set as the reference level; therefore, significant coefficients indicate deviations from baseline values.

Given the large feature set (28 from PPG and EDA), only those features that showed a significant difference in at least one attempt are reported. Table II.2

TABLE II.2: Estimated attempt-wise deviations from baseline during the visual session (treatment-coded LMM). The intercept is the baseline mean for each feature; Es = fixed-effect estimate (difference from baseline), d = Cohen's d, P = $P - value$. Only attempts with significant deviations are shown; blank cells indicate non-significance.

Feature	Intercept	Attempt 1			Attempt 2			Attempt 3			Attempt 4			Attempt 5			Attempt 6			Attempt 7		
EDA	Es	Es	d	P	Es	d	P	Es	d	P	Es	d	P	Es	d	P	Es	d	P	Es	d	P
SC Area	215.2	299.9	2.15	<.001	296.4	2.12	<.001	300.3	2.15	<.001	285.5	2.04	<.001	299.4	2.14	<.001	312.1	2.23	<.001	311.7	2.23	<.001
SCR Area	8.6	16.604	1.75	<.001	11.0	1.16	<.001	11.5	1.22	<.001	9.7	1.02	<.001	15.1	1.59	<.001	12.9	1.36	<.001			
#No. Peaks	8.5	10.8	2.46	<.001	9.6	2.18	<.001	10.2	2.33	<.001	9.3	2.11	<.001	9.6	2.19	<.001	10.8	2.44	<.001	7.7	1.76	<.001
Rising Time	5.9	-3.4	-1.19	<.001	-3.9	-1.34	<.001	-3.7	-1.28	<.001	-3.9	-1.34	<.001	-3.9	-1.36	<.001	-4.3	-1.48	<.001	-4.0	-1.39	<.001
Decay Time	12.2	-7.9	-1.96	<.001	-7.9	-1.98	<.001	-8.2	-2.03	<.001	-8.4	-2.10	<.001	-8.1	-2.01	<.001	-9.0	-2.25	<.001	-8.2	-2.05	<.001
STD SCR	0.071	0.120	1.15	<.001	0.089	0.85	.001	0.074	0.70	.007	0.064	0.61	.025	0.102	0.97	.001	0.089	0.85	.013			
SCR power	0.072	0.125	0.61	.019																		
HRV	Es	Es	d	P	Es	d	P	Es	d	P	Es	d	P	Es	d	P	Es	d	P	Es	d	P
Mean	810.8	-61.0	-1.66	<.001	-55.7	-1.52	<.001	-50.2	-1.37	<.001	-64.8	-1.77	<.001	-59.8	-1.63	<.001	-51.9	-1.42	<.001	-57.7	-1.58	<.001
Median	810.1	-58.3	-1.69	<.001	-54.8	-1.59	<.001	-48.6	1.41	<.001	-62.1	-1.80	<.001	-60.8	-1.76	<.001	-52.7	-1.53	<.001	-51.2	-1.48	<.001
HR	77.2	5.6	1.62	<.001	4.8	1.39	<.001	4.3	1.25	<.001	6.0	1.75	<.001	5.4	1.56	<.001	4.6	1.34	<.001	4.5	1.31	<.001
VLF power	7.6 $\times 10^{-4}$	-5.9 $\times 10^{-4}$	-2.05	<.001	-6.2 $\times 10^{-4}$	-2.14	<.001	-6.2 $\times 10^{-4}$	-2.14	<.001	-6.9 $\times 10^{-4}$	-2.37	<.001	-6.8 $\times 10^{-4}$	-2.35	<.001	-6.9 $\times 10^{-4}$	-2.38	<.001	-6.9 $\times 10^{-4}$	2.37	<.001
SDRR/ RMSSD	1.393	-0.228	-0.77	.003	-0.261	-0.88	<.001	-0.189	-0.64	.013	-0.256	-0.87	.002	-0.261	-0.88	.003	-0.224	-0.76	.027			

Es = Estimate (difference from baseline intercept); d= Cohen's d; P = p-value.

TABLE II.3: Estimated attempt-wise deviations from baseline during the code session (treatment-coded LMM). The intercept is the baseline mean; Es = fixed-effect estimate (difference from baseline), d = Cohen's d, P = $P - value$. Only attempts with significant deviations are shown; blank cells indicate non-significance.

Feature	Intercept	Attempt 1			Attempt 2			Attempt 3			Attempt 4		
EDA	Baseline	Es	d	P	Es	d	P	Es	d	P	Es	d	P
SC Area	215.2	287.5	1.81	<.001	307.3	1.93	<.001	313.7	1.97	<.001	258.8	1.63	.014
SCR Area	8.6	10.6	1.15	<.001	9.4	1.02	<.001	9.2	1.00	<.001			
#No. Peaks	8.5	8.8	2.46	<.001	8.1	2.27	<.001	9.4	2.64	<.001	9.5	2.66	<.001
Rising Time	5.9	-3.7	-1.13	<.001	-3.7	-1.12	<.001	-3.6	-1.09	<.001			
Decay Time	12.2	-7.6	-1.57	<.001	-7.7	-1.59	<.001	-7.9	-1.63	<.001	-8.0	-1.66	.009
STD SCR	0.071	0.093	1.17	<.001	0.078	0.98	<.001	0.076	0.96	<.001			
HRV	Baseline	Es	d	P	Es	d	P	Es	d	P	Es	d	P
Mean	810.8	-42.0	-1.14	<.001	-42.9	-1.16	<.001	-52.3	-1.42	<.001			
Median	810.1	-41.1	-1.11	<.001	-40.8	-1.10	<.001	-50.6	-1.37	<.001			
HR	77.2	3.2	0.97	<.001	3.1	0.95	<.001	4.4	1.35	<.001			
VLF_power	7.6 $\times 10^{-4}$							-3.8 $\times 10^{-4}$	-0.76	.003			
Skewness	0.130	0.589	0.53	.038									
HF/LF	1.159				-0.45	-0.72	.005	-0.397	-0.63	.013			

Es = Estimate (difference from baseline intercept); d= Cohen's d; P = P-value.

summarizes the features that significantly diverged from baseline during the visual instruction session. The second scenario followed the same procedure for the code-based instruction session. Table II.3 lists the features that showed significant attempt-level deviations from baseline in this condition.

The intercept values in Tables II.2 and II.3 represent the average baseline for each feature and serve as the reference level in the LMMs. Changes across attempts were evaluated using fixed-effect estimates: positive values indicate increases relative to the baseline, and negative values indicate decreases. The magnitude of these

estimates reflects the extent of the change and provides insight into how cognitive load evolves across repeated attempts. Additionally, Cohen's d was reported to capture the practical significance of the observed effects.

In the EDA domain, seven features reached significance in the first scenario and six in the second. Across both instruction types, skin conductance (SC) area, skin conductance response (SCR) area, and number of peaks consistently increased from baseline to task attempts. These upward trends indicate that these features are sensitive markers of heightened physiological arousal during task performance.

In contrast, SCR rise time and decay time decreased significantly from baseline to the attempts in both sessions, reflecting more frequent peaks that limited full rise or recovery of the signal. The standard deviation (STD) of the SCR increased across attempts in both sessions, indicating greater variability and stronger fluctuations linked to task-related arousal. The power spectral density (PSD) of the SCR was significant only in the first attempt of the visual instruction session, showing an initial increase from baseline, but it did not respond during the attempts of code-based instructions.

For heart rate variability (HRV) features, clear changes from baseline were observed across repeated attempts. The mean and median RR intervals decreased, indicating higher sympathetic activity and lower parasympathetic control during task performance. At the same time, heart rate (HR) increased, supporting the RR findings and indicating stronger cardiovascular activation. Very-low-frequency (VLF) power showed a steady decrease across all seven attempts in the visual session. However, in the code session, the drop appeared only at Attempt 3. These results demonstrate that HRV features can track changes in physiological responses as participants repeat the task.

The SDRR/RMSSD ratio decreased significantly across attempts in the visual session. This suggests that this measure is more sensitive to repeated attempts with physical movement than to purely cognitive demands. In contrast, no significant change was observed in the attempts of the code session. RR interval skewness increased only in the first attempt of the code session but showed no responsiveness in the visual session. This suggests a higher sensitivity to the repeated execution of mental tasks. The HF/LF ratio decreased during Attempts 2 and 3 of the code session, indicating sensitivity to cognitive load at specific stages of repetition.

II.4.2 Scenarios 3 and 4

In this section, the focus is shifted from format-level averages to attempt-wise dynamics within each session. Specifically, Scenarios 3 and 4 analyze how physiological (EDA, HRV) and kinematic features evolve across successive repetitions of the same block, using LMMs to estimate attempt-by-attempt changes.

In Scenario 3, the analysis tests whether features track learning and adaptation across repeated attempts within the visual session. Using LMMs, pairwise attempt contrasts were run to quantify change from one attempt to the next: Attempt 1 vs. 2–7, Attempt 2 vs. 3–7, ..., Attempt 6 vs. 7. Only features showing at least one significant pairwise difference ($p < .05$) are reported. Results are summarized as heatmaps (Figure II.8), which display the fixed-effect estimates (E), Cohen's d, and red text for significant cells. This provides a compact view of how the features evolve across repeats.

The duration of each attempt was first inspected as a behavioral marker of learning. Only the transition from Attempt 1 to later attempts (2–7) showed significant decreases. No other adjacent or non-adjacent contrasts among Attempts 2–7 were significant. This pattern indicates a rapid performance gain after the first exposure, followed by a stable plateau.

In terms of physiology, seven features showed at least one significant attempt-to-attempt change within the visual session. Three were HRV frequency-domain features, whereas time-domain HRV did not yield significant pairwise differences. VLF power decreased from Attempt 1 to Attempt 4 and then continued to trend downward without additional significant contrasts. In contrast, HF power and total HRV power displayed fluctuating trajectories with multiple significant increases across pairs.

On the EDA side, the SCR area fluctuated across attempts, with significant decreases from Attempt 1 to some later attempts, but not consistently across all. SC area, however, increased across most of the session attempts and produced the most frequent significant increases, marking it as a robust indicator of task-related arousal during repetition. Two additional EDA features, SCR rise time and SC standard deviation, both decreased during the transitions from Attempt 1, indicating faster and more frequent responses, as well as reduced variability, as participants became accustomed to the task (Figure II.8(a)).

To complement the pairwise heatmap, Figure II.9 plots attempt-wise trajectories as estimated marginal means (±95% confidence intervals CIs) from the LMM for the attempts of the visual session. These curves make the practice dynamics explicit: the sharp drop in duration after Attempt 1, the downward trend in VLF power, the oscillatory increases in HF/total HRV power, and the monotonic rise in SC area. Together, the trajectories provide an intuitive view of how physiological features evolve across repeated executions of the same task.

To extend beyond EDA and HRV, wrist kinematics were investigated to determine whether they exhibit similar attempt-wise adaptation. Eleven acceleration features differed significantly in at least one pairwise contrast (see Figure II.8(b)). Most of the effects clustered in the early transitions, particularly from Attempt 1 to subsequent attempts and from Attempt 2 onward. This indicates a rapid adjustment of motor after initial exposure.

The majority of significant changes were increases relative to their reference attempts, consistent with faster or more forceful movements as the task became more familiar. However, two features (kurtosis and entropy) decreased, suggesting that motion became more regular and less bursty with practice. Together, these kinematic patterns reinforce the study's goal: features capture within-session learning and adaptation across repeated executions of the same block. To complement the acceleration heatmap, Figure II.10 shows attempt-wise trajectories (LMM estimated marginal means ±95% CIs) for wrist-kinematic (acceleration) features in the visual session. These curves provide a clear, intuitive view of motor adaptation over repeated executions of the same block.

For Scenario 4, the LMM analysis was repeated for the code-based session to test whether the features tracked attempt-wise adaptation (see Figure II.11). Attempt duration dropped significantly from Attempt 1 to all subsequent attempts, and again from Attempt 2 to 3, but no other contrasts were significant. This indicates that, in the more cognitively demanding code condition, participants reached a stable execution pace after two repetitions, consistent with rapid early learning followed by a plateau.

Building on this timing pattern, the physiological responses were next examined. Fewer features showed attempt-wise changes in the code session than in the visual session (Figure II.11(a)). Three HRV time-domain features, HR, RR kurtosis, and RR skewness, were affected, alongside two EDA features, SC area and its standard

(a) Time, HRV, and EDA.

(b) Wrist acceleration.

FIGURE II.8: Pairwise LMM contrasts across visual-session attempts. Heatmaps display fixed-effect estimates (E) and Cohen's d; red text marks $p < .05$.

deviation. HR showed a fluctuating trend in the transition from one attempt to the subsequent one, with significant increases from Attempt 1 to 3 and 2 to 3. RR kurtosis and RR skewness both significantly declined from Attempt 1 to 3, indicating a more regular RR distribution as practice progressed.

On the electrodermal side, SC area increased from Attempt 1 to 2 and 1 to 3, whereas SC standard deviation decreased over the same transitions. Together, these results point to a short adaptation phase during progressive attempts at code instructions, stronger cardiovascular activation, larger SC output, and more regular cardiac variability, followed by stabilization.

Eight acceleration features showed attempt-wise change (Figure II.11(b)), fewer than in the visual session, consistent with lower motor engagement under code instructions. Most significant contrasts clustered early, from Attempt 1 and Attempt 2 to later trials, indicating a short adaptation phase. Estimated effects were predominantly positive, reflecting increased movement once the mapping from codes to actions became familiar. In contrast, kurtosis and skewness decreased across their significant contrasts, suggesting that wrist motion became more regular and less bursty with practice. Overall, the kinematic results support the study's aim:

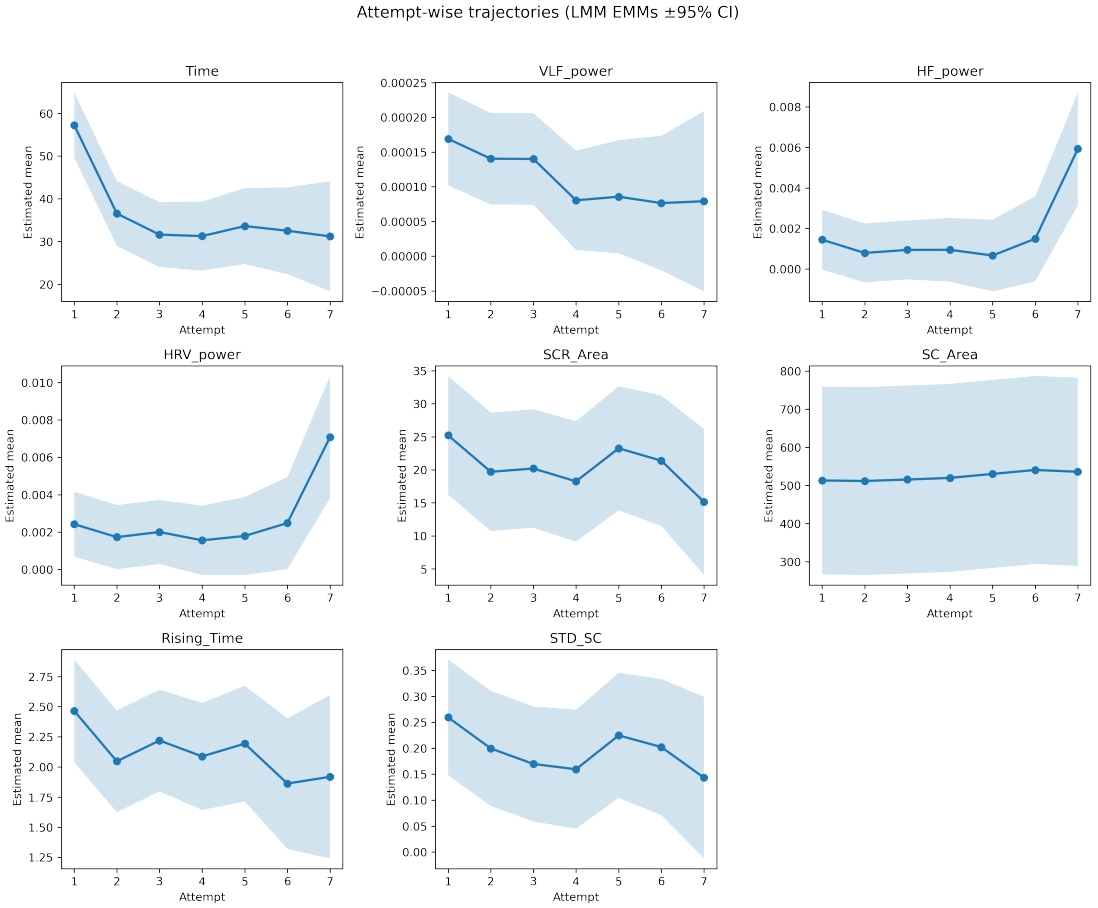


FIGURE II.9: Attempt-wise trajectories of duration, HRV, and EDA features during the attempts of the visual instruction session. Values represent estimated marginal means from LMMs, along with their 95% confidence intervals.

features capture within-session learning dynamics even in the more cognitively demanding code condition.

To complement the heatmaps, Figure II.12 plots attempt-wise trajectories (LMM estimated marginal means ±95% CIs) for the HRV, EDA, and acceleration features in the code session. The curves visualize the same early-adaptation pattern, larger changes after the first two attempts and stabilization thereafter.

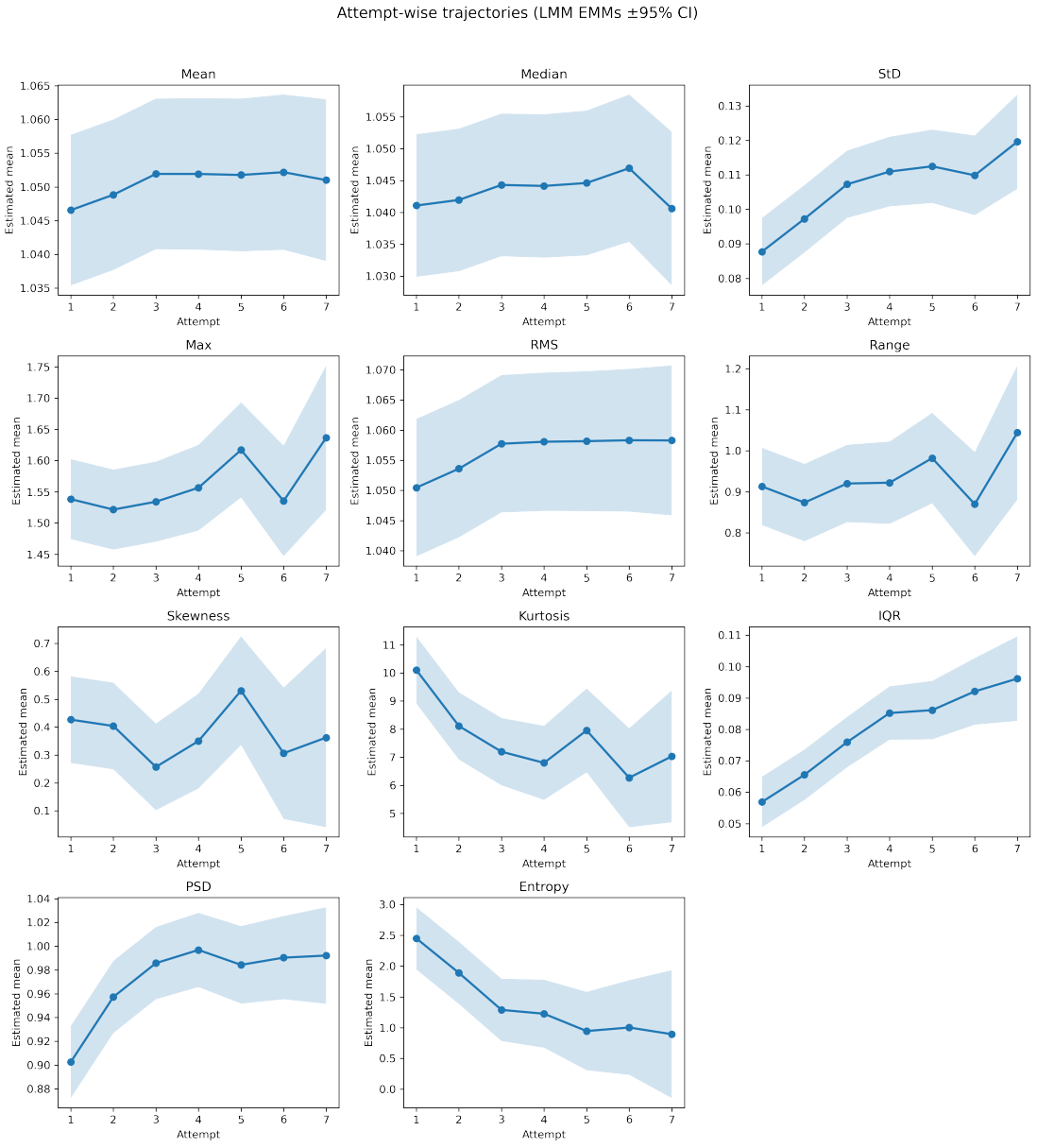
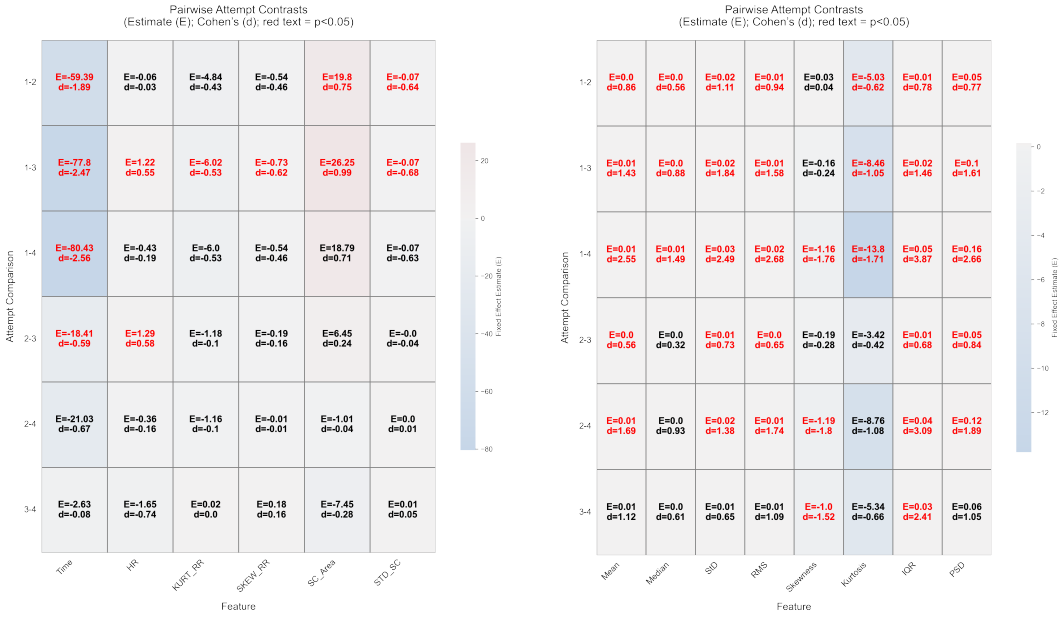


FIGURE II.10: Attempt-wise trajectories of acceleration features during the attempts of the visual instruction session. Values represent estimated marginal means from LMMs, along with their 95% confidence intervals.



(a) Time, HRV, and EDA.

(b) Wrist acceleration

FIGURE II.11: Pairwise LMM contrasts across code-based-session attempts. Heatmaps display fixed-effect estimates (E) and Cohen's d; red text marks $p < .05$.

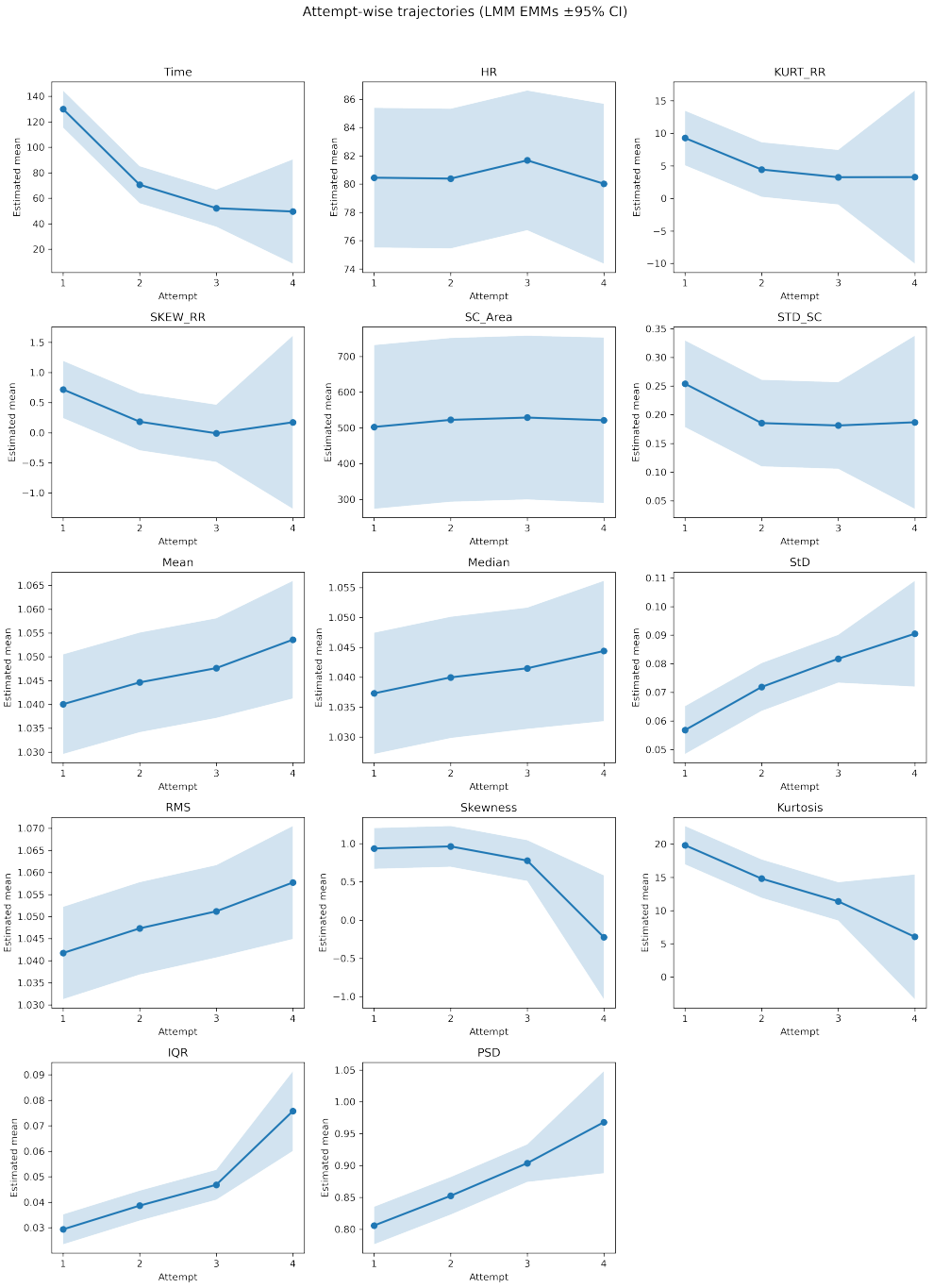


FIGURE II.12: Attempt-wise trajectories of duration, HRV, EDA, and acceleration features during the attempts of the code instruction session. Values represent estimated marginal means from LMMs, along with their 95% confidence intervals.

II.5 Discussion

II.5.1 Summary of the Main Findings

The goal was to test whether physiological and wrist-kinematic features can track cognitive-load dynamics within a session, as individuals repeat the same task under two instruction formats. The results support this goal.

First, relative to the baseline, both formats produced clear task-engagement signatures: SC area, SCR area, and peak count increased, SCR rise and decay times decreased, mean/median RR intervals decreased, and HR increased. These effects show heightened sympathetic activation during task execution, independent of instruction style.

Second, the attempt-wise analyses revealed learning and adaptation within sessions. Duration dropped sharply after the first attempt in both formats (and stabilized after Attempt 2 in the code session). On the physiology side, the SC area rose across repeats, while the SCR rise time and SC variability tended to fall after the first attempt, patterns consistent with faster responding. In HRV, VLF power declined most clearly in the visual session, whereas HF and total HRV power showed oscillatory increases across attempts. Wrist kinematics also changed mainly in the early transitions (Attempts 1 to 2 and 2 to later), with kurtosis/entropy (or skewness) generally decreasing, suggesting that in most cases movement became more regular as the task became familiar.

Together, these results demonstrate that a compact set of physiological and kinematic features can capture changes in cognitive load driven by practice within a single session. This allows the distinction between the effects of learning and adaptation and mere differences between instruction formats.

II.5.2 Interpretation of Physiological Responses

The aim was to test whether physiology can track cognitive-load dynamics within a session as participants repeat the same task. The results show a clear shift toward sympathetic activation during task execution relative to baseline. Critically, these changes evolved across attempts in a way consistent with learning and adaptation.

Prior work shows that higher cognitive demand increases both the frequency and amplitude of SCRs [97]. The findings are consistent with this. In both instruction formats, SCR peaks became more frequent and larger than at baseline, and SCR rise and decay times were shorter. SC and SCR areas also increased. Together, these changes indicate stronger sympathetic activation and greater task engagement. The shorter rise and decay times suggest a faster and more intense autonomic response to successive task events.

During task attempts, HR increased and RR intervals shortened in both sessions, which is a typical response to a higher mental workload [98]. Frequency-domain features exhibited a complementary pattern. VLF power decreased during the task, with the clearest drop in the attempts of the visual session. This is consistent with reports of VLF suppression under cognitive load [99, 100]. In the code session attempts, where the demand was higher, the HF/LF ratio decreased in the initial attempts, indicating a shift toward sympathetic dominance [101]. Selectivity by attempt also appeared in time-domain features: the SDRR/RMSSD ratio was sensitive in the visual session, whereas RR-interval skewness emerged only in the first code attempt. This suggests an initial cognitive challenge that did not persist.

The two formats imposed different demands on the participants. The visual session focused on hand/motor coordination, while the code session required more parsing and working memory effort. Consistent with this, the attempt-wise physiological patterns varied. Figure II.13 shows, for each shared feature in Tables II.2 and II.3, the attempt-wise estimated means from the LMM on the original measurement scale, with the baseline mean overlaid as a reference. The curves show that the same features follow different trajectories under the two instruction formats. This supports the study's goal: EDA and HRV features do not simply differ from rest, they track how load evolves within a session as participants repeat the same task and settle into a stable pattern.

II.5.3 Attempts and Learning Effects on the Features

Repeated executions revealed clear, attempt-wise learning and adaptation, with performance improving primarily in terms of task completion time. In the visual session, performance improved immediately. Completion time dropped sharply from Attempt 1 to Attempt 2 and then stayed low, indicating a rapid shift to an efficient routine. In the code-based session, the same stabilization occurred after

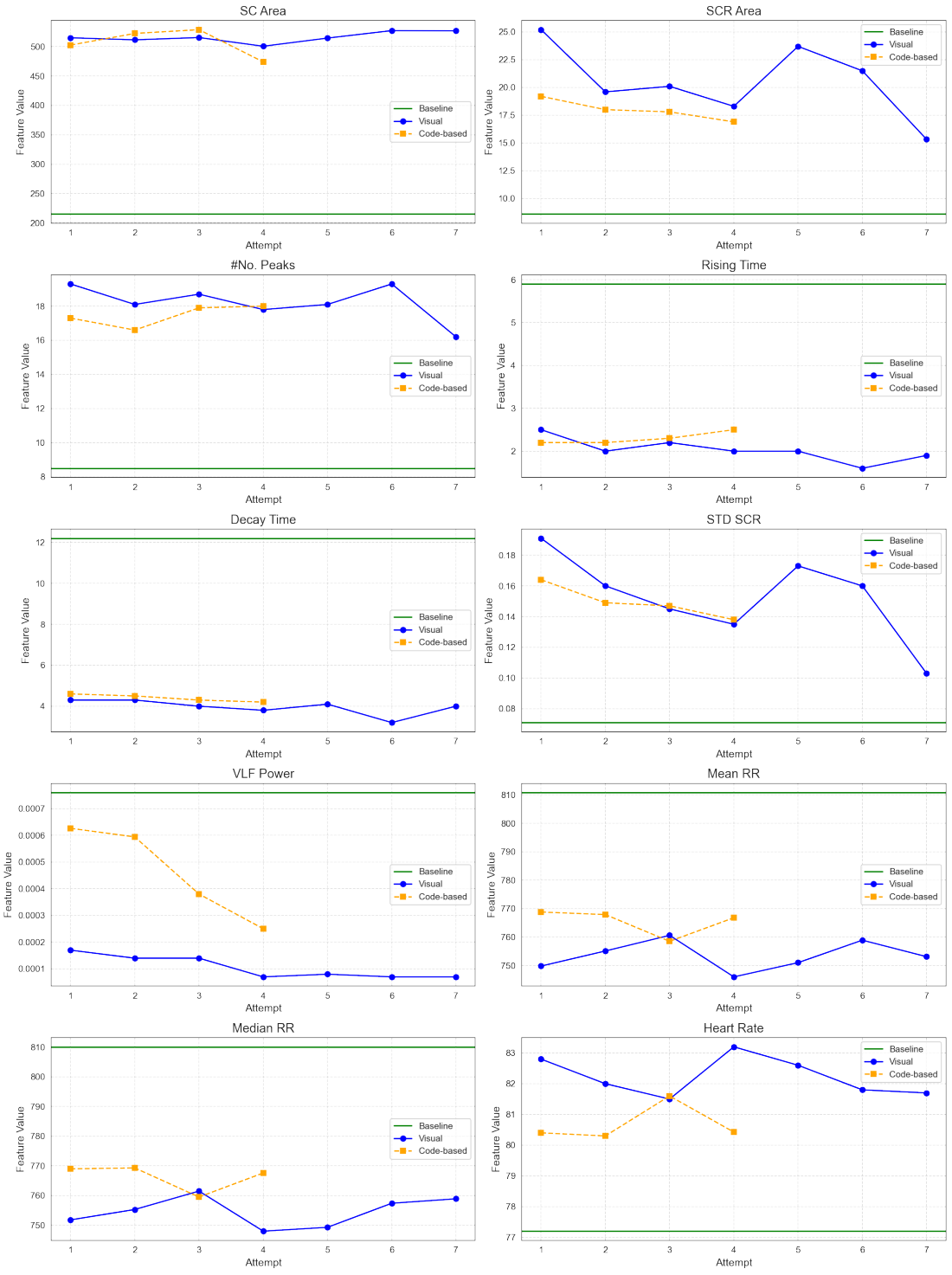


FIGURE II.13: Attempt-wise estimated means from the LMM on the original measurement scale for EDA and HRV features that are common to both sessions. Orange and blue lines show visual and code-based trajectories, the horizontal green line marks the baseline mean (reference). Values illustrate how feature levels diverge from baseline across repeats under each instruction.

two repeats (a further significant drop from Attempt 2 to Attempt 3), which is consistent with the higher interpretive demand. This pattern is consistent with the broader performance results reported in Chapter III.3, where Table III.3.6 and Figure III.3.7(b) show a significant decrease in mean task completion time in the visual session compared to the code-based session. These patterns align with the classic power law of practice, which states that the most significant improvements in performance occur in the early stages of practice, followed by slower, more gradual gains [93, 102].

More features were attempt-sensitive in the visual session than in the code session. Participants completed more repeats and executed more hand movements under visual guidance, and this higher motor throughput was mirrored by stronger changes in EDA, HRV, and kinematic features across attempts. In contrast, code instructions yielded fewer repeats and less movement, with correspondingly fewer attempt-wise physiological changes. This is consistent with the performance results in Chapter III.3, where Table III.3.6 and Figure III.3.7(a) show a significant increase in task repetitions in the visual session compared to the code-based session. This pattern supports the hypothesis that when the task allows for more repetitions, the features more clearly track learning and adaptation across attempts. This is consistent with prior evidence that visual instructions reduce cognitive load while enabling more task execution [103]).

Physiological arousal is strongly influenced by movement, which matters for attempt-wise tracking. Prior work shows HRV responses are amplified when mental tasks follow physical activity [104], and that a purely physical task can elicit arousal comparable to a combined physical and mental task [105]. In this study, visual instructions enabled more frequent and vigorous movements, which produced larger EDA/HRV changes across repeats (see Figure II.13). This helps explain why more features were attempt-sensitive in the visual session. The format reduced cognitive demands sufficiently to permit faster throughput, and the resulting motor activity amplified the physiological indicators used to track learning and adaptation across attempts.

HRV showed limited discrimination across repeats in the visual session. Only three frequency-domain features were attempt-sensitive: VLF power (one contrast, Attempt 1 vs 4) and HF power plus total HRV power (six contrasts). It is worth highlighting that those HF/total-power effects mainly involved comparisons to Attempt 7, where only 7 participants contributed data (23.3%), which inflates

standard errors and warrants caution. Overall, no HRV feature displayed a clear monotonic trend across attempts (Figures II.8(a), II.9). This pattern is plausible: the assembly attempts were short, and short-window HRV indices may be relatively insensitive to incremental, attempt-to-attempt changes in this setting.

In the code-based session, significant HRV effects appeared only in time-domain measures and only after Attempt 2 (Figure II.11(a)). This pattern suggests an initial learning phase during the first two trials, followed by more automatic execution from Attempt 3 onward. The later HRV changes likely reflect greater efficiency and steadier motor activity as the task became familiar.

In the visual session, four EDA features changed significantly across repeats (Figure II.8(a)). Three of these features—SCR area, SCR rise time, and SC standard deviation—showed their significant differences mainly when comparing Attempt 1 to later attempts, with little additional change beyond Attempt 2, suggesting early stabilization. SC area was the most informative feature; it yielded significant contrasts in about half of the pairwise tests over the first four attempts. This indicates sustained modulation of arousal as participants practiced. The lack of effects at Attempts 6–7 likely reflects a stable activity level once the task became familiar. Overall, SC area provided the clearest attempt-wise tracking feature, while other EDA measures captured primarily the initial adjustment after first exposure.

For EDA in the code-based session, only SC area and SC standard deviation showed attempt-wise changes (Figure II.11(a)). Both differed when comparing Attempt 1 with Attempts 2 and 3, but not beyond that, indicating an early adjustment in EDA after the initial exposure. The arousal is highest on first contact with the code format (decoding plus execution), then settles as the task becomes familiar. It is contended that limited familiarization and the session's time constraints prevented participants from reaching a lower-load, higher-throughput execution regime (as observed in the visual session), which likely reduced the magnitude of contrasts in the extracted features.

In summary, code-based instruction produced an initial drop in arousal after the first attempt, after which the autonomic state remained stable. This stability likely reflects the continued balance of mental decoding and physical assembly demands, which kept arousal at a moderate level without further rise or decline.

Kinematic data provided strong evidence of motor adaptation across repeated attempts in both sessions. In the visual condition, 11 acceleration-derived features showed significant changes, compared to 8 in the code condition. Overall, kinematic features were more effective at distinguishing between attempts than EDA or HRV features. Most significant contrasts clustered in the early transitions, particularly from Attempts 1 and 2 to later repetitions. This indicates that hand movements stabilized after the third attempt. The interquartile range (IQR) was the most sensitive feature, capturing systematic changes across multiple early pairwise comparisons. In the code session, a similar pattern emerged, with effects concentrated in the first two attempts. However, the smaller number of completed repetitions limited the ability to identify a clear stabilization point.

Taken together, the attempt-wise results show rapid early adaptation (strong changes from Attempts 1–2) followed by stabilization of both physiology and kinematics. This pattern is consistent with learners settling into a steady execution rhythm after the first exposures.

From a Cognitive Load Theory perspective, repeating the same task shifts working memory resources. Demands that initially went to intrinsic/extraneous processing are reallocated toward schema consolidation and automation (germane load). As practice progresses, acute effort features can plateau across attempts, while germane processes remain engaged as participants refine and automate the procedure.

Finally, the conservative size of some effects likely reflects the simplicity of the assembly-like task. In more complex industrial stations involving tool use, bimanual coordination, and time pressure, sharper attempt-level differentiation in both physiological and kinematic features would be expected. The findings, therefore, provide a lower-bound estimate of detectability and motivate testing the same attempt-resolved framework in richer, real-world workflows.

Additionally, because cognitive load is understood as a multi-faceted construct reflecting the combined demands a task places on the operator [29], relying solely on HRV or EDA may not fully capture its dynamics. To detect earlier, rapid shifts in effort that may not yet influence autonomic responses, complementary measures such as eye-tracking/pupillometry or cortical signals (EEG, fNIRS) would be useful.

II.6 Conclusion

It was examined whether physiological (EDA, HRV) and kinematic (wrist acceleration) features can track attempt-by-attempt changes during a repeated assembly-like task under two instruction formats. Using LMMs aligned to true repetition boundaries, clear rest-to-task shifts were found: EDA and HRV reliably departed from baseline under both formats (such as higher SC/SCR activity, higher HR, and shorter RR). Within sessions, attempt duration dropped sharply after the first trial, suggesting rapid learning.

For attempt-wise dynamics, signals differed in sensitivity. EDA showed meaningful within-session changes, most consistently a progressive rise in SC area across repeats in the visual session, while HRV changes were modest and inconsistent over short attempts. In contrast, kinematic features best captured early adaptation: many acceleration features changed between the first two repetitions and then plateaued, indicating stabilization of motor execution.

Overall, the results support the central aim that multimodal features can index learning and adaptation within a single session. Kinematics best reflect early practice effects, and EDA offers complementary arousal markers. HRV remains reliable for rest–task differences but is less sensitive to short, within-session changes. These findings motivate multimodal monitoring aligned with attempts and point to future work in real industrial settings with richer sensing and balanced repetitions for real-time cognitive load tracking.

Thesis III

Practical Evaluations of Cognitive Load: Applied Studies in Awareness, Switching, Instructions, and HRC

Chapter III.1

Situational Awareness Monitoring via Reaction Times and Physiological Data

III.1.1 Introduction

Workers in modern industrial environments frequently manage multiple responsibilities during physical manufacturing tasks, such as monitoring production systems, fine-tuning equipment, and ensuring that quality standards are met. These settings present unique challenges that demand high situational awareness and effective decision-making to reach the targeted goals [106, 107]. Managing multiple tasks in such situations may cause distractions and overload cognitive abilities, potentially compromising overall performance [108]. Understanding how to optimize work in these environments is critical for designing better systems and improving workplace efficiency. This requires a deep understanding of how multitasking affects not only performance but also human factors, such as stress.

However, studying situational awareness and attention in industrial contexts poses significant challenges. These constructs are often assessed using questionnaires like Situational Awareness Global Assessment Technique (SAGAT) [109], which are performed while the task is at rest. This approach may not fully capture the scope of real-time conditions in manufacturing, where errors tend to manifest as singular events such as failing to identify a quality-related problem, rather than patterns that can be easily aggregated across extended periods.

To address this limitation, reaction times are proposed to be investigated in combination with physiological measures. Reaction times have already been used as a metric to assess participants' situational awareness in prior research [110, 111], as they provide an indicator of perceiving and processing information in concurrent multitasking settings. In this regard, longer reaction times during dual tasking could indicate that one task monopolizes attentional resources. Physiological signals, such as HRV and EDA, could offer complementary measures of physiological response to various events. Physiology has already been used to study awareness in such settings [112] and has been shown to provide a proxy for psychosocial factors such as cognitive stress or time pressure.

In this chapter, an exploratory study of attentional dual-tasks in manufacturing-like settings is introduced. An experiment was performed with 12 participants that involved screwing as a primary task, while a Go/No-Go Test[113] was conducted as a secondary task, which required a high level of situational awareness. Measures of reaction time in the Go/No-Go Test were obtained as a proxy of

attentional demand, along with physiological responses from the ECG and EDA during task execution, as well as wrist acceleration as a task-related measure. The data were evaluated using event-based analyses, and the results indicated a potential effect of physiological arousal, as well as the timing of the stimulus with respect to the assembly task, on the reaction times. These findings provide a new paradigm for studying attention and situational awareness in manufacturing tasks using an event-based approach and will benefit future research in human factors engineering.

III.1.2 Multitasking and attentional resource limits

Multitasking can take several forms depending on the execution setup. It can involve simultaneous execution of more than one task [114], alternating execution via frequent task switching [115], or hybrid patterns that combine concurrent processing with rapid switching [116, 117]. In industrial environments such as assembly lines, control rooms, and human–machine workstations, multitasking is often unavoidable because operators must concurrently execute manual actions, monitor processes, and respond to intermittent events. These requirements place sustained demands on limited-capacity systems, particularly attention and working memory, and can directly affect the operator's ability to maintain situational awareness and avoid errors.

A central theoretical account for why multitasking becomes difficult is that tasks compete for shared cognitive resources. According to Wickens' Multiple Resource Theory, dual-task performance depends on the cognitive resources required for these tasks [118]. Even though previous research has indicated that separate perceptual modalities, such as sight and hearing, operate with their own independent attentional capacities [119, 120], more recent work indicates that tasks requiring central attention can still share a common bottleneck, such that both unimodal and cross-modal dual tasks can impose comparable costs when task difficulty increases [121]. Extending this view, Threaded Cognition proposes that multiple cognitive "threads" may run in parallel, but they are constrained by shared resources such as attention and working memory; as a consequence, performance costs can emerge

from time-sharing and resource contention even when tasks appear separable at the perceptual level [122].

In this thesis, these theories provide the conceptual basis for interpreting the dual-task manipulation as an increase in central attentional demand. Importantly, physiological measures (HRV- and EDA-derived features) are not treated as direct readouts of attention itself, but as indirect correlates of load-related state change that can accompany increased resource competition. Wrist kinematics are used in parallel to account for concurrent physical engagement and to support interpretation when physiological changes may be influenced by movement.

III.1.3 Situational Awareness and Physiology

Situational awareness encapsulates the basic principles of gathering and analyzing data in order to apply a situational strategy in dynamic fields [123]. Since two decades, industrial fields have shown a significant increase in automation processes and task management and control tools [124, 125]. Given the growing complexity and speed of production in current manufacturing, enhancing situational awareness for the operator has become crucial [126]. A primary responsibility of industrial workers is to maintain the seamless functioning of processes while following safety and operational standards. To achieve this objective, operators must consistently keep themselves updated on both procedural and environmental changes and advancements. In this regard, a common interpretation of the term "situational awareness" in industrial settings is being aware of a single circumstance [125].

In cognitive engineering, Endsley defined situational awareness as "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future" [116]. This model highlights the cognitive processes that enable individuals to maintain awareness and make informed decisions in real-time. Over time, various definitions have built upon this model. For example, Dominguez highlights perception's active and cyclical nature, while Bedny and Meister frame situational awareness as a reflection of the situation's potential features [112]. However, these

approaches largely revolve around qualitative interpretations of situational awareness. To truly grasp how situational awareness shapes task performance in practical settings, assessment methods capable of measuring individual's situational awareness are needed.

There are two categories into which the techniques of measuring situational awareness fall. Direct measurements depend on gathering individual's perception and understanding of the situation by questioning and observation. In contrast, indirect methods have been based on monitoring indirect measures, such as behavior, physiological signals, and task performance [112]. However, direct assessment presents a challenge and has significant limitations [127, 128]. Despite the vast number of studies that have focused on monitoring cognitive load, fatigue, and stress using physiological signals [129, 130], there remains a deficiency in these studies that address the indirect assessment of situational awareness through physiological signals [112]. Accordingly, in this thesis, physiological markers are interpreted as indirect indicators of cognitive load and are examined only for their potential relationship to situational awareness, rather than as direct measures of situational awareness.

Based on the systematic review conducted by Zhang et al., in 2023, which collected studies comparing direct and indirect measures of situational assessment and physiological signals, it was found that eye tracking was the most commonly used technique for indirect assessment, and there was a strong correlation between this physiological measure and subjective measures of situational awareness [112]. Additionally, HRV was found to be the second most popular physiological measure for monitoring situational awareness, with a high focus on the heart rate (HR) as the most commonly used feature for predicting situational awareness [131, 132]. However, not all of these works showed a statistically significant relationship between direct and indirect assessment based on HR. Researchers have also investigated electroencephalography (EEG), respiratory rate, or EDA as physiological measures for indirect assessment of situational awareness. Studies that used EEG found a positive correlation between this physiological measure and situational awareness [133, 134, 135], and a study presented in [136] observed a positive correlation between the mean of EDA and situational awareness.

Still, the number of studies that investigate situational awareness using physiology is limited, and a significant portion of them focused on the aviation domain [112].

A gap is observed in empirical research addressing the indirect assessment of situational awareness in industrial settings. Furthermore, prior research mostly studied aggregated metrics, such as the mean of the physiological signals. As errors in manufacturing typically occur as singular events that might not be well suited for aggregated metrics, an event-based approach to the analysis is considered potentially beneficial for understanding task awareness.

III.1.4 Experiment

The exploratory study involved a concurrent multitasking setup to simulate the nature of industrial work. The manual work task involved tightening screws to a specific level according to instructions. During the manual task, participants had to simultaneously react to a Go/No-Go test, which was intended to emulate secondary awareness demands imposed by industrial environments, such as in monitoring machine states. During the experiment, reaction times to the attentional test, acceleration from wrist sensors, and physiological measures were recorded, and an event-based analysis was performed to investigate whether these variables correlate with increased reaction times.

III.1.4.1 Participants

This study recruited 12 participants, including 9 males and 3 females with academic backgrounds (university students and researchers), to take part in a controlled multitasking/situational-awareness experiment. The study examined how dual-task demands influence situational awareness, operationalized via Go/No-Go response behavior, including reaction time and miss rate, and how these changes co-occur with HRV/EDA and wrist-motion features as indirect, load-related correlates. Participants were approached through the University of Pannonia community (mailing lists and laboratory announcements) and enrolled on a voluntary basis. The participants' ages ranged from 20 to 49 years ($M = 33.33$ and $SD = 8.659$). All participants gave their written consent prior to the study, and the experiment was approved by the local ethical commission. A priori sample size estimation was not conducted due to the lack of closely comparable effect-size and variance estimates.

III.1.4.2 Experiment design

Figure III.1.1 illustrates the experiment setup, which involved two simultaneous tasks. The Go/No-Go test, an attentional demand task, was used, and the participants' reaction times were tracked as a proxy for situational awareness. The task involved a blank white screen that would change to a full-screen display of brown, grey, or black colour. Designed to require sustained attention, the task lasted for a total of 15 minutes and was divided into 15-second blocks. During each block, the white screen would change to between Go and No-Go stimuli. The Go stimulus consisted of a black screen displayed for one second, during which participants were required to respond by pressing a pedal with their foot. The No-Go stimuli, which consisted of full-screen displays in brown and grey, were each presented alternately for durations of either one or three seconds. When the brown display was presented for one second, the grey display would subsequently appear for three seconds, and vice versa. Participants were instructed not to respond during these No-Go stimuli. The display screen was positioned directly in front of the participants at a distance of 50 centimeters to ensure consistent viewing conditions (label **H** in Figure III.1.1).

The manual task involved participants performing screw tightening in sequence to one of three specific marks placed on their stems, as shown in the zoomed-in part of Label **E**. An assembly table (Label **D**) with 80 screw positions was prepared, organized into eight columns and ten rows. To guide the participants, a printed instruction sheet (Label **G**) was used, incorporating simple arithmetic operations, addition, subtraction, multiplication, and division, which indicated one of the mark to which participants had to tight the screw to.

The number of digits in the instructions varied in each row. For the first four screws, the instructions consisted of one arithmetic operation involving two single-digit numbers. For the other four screws in the row, the instructions included two arithmetic operations with three single-digit numbers. These arithmetic operations produced results limited to one of three values: 1, 2, or 3. Based on the calculated outcome, participants had to tighten each screw to the corresponding mark on its stem—the first mark for a result of 1, the second for 2, and the third for 3. Timestamps for both conditions were recorded, enabling the investigation of whether participants adapting to different work instructions influenced the secondary reaction test.

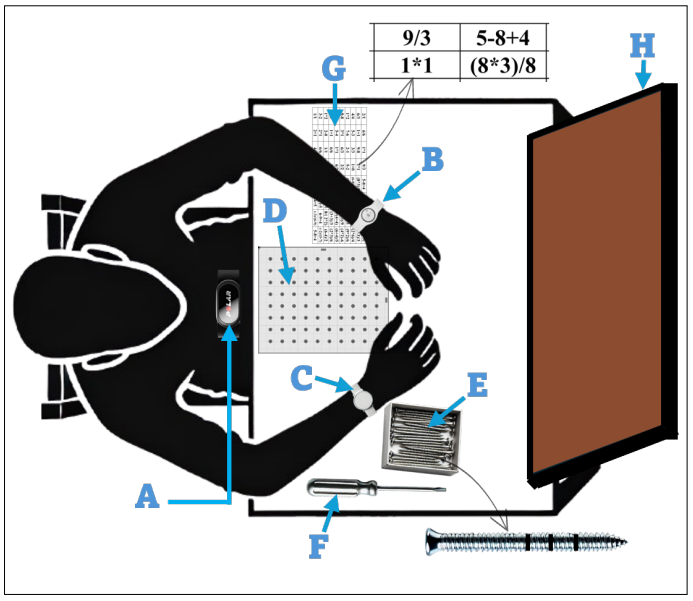


FIGURE III.1.1: The complete experiment setup is described as follows: **Polar H10 sensor (A)**: to capture the ECG. **Embrace plus sensor (B)**: to record the EDA and hand motion of the non-dominant hand. **Metamotion accelerometer (C)**: to monitor the dominant hand motion. **Manual assembly table (D)**: a table including 80 screws placeholders. **Screws (E)**: used in the experiment with three signs on each one. **Screwdriver (F)**. **Assembly instructions (G)**. **Display (H)**: to present the Go/No-Go task.

III.1.4.3 Recorded Measures

Wearable sensors were used to capture physiological signals and acceleration data, which served as a proxy for task engagement. The chest-strapped Polar H10 sensor (Label **A**) was utilized to record ECG signals. Participants also wore the Embrace Plus device from Empatica on the wrist of their non-dominant hand to measure electrodermal activity EDA and hand acceleration in three dimensions. On the wrist of the dominant hand, a Metamotion accelerometer sensor was used to capture hand motion. The performance in the Go/No-Go test was measured through participants' reaction times. For each Go stimulus, the timestamps of both the stimulus presentation and the participants' reaction to the stimulus were saved. Overall, the dataset consisted of 60 stimuli per participant, resulting in 720 events for the subsequent analysis.

III.1.5 Methods

III.1.5.1 Signal Preprocessing

First, the recorded measures were post-synchronised by matching the timestamps, and the recordings were manually validated. The preprocessing of the electrodermal activity was performed using the NeuroKit2 toolkit [137]. The signal was decomposed into its tonic and phasic components using a median smoothing filter, and peak detection was performed on the phasic signal. For the accelerometer data, the total magnitude of acceleration was computed as the square root of the sum of the squares of the components along each axis. The remaining signals were processed in raw form.

III.1.5.2 Event-Based Feature Extraction

The event-based analysis consisted of extracting windows around the time of stimulus presentation and was divided into two different metrics. For the physiological data, the interest was in determining whether the stimuli manifested in a physiological correlate; therefore, a 10-second segment was extracted starting from the onset of each Go stimulus. The extracted segment of the ECG data was processed using the HeartPy library [138]. The ECG signal was first filtered using a notch filter around the cut-off frequency of 0.01 Hz, followed by peak extraction provided by the HeartPy algorithm, with cut-off frequencies outside of $65 - 120$ beats per minute, which was determined by exploratory data analysis. The processing resulted in three HRV measures, such as mean heart rate in the segment, mean interbeat-intervals, and root mean square of successive differences. For the tonic and phasic components of EDA, common statistical measures used in time series analysis, such as mean, standard deviation, and peak-to-peak range, were computed.

The accelerometer data was extracted for three seconds prior to the stimulus, and was aimed to serve as a proxy for whether the Go-stimulus was timed during task engagement. In particular, interest was directed toward whether presenting the stimulus while participants were engaged in the screwing task would have an effect on the reaction time. For both the dominant and non-dominant hands, statistical

features commonly used in time series analysis, such as mean, standard deviation, and kurtosis, were computed.

The final processing consisted of removing missing values due for segments for which features could not be reliably computed, and data normalization to account for potential individual differences and increase generalizability. The normalization was performed by aggregating all events for each participant, and performing z-score normalization by subtracting mean and dividing by standard deviation per feature. Finally, all outliers in the reaction times that could potentially skew the analysis results were removed. This was performed by removing all missed stimuli and removing reaction times beyond 3σ of the distribution per participant. Overall, the final dataset consisted of 558 events, and 39 features for each event.

III.1.5.3 Analysis and Feature Selection

For the analysis, a linear regression model was chosen due to its interpretability, as the aim of the exploratory study was to better understand the coefficients related to longer reaction times. To enhance statistical power, recursive feature elimination was performed to identify the five most prominent features, with a step size of one.

Finally, the regression model was run with the top five features, and the model was tested for normality of residuals using QQ plots of residuals and residuals versus predicted values to assess linearity. No substantial deviations were found; hence, the model was assumed to be robust.

III.1.6 Results and Discussion

To investigate which physiological and task-related metrics correlated with increased reaction times, recursive feature elimination was performed, resulting in the acceleration of the dominant hand and the tonic component of the EDA emerging as the most prominent predictors. The top five selected covariates are presented in Table III.1.1. The resulting model was significant with a p-value of $< .001$, $F = 5.238$, however, the regression performance turned out to be low ($R = 0.213$, $R^2 = 0.046$). Although the model was significant, the low regression performance

Table III.1.1: Overview of the top five selected features, which were included as covariates in the final regression model.

Covariate	Standardised Coefficient	t	p-value
Intercept	-	−1.805	0.072
Mean Tonic EDA	−0.535	−2.184	0.029
Standard Deviation of Tonic EDA	0.289	1.663	0.097
Interquantile range of Tonic EDA	0.354	2.971	0.003
Standard Deviation of Primary Hand Acceleration	−0.345	−3.520	$< .001$
Peak-to-Peak Amplitude of Primary Hand Acceleration	0.255	2.597	0.010

indicates only a potentially small statistical effect between the collected variables and the reaction times.

The coefficient analysis resulted in all variables with the exception of standard deviation of the tonic EDA signal to be significant. Notably, a negative coefficient of the mean tonic EDA ($\beta = -0.535$, $p = 0.029$) on reaction times was observed, indicating that increased mean tonic EDA led to faster reactions. In contrast to that, the difference of the 25th and 75th interquartile range ($\beta = 0.289$, $p = 0.003$) impacted the reaction performance negatively.

Interestingly, the standard deviation of the acceleration from the primary working hand was found to correlate with lower reaction times ($\beta = -0.345$, $p < .001$), while the difference in peak-to-peak amplitude was positive ($\beta = 0.255$, $p = 0.010$). This effect was interpreted as resulting from the repetitive pattern of the screwing task causing a higher standard deviation, which would therefore correlate with participants being engaged in the screwing task. Other tasks such as fetching screws could have lead to a higher peak to peak change. In summary, these effects were interpreted as participants being able to react faster during the actual manual task, potentially due to the low attentional requirements of the repetitive screwing employed in the experiment.

The significant negative relationship that was observed between mean tonic EDA and reaction times ($\beta = -0.535$, $p = 0.029$), suggests that higher levels of tonic EDA were associated with faster reactions. One potential explanation for this effect is that elevated mean tonic EDA may reflect increased physiological arousal or stress levels. Previous research papers have established a link between stress and enhanced cognitive and physical performance under certain conditions, often referred to as the "stress-performance curve" or Yerkes-Dodson law. Moderate stress levels could increase vigilance and attention readiness, thereby enabling

faster responses to external stimuli, such as the Go/No-Go test used in the experiment. Also, this result is aligned with the study presented in [136], as the authors observed a positive correlation between the mean of EDA and situational awareness.

However, it is also important to consider the variability in EDA. The interquartile range difference ($\beta = 0.289$, $p = 0.003$) had a negative impact on reaction times, indicating that unstable arousal levels may disrupt attentional focus and situational awareness. This aligns with findings that inconsistent stress responses can lead to cognitive overload, hindering performance. The combination of these results underscores the complex role of physiological arousal, where consistent, elevated arousal might enhance reaction speed, while fluctuating arousal may impair it.

The potential link between high mean tonic EDA and stress provides an interesting avenue for further exploration. Stress, often viewed as detrimental, may under specific conditions enhance performance in manufacturing settings by increasing vigilance and focus. The observed effects of physiological arousal, stress, and task engagement have significant implications for task and environment design in manufacturing. Repetitive tasks that induce moderate physiological arousal may optimize situational awareness and response times, whereas tasks requiring frequent transitions or greater variability could disrupt attentional focus. These findings emphasize the need to balance task complexity and physiological demands to maintain optimal performance and safety in industrial environments.

The results of the exploratory analyses provide valuable insights into the relationships between physiological signals, task-related measures, and attentional demand in a dual-task manufacturing-like setting. Specifically, the acceleration of the dominant hand and the tonic component of EDA emerged as the most prominent predictors of reaction times to the Go/No-Go test.

While valuable insights were provided by the findings, several limitations must be acknowledged. One significant limitation is the potential floor effect observed in the screwing task, which was relatively simple and may not have adequately challenged participants' attentional or cognitive capacities. This simplicity could have resulted in an underestimation of the interactions between primary and secondary tasks, as participants were likely able to manage both with minimal cognitive strain. Future studies should incorporate tasks with varying levels of complexity

and cognitive demands to better understand how task difficulty influences dual-task performance and the associated physiological and attentional measures.

Additionally, the sample size was small, and the exploratory nature of the analyses limits the generalizability of the findings. Larger, more diverse participant cohorts and a broader range of task types and scenarios would be essential for validating these results and expanding their applicability. Future work should also consider more nuanced approaches to experimental design and task selection to avoid potential floor effects and ensure that findings reflect realistic manufacturing scenarios.

Future work could explore integrating eye-tracking as an additional measure of situational awareness. Eye-tracking is a widely used measure [112] that could provide deeper insights into visual attention and its effect on situational awareness. Another important direction for future research is investigating the potential impact of rising levels of automation, or human-robot teaming. Certain manual tasks might be offloaded as industrial settings become more automated [139]. These scenarios could reduce physical workloads but introduce additional cognitive demands from monitoring automated systems in real time. Additionally, human-robot or human-human teaming can redistribute operators' attention as they perform their tasks and coordinate, monitor, and communicate with teammates [140]. These kinds of collaboration may impact physiological arousal patterns and situational awareness demands differently than physical multitasking. For instance, shared tasks may reduce individual workload but increase the complexity of joint decision-making. Extending the current framework to systematically vary levels of automation and teaming might reveal whether the observed relationships between EDA, HRV, acceleration, and reaction times remain consistent or shift when workers rely on interactive or intelligent systems.

III.1.7 Conclusion

This study presents an experimental design and an exploratory analyses for investigating attentional demand and physiological responses in manufacturing-like dual-task scenarios. Through preliminary analyses, the acceleration of the dominant hand and the tonic component of EDA were identified as significant predictors of reaction times in a Go/No-Go task. Key findings include the association of higher

mean tonic EDA with faster reaction times, potentially reflecting the facilitating role of stress-induced arousal, and the impact of hand acceleration metrics, where repetitive manual tasks with predictable patterns supported quicker reactions. These results provide foundational insights into the interplay between physiological arousal, task engagement, and attentional performance. The implications extend to human factors research and manufacturing applications, emphasizing the importance of task design in optimizing performance. By creating tasks that balance physiological demands and attentional requirements, industrial systems can be better tailored to support human operators' capabilities.

The preliminary results provide a valuable resource for advancing research on attention, situational awareness, and physiological responses in industrial contexts. This study lays the groundwork for further exploration into human performance in dual-task environments, contributing to the development of human-centric and resilient systems in the context of Industry 5.0.

Chapter III.2

Balancing Cognitive Load via Task-Switching

III.2.1 Introduction

Work-related musculoskeletal disorders and stress-related conditions, including burnout, are widespread [141]. These health problems have a multi-factorial etiology and were found to be linked to sustained demands paired with inadequate recovery [142, 141]. Some of the psychosocial factors, like stress, have been demonstrated to interact with individuals' behaviours, their lifestyle, and how they interact with physical environments [142]. Routine restoration during the workday reduces acute fatigue and decreases the long-term risk of developing a disorder [141].

The allostatic perspective suggests that maintaining health and well-being depends on achieving a balance between demands at work and opportunities for recovery. This highlights the importance of mapping exposure patterns where mental and physical demands alternate in ways that allow one to recover from the other. This approach aligns with ergonomic principles, which emphasize that adequate variation is essential for a job to be considered sustainable and acceptable [143].

In practice, one way to achieve this balance between demands and recovery is through job rotation (also known as task-switching). This approach involves alternating workers between tasks that differ in their physical and mental load. This approach will make periods of high demand be followed by lighter tasks that allow for recovery [144].

Another line of research has examined how task switching is studied in controlled experimental settings. Three main paradigms are commonly distinguished. In *predictable switching* designs, the order of tasks follows a fixed sequence, such as ABAB or AABB, which allows participants to anticipate upcoming changes and partly prepare for them [115, 145]. In *cued switching*, each trial is preceded by a signal that specifies which task should be performed, with cues presented in a random sequence. This makes preparation dependent on cue timing rather than on a learned pattern [146]. The third approach is *voluntary switching*, where participants are free to choose which task to perform on each trial, usually under instructions to distribute their choices evenly [147, 145].

A well-known finding in task sequencing studies is the existence of cognitive cost (switch cost) [148]. When individuals shift from one task to another, they usually respond more slowly and make more errors than when they repeat the same task.

These costs are often used as a measure of cognitive flexibility: the smaller the costs, the more easily individuals can adapt to changing demands. What makes this phenomenon particularly important is that it persists even with practice or preparation time. This suggests that cognitive control is either limited in its ability to fully anticipate the next task or that individuals choose not to allocate all their resources to it [145, 149]. Studies in recent years have also shown a robust association between frequent switching and negative mood. This negative mood acts as a bridge from performance to mental workload, which can raise overall workload [150].

While evidence for task-switching effects on aspects such as job control or absence is mixed, it has been linked to greater job satisfaction. Moreover, rotating tasks can broaden skills, support productivity and quality, reduce turnover, and make organizations more adaptable [144].

Despite the widespread interest in job rotation, the optimal pairing of tasks and scheduling of breaks to balance demand and recovery remains unclear [151]. Most prior studies have aimed to ease physical fatigue by inserting non-physical tasks or reorganizing rest breaks between blocks of physical tasks [152, 153, 154]. In a neutral way, the complementary question is asked: *Does interleaving brief physical work between mental tasks change cognitive load?* No directional assumption is made. One view is that alternating tasks provides the attention system with brief recovery periods. Another view is that focusing on one task builds set-specific efficiency and reduces switching costs. Therefore, an alternative schedule (M–P–M–P; **M** = **M**ental task block, **P** = **P**hysical task block) is compared with a continuous schedule (M–M–P–P, or P–P–M–M) within the same participants.

III.2.2 Experiment

III.2.2.1 Participants

Twenty-six university students (14 male and 12 female) with diverse academic and demographic backgrounds were recruited via an official university announcement and enrolled through an online registration form. Their ages ranged from 21 to 41 years ($M = 25.7$, $SD = 4.6$). The study examined how task sequencing (alternating vs continuous) influences cognitive load, assessed via ISA and

NASA-TLX, task-output performance metrics (mental accuracy/errors and physical throughput), and load-related physiological/kinematic markers (HRV, EDA, wrist acceleration) as reported in this chapter. All participants received written information about the study and provided informed consent. The protocol was approved by the University of Pannonia IRB (approval no. 4/2025.(02.28.)). Each participant received 5,000 HUF as compensation. No a priori sample size estimation was conducted; this limitation is acknowledged given the practical constraints typical of controlled, sensor-based experiments.

III.2.2.2 Experiment design

The experiment involved two types of tasks (Figure III.2.1): mental arithmetic and physical assembly. During the mental task, two reference tables were displayed on the desk. One table mapped uppercase letters A–Z to digits 0–9, and the other mapped lowercase letters a–z to digits 0–9. Below these tables was a worksheet with problems to solve. Each line had two four-letter groups and an operator. For example: $(PToG)/(BWcZ)$, $(AqWy) + (CgJk)$, and $(QlOi) * (HjtV)$. For each problem, participants looked up the digit corresponding to each letter and used their working memory to add the four digits within each set of brackets. Then, they wrote down the sum of each bracket before applying the given operator to the two sums. Finally, participants recorded the overall result on the sheet. Participants were immediately notified of calculation errors at the end of the line and asked to correct them. This procedure prevented them from mistakenly believing they had performed well when they had not actually engaged in the intended mental load. The physical task was simple: Participants used a screwdriver to tighten screws on a prepared board.

Each task block lasted four minutes. First, all participants completed a training phase to become familiar with both tasks and the repetition procedure. The actual experiment consisted of two sessions, each with four blocks (two mental and two physical). In the continuous session, the tasks were grouped together (M–M–P–P or P–P–M–M). In the alternating session, the tasks switched back and forth (M–P–M–P). To avoid order effects, the order of sessions was counterbalanced across participants, and within the continuous session, the sequence type (M–M–P–P vs. P–P–M–M) was balanced as well. To reduce learning effects

further, separate letter–digit tables were prepared for each session so participants would not repeat the same material.

After each mental block, the participants rated their perceived workload using the Instantaneous Self-Assessment (ISA) scale. At the end of each session, they completed the NASA-TLX questionnaire to evaluate their overall workload.

The collected data included both physiological and performance measures. Physiological signals were recorded using two wearable sensors: a chest-strapped Polar H10 for electrocardiogram (ECG) and an Empatica Embrace Plus on the dominant wrist for electrodermal activity (EDA) and kinematic data. Performance was assessed separately for each task. For the mental task, The number of completed correct lines and the number of errors per block were measured. For the physical task, performance was assessed by counting the number of screws tightened in each block.

All physiological signals were synchronized and segmented based on the timestamps of each four-minute task block and the two-minute baseline. From the ECG recordings, 19 HRV features were extracted, covering time-domain, frequency-domain, and non-linear measures. For the EDA, continuous decomposition analysis (CDA) was applied to separate the signal into its tonic (SCL) and phasic (SCR) components, from which 11 features were computed. For the kinematic data (ACC), the magnitude of the three acceleration axes (X, Y, Z) was calculated, and 12 statistical features were derived from it.

III.2.3 Results and Discussion

III.2.3.1 Subjective and Performance Data Analyses

The analyses were initiated with the session-level NASA-TLX and its six subscales using a paired linear mixed model (LMM). This model was used to test whether the order of tasks within a session affects perceived workload. Results are shown in Table III.2.1, and none of the session-level comparisons were significant. However, the point estimates suggest a slight tendency toward lower workload in the alternating schedule for the global score (Alt-Cont = -2.73) and for several subscales (Mental, Physical, Temporal, and Frustration), while Effort was slightly higher

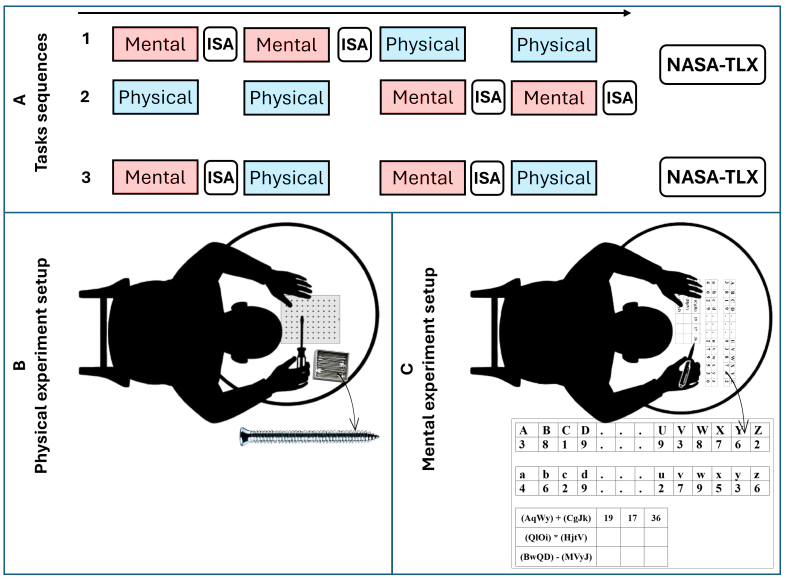


FIGURE III.2.1: Experimental design and tasks. (**A**) The top panel shows the task sequences: continuous (M–M–P–P or P–P–M–M) and alternating (M–P–M–P). The bottom panels illustrate the two task types: (**B**) physical assembly (screw tightening), and (**C**) mental arithmetic (letter–digit mapping with basic operations).

TABLE III.2.1: NASA-TLX contrasts (Alternative-Continuous)

Outcome	**Contrast**	**Estimate**	**P-value**
Work_Load	Alt-Cont	-2.73	0.33
Mental	Alt-Cont	-1.73	0.61
Physical	Alt-Cont	-4.03	0.31
Temporal	Alt-Cont	-4.80	0.26
Performance	Alt-Cont	-0.76	0.76
Effort	Alt-Cont	3.65	0.37
Frustration	Alt-Cont	-4.80	0.27

(Alt-Cont = +3.65) and Performance remained minimal. Overall, the workload appeared similar across schedules, with only modest, nonsignificant trends that may reflect minor opposing shifts across subcomponents. These patterns motivate examining block-level measures (ISA, performance, and physiology), where short-lived effects are more likely to appear.

Building on the session-level results, the block-level measures, including the ISA, task performance, and the extracted physiological features, were subsequently examined. To analyze these outcomes, paired linear mixed models with random intercepts for participants were also used. This approach was applied to the ISA ratings, the number of correct and incorrect responses in the mental task, the rate

TABLE III.2.2: Contrasts for ISA and performance across sequencing conditions.

Contrast	Est_ISA	P_ISA	Est_Corr	P_Corr	Est_Err	P_Err	Est_Phy	P_Phy
LikeForLike_1st: Alt-Cont	-0.045	0.792	0.035	0.639	-0.068	0.184	0.001	0.998
LikeForLike_2nd: Alt-Cont	0.376	**0.029**	0.059	0.426	-0.12	**0.023**	0.429	0.138
Order_in_Cont: 2nd-1st	-0.202	0.229	0.078	0.291	0.010	0.849	0.032	0.913
Order_in_Alt: 2nd-1st	0.219	0.195	0.102	0.167	-0.038	0.447	0.460	0.112
SwitchCost: (AltΔ) - (ContΔ)	0.421	0.075	0.025	0.818	-0.048	0.502	0.428	0.295

Alt = alternating sequence; Cont = continuous sequence; Est = estimate; Corr = correct responses (rate); Err = errors (rate); Phy = physical completion rate; p = p-value.

of physical task completion, and the physiological features extracted from ECG, EDA, and kinematic data.

Three contrasts were tested: like-for-like (Alternative vs. Continuous at the same block order), order (2nd – 1st within each condition), and switch-cost (the difference in the 1st-to-2nd change between conditions). The results are shown in Table III.2.2. For ISA, the 1st like-for-like contrast showed no difference, while the 2nd like-for-like contrast showed a small increase in perceived load for the Alternative sequence. This suggests that a mental block after switching from a physical block can feel slightly more demanding. The order contrasts were not significant in either condition (no reliable 1st to 2nd change within Continuous or Alternative conditions). The switch-cost contrast showed a small, non-significant increase for Alternative relative to Continuous. Overall, interleaving did not change session-level workload, but it may raise perceived load within specific blocks.

For the task performance metrics, the number of correct operations was first tested. None of these comparisons were significant. This likely reflects the small sequence effects on speeded correctness under the controlled difficulty level and the limited number of blocks (only two per session), during which participants quickly established a consistent routine for both schedules.

For the number of incorrect responses, the pattern matched the ISA, in which the second like-for-like contrast reached significance, but the direction differed. Errors were lower in the alternating sequence, where the ISA was slightly higher. The other contrasts were not significant. Finally, for the rate of physical task completion, all pairwise contrasts were non-significant, indicating that interleaving versus blocking did not affect short-term throughput in the screw-tightening task.

TABLE III.2.3: HRV, EDA, and kinematic features with at least one significant effect in mental-block contrasts (percent change and p-values).

Feature	Est_1st Alt-Cont	P	Est_2nd Alt-Cont	P	Est_Cont 2nd-1st	P	Est_Alt 2nd-1st	P
HF/LF	11.52	**0.009**	7.323	0.094	0.517	0.902	-3.270	0.432
LF/HF	-11.43	**0.046**	-6.266	0.288	-3.228	0.590	2.413	0.695
RMSSD	15.02	0.078	17.22	**0.045**	-1.867	0.812	0.015	0.998
SD1	14.88	0.077	17.04	**0.045**	-1.821	0.815	0.025	0.997
SD2	1.106	0.832	10.80	**0.048**	-3.235	0.527	6.050	0.259
SDRR	3.724	0.503	13.10	**0.024**	-3.157	0.557	5.604	0.318
SDRR/RMSSD	-5.704	**0.025**	-1.629	0.531	-0.834	0.749	3.450	0.196
SDSD	15.02	0.078	17.21	**0.045**	-1.858	0.813	0.016	0.998
Area_EDA	-14.34	0.092	-2.055	0.821	13.35	0.172	29.62	**0.004**
Area_SCL	-13.83	0.108	-1.609	0.861	14.06	0.156	30.24	**0.004**
Power_SCL	-0.731	0.838	5.015	0.173	2.298	0.527	8.220	**0.028**
Kurtosis_ACC	-37.97	**0.025**	-9.938	0.623	-4.375	0.833	38.84	0.123

Est = Estimate: % change ($\approx 100(\exp(\Delta) - 1)$); Alt = Alternating sequence; Cont = Continuous sequence; P = P-Value; 1st: First mental block in the session; 2nd: Second mental block in the session.

III.2.3.2 Physiological and Kinematic Data Analyses

The analyses began by preprocessing all EDA, HRV, and kinematic features with a natural-log transform using $\log 1p(\text{feature}) = \ln(1 + \text{feature})$; this reduces right-skewness and stabilizes variance. After the transformation, the LMM was fitted with a random intercept for participants. Due to space limitations, only the results for mental blocks are reported: (1) like-for-like contrasts (Alternative vs. Continuous at the same mental block order), and (2) within-condition order contrasts (2nd - 1st mental block).

To facilitate interpretation, the model estimates were back-transformed from the log scale to percent change, which is computed as $100 \times (\exp(\Delta) - 1)$. This represents the percent difference between conditions on the original measurement scale. Of the 42 extracted features, only those showing at least one significant effect in the planned mental contrasts are presented. The results are summarized in Table III.2.3.

In the like-for-like analysis of the first mental block, three HRV features showed significant differences between the two sequences, while EDA remained unchanged. Compared to the continuous condition, the alternating sequence exhibited higher HF/LF and lower LF/HF values. These opposite shifts both indicate stronger parasympathetic activity and reduced sympathetic drive [155]. The SDRR/RMSSD ratio also decreased, following the same trend and further supporting a parasympathetic predominance [156]. Together, these changes suggest that participants were in a calmer physiological state during the alternating condition and were

possibly experiencing lower arousal and cognitive effort. The decrease in Kurtosis_ACC supports this interpretation, as smoother wrist movements indicate more stable behavior and less activity. When interpreting this difference, consider that the alternating session always began with a mental block, while half of the participants began with a physical block and the other half began with a mental block after two physical blocks. This may justify the stronger vagal influence in the alternating session.

In the second like-for-like (2nd mental block), five time-domain HRV measures were higher in the alternating sequence: RMSSD, SD1, SD2, SDRR, and SDSD. These metrics reflect short- and longer-term variability driven largely by vagal control [24]. The pattern again indicates a more parasympathetic state when the mental block follows a physical block. In this setting, that means lower physiological arousal during the mental work, even though the ISA for this block was slightly higher. Physiologically, the alternation appeared "calmer," while subjectively, it felt a bit harder.

For order effects, none of the features showed a significant change between the first and second mental blocks within the continuous session. In contrast, in the alternating condition, three EDA features (Area_EDA, Area_SCL, and Power_SCL) increased in the second mental block. Since EDA reflects purely sympathetic activity, these rises indicate greater sympathetic arousal [157]. The same pattern appeared in the continuous condition, but was weaker and not statistically significant. Overall, this trend suggests that repeatedly switching between task types may gradually increase sympathetic activation, even when cognitive performance remains stable.

Overall, the like-for-like contrasts showed that HRV features were more strongly associated with parasympathetic activity in the alternating condition, suggesting that short physical tasks before mental work helped reduce physiological arousal. The NASA-TLX and its subscales showed the same general trend toward lower perceived workload in the alternating condition, although these differences were not statistically significant. In contrast, within the alternating session, the EDA results indicated higher sympathetic activation during the second mental block, pointing to an increase in physiological effort with continued switching. Since each session included only two mental blocks, these findings provide initial but not conclusive evidence. A larger sample size and more task repetitions would

be necessary to confirm whether alternating between task types truly supports recovery or leads to cumulative strain over time.

III.2.4 Conclusion

Interleaving brief physical work with mental tasks showed a general trend toward lower perceived cognitive load in the alternating condition, though these differences were not statistically significant. At the block level, the second mental block in the alternating sequence showed slightly higher perceived load compared to the continuous one. The HRV data revealed a shift toward greater parasympathetic activity in the alternating condition, which suggests reduced physiological arousal. However, EDA patterns indicated higher levels of sympathetic activation later in the alternating session. Together, these results suggest that alternating task types can influence the balance between recovery and arousal during mental work. However, since the experiment included only two mental blocks and a small sample size, further studies with more participants and repetitions are needed to confirm these effects and guide future task rotation designs.

Chapter III.3

From Perception to Precision: How Instruction Format Shapes Workload and Accuracy

III.3.1 Introduction

The transition toward Industry 4.0 and 5.0 has marked the end of Tayloristic industrial production, a system in which tasks are broken into small, standardized steps to maximize efficiency. Modern industrial settings are now distinguished by higher complexity and greater flexibility [158]. Manual assembly is not exempt from these transitions through reducing production depth and increasing reliance on suppliers, and small and more diverse batches [159, 158]. This shift leads to less predictability and routine for assembly workers. This uncertainty has increased workers' workloads and put more pressure on designers to design efficient assembly instructions.

Human operators in these environments face multifaceted challenges, intensified by the rise in product variants that require precise cognitive engagement. Supporting these operators effectively involves not only enhancing the clarity and accessibility of work instructions but also customizing these instructions to reduce cognitive load [86, 13]. Given these escalating complexities and the imperative for human-centric approaches, re-assessing conventional work instructions emerges as a vital step to maintain productivity, reduce errors, and manage operator strain in increasingly dynamic manufacturing scenarios [160, 86].

In the industrial setting, poorly designed instructions can significantly undermine productivity, increase the likelihood of errors, and lower overall job satisfaction. Moreover, the detrimental economic and social consequences of poor instruction have been extensively documented, resulting in reduced levels of customer satisfaction, increased operational costs, and inefficient decision-making processes [161]. This highlights the necessity for companies to prioritize high-quality information in their operational instructions [162, 163, 164, 161].

Although numerous studies have explored the benefits of simplified or digital work instructions—such as textual guides or augmented reality (AR)-based solutions [165, 166]—these approaches often do not systematically validate the objective metrics with the subjective experience of workers based on the utilized instructions. Furthermore, research that integrates subjective questionnaires and objective physiological metrics to comprehensively evaluate worker cognitive load and efficiency based on work instructions remains limited. This gap is particularly pressing in modern assembly environments, where rising task complexity calls for

instruction designs that are both cognitively considerate and operationally effective. To address this gap, the present chapter systematically compares two distinct instructional approaches—code-based and visual-based—within an assembly-like scenario. This chapter utilizes the same raw data employed in Thesis II.

It is hypothesized that code-based instructions, which use alphanumeric identifiers to guide assembly, impose a higher subjective cognitive load due to the increased mental effort required for code interpretation and spatial translation. This elevated effort is expected to be accompanied by effort- and arousal-related physiological modulation, reflected in changes in HRV- and EDA-derived features during task execution. In contrast, visual-based instructions are expected to reduce cognitive load by providing more intuitive, image-based guidance, thereby lowering the mental effort required for interpretation. However, their simplicity may lead to greater physical engagement, such as more frequent hand movements and task repetitions, which could also influence physiological measures, particularly through movement-related arousal and autonomic activation.

To examine these hypotheses, both subjective and objective indicators of cognitive load and performance are evaluated. Subjective measures include the NASA Task Load Index (NASA-TLX) and the short Dundee Stress State Questionnaire (short DSSQ [167]), while objective measures include physiological signals (GSR, HRV derived from recorded PPG data, and dominant-hand acceleration) and task performance metrics.

The short DSSQ focuses on three key psychological states: *engagement*, *distress*, and *worry*. Task *engagement* refers to the individual's energy level, personal concentration, and task motivation, indicating how strongly someone applies themselves toward achieving goals. Low task engagement is characterized by low energy, reduced motivation, and easy distraction, often manifesting as fatigue. *Distress*, on the other hand, is associated with negative emotional states; it reflects an overload of processing capacity that leads to feelings of lost control and reduced capability. Finally, *worry* involves negative self-assessments and intrusive thoughts that distract from task performance by shifting focus to the personal relevance of the task [168].

To evaluate performance, three key metrics were considered: *Task Completion Time* (TCT), *Number of Task Repetitions* (NTR), and *Assembly Precision* (AP). The TCT represents the duration required for participants to complete the task;

shorter completion times indicate clearer and more efficient instruction comprehension, whereas longer times may suggest cognitive overload or confusion [169]. The NTR reflects how often participants could repeat the task within a session, providing insight into the extent of procedural mastery and ease of instruction use [170, 171]. The AP captures the accuracy of block placement, quantified through video-based tracking of ArUco marker positions, where lower variance indicates greater precision [161, 172, 173]. Together, these indicators offer complementary perspectives on performance efficiency and execution quality under different instructional approaches.

Building on the framework proposed in Thesis I, the experiment was designed to maintain a constant level of intrinsic task complexity, each assembly involving the same number of pieces, while manipulating extraneous cognitive load through two types of instructions: code-based and visual-based. This approach enables a focused investigation into how instructional design influences both cognitive demand and operational performance. Accordingly, this chapter addresses the following research question: *How do changes in instructional design affect the alignment between subjective workload (NASA-TLX, short DSSQ), objective performance (TCT, NTR, AP), and physiological responses?*

III.3.2 Methodology

The detailed description of participants, work instruction development, and experimental design has been presented in Thesis II, Section II.2. The present section therefore focuses on the specific setup and additional methodological elements relevant to the performance and subjective analyses conducted in this chapter.

Figure III.3.1 illustrates the experimental setup used in this study. Each participant was seated at an adjustable workstation and equipped with physiological sensors for data collection, as described in Thesis II, section II.2.3. To monitor the Assembly Precision (AP) performance metric, ArUco markers—square, barcode-like black-and-white stickers—were attached to each assembly piece. These markers enabled a computer vision algorithm to track and quantify the positional accuracy of the constructed patterns through continuous video monitoring. This approach allowed objective assessment of precision and consistency in the participants' assembly performance.

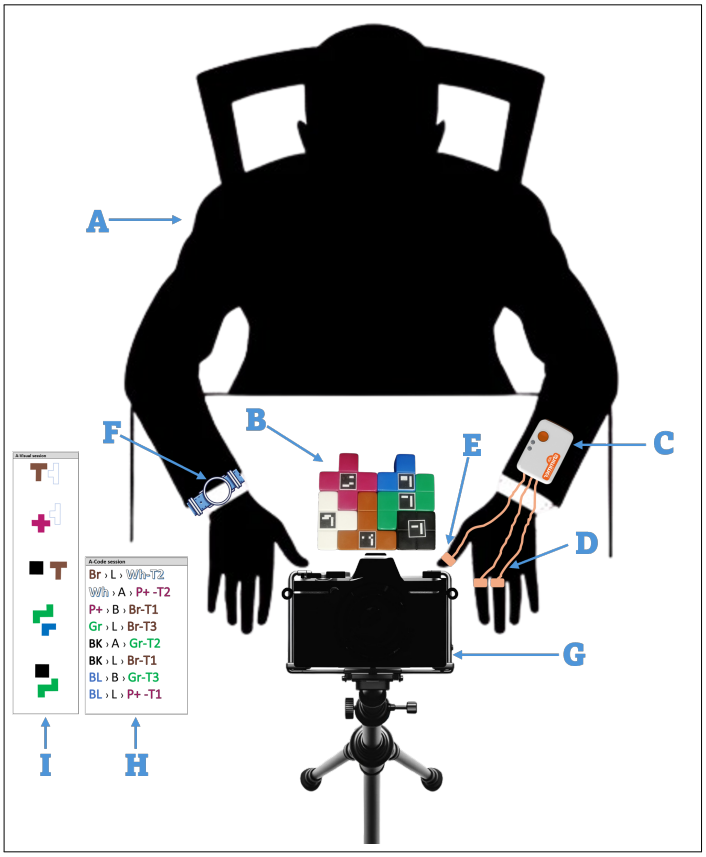


FIGURE III.3.1: This figure illustrates the comprehensive setup used in this experiment: **Participant (A)**: The participant sits on a chair facing a table where the tasks take place. **Building Blocks (B)**: Displayed on the table are the building blocks used in the experiment, each tagged with an ArUco marker to identify them during the tasks. **Shimmer3 Sensor (C)**: This sensor is attached to the arm of the participant's non-dominant hand to monitor the physiological signals (GSR and PPG). **GSR Electrodes (D)**: These electrodes are fixed to the proximal phalanx of the index and middle fingers of the non-dominant hand to measure skin conductance. **PPG Electrode (E)**: Positioned on the thumb's distal phalanx of the non-dominant hand, this electrode monitors the PPG signal. **Metamotion Sensor (F)**: an accelerometer worn on the dominant hand's wrist; this sensor tracks the participant's physical motion while engaging in the tasks. **Video Camera (G)**: This camera is mounted on a stand to capture a top-view of the task area. It records the activities during the experiment. **Code-based Instructions (H)**: A sample of code-based instructions provided to participants for task guidance. **Visual-based Instructions (I)**: This is a sample of visual instructions used to direct participants in the experiment.

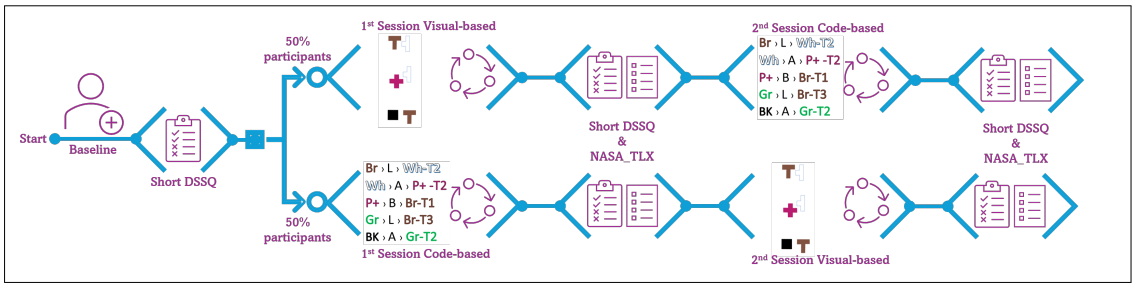


FIGURE III.3.2: Experimental sequence and counterbalanced session order for visual- and code-based tasks.

Each participant completed both visual-based and code-based assembly sessions. To counterbalance task difficulty and minimize order effects, half of the participants began with the visual-based instructions, while the remaining half started with the code-based ones. Four unique assembly configurations (labeled 1–4), each composed of six distinct pieces, were developed to ensure task variability and balanced exposure across instructional conditions. Every participant completed two assembly sessions using visual-based instructions and two using code-based instructions.

Each session lasted at least five minutes, during which participants were required to complete each assigned pattern at least three times to allow for learning curve analysis. In cases where the minimum number of task repetitions (NTR) was not achieved within this period, the session was extended until the required repetitions were completed. Throughout each session, time-stamped data were continuously recorded to capture task progress and performance. Since physiological signals are highly sensitive to motion [174], participants were instructed to use only their dominant hand during the assembly task to ensure consistency and minimize unwanted movement artifacts. The complete experimental flow and participant distribution across conditions are illustrated in Figure III.3.2.

Before starting each session, participants completed a short training phase tailored to the corresponding instructional format (visual or code-based) to ensure familiarity with the task. The experiment began with a three-minute baseline recording of physiological signals, followed by the administration of the pre-DSSQ to assess initial stress levels. After completing the first session, participants filled out the post-DSSQ and NASA-TLX questionnaires to report their perceived workload and stress. The same post-session assessment procedure was repeated after the second session to capture changes in subjective states across instructional conditions.

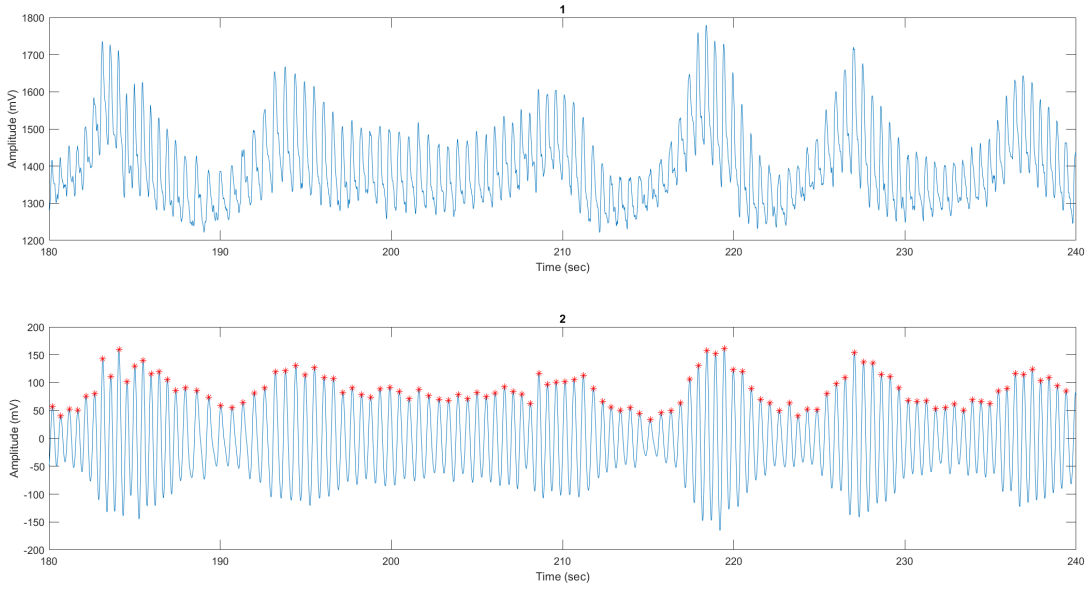


FIGURE III.3.3: (1) A sample of raw PPG signal for 60 seconds, (2) The same sample of the PPG signal after being filtered and removing the DC offset and low-frequency trend with its detected peaks

III.3.2.1 Data Preprocessing

The low-frequency trend noise accompanies most of the recorded PPG signals, complicating direct HRV extraction. Mean correction was first applied by removing the DC level offset to ensure that the signals oscillated around the zero baseline. Following this step, the Savitzky–Golay filter was implemented to remove low-frequency trend noise, and the peak detection technique was then applied to extract the HRV. Figure III.3.3 shows a sample of a 60-second PPG signal before and after removing the DC offset and low-frequency trend. To increase the data size, a 60-second segmentation window was applied to the filtered signals. HRV data were extracted for each 60-second window by calculating the variation between consecutive detected peaks on the time axis. Nineteen features were extracted from the HRV for each window. These features are already described in Thesis II, and presented in the Appendices, Table A.9.

For processing the GSR data, the Matlab-based Ledalab software V3.4.9 was utilized, which uses a standard deconvolution algorithm to separate the SC into its two components $SC = SC_{\text{tonic}} + SC_{\text{phasic}}$ [94]. Before starting the separation process, a built-in adaptive smoothing filter was applied to the signals to remove noise. The separation process was then initiated by applying continuous decomposition

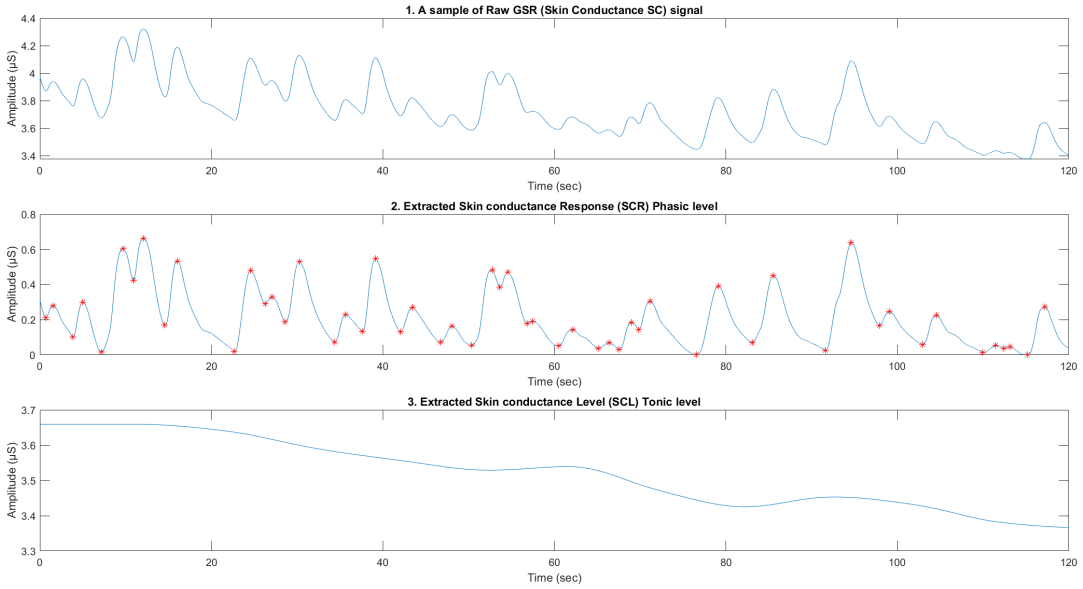


FIGURE III.3.4: (1) A sample of the recorded GSR for 120 seconds; (2) The extracted SCR with its peaks and bottoms for feature extraction; and (3) The extracted SCL.

analysis (CDA). A 60-second segmentation was also employed on the extracted signals to ensure consistency in sample size with the previous HRV measurements. From the GSR and its two components, nine features were extracted, which are described in Thesis II, and presented in the Appendices, Table A.8. Figure III.3.4 shows a sample of 2 minutes of GSR recording with its two components (SCR and SCL).

For the acceleration data recorded from the sensor on the dominant hand's wrist, three data axes were captured: X, Y, and Z. Consistent with the previous physiological data, a 60-second segmentation was applied to the acceleration signals. The resultant of these three axes was calculated, and six features were extracted from each axis, resulting in a total of 24 features. These features included mean, median, standard deviation, minimum, maximum, and signal power.

Finally, the assembly precision (AP) of each constructed pattern was evaluated through a customized algorithm. This algorithm processes video-captured images and analyzes the placement and orientation of each piece via ArUco markers. It calculates the Euclidean distance between the centers of the markers in the constructed pattern and compares it against a reference, whereby variances are determined as a measure of standard deviation. A higher value of the standard deviation indicates lower AP, while a lower value suggests higher AP.

III.3.3 Results

III.3.3.1 Subjective Data Analyses

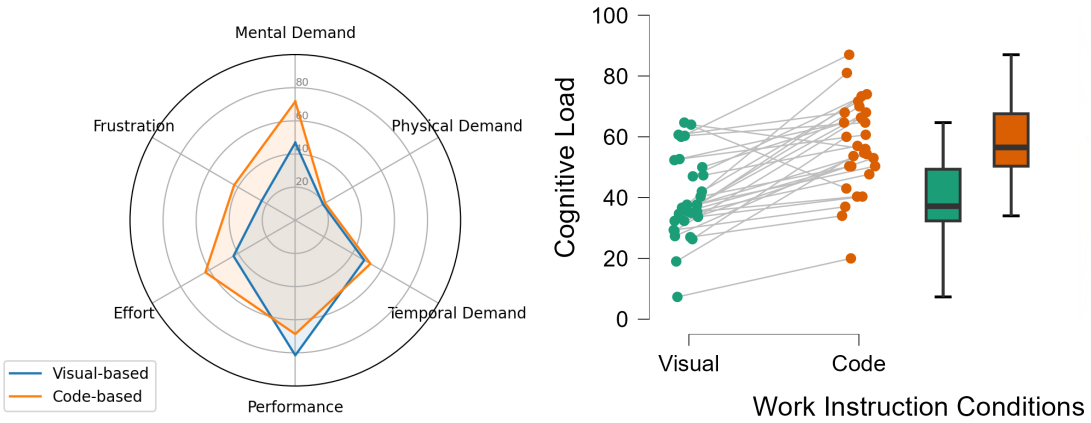
In this subsection, the subjective data collected from participants during the three sessions of the experiment are analyzed. Two questionnaires were utilized: the NASA-TLX and the short version of the DSSQ. These questionnaires capture the perceptions of the participants after each session of the experiment.

III.3.3.1.1 NASA-TLX questionnaire

The NASA-TLX expresses six categories as percentages: mental demand, physical demand, temporal demand, performance, effort, and frustration. Figure III.3.5(a) is a radar chart to visually compare these categories between the code- and visual-based instructions sessions. It shows that the code-based instructions induced higher levels of mental demand, frustration, and effort compared to the visual-based instructions. The statistical paired t-test confirmed significant differences between them, with $p - values < 0.001$ and effect sizes of -1.522 for mental demand, -0.788 for frustration, and -0.913 for effort.

These findings indicate that code-based instructions were more mentally demanding and frustrating, requiring more effort to decipher than visual-based instructions. Additionally, the results showed slightly higher levels of both physical and temporal demands for the code-based instructions compared to the visual-based instructions. However, these differences were not statistically significant. The $p - value$ for physical demand was 0.775 with an effect size of -0.082, and for temporal demand, the $p - value$ was 0.339 with an effect size of -0.177. This suggests that participants did not feel rushed by time constraints, but they were more challenged by aspects related to their limited working memory.

Finally, the nature of the NASA-TLX scale interprets the "performance" dimension in the opposite direction of the other five categories, yet assigns its weight in the same direction as the others. This means that a higher perceived performance results in a lower NASA-TLX score, contributing to a lower overall cognitive load. For visual clarity in the radar chart, the performance weight was assigned in the reverse direction to the other categories to reflect each participant's self-perceived



(a) NASA-TLX subscales. (b) Overall cognitive load.

FIGURE III.3.5: Subjective workload under visual vs. code instructions. (a) Mean NASA-TLX category scores. (b) Paired participant cognitive-load scores with boxplots summarizing each condition.

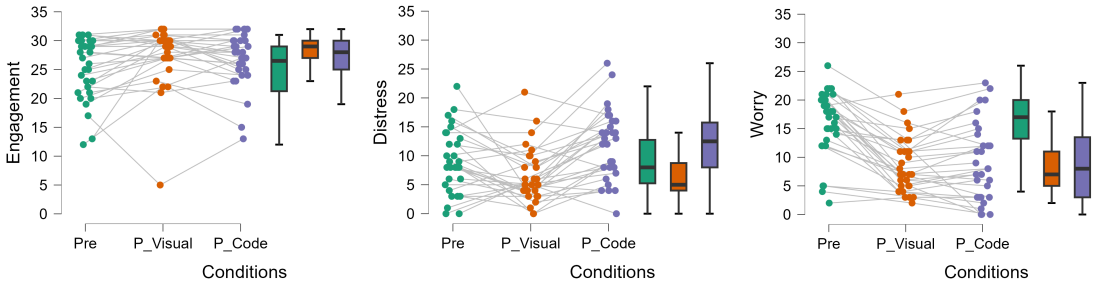
performance. According to the radar chart and the paired t-test, participants reported higher perceived performance with the visual-based instruction compared to the code-based instruction, with a significant difference ($P-value < 0.001$ and Effect Size $= -0.852$).

The cognitive loads (CLs) for the visual-based (Visual) and code-based (Code) sessions were compared in Figure III.3.5(b). These CLs were calculated from the NASA-TLX categories. The figure presented a combination of individual data points (personal CLs) with paired lines and box plots. The lines connecting the dots across the two sessions indicate the shift in CL for each participant from "Visual" to "Code", highlighting a general increase in CLs in the code-based session.

The box plot shows the distribution of CLs in both sessions, with a higher median and wider interquartile range in the code-based session. This suggests more variability and a higher overall CL. The mean values of the two sessions align with these box plots, where the visual-based session gave a $Mean$ of 39.84 with SD of 13.74 compared to the code-based session, which gave a $Mean$ of 57.24 with SD of 14.71. The paired t-test analysis supported these observations. It showed that CL increased significantly from the visual-based session to the code-based session ($P < 0.001$), with an effect size of -1.182. This indicates that code-based instructions, compared to visual-based ones, place a significantly higher cognitive demand on participants. The statistical results for the comparisons of cognitive

TABLE III.3.1: NASA-TLX Comparison Between Instructions

NASA Visual	NASA Code	Normality p-value	**Test**	**Z**	Effect Size	t-test p-value
Cognitive load	Cognitive load	0.208	Student	N/A	-1.182	< 0.001
Mental demand	Mental demand	0.150	Student	N/A	-1.522	< 0.001
Physical demand	Physical demand	< 0.001	Wilcoxon	-0.305	-0.082	0.775
Temporal demand	Temporal demand	0.575	Student	N/A	-0.177	0.339
Performance	Performance	0.114	Student	N/A	-0.852	< 0.001
Effort	Effort	0.942	Student	N/A	-0.913	< 0.001
Frustration	Frustration	0.125	Student	N/A	0.788	< 0.001



(a) Engagement. (b) Distress. (c) Worry.

FIGURE III.3.6: Short-DSSQ states across sessions. Each dot is a participant; grey lines connect within-subject changes from Pre to Post-Visual (P_Visual) and Post-Code (P_Code). Boxplots at right summarize distributions.

load and each NASA-TLX dimension between the visual-based and code-based sessions are summarized in Table III.3.1.

III.3.3.1.2 Short DSSQ questionnaire

Participants in this study completed the DSSQ three times under the following conditions: pre-experiment, post-visual-based session, and post-code-based session. The scores for each of the three psychological states were in the range of $(0 - 32)$. Starting with the first psychological state, engagement, the descriptive statistics scores revealed variations across the sessions. Prior to the experiment (Pre), the *Mean* engagement score was 25.06, with *SD* of 5.33. Following the visual-based session (P_Visual), the *Mean* engagement score increased to 27.43, accompanied by *SD* of 5.21. Following the code-based session (P_Code), the *Mean* engagement score decreased slightly to 26.83, with *SD* of 4.75. Figure III.3.6(a) displays the engagement scores of each individual. The lines connecting them across the sessions is to illustrate changes in it while the Boxplots summarize the distribution of scores within each session.

To examine whether the study sessions had a significant effect on task engagement, the Repeated Measures (RM) ANOVA was implemented. To verify the assumption of sphericity, Mauchly's test was applied, which revealed a violation of assumptions with a $p - value$ of 0.043. Therefore, the Huynh–Feldt correction was applied to account for this violation. The corrected RM ANOVA identified significant differences in engagement scores across the sessions, with a significant $p - value$ of 0.012.

Following the main RM ANOVA test, a Post-Hoc analysis was also applied. A significant increase in engagement from the pre-experiment to the post-visual-based session was observed, as evidenced by a significant $p - value$ of 0.041 and an effect size of -0.464. Although the increase in engagement from the pre-experiment to the post-code-based session had a p-value of 0.059 with effect size of -0.346, it did not meet the conventional significance threshold of 0.05. Finally, a $p - value$ of 0.323 and an effect size of 0.118 indicated no significant differences between the post-visual-based and post-code-based sessions. This suggests that both types of instructions managed to sustain close levels of engagements. These results indicate that while both instructional methods effectively boosted engagement compared to the baseline, the visual-based instructions proved particularly effective, as reflected in higher mean engagement score.

The distress scores of participants across the three sessions were also examined. There was a notable variation in these scores. Initially, before the experiment (Pre), the $Mean$ distress score was 9.00, with SD of 5.52. After the visual-based session (P_Visual), the distress scores decreased to a $Mean$ of 6.56 with SD of 4.76. However, following the code-based session (P_Code), the $Mean$ distress score increased to 11.83, with SD of 6.02 (see Figure III.3.6(b)).

To determine whether these changes in distress scores were statistically significant, the RM ANOVA test was applied, followed by Post-Hoc tests. As before, Mauchly's test was applied to check for sphericity. The test results showed no violations ($p - value = 0.792$), indicating that a standard RM ANOVA could be used without adjustments. The ANOVA results indicated that there were indeed significant differences in distress scores across the sessions, with a highly significant $p - value < 0.001$.

In the Post-Hoc tests of the ANOVA results, the changes in distress scores between the sessions were examined in greater detail. A significant decrease in distress from

the pre-experiment to the post-visual-based session was found, with a $p - value$ of 0.030 and an effect size of 0.446. Furthermore, the transition from the pre-experiment to the post-code-based session revealed a similar significant increase in distress, with a $p - value$ of 0.030 and an effect size of -0.519. Most notably, the transition from the post-visual-based session to the post-code-based session marked a substantial increase in distress levels, with a $p - value < 0.001$ and a high effect size of -0.964.

Distress, which is linked to negative emotional states, was initially high, as demonstrated by the pre-experiment mean score, indicating significant initial stress among participants. However, after using visual instructions, there was a noticeable drop in distress levels, indicating a sense of relief. In contrast, the distress levels increased sharply after the code-based session, suggesting that this instruction significantly heightened the negative emotional state, which is related to the overload of processing capacity. This pattern demonstrates the substantial impact that different instructional designs can have on participants' psychological stress. The marked differences between the visual and code-based sessions highlight the need to carefully consider the type of instructional material used and its potential psychological effects on individuals.

Next, the final psychological state, Worry, was examined. The worry scores changed notably across sessions. At the pre-experiment (Pre), the $Mean$ worry score was quite high, at 15.86, with $SD = 5.85$. After the visual-based session (P_Visual), this score significantly dropped to 8.60 ($SD = 4.86$), showing a large reduction in worry. However, after the code-based session (P_Code), the $Mean$ worry score increased slightly to 9.23, with $SD = 6.88$ (see Figure III.3.6(c)). To validate these observations, the assumption of sphericity was first examined using Mauchly's test, which indicated a violation ($p-value = 0.037$). Consequently, the Huynh–Feldt correction was applied before proceeding with the RM ANOVA. This analysis confirmed that there were significant differences in worry scores across the sessions, with a highly significant $p - value < 0.001$.

In the Post-Hoc tests of the ANOVA, worry greatly decreased from the pre-experiment to the post-visual-based session, with a $p - value < 0.001$ and a large effect size of 1.226. Similarly, worry significantly decreased from the pre-experiment to the post-code-based session, with a $p - value < 0.001$ and an effect size of 1.120. However, the worry scores did not significantly change from the post-visual-based to the post-code-based session ($p - value = 0.397$, effect size =

Table III.3.2: Repeated Measure ANOVA of DSSQ Variables (Engagement, Distress, and Worry) with Post-Hoc Comparisons

DSSQ states	Mauchly p-value	Sphericity Correction	ANOVA p-value	ANOVA Post-Hoc		
				Cases	p-value	Effect Size
Engagement	0.043	Huynh-Feldt	0.012	Pre vs. P_Visual	0.041	-0.464
				Pre vs. P_Code	0.059	-0.346
				P_Visual vs. P_Code	0.323	0.118
Distress	0.792	None	< 0.001	Pre vs. P_Visual	0.030	0.446
				Pre vs. P_Code	0.030	-0.519
				P_Visual vs. P_Code	< 0.001	-0.964
Worry	0.037	Huynh-Feldt	< 0.001	Pre vs. P_Visual	< 0.001	1.226
				Pre vs. P_Code	< 0.001	1.120
				P_Visual vs. P_Code	0.397	-0.107

Note: Pre: Pre-experiment, P_Visual: Post Visual-based session, and P_Code: Post Code-based session.

-0.107). This suggests that the code-based session did not negatively affect the initial reduction in worry.

These findings show that the way instructions are designed can greatly affect worry, which is linked to negative self-assessments. The large decrease in worry scores after the visual-based session suggests that this method can effectively reduce worry, helping individuals focus better and feel more comfortable. On the other hand, the slight increase in worry after the code-based session, although not significant compared to the visual session, shows that certain instructional methods might make anxiety worse under specific conditions. The visual representation of these results is shown in Figure III.3.6(c).The statistics for the DSSQ's three variables—engagement, distress, and worry—along with the Post-Hoc analysis, are presented in Table III.3.2.

III.3.3.2 Objective Data Analyses

Starting with the recorded physiological data, nine GSR features and nineteen HRV features were extracted, as listed in the Appendices, Tables A.8 and A.9, respectively. Feeding all of these physiological features for classification purposes or even statistical analyses can lead to poor accuracy and precision because some of these features could be highly correlated while others may not show a high contribution to predicting the target. Based on this criterion, it is inevitable to implement a feature selection technique before classification processes.

Wrapper methods are the most effective for feature selection, according to Rezaei and Jabbari [175]. A feature selection technique belonging to the wrapper methods, namely backward elimination, was implemented. This technique is based on

employing the entire set of features in the first step and gradually iterating and removing the features. Each iteration removes the feature that contributes the least to the target. This process continues as long as the model improves with feature removal.

Due to the nature of the experimental design, the data from the three sessions were not equally sized. Repeated analyses with varying sizes contravene standard statistical analyses such as the RM ANOVA and paired t-test. However, logistic regression analysis can be used for this purpose. The rationale for utilizing logistic regression lies in its ability to provide classification properties, in addition to displaying the contribution of each feature to the target along with its p-value. The 28 extracted features were fed into the model using the backward elimination method to iteratively select the most significant ones. Following the designed sessions of this experiment (refer to Figure III.3.2), comparisons were made across all three sessions, as well as between each session and the other two separately, similar to the Post-Hoc tests in the subjective analyses.

The aim was to provide a comprehensive overview of the impact of the type of work instruction on physiological features. A multinomial logistic regression model was used, setting the baseline session as the reference category for the visual- and code-based sessions. This means that the features of both instruction sessions will be compared to the baseline session. The fitness of the model was assessed using the Chi-Square test, which revealed a significant improvement over the null model with Chi-Square = 270.503, $d = 30$, and a $p-value < 0.001$. The backward elimination method removed 14 features and selected the top 15 contributing features, resulting in the optimal model classification parameters. Table III.3.3 presents a list of the selected features, as well as their coefficient magnitudes and p-values.

Three binary logistic regression models were also implemented to compare each session with the others and identify which features contributed significantly to the target. The models' fitness was once again assessed using the Chi-Square test. The results revealed a significant improvement in the models compared to the null models without predictors. The visual-based vs. baseline model has shown Chi-Square = 140.358, $d = 12$, and a $p - value < 0.001$, the code-based vs. baseline model has shown Chi-Square = 176.691, $d = 13$, and a $p - value < 0.001$, and the code-based vs. visual-based model has shown Chi-Square = 91.401, $d = 14$, and a $p - value < 0.001$. Table III.3.4 presents the results of the selected features with their coefficient magnitudes and significance evaluation parameter, p-values.

Table III.3.3: Estimated parameters for selected features using multinomial logistic regression with baseline session as reference

Condition	**Features**	**B**	**p-value**	**Condition**	**B**	**p-value**
	SC Area	0.007	< 0.001		0.007	< 0.001
	#No.Peaks	0.295	< 0.001		0.133	0.049
	Rising Time	0.238	0.11		0.338	0.017
	Decay Time	- 0.0001	0.957		-0.228	0.006
	STD SC	-17.15	< 0.001		-13.008	< 0.001
	STD SCR	29.078	< 0.001		27.274	< 0.001
	SCR Power	-0.215	< 0.001		-0.389	< 0.001
Visual-based	MEAN_RR	-0.018	0.006	Code-based	-0.028	< 0.001
	SDRR	-0.147	< 0.001		-0.015	0.634
	SDSD	0.139	< 0.001		0.049	0.114
	SDRR_RMSSD	2.708	0.028		0.068	0.953
	HR	-0.218	0.01		-0.344	< 0.001
	pNN25	-0.019	0.301		-0.059	< 0.001
	LF_HF	-0.004	0.034		-0.006	0.001
	HF_LF	0.544	0.784		1.809	0.362

The reference category is the Baseline session; B: Coefficient Magnitudes

Table III.3.4: Estimated parameters for selected features using three binary logistic regression models.

Condition 1	**Features**	**B**	**p-value**	**Condition 2**	**Features**	**B**	**p-value**	**Condition 3**	**Features**	**B**	**p-value**
	SC Area	0.006	0.003		AreaSCR	0.465	< 0.001		#No.Peaks	-0.158	< 0.001
	#No.Peaks	0.164	0.005		#No.Peaks	0.165	0.010		Decay Time	-0.200	0.032
	STD SC	-13.921	0.001		STD SC	-9.952	0.006		STD SC	8.201	0.001
	STD SCR	22.950	< 0.001		SCR Power	-3.746	0.005		STD SCR	-8.528	0.016
	SCR Power	-0.177	0.006		MEAN_RR	-0.031	< 0.001		SCR Power	-1.177	0.016
Visual-based	RMSSD	0.091	< 0.001	Code-based	SDSD	0.092	< 0.001	Code-based	MEAN_RR	-0.078	0.008
vs.	MEDIAN_RR	-0.011	0.027	vs.	HR	-0.753	0.003	vs.	SDSD	-0.056	< 0.001
Baseline	SDRR	-0.054	0.003	Baseline	pNN25	-0.047	0.034	Visual-based	SDRR_RMSSD	-2.664	< 0.001
	HR	-0.129	0.025		pNN50	-0.095	0.020		HR	-0.546	0.005
	pNN25	-0.048	0.023		KURT_RR	-0.093	0.007		pNN50	-0.051	0.003
	KURT_RR	-0.072	0.087		VLF	0.000	0.048		SD2	0.076	< 0.001
	LF_HF	-0.003	0.117		LF_HF	-0.005	0.029		LF	0.000	0.015
									HF_LF	1.543	0.002

Baseline is the reference category for Conditions 1 and 2, while Visual-based is the reference category for Condition 3; B: Coefficient Magnitudes

Negative coefficients suggest that as the predictor (a specific feature) increases, the likelihood of the outcome being in the respective condition (target) decreases compared to the reference category.

By comparing the three sessions in this study, the average of the models' performance parameters—*accuracy*, *precision*, and *recall*—was calculated and presented in Table III.3.5. The following abbreviations were used for each session: B for Baseline (Pre-experiment), V for Visual-based instruction, and C for Code-based instruction. The high-performance classification metrics (B—C, B—V) showed that the models found a clear boundary between the baseline session (Pre-experiment) and both the visual-based and code-based sessions. This pattern aligns with the trends observed in the ANOVA and Post-Hoc tests of the DSSQ

TABLE III.3.5: Average of performance metrics of logistic regression models under various conditions based on the physiological features (GSR and HRV).

Classifier	Conditions	Accuracy	Precision	Recall
Multinomial Logistic Regression	B — V — C	78.04	67.79	63.12
Binary Logistic Regression	B — V	83.88	82.99	82.2
	B — C	**90.42**	**89.97**	**85**
	V — C	75.51	74.92	71.95

B: Baseline (Pre-experiment), V: Visual-based session, and C: Code-based session

psychological states: engagement, distress, and worry. Although all selected features of condition 3 (Code based vs. Visual based) in Table III.3.4 demonstrated significant differences, model V—C in Table III.3.5 exhibited the lowest performance metrics when compared to the other binary classifier models. This is somewhat consistent with the previous statistical tests that compare these two sessions within the subjective DSSQ states.

To analyze participants' performance, three key metrics were considered: the Number of Task Repetitions (NTR), Task Completion Time (TCT) (with a minimum of five minutes), and the Assembly Precision (AP). This precision was represented by the average of the standard deviation (SD) of Euclidean distances between the centers of the building pieces, derived from video-captured images; lower SD values indicated better AP. The camera setup was calibrated, and trials within each instruction session were averaged to minimize potential algorithmic inaccuracies. Shapiro–Wilk tests were conducted to assess the normality of the data, followed by paired t-tests to evaluate differences between sessions.

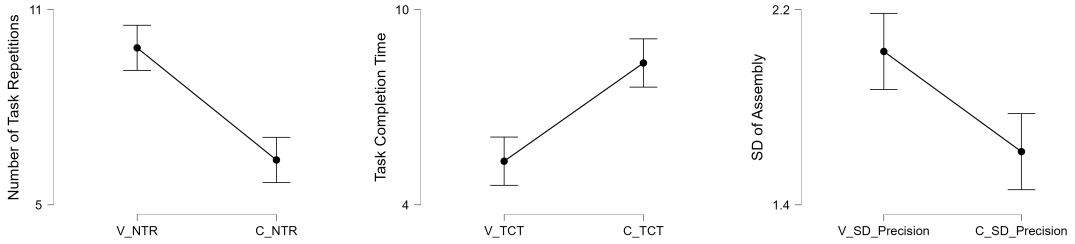
Table III.3.6 presents the main parameters extracted from these tests. Significant differences were observed in the three parameters between the two instruction sessions, with $p - values < 0.001$. Figure III.3.7 presents the descriptive plots of these three parameters as means with their 95% confidence intervals. Figure III.3.7(a) presents the means of the NTR during the two sessions of the work instructions. In the code-based session, most of the participants found themselves stuck at the minimum number of iterations ($Mean = 6.379$), whereas they showed a greater capability to repeat the task in the visual-based session ($Mean = 9.828$).

Despite the higher number of repetitions in the visual-based session, the majority of participants did not exceed the allocated time for this session and showed a mean of 5.342 minutes compared to 8.363 minutes in the code-based session (refer to Figure III.3.7(b)). These results are aligned with the subjective results from

TABLE III.3.6: Comparison of participant performance metrics between Visual- and Code-based sessions using paired t-tests.

Measure 1	**Measure 2**	Shapiro-Wilk p-value	**Test Type**	**Z**	Effect Size	t-test p-value
V_NTR	C_NTR	0.027	Wilcoxon signed-rank	4.286	1.00	< 0.001
V_TCT	C_TCT	0.018	Wilcoxon signed-rank	-4.573	-0.972	< 0.001
V_SD_Precision	C_SD_Precision	0.3	Student	N/A	0.709	< 0.001

V_NTR and C_NTR are the Number of Task Repetitions in the visual- and code-based sessions, respectively. V_TCT and C_TCT are the Task Completion Time (in minutes) of the visual- and code-based sessions, respectively. V_SD_Precision and C_SD_Precision are the SDs that reflect the Assembly Precision (AP) for the visual- and code-based sessions, respectively.



(a) V_NTR vs. C_NTR. (b) V_TCT vs. C_TCT. (c) V_SD vs. C_SD.

FIGURE III.3.7: Comparison of task performance metrics between visual-based (V) and code-based (C) instruction sessions, showing (a) number of task repetitions, (b) task completion time, and (c) assembly precision (SD of assembly). Error bars represent 95% confidence intervals.

NASA-TLX, where participants showed a significant increase in the cognitive load from visual-based instructions to code-based instructions. Figure III.3.7(c), however, showed intriguing results with lower SD values for code-based instructions (which means better assembly precision AP) compared to visual-based instructions. This does not align with the subjective results of the NASA-TLX performance category, where participants evaluated themselves as having better visual instructions performance.

The higher number of task repetitions NTR and lower task completion time TCT in the visual-based session indicate that participants made more hand movements in this session compared to the code-based session. Table III.3.7 shows the t-test analyses of the hand movements in the three coordinates (X, Y, Z) during the two instruction sessions. The p-values from these analyses (< 0.05) indicate significant differences in hand movement across the three coordinates during the visual-based session compared to the code-based session, highlighting variations in NTR and TCT between the two sessions.

Combining acceleration features with the physiological features (GSR and PPG) will establish a clear boundary between the two instruction sessions. Feeding these

TABLE III.3.7: Results of t-test analyses comparing the mean accelerometer values across (X, Y, Z) during the visual-based and code-based sessions.

Visual-based	**Code-based**	**Test Type**	**Z**	**p-value**
meanX	meanX	Wilcoxon signed-rank	2.411	0.016
meanY	meanY	Student	N/A	0.002
meanZ	meanZ	Wilcoxon signed-rank	2.038	0.042

meanX, meanY, and meanZ are the mean values of the accelerometer data at the three coordinates X, Y, and Z, respectively.

52 features into the binary logistic regression model with the backward elimination method has produced promising results. The fitness of the model was again assessed using the Chi-square test, which revealed a significant improvement over the null model, with Chi-Square = 368.234, $d = 27$, and a $p - value < 0.001$. The model showed excellent performance metrics with average $accuracy = 92.91$, $precision = 92.67$, and $recall = 92.35$. These values outperform the model performance in Table III.3.5 V—C condition.

III.3.4 Discussion

This study investigated the impact of work instruction methods on the human cognitive load and their operational efficiency. In a controlled, assembly-like scenario inspired by industrial tasks, the study used two work instructions—visual-based and code-based—and a range of subjective and objective assessment methods. The alignment between subjective and objective evaluation methods was also examined to enhance the accuracy of conclusions by providing context for physiological responses and validating the experimental conditions.

The findings revealed that code-based instructions imposed a higher subjective cognitive load on participants compared to visual-based instructions. This aligns with Cognitive Load Theory (CLT), which posits that extraneous cognitive load—stemming from the way information is presented—can hinder learning and performance [12].

The results are also consistent with previous studies indicating that visual aids can enhance comprehension and reduce cognitive load in assembly tasks. For instance, Li et al. (2018) found that supporting the work instructions with pictures can reduce the cognitive load and improve task performance compared to the traditional text instructional methods [86]. Similarly, the study suggests that

visual-based instructions lead to faster task completion and higher task repetition rates. This is likely due to the reduced mental effort required to interpret the instructions, as visual-based instructions are less abstract and easier to interpret than code-based instructions. Furthermore, Vanneste et al. (2024) demonstrated that augmented reality (AR) visual instructions led to lower assembly times and a lower perceived physical effort compared to traditional methods [165]. This supports the idea that technologically advanced visual aids can further enhance the effectiveness of work instructions, which aligns with the findings of this study on the superiority of visual-based instructions in most cases.

Taking each of the six categories in the NASA-TLX and comparing them between the two instructional sessions has produced profound results. The t-test analyses of each pair of the six categories within the NASA-TLX have shown a significant increase in mental demand, frustration, and effort in the code-based session. While there was a slight increase in physical demand in the code-based session, there was no significant increase. Conversely, the t-test analyses of hand movements in the X, Y, and Z coordinates, as well as the NTR, indicated higher means and significant differences in the visual-based session compared to the code-based session. As the hand movements were not exertive, participants focused on their goal of repeating the task during the visual-based session, where a higher repetition rate was intended to lead to better outcomes. Consequently, they did not perceive the task as physically demanding when filling out the Physical Demand category of NASA-TLX.

However, body movement significantly impacts physiological signals due to the alterations in the autonomic sympathetic arousal resulting from increased energy expenditure [176]. Subjectively, the code-based instructions were more cognitively demanding. These instructions were also expected to influence the objective physiological data. On the other hand, although visual-based instructions posed subjectively lower cognitive demands, their straightforward nature objectively led to more hand movements. The higher hand movements were expected to impact the physiological responses. These were clearly reflected in the performance metrics of the logistic regression models in Table III.3.5. Classifying the code-based instruction session from the baseline session yielded the highest performance metrics, followed by the comparison of the visual-based to the baseline. Both cognitively demanding tasks, and tasks involving body movements significantly influence physiological signals, justifying this. Simultaneously, when the code-based

and visual-based sessions were classified, the logistic regression models displayed relatively low performance metrics because both tasks influenced the physiological signals.

The low-performance metrics for classifying the two sessions based on physiological data do not necessarily indicate a lack of alignment between the subjective and objective data metrics. However, they do imply that differentiating operators' conditions using the objective physiological data may not be entirely reliable, especially in scenarios combining cognitive and physical tasks. On the other hand, supporting the physiological features (GSR and HRV) by the acceleration features has provided a clear boundary between the two instruction sessions. This is due to the higher levels of hand movements in the visual-based session. This supports the use of these kinds of data in conjunction with other objective data to support operator condition analyses.

Performance was also used as a metric in this study. Participants were informed about the criteria for evaluating their performance prior to the experiment. This metric was analyzed in two ways: subjectively using the NASA-TLX, and objectively using the parameters in Table III.3.6—the number of task repetitions (NTR) within the given time, the task completion time (TCT), and the assembly precision (AP), represented by the standard deviation (SD) of the Euclidean distances between the assembled pieces.

Participants subjectively rated their performance significantly higher in the visual-based instruction session. Participants seem to prioritize the possibility of repeating the task beyond its lower limit, disregarding the precision of their work. This higher repetition number gave them a sense of achieving their task with high performance in the visual-based session compared to the code-based session. The objective performance metrics aligned the subjective rate with respect to the NTR and TCT, as shown in Table III.3.6 and Figures III.3.7(a) and III.3.7(b). In most cases, participants repeated the task significantly more during the visual-based session without exceeding the allocated time.

However, while the industry aims to increase production batches with short production times, it does not overlook the importance of product quality. In this study, the SD of the assembly process represented this metric. Due to their increased focus on the NTR, participants did not pay as much attention to their assembly precision. This resulted in a higher SD for the visual-based session,

indicating lower AP compared to the code-based instruction session, where participants thoughtfully assembled each piece without rushing through the process (see Figure III.3.7(c)). This suggests that while visual instructions may enhance speed and reduce perceived effort, they may inadvertently encourage less attention to detail. This objective metric is primarily not aligned with the subjective performance in the NASA-TLX. This contrast highlights the strengths and limitations of each measurement approach: subjective tools, such as NASA-TLX, can capture perceived workload or satisfaction, but they might miss more complex aspects of actual task performance. Objective measures like assembly precision provide quantifiable outcomes but do not fully account for internal states such as confidence or perceived effort.

The mismatch between the objective and the subjective metric of performance might be understood within the framework of the Dynamic Model of Sustained Attention and Stress [177]. According to this model, individuals adjust their attention and effort allocation dynamically based on perceived task demands, available cognitive resources, and stress level. Thus, when task demands decrease, attention can become less focused, leading to a drop in task performance. Therefore, while visual-based instructions optimize speed and effort, they may not sufficiently maintain the level of attention needed for precision.

These findings highlight the importance of achieving the optimal cognitive load during tasks. Although visual instruction can reduce cognitive load and increase assembly speed, it may not result in optimal cognitive performance in terms of precision. This indicates that while visual instruction may lower cognitive load and enhance efficiency, it might compromise attention and precision. Thus, in high-stakes or precision-demanding tasks, a certain level of cognitive load might be necessary to ensure attention and accurate performance. Therefore, instruction design should consider not just reducing cognitive load but also achieve optimal cognitive load that supports both efficiency and precision, optimizing overall task performance.

Future research can explore hybrid or modified methods to mitigate this trade-off. For instance, adaptive or context-sensitive instructions could primarily use visual aids for most assembly steps, yet incorporate code-based details during critical high-precision tasks. Alternatively, layered instructions—where a simple visual overview is supplemented by optional, more detailed code-based guidance—could preserve the clarity of visual methods while ensuring precision where it is needed.

Such approaches might achieve a more optimal balance between efficiency and precision without overstressing the operator's cognitive resources.

III.3.5 Limitations and Future Research

This study faced several limitations that suggest directions for future research. One key limitation of this study is that the experiment was conducted in a controlled laboratory environment, which does not fully mirror the complexity of real-world industrial settings. In actual production lines, factors like noise, teamwork, multitasking, and real-time pressures can substantially influence the operators' cognitive load and performance. Consequently, the results presented here should be interpreted as foundational insights rather than direct predictions of on-site outcomes. Nonetheless, the findings highlight the importance of minimizing extraneous cognitive load in the design of effective work instructions. Future research may further validate these insights by integrating realistic workplace parameters, such as time constraints, loud machinery, and group-based tasks, into experimental protocols.

Another limitation of this study is the sample representativeness, where the participant pool was restricted to university students and researchers. While this homogeneous sample allowed for consistent baseline characteristics, it may not adequately represent the demographic and experiential diversity of industrial workers. Therefore, caution is warranted in generalizing the findings to actual industrial environments. In future work, the sample will be broadened to include operators from various industrial settings. This expanded approach will help validate the current results and further refine guidelines for optimal instructional design.

Additionally, physiological sensor placement on the non-dominant hand limited task execution to one-handed, which was another limiting factor in this study. This constraint potentially affected both the pace and strategies used, reducing the ecological validity of the findings. In future studies, less intrusive sensor placements (e.g., wearable wristbands, arm and chest straps, or forehead sensors) will be adopted to enable two-handed operation and better replicate industrial conditions.

Moreover, the study utilized a limited set of physiological signals (GSR and PPG). Incorporating a broader array of biosignals, such as eye tracking, body motion or posture tracking, electromyography (EMG), electroencephalography (EEG), and electrooculography (EOG), can provide deeper insights into the cognitive and

physical states of workers, offer more robust support for the study's hypotheses, or even provide a different point of view. Furthermore, while the sample size of 30 participants was substantial, future studies can expand it to enhance the statistical power and generalizability of the findings.

Finally, the five-minute time limit for the instruction sessions, which could only be extended if the specific pattern was not repeated three times, restricted most participants to completing the code-based session only three times. This limitation prevented the examination of the full learning curve. It is possible that with more practice, participants could become more efficient with code-based instructions, potentially improving task performance over time.

III.3.6 Conclusion

In this study, it was found that visual-based instructions significantly reduce cognitive load and improve certain operational aspects, such as shorter TCT and higher NTR, compared to code-based instructions. However, the findings show a clear divergence between participants' subjective ratings of performance through the NASA-TLX and the objective performance metric, assembly precision AP. While subjective measures are valuable for gauging perceived workload and emotional states, they can be influenced by factors like self-efficacy and momentary satisfaction. Conversely, the objective precision metric provides a direct measure of actual task outcomes but may overlook internal experiences of strain. As a result, high subjective performance scores did not always correspond to high objective precision.

The study suggests that simple and direct instructions (visually based in this study) can enhance some of the operational aspects and reduce cognitive load, demonstrating that these kinds of instructional strategies are particularly beneficial in environments where quick task execution is critical. On the other hand, for tasks that require high precision and meticulous attention to detail, instructions that require deep thinking (code-based in this study) may be more appropriate. This discrepancy underscores the importance of a multi-method approach.

Chapter III.4

Attentional Multitasking in Human–Robot Collaboration

III.4.1 Introduction

Collaborative robots (cobots) have been designed to work alongside human operators and promise productivity gains and increased flexibility [178]. The integration of cobots in manufacturing and assembly prompted a new line of research on human-robot collaboration (HRC), where humans work together with robots in a shared workspace. Prior works often exemplify the value of HRC as the optimal combination of automation to handle physical tasks while utilising cognitive skills of the human agent [179, 180]. In theory, the robot taking over labourious manual tasks should free the human to engage in additional cognitive tasks. Therefore, multitasking in HRC has been proposed to further increase productivity [181]. Such applications can include simultaneous HRC and quality control, or working together with multiple cobots at the same time. However, multitasking requires the operator to split their attention among multiple tasks, which can in turn increase the load on the operator and reduce efficiency. Although cognitive load has already been studied in HRC [182], research on attentional factors is still lacking. Overall, depleting attentional resources through improper design of HRC applications might result in overseeing errors or reduced awareness of safety risks.

In this chapter, the feasibility of attentional multitasking in collaborative human-robot assembly is investigated. An exploratory study was performed in which participants carried out a wire harnessing task with a cobot while simultaneously engaging in a parallel attention-demanding task through a Go/No-Go test. To evaluate the effects of multitasking, the Go/No-Go test was designed with two levels of difficulty in terms of attentional demands. A user study was conducted with 16 participants, and quantitative metrics on task performance and response rates, as well as qualitative feedback, were gathered to evaluate the ability to engage in secondary attentional tasks.

The experiment suggests that multitasking scenarios may lead to higher cycle times and potentially even increased errors, and operators might be prompted to adapt to the attentional load by prioritising only one of the tasks. Although multitasking in HRC is considered feasible, concerns are raised about potential effects on productivity, and future research is encouraged to focus on designing HRC applications that do not deplete attentional resources.

III.4.2 Research Background

III.4.2.1 Multitasking

Multitasking is defined in cognitive psychology as the capacity to manage more than one task at a time. Multitasking can be simultaneous execution of more than a single task[114], a switching between multiple tasks that execute in parallel [115], or a combination of both in which frequent switching of attention is required. This capacity is essential in industrial environments, as individuals often handle multiple tasks, impacting both their cognitive workload and task performance. Theoretical frameworks, such as multiple resource theory [118], propose that the effectiveness of multitasking depends on the cognitive resources needed for the activities. There is evidence that separate perceptual modalities follow independent attentional capacities [183, 119, 120], but recent results also suggest tasks requiring central attention based on perceptual input nevertheless share attentional resource; both unimodal and bimodal dual-tasks lead to increased overall load, with equivalent costs following increased task difficulty[121].

The theory of threaded cognition [122] goes even further, suggesting that when performing multiple tasks, several mental processes run in parallel, with limitations imposed by resources such as attention and working memory. These theories explain how competing for limited mental resources when performing multiple tasks might make it difficult to focus on each task at a high enough level to ensure good performance [184, 185, 186], with possible detrimental effects leading to errors and accidents [187, 188].

III.4.2.2 Attention and Awareness in HRC

The demands of attention and cognitive load are significant factors that impact operators' situational awareness in industrial environments [189, 190]. The concept of situational awareness, as defined by Endsley, refers to the "perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future" [191]. In the context of HRC, this translates to an understanding of the cobot's actions,

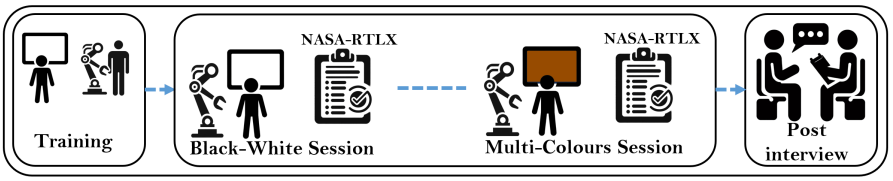


FIGURE III.4.1: Overview of the setup and the procedure of the study. The task order was counterbalanced, with 8 participants starting with the black-white condition, and 8 starting with the multi-colour session.

intentions, and the overall state of the manufacturing process by the human operator. Prior research has highlighted the significance of operator awareness for task performance, safety, and productivity in HRC. Liu and Wang conducted a study investigating the impacts of awareness on manufacturing work, demonstrating that enhanced awareness leads to more efficient and safer interactions with robots in a manufacturing environment [192]. Additionally, the influence of cognitive fatigue on task performance and situational awareness in HRC was investigated in ref [193]. Their findings suggest that attentional demands have a substantial effect on cognitive fatigue, which in turn affects situational awareness. A study performed in ref [194] investigated the correlates of visual attention measured by an eye tracker, situational awareness, and performance in HRC. Their results highlight the significance of attentional factors, with visual attention metrics being the main predictor for performance and awareness in HRC. In multitasking scenarios, operators are required to continuously adapt to changing tasks and the robot behavior, hence posing a high load on their attention and awareness. However, it is still unclear how attentional demands impact the error rates and the workload imposed on the operator in multitasking HRC scenarios.

III.4.3 Experiment

This experiment was designed to divide participants' attention, providing a realistic assembly scenario where the participant must balance the attentional load of the main task while also responding to the demands of the secondary task to simulate multitasking. The main task involved working on a wire harnesses in collaboration with a UR5e cobot, while the secondary task involved a Go/No-Go test to impose increased attentional demands. The experimental study was carried out at the Industry 5.0 laboratory at the University of Pannonia [195].

III.4.3.1 Participants

The study investigated attentional multitasking during collaborative assembly with a cobot, focusing on task completion time, main-task errors, secondary-task miss and error rates, and perceived workload and effort (NASA RTLX). A voluntary cohort of 16 participants was recruited from the university environment (students and researchers) and enrolled based on willingness to participate. The sample included 12 males and 4 females, aged 21 to 42 years ($M = 30$, $SD = 6.16$). All participants received written study information and provided written informed consent prior to participation. A priori sample-size estimation was not conducted; this is acknowledged given the exploratory nature of the study and the limited prior evidence needed to specify an expected effect size for power planning.

III.4.3.2 Design and Procedure

To assess how multitasking affects HRC, two levels of difficulty were introduced for the secondary task while the main task was kept constant. The study used a within-subjects design, where participants performed two sessions, one for each difficulty level of the secondary task (Figure III.4.1). To mitigate order effects, the sequence of the sessions was counterbalanced for each participant. Moreover, training sessions were introduced before the experiment to further reduce potential order effects. First, the participants were introduced to the secondary task, which required participants to react to changes in screen colour through pressing a pedal on the floor. The frequency of these changes mirrored the conditions of the actual experiment; however, the training task was designed in a different manner than the task encountered in the experiment. Subsequently, the participants were introduced to the main HRC task, which involved wire harnessing in collaboration with a cobot. During this phase of training, participants practiced the wire harnessing procedure and interacting with the cobot. Besides order effects, the purpose of these training sessions was to ensure that participants were comfortable with the tasks, and minimise potential differences in dexterity.

Main task: During the main task, the cobot held a cylindrical hub that served as the central component of the task. The cobot swiftly rotated the hub to indicate the next assembly step to the operator, assisting in the assembly process. This rotation aligned the terminal block connectors, making it easier for the participant

to continue the activity without any interruptions. The assembly included 24 wires with 12 emerging from the rear side and 12 from the front side of the cylindrical hub. These wires were to be connected depending on their label. The label was indicated by a combination of symbols and shapes. The labels simulated part numbers, which are commonly used in industrial assembly to recognize corresponding assembly components. To introduce a challenging aspect to the experiment, the colours of the wires were intentionally varied.

The task involved three main steps: (1) Selecting one of the wires extending from the back of the cylindrical hub and attaching it to the terminal block connector using a screwdriver. (2) Identifying the correct counterpart wire based on the symbol and shape combination and attaching it to the terminal block connector using the screwdriver. (3) Verifying the integrity of the connection.

The possible mistakes in the primary task included mismatching of symbols and shapes, unintentional insertion into an adjacent terminal block connector rather than the designated one, insecure wire connections susceptible to detachment, or the total absence of a wire connection.

Secondary task: While the participants were engaged in the main task, a secondary task was introduced to evaluate their attentional capacity. The secondary task was deployed in form of a Go/No-Go test on a screen positioned in front of the participants. To evaluate the influence of attentional multitasking, two conditions with different levels of attentional load were designed. The first condition was designed as a Go/No-Go test with a lower level of difficulty in terms of attentional demands. The test consisted of a white screen, with a black screen randomly appearing for two seconds within 15-second intervals. The timing of the stimulus was not predetermined and did not follow a regular pattern during the session, making it unpredictable. Hence, participants needed to constantly focus on the attentional test, while also focusing on the human-robot assembly. Each time when the stimulus in form of the black screen was presented, the participant was required to perform a reflexive action by pressing a pedal with their foot. This task assessed participants' ability to sustain attention and respond effectively within the requirements of the main task.

In the second condition, the same test was implemented but with a higher level of attentional demand, primarily influenced by the increased frequency of stimuli

presentation. Three different colours — grey, brown, and white — served as No-Go stimuli, with the Go stimulus again being a black screen. Each of the No-Go colours appeared briefly and randomly for two seconds. Once every 15 seconds, a Go stimulus was presented, with random timing. This experiment required the participants to sustain a higher level of attention on the frequency of the stimuli, as they had to differentiate between Go and No-Go stimuli. The responses were, again, registered via a pedal on the floor.

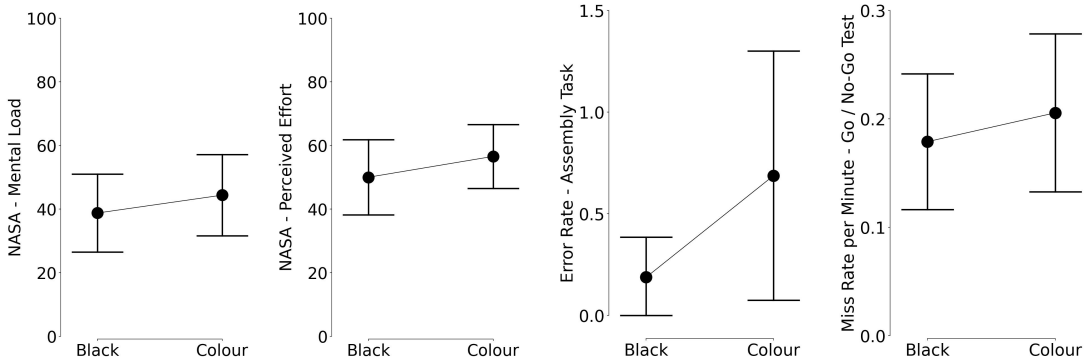
The possible mistakes in the secondary task included failing to react to the desired colour, or incorrectly responding to a different colour. These possible mistakes were tracked by a code used to build the secondary task. Counterbalancing the conditions was achieved by splitting the participants into two groups, with eight participants starting the experiment with the simpler (black) condition, and eight participants starting with the harder (coloured) condition (Figure III.4.1).

III.4.3.3 Data and Analysis

A quantitative analysis was performed on self-reported data collected through post-experiment questionnaires, as well as on objective metrics. The objective metrics included task completion time (TCT), the number of errors in the main task, the number of times the participant missed the stimulus in the secondary attentional tasks, and the number of errors in the secondary task. The subjective metrics were collected through the NASA RTLX questionnaire. The analysis was conducted using a two-tailed paired samples t-test. Due to the exploratory nature of the study and the low sample size, descriptive statistics were also reported as supplementary metrics.

Additionally, the quantitative results were complemented by a qualitative analysis conducted through post-interviews. Participant selection for the interviews was based on a manual data analysis, through which seven participants were identified and asked further questions regarding their perception of the differences between the two conditions.

For this study, it was assumed that the two conditions would exhibit a difference in mental workload measured by NASA RTLX. This hypothesis was based on the theory of threaded cognition, which suggests that increased attentional demands contribute to higher cognitive load. Additionally, it was hypothesized that there



(a) Mental load. (b) Perceived effort. (c) Assembly errors. (d) Miss rates.

FIGURE III.4.2: Comparative plots of the mean and the confidence intervals ($CI = .95$) for the two conditions and per variable.

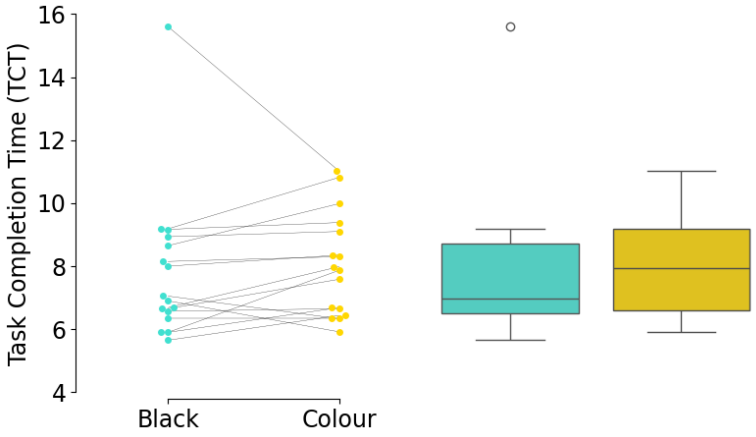


FIGURE III.4.3: Swarmplot and boxplot of the differences in task duration per condition.

would be differences in the objective metrics, while the overall workload would remain statistically equal.

III.4.4 Results

III.4.4.1 Quantitative Analysis

First, the differences in the perceived mental load between the two conditions were analyzed. Student's t-test showed a weakly significant effect with $t(15) = -2.087$, $p = .054$, and Cohen's d of -0.522. The analysis also revealed a weakly significant difference in perceived effort between the two conditions ($t(15) = -1.787$, $p = .094$, Cohen's d $= -0.447$). The descriptive plots of the two variables are

depicted in Figures III.4.2(a) and III.4.2(b). This indicates that the colour condition was perceived as cognitively harder, with a small to medium effect size. No other subjective metrics from the NASA RTLX questionnaire showed a significant difference. This correlates with the initial expectation, as the experiment was controlled for increased attentional load, while the overall workload for the task remained the same.

Regarding the performance metrics, assembly errors were analysed using the Wilcoxon signed-rank test (Shapiro-Wilk test, $p - value < .05$), which yielded no significant differences ($Z = -1.618$, $p = .134$). Although not significant, as the majority of the participants did not perform any errors in the assembly task, the descriptive statistics (Figure III.4.2(c)) indicate a possible trend towards more errors during the secondary task in the colour condition. To analyze the performance of the secondary tasks, the miss rate, i.e. the number of times the participant did not react to the Go-stimulus, was investigated. As the number of stimuli in the experiment depended on task duration, the variable was scaled by the experiment duration to ensure comparability. The test did not reveal any significant differences ($t(15) = -1.117$, $p = .282$). The descriptive results of both errors rates are depicted in Figures III.4.2(c) and III.4.2(d).

Finally, the effect of the two conditions on task completion time (TCT) was analyzed. As the black condition contained a strong outlier and the Shapiro–Wilk test yielded a $p - value < .05$, the Wilcoxon signed-rank test was again applied to analyze the data. The results displayed a Z-score of -1.647 ($p = .105$). Despite non-significant results, the effect size of -0.483 (rank-biserial correlation) and the descriptives provided in Figure III.4.3 indicate a trend towards more time needed for the colour condition.

III.4.4.2 Qualitative Analysis

One week after the experiment, selected participants were invited for a post-interview. This included random participants, selection based on observation during the experiment, or participants whose data contradicted the trend, such as in the case of the participant with a very long task duration in the black condition (Figure III.4.3). The statements from the participants were analyzed, and the data were clustered into three themes:

- *Strategy Adaptation* - the split of the attention forced some participants to reduce their focus on one task and prioritise the other to avoid mistakes. This was, for example, the explanation for the outlier TCT in Figure III.4.3. In this case, the participant was overly focused on the secondary task not to miss the stimuli, which lead to a high completion time of the main task.

- *Learning Effect* - three participants mentioned that they got more comfortable with the experiment over the time, and thus they made less mistakes and also perceived the second session as easier, disregarding whether they experienced the black or colour condition in their second experiment.

- *Subjectivity and Fatigue* - some participants perceived one of the conditions easier, despite making more mistakes than in the condition they perceived as harder. Moreover, two participants gave different statements with regard to their perception of the two conditions, with the responses from the interview contradicted their NASA metrics. Through further questions, they attributed this to fatigue.

III.4.5 Discussion

Prior works on HRC in manufacturing often assume that humans can effortlessly transition to cognitive tasks while robots assist them with labourious manual tasks. However, engaging in collaboration with a robot requires attention, which can deplete attentional resources required for other tasks. The impact of attentional multitasking in HRC on productivity was studied, and the potential disadvantages were explored. Although not statistically significant due to the exploratory nature of the study and low sample size, the results indicate that workers are required to split their attention to manage multiple tasks. This can potentially reduce productivity, as evidenced by a trend towards more errors in assembly tasks. Additionally, while the miss rate in secondary tasks did not show significant differences, the descriptive results suggest that attentional multitasking could transition into workers overseeing quality-related issues and their awareness of the environment and work can be reduced. In turn, this can pose safety related issues. Moreover, potential implications for task completion time were observed, with a trend toward increased duration in conditions where the secondary task posed increased cognitive demands. These findings underscore the importance

of considering the impacts of attentional requirements in HRC to optimize task performance and efficiency.

The results also show that, when multitasking in HRC settings, participants may adapt their strategy and prioritise one task over another, leading to more errors in the respective task. It is speculated that this may be either due to the inability to split attention between multiple tasks or a preference for maximizing efficiency in one task while sacrificing efficiency in another. Preferred strategy for multitasking tends to converge towards that of minimal interference to either task, but the process of finding the optimal solution is not automatic, and is not always observed across all participants [196]. Participants' preferred strategy itself may have impact on performance and effort, regardless of task prioritization instructions[197].

The results of the experiments are aligned with the threaded cognition theory, which suggests that cognitive capacity might be reduced due to HRC depleting their attentional resources. It is important to consider real-life examples of such attentional multitasking such as concurrent HRC and quality control, supervision of and collaboration with multiple robots, or being able to flexibly react to short-term interruptions or a problem in the process. Instead of dichotomously adding up the physical capabilities of the robot with the cognitive skills of the operator, humans collaborating with robots are proposed to be viewed as a blend of cognitive constraints influenced by the robot's physical actions. This perspective challenges the notion of clear-cut boundaries between the agents and calls for a reevaluation of claims related to HRC and productivity gains. Designing HRC applications requires a more nuanced understanding of how attentional and cognitive resources are used, yet, empirical research on such applications in manufacturing is missing.

Finally, two design implications are presented based on the experimental study. First, it is emphasized that the integration of collaborative robots should not lead to the exploitation of the limits of cognitive resources by placing additional tasks on workers. This study indicates that potential consequences might include increased cognitive load and perceived effort, as well as reduced productivity. Still, multitasking can be feasible in opportune moments, such as at times when the engagement in the HRC task is reduced. Transferring this into HRC applications requires an improved communication of robot's intent to the human, indicating when the human agent can reduce their awareness of the robot movement and focus their attention on other tasks. Alternatively, multitasking in HRC could be supported by attention management systems, which have been shown to limit

attention fragmentation [198]. Second, the design of HRC applications involving parallel, simultaneous multitasking is discouraged, and a sequential approach is proposed instead. This involves a clearer definition of task boundaries, for instance, by task allocation and scheduling algorithms, mitigating the risk of having to prioritise one task over the other.

III.4.5.1 Limitations and Future Work

The primary constraint in this experiment was the limited sample size due to the exploratory nature of the study. Given the relatively low effect size, there is need to adequately test the generalizability of the results. As such, the findings are intended to be reevaluated with a larger sample size study. The results also indicate the presence of a possible floor effect. This suggests that the main task may not have been sufficiently difficult to fully evaluate the whole spectrum of multitasking abilities. Performance effects from multitasking are more apparent when the main task is deemed difficult by the participants[199]. As such, the main task difficulty and study design will be further refined to better emulate multitasking in real-life manufacturing conditions.

Furthermore, based on the results, there may be evidence of an interaction between task completion time and errors in both the main and secondary tasks. This might provide additional insights into how performance metrics are interrelated in multitasking scenarios. However, the study design did not allow this investigation to be performed. For example, manipulation of stimulus onset asynchrony and task order in dual-tasks have been reported to affect response times, while having no effect on error rates [200]. It should be noted however that a common criticism of dual-task design lies in the difficulty of replication due to specificity and customized nature of instruction and task design; parameters such as the choice of modality in each task can influence observed cognitive load, leading to different findings between studies investigating similar qualities [201]. To ensure replicability of the results, clear definition of both tasks' modalities as well as correct definitions of assumptions on measurement of cognitive load is important. Additionally, future studies should compare these multitasking scenarios with non-multitasking conditions to isolate the specific contributions of concurrent task management to cognitive load.

III.4.6 Conclusion

In this work, attentional factors in industrial HRC were explored. Through the experimental design, the impact of attentional multitasking with two different levels of difficulty on productivity was observed. The results show that secondary tasks with increased attentional requirements may lead to decreased productivity in the form of higher error rates and longer task durations. The findings underscore the importance of considering attentional demands in the design and implementation of collaborative human-robot applications in manufacturing. Moving forward, this exploratory study aimed to address a gap in empirical research on human factors in HRC. Design recommendations are presented for developing applications aimed at mitigating potential productivity decreases and preventing increased cognitive demands on workers. Further empirical research on industrial HRC is encouraged to deepen understanding of this field and to shape applications that benefit workers.

Conclusions of the Dissertation

1 Aim and Scope

This Dissertation investigates how to model, sense, and influence cognitive load in manual and collaborative assembly. It builds an assistance architecture that links task demands and operator state to concrete support actions. Then, it tests physiological tracking with careful signal processing, windowing, feature engineering, and validation. The approach is then applied in four studies examining attention during dual tasks, the sequencing of mental and physical blocks, the effects of instruction format on workload and accuracy, and multitasking with a robot partner. This research addresses practical questions such as how to read the body in real time, how to relate those readings to task design, and how to adjust support to enhance performance and worker ergonomics. The objectives are twofold: (1) to design and implement the assistance architecture, and (2) to test whether heart activity, skin conductance, and motion features can track dynamic load shifts in real time. These objectives answer RQ1–RQ6 across three layers: architecture (inputs, state estimation, and decision rules); methods (signal processing, windowing, features, and validation); and applications (the four studies in Thesis III).

2 Summary of Findings

Thesis I

I developed an extended conceptual Human Asset Administration Shell (HAAS) model to manage cognitive load in human-in-the-loop systems.

Thesis I treats cognitive load in human-in-the-loop systems as a quantity that can be estimated, compared to task requirements, and influenced through specific support actions. It introduces a conceptual extension of the Human-AAS (HAAS) that connects four main modules within a single assistance architecture: physiology, worker characteristics, task type and level, and environment.

From these input modules, the HAAS derives two core quantities. The first is the evaluated cognitive load, estimated from physiological signals. The second is the required cognitive load, derived from task and context descriptors. Person-specific

thresholds are set from neutral baselines, and an OODA loop compares evaluated and required load in real time. When a mismatch is detected, the system can, in principle, adapt task pacing, instruction format, difficulty, or selected aspects of the surroundings.

The HAAS extension follows Asset Administration Shell conventions, so that submodels, identifiers, and properties are explicitly defined and interoperable with digital-twin infrastructures. This thesis therefore answers the first research question by providing a feasible conceptual and architectural model for cognitive load management in manual and collaborative assembly.

Thesis II

I designed and validated a physiological and kinematic analysis pipeline to track attempt-level changes in cognitive load during repeated assembly-like tasks.

Thesis II evaluates whether physiological signals can track dynamic changes in cognitive load when operators repeat assembly-like tasks. It focuses on signal processing, windowing, feature extraction, and validation for HRV, EDA, and wrist acceleration.

Short analysis windows aligned to task attempts are used to capture within-session variation. Heart activity and skin conductance show consistent shifts between baseline and task and across repetitions. Wrist acceleration provides contextual information by indicating segments dominated by hand movement. The analyses show that multimodal features from EDA and HRV, supported by acceleration, can distinguish baseline from task and remain sensitive to within-session changes. At the same time, reliability is reduced during periods with substantial hand motion, which motivates using acceleration to flag motion-heavy segments and support quality control.

Together, these methodological results answer the second research question. They show that real-time tracking of cognitive load is feasible, but requires careful preprocessing, windowing choices, person-specific normalization, and motion-aware interpretation. These methods and constraints are then reused and extended in the application studies of Thesis III.

Thesis III

I implemented the assistance architecture and multimodal physiological methods in four manufacturing-like studies to quantify how dual tasks, task sequencing, instruction formats, and human–robot collaboration influence worker attention, workload, and performance.

Thesis III applies the assistance concept and a multimodal physiological and kinematic analysis framework in four experimental studies. Each chapter examines a different aspect of cognitive load and attention in manufacturing-like settings and contributes to one or more of the remaining research questions.

Situational awareness and dual task attention

The first study, in Chapter III.1, investigates event-based awareness in a dual-task environment that combines screw tightening as a primary task with a Go/No-Go test as a secondary task. Reaction time in the Go/No-Go test is used as a proxy for attentional demand and situational awareness, together with ECG, EDA, and wrist acceleration during task execution.

Recursive feature elimination and linear regression identify features from dominant-hand acceleration and tonic EDA as important predictors of reaction time. Higher mean tonic EDA is associated with faster responses, which is interpreted as a facilitation effect of moderate arousal. In addition, acceleration features help distinguish segments where participants are likely engaged in the repetitive screwing movement from other hand actions, such as fetching screws. These findings answer the third research question by showing that specific EDA and kinematic features can capture momentary shifts in attention in a dual-task, manufacturing-like scenario and could be integrated into the evaluation part of the HAAS model.

Task sequencing and the balance between recovery and arousal

The second study, in Chapter III.2, examines how the sequencing of mental and physical tasks influences subjective workload, recovery, and physiological arousal. Participants perform mental arithmetic and screw tightening in two schedules. In the continuous schedule, tasks are grouped (mental–mental–physical–physical

or physical–physical–mental–mental). In the alternating schedule, mental and physical blocks alternate (mental–physical–mental–physical). Subjective workload is measured with NASA-TLX and the Instantaneous Self-Assessment of workload, while HRV, EDA, and kinematics are analyzed at the block level using linear mixed models.

On the subjective side, session-level workload does not differ significantly between schedules, but block-level ratings show a small increase in perceived load for the second mental block in the alternating condition. HRV features show a pattern of stronger parasympathetic dominance in the alternating sequence, especially in the second mental block, which suggests reduced physiological arousal when mental work follows short physical activity. In contrast, EDA features in the alternating sequence rise later in the session, indicating increased sympathetic activation with continued switching. These results address the fourth research question. They show that an alternating sequence can shift the balance between recovery and arousal, but that repeated switching may also build up sympathetic activation. The study highlights that rotation design needs to consider both the beneficial and the potentially costly effects of alternation.

Instruction format, workload, and precision

The third study, in Chapter III.3, examines how work instruction design influences subjective workload, performance, and physiological responses in an assembly-like task based on Make 'N' Break Extreme pieces. Participants complete sessions with visual-based instructions and with code-based instructions. Subjective workload is assessed with NASA-TLX and the short DSSQ. Objective performance is evaluated through task completion time, number of task repetitions, and assembly precision derived from ArUco-based tracking. Physiological data include HRV, skin conductance measures from Ledalab decomposition of EDA, and acceleration, with features analyzed using logistic regression and feature selection.

The visual-based instructions lead to lower overall cognitive load, lower mental demand, effort, and frustration, as well as shorter task completion times and more repetitions within the allotted time. The code-based instructions, while more demanding and associated with higher distress, result in better assembly precision. Subjective performance ratings favor the visual-based condition, but the objective precision metric shows the opposite pattern. This divergence shows

that self-assessed performance does not necessarily reflect the detailed quality of the output. On the physiological level, multinomial and binary logistic regression models show that HRV and EDA features can reliably separate baseline from task conditions, and that adding acceleration features allows the model to clearly distinguish between visual and code-based sessions.

These findings address the fifth research question. They show that instruction design can shift the balance between speed, repetition rate, and precision. For assistance architectures based on HAAS, this means that instruction format should be represented as a controllable element in the task submodel, and that multiple subjective and objective metrics are needed when deciding on adaptations.

Attentional multitasking in human-robot collaboration

The fourth study, in Chapter III.4, examines attentional factors in collaborative human–robot assembly. Participants perform a wire harnessing task with a cobot while also responding to a Go/No-Go test with two levels of attentional demand. Objective measures include task completion time, assembly errors, and miss rates in the secondary task. Subjective workload is measured with NASA-RTLX, and semi-structured post-interviews provide qualitative insights.

Quantitative analyses show small, non-significant increases in perceived mental load and effort in the higher-demand condition, together with tendencies toward longer task duration and more errors. The small sample size and possible floor effects in the primary task limit the strength of these effects. The qualitative reports indicate that participants often adapt their strategy by prioritising one task over the other, that learning effects shape perceived difficulty across sessions, and that fatigue can weaken the link between subjective ratings and observed performance.

These results address the sixth research question. They suggest that adding a secondary attentional task in human–robot collaboration can increase cognitive effort and may reduce productivity, even when the main task appears relatively simple. The study argues that collaborative robot applications should not assume that the robot simply frees the worker for additional cognitive tasks. Instead, robot behavior and secondary tasks should be scheduled and communicated in

ways that respect limited attentional resources and avoid pushing operators into unsafe or inefficient multitasking strategies.

Bibliography

[1] David Romero, Peter Bernus, Ovidiu Noran, Johan Stahre, and Åsa Fast-Berglund. The operator 4.0: Human cyber-physical systems & adaptive automation towards human-automation symbiosis work systems. In *IFIP international conference on advances in production management systems*, pages 677–686. Springer, 2016.

[2] Marina Crnjac Zizic, Marko Mladineo, Nikola Gjeldum, and Luka Celent. From industry 4.0 towards industry 5.0: A review and analysis of paradigm shift for the people, organization and technology. *Energies*, 15(14):5221, 2022.

[3] Sihan Huang, Baicun Wang, Xingyu Li, Pai Zheng, Dimitris Mourtzis, and Lihui Wang. Industry 5.0 and society 5.0—comparison, complementation and co-evolution. *Journal of manufacturing systems*, 64:424–428, 2022.

[4] Joel Alves, Tânia M Lima, and Pedro D Gaspar. Is industry 5.0 a human-centred approach? a systematic review. *Processes*, 11(1):193, 2023.

[5] Baicun Wang, Huiying Zhou, Xingyu Li, Geng Yang, Pai Zheng, Ci Song, Yixiu Yuan, Thorsten Wuest, Huayong Yang, and Lihui Wang. Human digital twin in the context of industry 5.0. *Robotics and Computer-Integrated Manufacturing*, 85:102626, 2024.

[6] Tamás Ruppert, Szilárd Jaskó, Tibor Holczinger, and János Abonyi. Enabling technologies for operator 4.0: A survey. *Applied sciences*, 8(9):1650, 2018.

[7] Julian Müller. Enabling technologies for industry 5.0. *European Commission*, pages 8–10, 2020.

[8] D Battini, M Faccio, A Persona, and F Sgarbossa. New methodological framework to improve productivity and ergonomics in assembly system design. *International Journal of industrial ergonomics*, 41(1):30–42, 2011.

[9] Christoph H Glock, Eric H Grosse, Hamid Abedinnia, and Simon Emde. An integrated model to improve ergonomic and economic performance in order picking by rotating pallets. *European Journal of Operational Research*, 273(2):516–534, 2019.

[10] W Patrick Neumann and Jan Dul. Human factors: spanning the gap between om and hrm. *International journal of operations & production management*, 2010.

[11] Fabio Sgarbossa, Eric H Grosse, W Patrick Neumann, Daria Battini, and Christoph H Glock. Human factors in production and logistics systems of the future. *Annual Reviews in Control*, 49:295–305, 2020.

[12] John Sweller. Cognitive load theory. In *Psychology of learning and motivation*, volume 55, pages 37–76. Elsevier, 2011.

[13] Melina Klepsch and Tina Seufert. Understanding instructional design effects by differentiated measurement of intrinsic, extraneous, and germane cognitive load. *Instructional Science*, 48(1):45–77, 2020.

[14] Antonio Luque-Casado, José C Perales, David Cárdenas, and Daniel Sanabria. Heart rate variability and cognitive processing: The autonomic response to task demands. *Biological psychology*, 113:83–90, 2016.

[15] William Romine, Noah Schroeder, Tanvi Banerjee, and Josephine Graft. Toward mental effort measurement using electrodermal activity features. *Sensors*, 22(19):7363, 2022.

[16] Shuvodeep Saha, Komal Jindal, Divya Shakti, Suman Tewary, and Viren Sardana. Chirplet transform-based machine-learning approach towards classification of cognitive state change using galvanic skin response and photoplethysmography signals. *Expert Systems*, 39(6):e12958, 2022.

[17] Alessandro Leone, Gabriele Rescio, Pietro Siciliano, Alessandra Papetti, Agnese Brunzini, and Michele Germani. Multi sensors platform for stress monitoring of workers in smart manufacturing context. In *2020 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, pages 1–5. IEEE, 2020.

[18] Andrea Lucchese, Antonio Padovano, and Francesco Facchini. Comprehensive systematic literature review on cognitive workload: Trends on methods, technologies and case studies. *IET Collaborative Intelligent Manufacturing*, 7(1):e70025, 2025.

[19] E Arthur. Physiological metrics of mental workload: A review of recent progress. 1990.

[20] Pieter Vanneste, Annelies Raes, Jessica Morton, Klaas Bombeke, Bram B Van Acker, Charlotte Larmuseau, Fien Depaepe, and Wim Van den Noortgate. Towards measuring cognitive load through multimodal physiological data. *Cognition, Technology & Work*, 23:567–585, 2021.

[21] HM De Morree, BM Szabó, G-J Rutten, and WJ Kop. Central nervous system involvement in the autonomic responses to psychological distress. *Netherlands Heart Journal*, 21:64–69, 2013.

[22] GV Portnova, KM Liaukovich, LN Vasilieva, and EI Alshanskaia. Autonomic and behavioral indicators on increased cognitive loading in healthy volunteers. *Neuroscience and Behavioral Physiology*, 53(1):92–102, 2023.

[23] Mark S Schwartz and Frank Andrasik. *Biofeedback: A practitioner's guide*. Guilford Publications, 2017.

[24] Fred Shaffer and Jay P Ginsberg. An overview of heart rate variability metrics and norms. *Frontiers in public health*, 5:258, 2017.

[25] Riccardo Gervasi, Matteo Capponi, Luca Mastrogiacomo, and Fiorenzo Franceschini. Analyzing psychophysical state and cognitive performance in human-robot collaboration for repetitive assembly processes. *Production Engineering*, 18(1):19–33, 2024.

[26] Francisco Hernando-Gallego, David Luengo, and Antonio Artés-Rodríguez. Feature extraction of galvanic skin responses by nonnegative sparse deconvolution. *IEEE journal of biomedical and health informatics*, 22(5):1385–1394, 2017.

[27] MARIANNA Madonna, LUIGI Monica, SARA Anastasi, and MARIO Di Nardo. Evolution of cognitive demand in the human–machine interaction integrated with industry 4.0 technologies. *Wit Trans. Built Environ*, 189:13–19, 2019.

[28] Andrea Lucchese, Francesco Facchini, Giorgio Mossa, Giovanni Mummolo, and Francesco Paolo Sisto. A numerical assessment of the influence of industry 4.0 technologies on the cognitive complexity of procedure-guided tasks. *International Journal of Simulation and Process Modelling*, 18(2):112–123, 2022.

[29] Eija Haapalainen, SeungJun Kim, Jodi F Forlizzi, and Anind K Dey. Psychophysiological measures for assessing cognitive load. In *Proceedings of the 12th ACM international conference on Ubiquitous computing*, pages 301–310, 2010.

[30] Jing Du, Qi Zhu, Yangming Shi, Qi Wang, Yingzi Lin, and Daniel Zhao. Cognition digital twins for personalized information systems of smart cities: Proof of concept. *Journal of Management in Engineering*, 36(2), 2020.

[31] Elias Montini, Vincenzo Cutrona, Niko Bonomi, Giuseppe Landolfi, Andrea Bettoni, Paolo Rocco, and Emanuele Carpanzano. An iiot platform for human-aware factory digital twins. *Procedia CIRP*, 107:661–667, 2022.

[32] Xiaochen Zheng, Jinzhi Lu, and Dimitris Kiritsis. The emergence of cognitive digital twin: vision, challenges and opportunities. *International Journal of Production Research*, pages 1–23, 2021.

[33] Michael Jacoby, Branislav Jovicic, Ljiljana Stojanovic, and Nenad Stojanović. An approach for realizing hybrid digital twins using asset administration shells and apache streampipes. *Information*, 12(6):217, 2021.

[34] Tasnim A Abdel-Aty, Elisa Negri, and Simone Galparoli. Asset administration shell in manufacturing: Applications and relationship with digital twin. *IFAC-PapersOnLine*, 55(10):2533–2538, 2022.

[35] Elias Montini, Niko Bonomi, Fabio Daniele, Andrea Bettoni, Paolo Pedrazzoli, Emanuele Carpanzano, and Paolo Rocco. The human-digital twin in the manufacturing industry: Current perspectives and a glimpse of future. *Trusted Artificial Intelligence in Manufacturing: A Review of the Emerging Wave of Ethical and Human Centric AI Technologies for Smart Production*, pages 132–147, 2021.

[36] Dale Sparrow, K Kruger, and A Basson. Human digital twin for integrating human workers in industry 4.0. In *Proceedings of the International Conference on Competitive Manufacturing. Stellenbosch, South Africa*, 2019.

[37] Iris Graessler and Alexander Pöhler. Integration of a digital twin as human representation in a scheduling procedure of a cyber-physical production system. In *2017 IEEE international conference on industrial engineering and engineering management (IEEM)*, pages 289–293. IEEE, 2017.

[38] Barbara Rita Barricelli, Elena Casiraghi, Jessica Gliozzo, Alessandro Petrini, and Stefano Valtolina. Human digital twin for fitness management. *Ieee Access*, 8:26637–26664, 2020.

[39] Tolga Erol, Arif Furkan Mendi, and Dilara Doğan. The digital twin revolution in healthcare. In *2020 4th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*, pages 1–7. IEEE, 2020.

[40] Pertti Saariluoma, Jose Cañas, and Antero Karvonen. Human digital twins and cognitive mimetic. In *International Conference on Human Interaction and Emerging Technologies*, pages 97–102. Springer, 2020.

[41] MW Lauer-Schmaltz, P Cash, JP Hansen, and A Maier. Designing human digital twins for behaviour-changing therapy and rehabilitation: a systematic review. *Proceedings of the Design Society*, 2:1303–1312, 2022.

[42] Emily A Holmes and Corin Bourne. Inducing and modulating intrusive emotional memories: A review of the trauma film paradigm. *Acta psychologica*, 127(3):553–566, 2008.

[43] Emilie M Roth and Ann M Bisantz. Cognitive work analysis. *The Oxford handbook of cognitive engineering*, pages 240–260, 2013.

[44] Raya Fidel and Annelise Mark Pejtersen. From information behaviour research to the design of information systems: The cognitive work analysis framework. *Information Research: an international electronic journal*, 10(1):n1, 2004.

[45] David Romero, Johan Stahre, Thorsten Wuest, Ovidiu Noran, Peter Bernus, Åsa Fast-Berglund, and Dominic Gorecky. Towards an operator 4.0 typology: a human-centric perspective on the fourth industrial revolution technologies. In *proceedings of the international conference on computers and industrial engineering (CIE46), Tianjin, China*, pages 29–31, 2016.

[46] Chuck H Perala and Bruce S Sterling. Galvanic skin response as a measure of soldier stress. Technical report, Army Research Lab Aberdeen Proving Ground Md Human Research and Engineering ..., 2007.

[47] SR Dixon and CD Wickens. Automation reliability in unmanned aerial vehicle flight control: Evaluating a reliance-compliance model of automation dependence in high workload. *Human Factors*, 2003.

[48] Wolfram Boucsein. *Electrodermal activity*. Springer Science & Business Media, 2012.

[49] Rich Norman et al. Galvanic skin response & its neurological correlates. *Journal of Consciousness Exploration & Research*, 7(7), 2016.

[50] Renshang Gao, Atiqul Islam, Tom Gedeon, and Md Zakir Hossain. Identifying real and posed smiles from observers' galvanic skin response and blood volume pulse. In *International Conference on Neural Information Processing*, pages 375–386. Springer, 2020.

[51] David M Alexander, Chris Trengove, P Johnston, Tim Cooper, JP August, and Evian Gordon. Separating individual skin conductance responses in a short interstimulus-interval paradigm. *Journal of neuroscience methods*, 146(1):116–123, 2005.

[52] Soroosh Solhjoo, Mark C Haigney, Elexis McBee, Jeroen JG van Merrienboer, Lambert Schuwirth, Anthony R Artino Jr, Alexis Battista, Temple A Ratcliffe, Howard D Lee, and Steven J Durning. Heart rate and heart rate variability correlate with clinical reasoning performance and self-reported measures of cognitive load. *Scientific reports*, 9(1):14668, 2019.

[53] Chiara Cimini, David Romero, Roberto Pinto, and Sergio Cavalieri. Task classification framework and job-task analysis method for understanding the impact of smart and digital technologies on the operators 4.0 job profiles. *Sustainability*, 15(5):3899, 2023.

[54] Carl Benedikt Frey and Michael A Osborne. The future of employment: How susceptible are jobs to computerisation? *Technological forecasting and social change*, 114:254–280, 2017.

[55] Werner Kritzinger, Matthias Karner, Georg Traar, Jan Henjes, and Wilfried Sihn. Digital twin in manufacturing: A categorical literature review and classification. *Ifac-PapersOnline*, 51(11):1016–1022, 2018.

[56] Latif U Khan, Zhu Han, Walid Saad, Ekram Hossain, Mohsen Guizani, and Choong Seon Hong. Digital twin of wireless systems: Overview, taxonomy, challenges, and opportunities. *IEEE Communications Surveys & Tutorials*, 24(4):2230–2254, 2022.

[57] Eirik Halvdan Sølvberg Bratbak. Asset administration shell for life cycle management of safety systems. Master's thesis, NTNU, 2022.

[58] Niko Bonomi. *Exploiting human-in-the-loop characterization applied to AAS in the feld of Industry 4.0.* PhD thesis, University of Applied Sciences and Arts of Southern Switzerland (SUPSI), Innovative Technologies Department, 2022.

[59] Dale Eric Sparrow. *An Architecture for the Integration of Human Workers into an Industry 4.0 Manufacturing Environment.* PhD thesis, Stellenbosch: Stellenbosch University, 2021.

[60] P Adolphs, S Auer, H Bedenbender, et al. *The structure of the administration shell: Trilateral perspectives from france, italy and germany.* Ministry of Economy, Federal Ministry for Economic Affairs Finances and Ministero dello Sviluppo Economico Energy (BMWi), 2018.

[61] ecl@ss. `eclass.eu`, 2023. [Online; accessed 14-Febr-2023].

[62] National Center for O*NET Development. O*net online, 2024.

[63] Ada Chulef, Stephen Read, and David Walsh. A hierarchical taxonomy of human goals. *Motivation and Emotion*, 25:191–232, 09 2001.

[64] European Commission. European skills, competences, qualifications and occupations (esco), 2024.

[65] Lihua Xiao, Ronald Fayer, Una Ryan, and Steve J. Upton. cryptosporidium taxonomy: Recent advances and implications for public health. *Clinical Microbiology Reviews*, 17(1):72–97, 2004.

[66] Paul A. David. Knowledge, capabilities and human capital formation in economic growth. New Zealand Treasury Working Paper 01/13, New Zealand Treasury Working Paper, Wellington, 2000.

[67] Andrea Bettoni, Elias Montini, Massimiliano Righi, Valeria Villani, Radostin Tsvetanov, Stefano Borgia, Cristian Secchi, and Emanuele Carpanzano. Mutualistic and adaptive human-machine collaboration based on machine learning in an injection moulding manufacturing line. *Procedia CIRP*, 93:395–400, 2020. 53rd CIRP Conference on Manufacturing Systems 2020.

[68] Petr Marcon, Christian Diedrich, Frantisek Zezulka, Tizian Schröder, Alexander Belyaev, Jakub Arm, Tomas Benesl, Zdenek Bradac, and Ivo Vesely. The asset administration shell of operator in the platform of industry 4.0. In *2018 18th International Conference on Mechatronics - Mechatronika (ME)*, pages 1–5, 2018.

[69] Anwar Al Assadi, Christian Fries, Manuel Fechter, Benjamin Maschler, Daniel Ewert, Hans-Georg Schnauffer, Michael Zürn, and Matthias Reichenbach. User-friendly, requirement based assistance for production workforce using an asset administration shell design. *Procedia CIRP*, 91:402–406, 2020. Enhancing design through the 4th Industrial Revolution Thinking.

[70] Yan Zhao, Pengcheng Wen, Linting Bai, Xinzhi Liu, Zhonghua Wang, and Na Wu. Service-oriented intelligent ooda loop. In *2023 26th ACIS International Winter Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD-Winter)*, pages 182–187. IEEE, 2023.

[71] Nishant Uppal, Sushanta Kumar Mishra, and Neharika Vohra. Prior related work experience and job performance: Role of personality. *International Journal of selection and assessment*, 22(1):39–51, 2014.

[72] A Ghanbary Sartang, M Ashnagar, E Habibi, and S Sadeghi. Evaluation of rating scale mental effort (rsme) effectiveness for mental workload assessment

in nurses. *Journal of Occupational Health and Epidemiology*, 5(4):211–217, 2016.

[73] Mihaly Csikszentmihalyi. *Flow: The Psychology of Optimal Experience (P.S.)*. Harper Perennial Modern Classics, 2008.

[74] Lihui Wang. A futuristic perspective on human-centric assembly. *Journal of Manufacturing Systems*, 62:199–201, 2022.

[75] Anne Grethe Syversen, Martina Ortova, Godfrey Mugurusi, and Kristin H Hansen. A leap from operator 4.0 to operator 5.0: Antecedents, enablers, and barriers in human-centered manufacturing. In *IFIP International Conference on Advances in Production Management Systems*, pages 209–225. Springer, 2024.

[76] Yuqian Lu, Hao Zheng, Saahil Chand, Wanqing Xia, Zengkun Liu, Xun Xu, Lihui Wang, Zhaojun Qin, and Jinsong Bao. Outlook on human-centric manufacturing towards industry 5.0. *Journal of Manufacturing Systems*, 62:612–627, 2022.

[77] Dimitris Mourtzis, John Angelopoulos, and Nikos Panopoulos. Operator 5.0: A survey on enabling technologies and a framework for digital manufacturing based on extended reality. *Journal of Machine Engineering*, 22, 2022.

[78] Christian Bechinie, Setareh Zafari, Lukas Kroeninger, Jaison Puthenkalam, and Manfred Tscheligi. Toward human-centered intelligent assistance system in manufacturing: challenges and potentials for operator 5.0. *Procedia Computer Science*, 232:1584–1596, 2024.

[79] Tim Miller. Explanation in artificial intelligence: Insights from the social sciences. *Artificial intelligence*, 267:1–38, 2019.

[80] Cristina Iani, Daniel Gopher, and Peretz Lavie. Effects of task difficulty and invested mental effort on peripheral vasoconstriction. *Psychophysiology*, 41(5):789–798, 2004.

[81] OV Martynova, AO Roik, and GA Ivanitsky. Changes in some indices of the cardiovascular system in different mental tasks. *Human Physiology*, 37:673–678, 2011.

[82] JA Veltman and AWK9613226 Gaillard. Physiological workload reactions to increasing levels of task difficulty. *Ergonomics*, 41(5):656–669, 1998.

[83] Sylvia D Kreibig and Guido HE Gendolla. Autonomic nervous system measurement of emotion in education and achievement settings. In *International handbook of emotions in education*, pages 625–642. Routledge, 2014.

[84] Oscar Danielsson, Magnus Holm, and Anna Syberfeldt. Augmented reality smart glasses in industrial assembly: Current status and future challenges. *Journal of Industrial Information Integration*, 20:100175, 2020.

[85] Alessandra Papetti, Marianna Ciccarelli, Matteo Claudio Palpacelli, and Michele Germani. How to provide work instructions to reduce the workers' physical and mental workload. *Procedia CIRP*, 120:1167–1172, 2023.

[86] Dan Li, Sandra Mattsson, Omkar Salunkhe, Åsa Fast-Berglund, Anders Skoogh, and Jesper Broberg. Effects of information content in work instructions for operator performance. *Procedia Manufacturing*, 25:628–635, 2018.

[87] Marlon Antonin Lehmann, Ronny Porsch, and Christopher Mai. Assembly process digitization through self-learning assistance systems in production. In *Towards Sustainable Customization: Bridging Smart Products and Manufacturing Systems: Proceedings of the 8th Changeable, Agile, Reconfigurable and Virtual Production Conference (CARV2021) and the 10th World Mass Customization & Personalization Conference (MCPC2021), Aalborg, Denmark, October/November 2021 8*, pages 216–223. Springer, 2022.

[88] Wonse Jo, Ruiqi Wang, Go-Eum Cha, Su Sun, Revanth Krishna Senthilkumaran, Daniel Foti, and Byung-Cheol Min. Mocas: A multimodal dataset for objective cognitive workload assessment on simultaneous tasks. *IEEE Transactions on Affective Computing*, 16(1):116–132, 2024.

[89] Maximilian P Oppelt, Andreas Foltyn, Jessica Deuschel, Nadine R Lang, Nina Holzer, Bjoern M Eskofier, and Seung Hee Yang. Adabase: a multimodal dataset for cognitive load estimation. *Sensors*, 23(1):340, 2022.

[90] Debora P Salgado, Sheila Fallon, Yuansong Qiao, and Eduardo LM Naves. Wheelsimphysio-2023 dataset: Physiological and questionnaire-based dataset of immersive multisensory wheelchair simulator from 58 participants. *Data in Brief*, 54:110535, 2024.

[91] Anubhav Bhatti, Prithila Angkan, Behnam Behinaein, Zunayed Mahmud, Dirk Rodenburg, Heather Braund, P James Mclellan, Aaron Ruberto,

Geoffery Harrison, Daryl Wilson, et al. Clare: Cognitive load assessment in realtime with multimodal data. *arXiv preprint arXiv:2404.17098*, 2024.

[92] Christoph Anders, Sidratul Moontaha, Samik Real, and Bert Arnrich. Unobtrusive measurement of cognitive load and physiological signals in uncontrolled environments. *Scientific Data*, 11(1):1000, 2024.

[93] Thomas J Palmeri. Theories of automaticity and the power law of practice. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 1999.

[94] Mathias Benedek and Christian Kaernbach. A continuous measure of phasic electrodermal activity. *Journal of neuroscience methods*, 190(1):80–91, 2010.

[95] Jason J Braithwaite, Derrick G Watson, Robert Jones, and Mickey Rowe. A guide for analysing electrodermal activity (eda) & skin conductance responses (scrs) for psychological experiments. *Psychophysiology*, 49(1):1017–1034, 2013.

[96] David OK. ECG Class for Heart Rate Variability Analysis. `https://www.mathworks.com/matlabcentral/fileexchange/84692-ecg-class-for-heart-rate-variability-analysis`, 2025. MATLAB Central File Exchange. Retrieved April 11, 2025.

[97] Paul Ayres, Joy Yeonjoo Lee, Fred Paas, and Jeroen JG Van Merrienboer. The validity of physiological measures to identify differences in intrinsic cognitive load. *Frontiers in psychology*, 12:702538, 2021.

[98] Evgeniia I Alshanskaia, Natalia A Zhozhikashvili, Irina S Polikanova, and Olga V Martynova. Heart rate response to cognitive load as a marker of depression and increased anxiety. *Frontiers in Psychiatry*, 15:1355846, 2024.

[99] Kenneth J Hunt and Jittima Saengsuwan. Changes in heart rate variability with respect to exercise intensity and time during treadmill running. *Biomedical engineering online*, 17(1):128, 2018.

[100] Harunobu Usui and Yusuke Nishida. The very low-frequency band of heart rate variability represents the slow recovery component after a mental stress task. *PloS one*, 12(8):e0182611, 2017.

[101] Stéphane Delliaux, Alexis Delaforge, Jean-Claude Deharo, and Guillaume Chaumet. Mental workload alters heart rate variability, lowering non-linear dynamics. *Frontiers in physiology*, 10:565, 2019.

[102] Allen Newell and Paul S Rosenbloom. Mechanisms of skill acquisition and the law of practice. In *Cognitive skills and their acquisition*, pages 1–55. Psychology Press, 1981.

[103] Abdulrahman K Eesee, Vera Varga, György Eigner, and Tamás Ruppert. Impact of work instruction difficulty on cognitive load and operational efficiency. *Scientific Reports*, 15(1):11028, 2025.

[104] Caroline Di Bernardi Luft, Emílio Takase, and David Darby. Heart rate variability and cognitive function: Effects of physical effort. *Biological psychology*, 82(2):186–191, 2009.

[105] Anne Garde, Bjarne Laursen, Anker Jørgensen, and Bente Jensen. Effects of mental and physical demands on heart rate variability during computer work. *European journal of applied physiology*, 87:456–461, 2002.

[106] Francis T Durso and Arathi Sethumadhavan. Situation awareness: Understanding dynamic environments. *Human factors*, 50(3):442–448, 2008.

[107] Jana Paulusová and Martin Paulus. Situational awareness in hmi of industrial plants. In *2018 19th International Carpathian Control Conference (ICCC)*, pages 512–517. IEEE, 2018.

[108] Kenneth M Jackson, Tyler H Shaw, and William S Helton. The effects of dual-task interference on visual search and verbal memory. *Ergonomics*, 66(1):125–135, 2023.

[109] Mica R Endsley. Situation awareness global assessment technique (sagat). In *Proceedings of the IEEE 1988 national aerospace and electronics conference*, pages 789–795. IEEE, 1988.

[110] Gemma F Briggs, Graham J Hole, and Jim AJ Turner. The impact of attentional set and situation awareness on dual tasking driving performance. *Transportation research part F: traffic psychology and behaviour*, 57:36–47, 2018.

[111] Mary L Courage, Aishah Bakhtiar, Cheryll Fitzpatrick, Sophie Kenny, and Katie Brandeau. Growing up multitasking: The costs and benefits for cognitive development. *Developmental Review*, 35:5–41, 2015.

[112] Ting Zhang, Jing Yang, Nade Liang, Brandon J Pitts, Kwaku Prakah-Asante, Reates Curry, Bradley Duerstock, Juan P Wachs, and Denny Yu. Physiological measurements of situation awareness: a systematic review. *Human factors*, 65(5):737–758, 2023.

[113] Franciscus Cornelis Donders. On the speed of mental processes. *Acta psychologica*, 30:412–431, 1969.

[114] Harold Pashler. 12 task switching and multitask performance. *Control of cognitive processes*, page 277, 2000.

[115] Robert D Rogers and Stephen Monsell. Costs of a predictible switch between simple cognitive tasks. *Journal of experimental psychology: General*, 124(2):207, 1995.

[116] Mica R Endsley. Toward a theory of situation awareness in dynamic systems. *Human factors*, 37(1):32–64, 1995.

[117] Abdulrahman K Eesee, David Kostolani, Taeho Kang, Sebastian Schlund, Tibor Medvegy, János Abonyi, and Tamás Ruppert. May i have your attention?! exploring multitasking in human-robot collaboration. *IFAC-PapersOnLine*, 58(19):241–246, 2024.

[118] Christopher D Wickens. Multiple resources and mental workload. *Human factors*, 50(3):449–455, 2008.

[119] Marvin M Chun, Julie D Golomb, and Nicholas B Turk-Browne. A taxonomy of external and internal attention. *Annual review of psychology*, 62:73–101, 2011.

[120] Roberto Arrighi, Roy Lunardi, and David Burr. Vision and audition do not share attentional resources in sustained tasks. *Frontiers in psychology*, 2:56, 2011.

[121] Daryl Fougnie, Jurnell Cockhren, and René Marois. A common source of attention for auditory and visual tracking. *Attention, Perception, & Psychophysics*, 80:1571–1583, 2018.

[122] Dario D Salvucci and Niels A Taatgen. Threaded cognition: an integrated theory of concurrent multitasking. *Psychological review*, 115(1):101, 2008.

[123] Alexander Ya Fridman and Boris A Kulik. Situation awareness in modeling industrial-natural complexes. In *Cyber-Physical Systems and Control*, pages 247–256. Springer, 2020.

[124] Paul M Salmon, Neville A Stanton, Guy H Walker, Daniel Jenkins, Darshna Ladva, Laura Rafferty, and Mark Young. Measuring situation awareness in complex systems: Comparison of measures study. *International Journal of Industrial Ergonomics*, 39(3):490–500, 2009.

[125] Salman Nazir, Simone Colombo, and Davide Manca. The role of situation awareness for the operators of process industry. *Chemical Engineering Transactions*, 26, 2012.

[126] Marta Lall, Hans Torvatn, and Eva Amdahl Seim. Towards industry 4.0: increased need for situational awareness on the shop floor. In *Advances in Production Management Systems. The Path to Intelligent, Collaborative and Sustainable Manufacturing: IFIP WG 5.7 International Conference, APMS 2017, Hamburg, Germany, September 3-7, 2017, Proceedings, Part I*, pages 322–329. Springer, 2017.

[127] Mica R Endsley. The divergence of objective and subjective situation awareness: A meta-analysis. *Journal of cognitive engineering and decision making*, 14(1):34–53, 2020.

[128] Paul M Salmon, Neville A Stanton, and Kristie Lee Young. Situation awareness on the road: review, theoretical and methodological issues, and future directions. *Theoretical Issues in Ergonomics Science*, 13(4):472–492, 2012.

[129] Firgan Feradov, Todor Ganchev, and Valentina Markova. Automated detection of cognitive load from peripheral physiological signals based on hjorth's parameters. In *2020 International Conference on Biomedical Innovations and Applications (BIA)*, pages 85–88. IEEE, 2020.

[130] Kristin J Heaton, James R Williamson, Adam C Lammert, Katherine R Finkelstein, Caitlin C Haven, Douglas Sturim, Christopher J Smalt, and Thomas F Quatieri. Predicting changes in performance due to cognitive fatigue: A multimodal approach based on speech motor coordination and electrodermal activity. *The Clinical Neuropsychologist*, 34(6):1190–1214, 2020.

[131] Alexander Kunze, Stephen J Summerskill, Russell Marshall, and Ashleigh J Filtness. Automation transparency: implications of uncertainty communication for human-automation interaction and interfaces. *Ergonomics*, 62(3):345–360, 2019.

[132] Ranjana K Mehta, S Camille Peres, Ashley E Shortz, Wimberly Hoyle, Melissa Lee, Gurtej Saini, Hong-Chih Chan, and Mitchell W Pryor. Operator situation awareness and physiological states during offshore well control scenarios. *Journal of Loss Prevention in the Process Industries*, 55:332–337, 2018.

[133] Lee Guan Yeo, Haoqi Sun, Yisi Liu, Fitri Trapsilawati, Olga Sourina, Chun-Hsien Chen, Wolfgang Mueller-Wittig, and Wei Tech Ang. Mobile eeg-based situation awareness recognition for air traffic controllers. In *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 3030–3035. IEEE, 2017.

[134] Ardaman Kaur, Rishu Chaujar, and Vijayakumar Chinnadurai. Effects of neural mechanisms of pretask resting eeg alpha information on situational awareness: a functional connectivity approach. *Human Factors*, 62(7):1150–1170, 2020.

[135] Di Catherwood, Graham K Edgar, Dritan Nikolla, Chris Alford, David Brookes, Steven Baker, and Sarah White. Mapping brain activity during loss of situation awareness: an eeg investigation of a basis for top-down influence on perception. *Human factors*, 56(8):1428–1452, 2014.

[136] Qun Wang, Damin Zhuang, Hengyang Wei, and Xiaoru Wanyan. Evaluation method of cockpit display interface based on situation awareness. In *2012 5th International Conference on biomedical engineering and informatics*, pages 528–531. IEEE, 2012.

[137] Dominique Makowski, Tam Pham, Zen J. Lau, Jan C. Brammer, François Lespinasse, Hung Pham, Christopher Schölzel, and S. H. Annabel Chen. NeuroKit2: A python toolbox for neurophysiological signal processing. *Behavior Research Methods*, 53(4):1689–1696, feb 2021.

[138] Paul Van Gent, Haneen Farah, Nicole Van Nes, and Bart Van Arem. Heartpy: A novel heart rate algorithm for the analysis of noisy signals.

Transportation research part F: traffic psychology and behaviour, 66:368–378, 2019.

[139] Surajit Bag, Shivam Gupta, and Sameer Kumar. Industry 4.0 adoption and 10r advance manufacturing capabilities for sustainable development. *International journal of production economics*, 231:107844, 2021.

[140] Kalin Stefanov and Jonas Beskow. A multi-party multi-modal dataset for focus of visual attention in human-human and human-robot interaction. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 4440–4444, 2016.

[141] Helena Jahncke, Staffan Hygge, Svend Erik Mathiassen, David Hallman, Susanna Mixter, and Eugene Lyskov. Variation at work: alternations between physically and mentally demanding tasks in blue-collar occupations. *Ergonomics*, 60(9):1218–1227, 2017.

[142] Ulf Lundberg. Psychophysiology of work: Stress, gender, endocrine response, and work-related upper extremity disorders. *American journal of industrial medicine*, 41(5):383–392, 2002.

[143] Svend Erik Mathiassen. Diversity and variation in biomechanical exposure: what is it, and why would we like to know? *Applied ergonomics*, 37(4):419–427, 2006.

[144] Jennie A Jackson, Marianne Sund, Griztko Barlari Lobos, Lars Melin, and Svend Erik Mathiassen. Assessing the efficacy of a job rotation for improving occupational physical and psychosocial work environment, musculoskeletal health, social equality, production quality and resilience at a commercial laundromat: protocol for a longitudinal case study. *BMJ open*, 13(5):e067633, 2023.

[145] Iring Koch and Andrea Kiesel. Task switching: Cognitive control in sequential multitasking. *Handbook of human multitasking*, pages 85–143, 2022.

[146] Nachshon Meiran. Reconfiguration of processing mode prior to task performance. *Journal of Experimental Psychology: Learning, memory, and cognition*, 22(6):1423, 1996.

[147] Andrea Kiesel, Marco Steinhauser, Mike Wendt, Michael Falkenstein, Kerstin Jost, Andrea M Philipp, and Iring Koch. Control and interference in task switching—a review. *Psychological bulletin*, 136(5):849, 2010.

[148] Tobias Egner and Audrey Siqi-Liu. Insights into control over cognitive flexibility from studies of task-switching. *Current Opinion in Behavioral Sciences*, 55:101342, 2024.

[149] Gesine Dreisbach and Jonathan Mendl. Flexibility as a matter of context, effort, and ability: evidence from the task-switching paradigm. *Current Opinion in Behavioral Sciences*, 55:101348, 2024.

[150] Muhammad Johan Alibasa, Rizka W Purwanto, Kalina Yacef, Nick Glozier, and Rafael A Calvo. Doing and feeling: relationships between moods, productivity and task-switching. *IEEE Transactions on Affective Computing*, 13(3):1140–1154, 2020.

[151] Susanna Mixter, Svend Erik Mathiassen, and David Hallman. Alternations between physical and cognitive tasks in repetitive work–effect of cognitive task difficulty on fatigue development in women. *Ergonomics*, 62(8):1008–1022, 2019.

[152] Rosimeire Simprini Padula, Maria Luiza Caires Comper, Emily H Sparer, and Jack T Dennerlein. Job rotation designed to prevent musculoskeletal disorders and control risk in manufacturing industries: A systematic review. *Applied ergonomics*, 58:386–397, 2017.

[153] Tessy Luger, Tim Bosch, Dirkjan Veeger, and Michiel de Looze. The influence of task variation on manifestation of fatigue is ambiguous–a literature review. *Ergonomics*, 57(2):162–174, 2014.

[154] Priscilla C Leider, Julitta S Boschman, Monique HW Frings-Dresen, and Henk F van der Molen. Effects of job rotation on musculoskeletal complaints and related work exposures: a systematic literature review. *Ergonomics*, 58(1):18–32, 2015.

[155] George E Billman. The lf/hf ratio does not accurately measure cardiac sympatho-vagal balance, 2013.

[156] Michael R Esco, Henry N Williford, Andrew A Flatt, Todd J Freeborn, and Fabio Y Nakamura. Ultra-shortened time-domain hrv parameters at rest and

following exercise in athletes: an alternative to frequency computation of sympathovagal balance. *European journal of applied physiology*, 118(1):175–184, 2018.

[157] Hugo F Posada-Quintero, John P Florian, Alvaro D Orjuela-Cañón, and Ki H Chon. Electrodermal activity is sensitive to cognitive stress under water. *Frontiers in physiology*, 8:1128, 2018.

[158] Dominic Bläsing, Sven Hinrichsen, and Manfred Bornewasser. Reduction of cognitive load in complex assembly systems. In *Human Interaction, Emerging Technologies and Future Applications II: Proceedings of the 2nd International Conference on Human Interaction and Emerging Technologies: Future Applications (IHIET–AI 2020), April 23-25, 2020, Lausanne, Switzerland*, pages 495–500. Springer, 2020.

[159] Xiaowei Zhu, S Jack Hu, Yoram Koren, and Samuel P Marin. Modeling of manufacturing complexity in mixed-model assembly lines. In *International Manufacturing Science and Engineering Conference*, volume 47624, pages 649–659, 2006.

[160] Shixin Jack Hu, Xiaowei Zhu, Hui Wang, and Yoram Koren. Product variety and manufacturing complexity in assembly systems and supply chains. *CIRP annals*, 57(1):45–48, 2008.

[161] Anders Haug. Work instruction quality in industrial management. *International Journal of Industrial Ergonomics*, 50:170–177, 2015.

[162] Deondra S Conner and Scott C Douglas. Organizationally-induced work stress: The role of employee bureaucratic orientation. *Personnel review*, 34(2):210–224, 2005.

[163] Salla Lind. Types and sources of fatal and severe non-fatal accidents in industrial maintenance. *International Journal of Industrial Ergonomics*, 38(11-12):927–933, 2008.

[164] John Oakland. Leadership and policy deployment: the backbone of tqm. *Total Quality Management & Business Excellence*, 22(5):517–534, 2011.

[165] Pieter Vanneste, Kim Dekeyser, Luis Alberto Pinos Ullauri, Dries Debeer, Frederik Cornillie, Fien Depaepe, Annelies Raes, Wim Van den Noortgate, and Sameh Said-Metwaly. Towards tailored cognitive support in augmented

reality assembly work instructions. *Journal of Computer Assisted Learning*, 40(2):797–811, 2024.

[166] Gaurav Garg, Roy Andersson, and Mauro Caporuscio. Digitalization of work instructions in production plant. In *Sustainable Production through Advanced Manufacturing, Intelligent Automation and Work Integrated Learning*, pages 325–334. IOS Press, 2024.

[167] G Matthews, AK Emo, and GJ Funke. A short version of the dundee stress state questionnaire. In *Twelfth Meeting of the International Society for the Study of Individual Differences, Adelaide, Australia*, 2005.

[168] Gerald Matthews. Stress states, personality and cognitive functioning: A review of research with the dundee stress state questionnaire. *Personality and Individual Differences*, 169:110083, 2021.

[169] Mengting Zhao, Dongyu Qiu, and Yong Zeng. How much workload is a 'good'workload for human beings to meet the deadline: human capacity zone and workload equilibrium. *Journal of Engineering Design*, 34(8):644–673, 2023.

[170] Jaakko Peltokorpi and Mohamad Y Jaber. A group learning curve model with motor, cognitive and waste elements. *Computers & Industrial Engineering*, 146:106621, 2020.

[171] Chia-Fen Chi, Chih-Chan Cheng, Yuh-Chuan Shih, I-Sheng Sun, and Tin-Chang Chang. Learning rate and subjective mental workload in five truck driving tasks. *Ergonomics*, 62(3):391–405, 2019.

[172] Bartosz Misiurek. *Standardized work with TWI: eliminating human errors in production and service processes*. CRC Press, 2016.

[173] Peyman Khaleghi, Hossein Akbari, Negin Masoudi Alavi, Masoud Motalebi Kashani, and Zahra Batooli. Identification and analysis of human errors in emergency department nurses using sherpa method. *International emergency nursing*, 62:101159, 2022.

[174] Gabriele Volpes, Simone Valenti, Giuseppe Genova, Chiara Barà, Antonino Parisi, Luca Faes, Alessandro Busacca, and Riccardo Pernice. Wearable

ring-shaped biomedical device for physiological monitoring through finger-based acquisition of electrocardiographic, photoplethysmographic, and galvanic skin response signals: Design and preliminary measurements. *Biosensors*, 14(4):205, 2024.

[175] Nima Rezaei and Parnian Jabbari. *Immunoinformatics of Cancers: Practical Machine Learning Approaches Using R.* Academic Press, 2022.

[176] Xiaoyong Ji, Heng Li, Zhuofan Lu, Zifeng Wang, and Xinyu Chai. Research on the electrodermal activity during walking and running. In *2019 4th International Conference on Control and Robotics Engineering (ICCRE)*, pages 179–183. IEEE, 2019.

[177] Peter A Hancock and Joel S Warm. A dynamic model of stress and sustained attention. *Journal of Human Performance in Extreme Environments*, 7(1):4, 2003.

[178] Fahad Sherwani, Muhammad Mujtaba Asad, and Babul Salam Kader K Ibrahim. Collaborative robots and industrial revolution 4.0 (ir 4.0). In *2020 International Conference on Emerging Trends in Smart Technologies (ICETST)*, pages 1–5. IEEE, 2020.

[179] George Michalos, Panagiotis Karagiannis, Nikos Dimitropoulos, Dionisis Andronas, and Sotiris Makris. Human robot collaboration in industrial environments. *The 21st century industrial robot: when tools become collaborators*, pages 17–39, 2022.

[180] Uqba Othman and Erfu Yang. An overview of human-robot collaboration in smart manufacturing. In *2022 27th International Conference on Automation and Computing (ICAC)*, pages 1–6. IEEE, 2022.

[181] Alejandro Chacón, Pere Ponsa, and Cecilio Angulo. Cognitive interaction analysis in human–robot collaboration using an assembly task. *Electronics*, 10(11):1317, 2021.

[182] Claudia Carissoli, Luca Negri, Marta Bassi, Fabio Alexander Storm, and Antonella Delle Fave. Mental workload and human-robot interaction in collaborative tasks: A scoping review. *International Journal of Human–Computer Interaction*, pages 1–20, 2023.

[183] David Alais, Concetta Morrone, and David Burr. Separate attentional resources for vision and audition. *Proceedings of the Royal Society B: Biological Sciences*, 273(1592):1339–1345, 2006.

[184] Niels A Taatgen, Ion Juvina, Marc Schipper, Jelmer P Borst, and Sander Martens. Too much control can hurt: A threaded cognition model of the attentional blink. *Cognitive psychology*, 59(1):1–29, 2009.

[185] Doug Rohrer and Harold E Pashler. Concurrent task effects on memory retrieval. *Psychonomic Bulletin & Review*, 10(1):96–103, 2003.

[186] Matthias Weigl, Andreas Müller, Nick Sevdalis, and Peter Angerer. Relationships of multitasking, physicians' strain, and performance. *Journal of patient safety*, 9(1):18–23, 2013.

[187] Steven H Appelbaum, Adam Marchionni, and Arturo Fernandez. The multitasking paradox: Perceptions, problems and strategies. *Management Decision*, 46(9):1313–1325, 2008.

[188] Barbara Metz, Nadja Schömig, and Hans-Peter Krüger. Attention during visual secondary tasks in driving: Adaptation to the demands of the driving task. *Transportation research part F: traffic psychology and behaviour*, 14(5):369–380, 2011.

[189] Alessandro Umbrico, Amedeo Cesta, and Andrea Orlandini. Enhancing awareness of industrial robots in collaborative manufacturing. *Semantic Web*, (Preprint):1–40, 2023.

[190] Matteo Lavit Nicora, Roberto Ambrosetti, Gloria J Wiens, and Irene Fassi. Human–robot collaboration in smart manufacturing: Robot reactive behavior intelligence. *Journal of Manufacturing Science and Engineering*, 143(3):031009, 2021.

[191] Mica R Endsley. Measurement of situation awareness in dynamic systems. *Human factors*, 37(1):65–84, 1995.

[192] Hongyi Liu and Lihui Wang. Collision-free human-robot collaboration based on context awareness. *Robotics and Computer-Integrated Manufacturing*, 67:101997, 2021.

[193] Sarah K Hopko, Riya Khurana, Ranjana K Mehta, and Prabhakar R Pagilla. Effect of cognitive fatigue, operator sex, and robot assistance on task performance metrics, workload, and situation awareness in human-robot collaboration. *IEEE Robotics and Automation Letters*, 6(2):3049–3056, 2021.

[194] Lucas Paletta, Amir Dini, Cornelia Murko, Saeed Yahyanejad, and Ursula Augsdörfer. Estimation of situation awareness score and performance using eye and head gaze for human-robot collaboration. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*, pages 1–3, 2019.

[195] Tamás Ruppert, András Darányi, Tibor Medvegy, Dániel Csereklei, and János Abonyi. Demonstration laboratory of industry 4.0 retrofitting and operator 4.0 solutions: Education towards industry 5.0. *Sensors*, 23(1):283, 2022.

[196] Menno Nijboer, Niels A Taatgen, Annelies Brands, Jelmer P Borst, and Hedderik van Rijn. Decision making in concurrent multitasking: do people adapt to task interference? *PloS one*, 8(11):e79583, 2013.

[197] Reinier J Jansen, René van Egmond, and Huib de Ridder. Task prioritization in dual-tasking: Instructions versus preferences. *PLoS One*, 11(7):e0158511, 2016.

[198] Christoph Anderson, Isabel Hübener, Ann-Kathrin Seipp, Sandra Ohly, Klaus David, and Veljko Pejovic. A survey of attention management systems in ubiquitous computing environments. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(2):1–27, 2018.

[199] Rachel F Adler and Raquel Benbunan-Fich. The effects of task difficulty and multitasking on performance. *Interacting with Computers*, 27(4):430–439, 2015.

[200] Juan Esteban Kamienkowski and Mariano Sigman. Delays without mistakes: Response time and error distributions in dual-task. *PLoS One*, 3(9):e3196, 2008.

[201] Shirin Esmaeili Bijarsari. A current view on dual-task paradigms and their limitations to capture cognitive load. *Frontiers in Psychology*, 12:648586, 2021.

[202] Miroslava Jindrová, Martin Kocourek, and Petr Telenskỳ. Skin conductance rise time and amplitude discern between different degrees of emotional arousal induced by affective pictures presented on a computer screen. *BioRxiv*, pages 2020–05, 2020.

[203] Chenjie Xia, Alexandra Touroutoglou, Karen S Quigley, Lisa Feldman Barrett, and Bradford C Dickerson. Salience network connectivity modulates skin conductance responses in predicting arousal experience. *Journal of cognitive neuroscience*, 29(5):827–836, 2017.

[204] Chong L Lim, Chris Rennie, Robert J Barry, Homayoun Bahramali, Ilario Lazzaro, Barry Manor, and Evian Gordon. Decomposing skin conductance into tonic and phasic components. *International Journal of Psychophysiology*, 25(2):97–109, 1997.

[205] Víctor Martínez-Cagigal. Sample Entropy. Mathworks. `https://www.mathworks.com/matlabcentral/fileexchange/69381-sample-entropy`, 2018.

[206] Joshua S Richman and J Randall Moorman. Physiological time-series analysis using approximate entropy and sample entropy. *American journal of physiology-heart and circulatory physiology*, 278(6):H2039–H2049, 2000.

[207] The MathWorks, Inc. pwelch (R2024b). `https://www.mathworks.com/help/releases/R2024b/signal/ref/pwelch.html`, 2024. Accessed: 2025-05-17.

[208] Christopher M DeGiorgio, Patrick Miller, Sheba Meymandi, Alex Chin, Jordan Epps, Steven Gordon, Jeffrey Gornbein, and Ronald M Harper. Rmssd, a measure of vagus-mediated heart rate variability, is associated with risk factors for sudep: the sudep-7 inventory. *Epilepsy & behavior*, 19(1):78–81, 2010.

[209] M Khawar Ali, Lijun Liu, Ji-Hong Chen, and Jan D Huizinga. Optimizing autonomic function analysis via heart rate variability associated with motor activity of the human colon. *Frontiers in physiology*, 12:619722, 2021.

[210] George E Billman. The effect of heart rate on the heart rate variability response to autonomic interventions. *Frontiers in physiology*, 4:222, 2013.

Appendices

Table A.1: Human Asset Administration Shell HAAS Header

Field Name	Type/Example	Description	The internal GUID
HAAS ID	URI, eCl@ss	Unique identifier for the HAAS.	
Title	String	Human Asset Administration Shell.	0658d398-d7ca-4f97-a6b2-cd0925bbfd8e
Date	Date	Date of creation or last modification.	
Contact	Email, Phone	Worker contact details	
Submodel 1	String	The Worker physiological parameter submodel.	9aba4948-8000-412d-b365-54084c5d8163
Submodel 2	String	The Worker characteristics submodel.	c290d893-66be-453a-9a8d-3f5b2c649445
Manifest	Link/Reference	A directory of key information derived from the properties of the submodels.	

Table A.2: Worker physiological parameters submodel

Name	Submodel 1		
GUID	9aba4948-8000-412d-b365-54084c5d8163		
Properties			
GUID	Name	Type	Description
35cd4db0-24a0-47a2-bb61-dea37e34b166	GSR peaks	Numeric stream	The number of peaks in the skin conductivity signal
b5d5635d-c59f-448f-bd7c-472bf9c12c73	GSR area	Numeric stream	The area under the skin conductivity signal
2c25d961-170e-48e9-a5e0-7a0d5e80f27a	GSR SCR area	Numeric stream	The area under the Skin Conductance Response SCR
b6605b89-a3a6-4c3e-843e-739b607ea51f	GSR rising time	Numeric stream	The average time from the baseline to the peak after the stimulus
a4dfe715-babd-42fe-98e0-d0dc9f2d61ff	GSR decay time	Numeric stream	The average time from the peak to the baseline
5c34cda8-b1f0-42ae-b0a3-327cd9a6ef68	HRV VLF	Numeric stream	Very Low Frequency power of HRV in the band (0.003-0.04) Hz
f7e01fc3-f9ed-467c-a588-60113bfa7874	HRV (HF/LF)	Numeric stream	Ratio of High-Frequency power (0.15-0.4) Hz to the Low-Frequency power (0.04-0.15) Hz
fcb355df-951e-4fd8-9439-a656959e7166	HRV SD1	Numeric stream	Poincaré plot descriptor (short-term HRV)
00806cd2-d88d-4d47-a1a1-5d988f817d06	HRV SD2	Numeric stream	Poincaré plot descriptor (long-term HRV)
9c2ae141-8ccf-43b5-b714-e4469432921e	HRV Skew	Numeric stream	Skewness of RR intervals in the HRV

Table A.3: Worker characteristics submodel

Name	Submodel 2		
GUID	c290d893-66be-453a-9a8d-3f5b2c649445		
Properties			
GUID	name	type	description
5d4a3d25-a63d-4924-9f1c-35d9d46f5bf0	Sex	Options	Biological classification
878660fc-9252-4900-a989-36f290503d73	Birthday	Date	Date of birth
95e85567-cdd5-4247-b37b-230d9a985222	Date	Date	Hiring date.
de854b37-bd07-413c-8b47-33457ecb9109	Job Experience	Options	Selection from Entry-level, Mid-level, Senior.
021817ec-cf00-4c1b-9652-580edead6df5	Worker condition	Options	Nervous, exhausted, happy, sad, etc.
33fdabdd-a6d3-47c1-a71f-4c5476f75533	Physical strength	Options	Selection from Low, Medium, High
7cbf7941-62b2-4cd6-9503-a6eaed5dd9eb	Stamina	Options	Selection from Low, Medium, High
d2741bc0-0971-4206-b9fb-3cdfda8875e7	Manual dexterity	Options	Selection from Basic, Intermediate, Advanced
85b40d59-bb17-48f6-92a8-d5b0f1d39915	Mental strength	Options	Selection from Low, Moderate, High
7f0013e7-f18b-450e-9eb9-cf727e0cfed5	Cognitive abilities	Options	Selection from Basic, Intermediate, Advanced
2e43e074-b2ab-4745-864a-ef94c4bcc7b7	Training and education	List	List of formal education, courses, certifications, and training programs

TABLE A.4: Asset Administration Shell Header of the Task types and characteristics

Field Name	Type/Example	Description	The internal GUID
AAS ID	URI, eCl@ss	Unique identifier for the AAS.	
Title	String	Task Asset Administration Shell.	76fe9622-2868-43f0-854f-0b9f91f13950
Date	Date	Date of creation or last modification.	
Task submodel	String	Task types and characteristics submodel	c8825b55-54c9-4137-ac61-273dc4ff4ca4
Manifest	Link/Reference	A directory of key information derived from the properties of the submodel.	

TABLE A.5: Task types and characteristics submodel

Name	Task submodel		
GUID	c8825b55-54c9-4137-ac61-273dc4ff4ca4		
Properties			
GUID	name	type	description
8e6f7a8e-27ba-47ff-81fc-b2997a6b8913	Routine Aspect	Options	Select between Routine, Non-routine
e2e60564-96ff-4866-9f6d-a66833ee8524	Nature of Task	Options	Select between Cognitive, Physical
ae84b07f-af41-4a52-958e-6da5e7ed7c5d	Social Aspect	Options	Select between Social, Individual
4e2fed17-1f66-4a16-88c6-cb5ce51fc633	Task Complexity	Options	Select between Low, Medium, High
9b056d18-8abc-4ca1-b3bc-97d1e13b63b2	Time pressure of Task	Time	Specific duration or deadline for task completion
22dc1ba6-723c-4e8a-9527-816ddb100bdc	Familiarity of Task	Numeric	Frequency of encountering the task (e.g., number of times per month)
b712cbd3-0564-4db5-b326-74cfa78d7ed1	Required skills level of Task	Options	The level of expertise needed (Beginner, Intermediate, Expert)
d5d58276-bd77-4bb2-9913-d7a8395aa682	Task Physical demands	Options	The level of physical exertion required (Light, Moderate, Heavy, Very Heavy)
6ef34463-196c-4445-9832-668dcc3bcd4c	Task Cognitive demands	Options	The level of cognitive exertion required (Simple, Moderate, Complex, Highly Complex)

TABLE A.6: Asset Administration Shell Header of the Environmental Conditions

Field Name	Type/Example	Description	The internal GUID
AAS ID	URI, eCl@ss	Unique identifier for the AAS.	
Title	String	Environmental condition Asset Administration Shell.	01adf4f2-2f32-472f-b893-6781f2a2cf8e
Date	Date	Date of creation or last modification.	
Environmental conditions submodel	String	Monitoring the worker's surrounding conditions	d8d345bf-2b1c-4491-abaa-b2087d3ac5ee
Manifest	Link/Reference	A directory of key information derived from the properties of the submodel.	

Table A.7: Environmental Conditions submodel

Name	Environmental conditions submodel		
GUID	d8d345bf-2b1c-4491-abaa-b2087d3ac5ee		
Properties			
GUID	name	type	description
3848608a-1e9d-40af-aad2-d2a362abc2ce	Noise level	Numeric	Ambient noise level
c05bff0a-7dad-4963-a483-a77e58f07676	Temperature	Numeric	Ambient temperature measured in degrees (°C or °F)
d2da3eb8-e527-498f-9b4e-0df652eb0116	Humidity	Numeric	Ambient relative humidity measured in percentage (%)

Table A.8: Description of features extracted from Electrodermal signals.

EDA Features	Description
SC Area	AreaSC $= \frac{100}{D}\int_0^T \mathrm{SC}(t)\,\mathrm{d}t$, where SC($t$) is the skin conductance at time t, D is the total duration of the recording for the specified repetition, and the factor of 100 serves as a scaling multiplier.
SCR Area	AreaSCR $= \frac{100}{D}\int_0^T \mathrm{SCR}(t)\,\mathrm{d}t$, where SCR($t$) is the skin conductance *response* at time t. SCR was found to correlate with higher sympathetic arousal, especially unpleasant arousal [202].
Number of Peaks	NoPeakes $= \frac{100 \times \#\mathrm{peaks}}{D}$, where #peaks is the total count of detected SCR peaks within the interval, and D again represents the total duration. Increased SCR peak frequency indicates heightened sympathetic reactivity to discrete stimuli [202, 203].
Average Rising Time [sec]	This function measures the average of the elapsed time from the moment when a stimulus manifests itself in the SCR to the moment when the signal reaches its peak value related to the same stimulus effect. Shorter SCR rise times correlate with greater sympathetic activation. This is because rapid sweat secretion kinetics reflect acute sympathetic discharge [202].
Average Decay Time [sec]	This feature captures how long it takes for the SCR to decline from its peak— produced by a specific stimulus—back to the baseline. Naturally, the decay time tends to be longer than the rising time due to the prolonged return process in skin conductance. Prolonged SCR decay may involve parasympathetic modulation of sweat gland recovery, although it primarily reflects sympathetic perspiration dynamics [204]
SCR Entropy	Entropy feature quantifies the complexity or irregularity present in the SCR signal. Higher entropy in the SC signals is indicative of sympathetic dominance, which is associated with irregular autonomic outflow patterns. In this work, the sample entropy algorithm developed by Víctor Martínez-Cagigal (2018) [205] was used, following the recommendations in Ref. [206].
SCR total power	The power spectral density (PSD) of the SCR signal was estimated using Welch's method [207].
STD SC	The standard deviation of the skin conductance (SC) signal.
STD SCR	The standard deviation of the skin conductance response (SCR).

TABLE A.9: Description of heart rate variability (HRV) features extracted from the PPG signal.

HRV Features	Description
RMSSD [msec]	The root mean square of successive differences between R-R intervals, is a time-domain metric of HRV. It reflects short-term, rapid fluctuations in heart rate and reflects parasympathetic activity [208, 209]. RMSSD $= \sqrt{\frac{\sum_{i=1}^{N-1}(RR_{i+1} - RR_i)^2}{N-1}}$, where N is the total number of RR intervals.
MEAN_RR	The average of interbeat intervals (in milliseconds), which is inversely correlated with heart rate and shows dual autonomic regulation, but predominantly reflects parasympathetic tone at higher values [210].
MEDIAN_RR	The median of all interbeat intervals (in milliseconds) which is less affected by outliers compared to the mean.
SDRR	Standard deviation of the RR intervals which reflects overall heartbeat variability.
SDSD	Standard deviation of the successive RR-interval differences.
SDRR/RMSSD	The ratio of SDRR to RMSSD and highlights the relationship between total HRV and short-term HRV.
Heart Rate (HR)	The instantaneous heart rate in beats per minute (bpm) reflects sympathetic-parasympathetic balance, with elevations indicating sympathetic dominance [210].
SD1	Poincaré plot short-term variability descriptor measures parasympathetic influence on beat-to-beat adjustments [209].
SD2	Long-term variability component in Poincaré plots.
pNN25	Percentage of successive RR differences differing more than 25 milliseconds.
pNN50	Percentage of successive RR differences differing more than 50 milliseconds.
KURT_RR	Kurtosis of the RR-interval distribution.
SKEW_RR	Skewness of the RR-interval distribution.
VLF_power	Spectral power in the Very Low Frequency range (0.003–0.04 Hz).
LF_power	Spectral power in the Low Frequency range (0.04–0.15 Hz).
HF_power	Spectral power in the High Frequency range (0.15–0.40 Hz).
Total_power	The sum of VLF, LF, and HF spectral components.
LF_HF_ratio	The ratio of LF power to HF power.
HF_LF_ratio	The inverse of LF_HF_ratio (i.e., HF power over LF power).

TABLE A.10: Extracted features from the acceleration magnitude signal.

Acc Features	Description
Mean	The average magnitude of the acceleration over the repetition duration.
Median	The middle value of the acceleration magnitudes. It offers a robust estimate of central tendency that is less affected by outliers.
StD	Standard Deviation: measures the variability or dispersion of the acceleration signal. It indicates how much the magnitude fluctuates over time.
Minimum	The lowest recorded magnitude during the segment shows the minimum intensity of movement.
Maximum	The highest recorded magnitude during the segment indicates peak movement intensity.
RMS	Root Mean Square: a measure of the power of the signal, combining both magnitude and variability. The higher RMS values generally reflect more intense or consistent motion.
Range	The difference between the maximum and minimum acceleration values.
Skewness	Describes the asymmetry of the distribution of the acceleration magnitudes. Positive skew indicates more small values and fewer high spikes, while negative skew indicates the opposite.
Kurtosis	Measures the "tailedness" of the distribution. The higher values indicate more frequent extreme acceleration events.
IQR	Interquartile Range: the range between the 25th and 75th percentiles of the acceleration data.
Zero-Crossing Rate	The rate at which the sign of the acceleration signal changes across samples.
PSD	Power Spectral Density: calculated using the Welch method, this feature estimates the total energy distributed across frequency components in the magnitude signal.
Entropy	Measure of the irregularity or complexity of the magnitude signal. Higher entropy indicates more unpredictable or variable movement. The sample entropy algorithm from Martínez-Cagigal (2018) [205] was used, based on the method of Richman and Moorman (2000) [206].