Assembly complexity and physiological response in human-robot collaboration: Insights from a preliminary experimental analysis

Robotics and Computer–Integrated Manufacturing 89 (2024) 102789

Contents lists available at ScienceDirect

Robotics and Computer-Integrated Manufacturing

journal homepage: www.elsevier.com/locate/rcim

Assembly complexity and physiological response in human-robot collaboration: Insights from a preliminary experimental analysis

Matteo Capponi, Riccardo Gervasi, Luca Mastrogiacomo, Fiorenzo Franceschini *

Politecnico di Torino, DIGEP (Department of Management and Production Engineering), Corso Duca degli Abruzzi 24, 10129 Torino, Italy

A R T I C L E I N F O

Keywords:
Assembly complexity
Human-robot collaboration
Physiological signals
Humans' well-being
Industry 5.0

A B S T R A C T

Industry 5.0 paradigm has renewed interest in the human sphere, emphasizing the importance of workers' well-being in manufacturing activities. In such context, collaborative robotics originated as a technology to support humans in tiring and repetitive tasks. This study investigates the effects of assembly complexity in Human-Robot collaboration using physiological indicators of cognitive effort. In a series of experiments, participants performed assembly processes of different products with varying complexity, in two modalities: manually and with cobot assistance. Physiological measures, including skin conductance, heart rate variability and eye-tracking metrics were collected. The analysis of physiological signals showed trends suggesting the impact of assembly complexity and cobot support. One key finding of the study is that a single physiological signal usually may not provide a complete understanding of cognitive load. Therefore, a holistic approach should be followed. This approach highlighted the importance of considering multiple measures simultaneously to accurately assess workers' well-being in industrial environments.

1. Introduction

The advent of Industry 5.0 defined a new era of manufacturing, characterized by a seamless integration of advanced technologies, such as robotics, artificial intelligence, and the Internet of Things (IoT), into traditional manufacturing environments [1]. Among these, human-robot collaboration (HRC) has emerged as a paradigm-shifting approach in which humans and robots work together to combine each other's capabilities: the precision and repeatability of the robot with the dexterity and flexibility of humans [2–4]. Although the benefits of HRC in terms of increased productivity and efficiency are well documented, understanding the impact of such collaboration on human well-being is critical for effective implementation and acceptance [5]. Recent studies began to explore the physiological responses of humans engaged in collaborative activities with robots, using parameters such as electrodermal activity, heart rate variability, and eye-tracking data to measure cognitive effort, and overall user experience [6–9]. These physiological measurements provide real-time information on the mental and emotional states of workers, providing a deep understanding of human factors in manufacturing and, more specifically, in human-robot collaboration processes. However, there is a gap in the literature regarding the influence of assembly complexity on these physiological responses. Assembly complexity, characterized by factors such as product variety, task organization, and variety of assembly sequences, has been widely studied in the scientific literature and is considered to be related to the effort required by humans to perform an assembly process [10,11]. Therefore, it can also potentially influence the cognitive load experienced by workers. Understanding this relationship is critical to optimize HRC environments and ensure not only productivity but also the well-being of workers. This paper aims to fill this gap by proposing an experimental campaign in which participants were asked to assemble different products of varying complexity in two modalities: manual and collaborative, that is, with the support of a cobot. Physiological signals such as electrodermal activity, heart rate variability and eye-tracking metrics were analysed in the experiment. Assembly complexity may increase cognitive load, with potential adverse effects on worker health and also on the quality of outputs. In addition, it is important to investigate how robots may actually support when assembly complexity changes. This knowledge could guide the development of more effective human-robot collaboration strategies aimed at improving workers' well-being and productivity. The paper is organised as follows: Section 2 provides a literature review on human factors in human-robot collaboration and on assembly complexity. Section 3 describes the material and methods of the experimental campaign. Section

* Corresponding author.
E-mail address: fiorenzo.franceschini@polito.it (F. Franceschini).

https://doi.org/10.1016/j.rcim.2024.102789
Received 29 February 2024; Received in revised form 19 April 2024; Accepted 10 May 2024

4 shows the main results of the study. Finally, section 5 and section 6 discuss respectively the main findings of the research and the key results obtained.

2. Literature review

In the last years, human well-being in manufacturing contexts has become a topic of primary concern [12,13]. Many studies underlined the importance of human factors to enhance companies' performances. Neumann and Dul [14] analysed 45 empirical studies that underlined the strong relationship between human effects (i.e., employee health, attitudes, physical workload a working life quality) and operation systems performance. Khamaisi et al. [15] developed a framework to assess user experience in industrial environments, using questionnaires and non-intrusive sensors to monitor stress-inducing activities that could affect worker performance. The analysis of human factors is even more crucial in human-robot collaboration processes [5]. Human-robot collaboration has been introduced as a novel manufacturing paradigm in which humans and robots cooperate sharing work-spaces and goals [16]. In such collaborative processes, humans and robots join their own capabilities, precision and repeatability for robots and flexibility and dexterity for humans. Khalid et al. [17] investigated the safety implications of using heavy robots in HRC, considering the physical and mental challenges of joint tasks. Kühnlenz et al. [7] studied the impact of different robot motion patterns on human workers by measuring heart rate variability and skin conductance. Galin and Meshcheryakov [3] analysed human-robot factors that affect the efficiency of HRC, highlighting the importance of emotional and cognitive elements. From another point of view, Colim et al. [18] formulated guidelines for the design of safe and ergonomic HRC workstations. Shirakura et al. [19] presented a human-robot collaboration system that optimise productivity by taking human workload into account. Lu et al. [20] investigated the mental stress caused by handover activities of a collaborative robot using both galvanic skin response and subjective ratings. The results suggested that handover activities can affect humans' mental stress levels. Moreover, Gervasi et al. [8] conducted a structured comparison between manual assembly and HRC, focusing on human-centred performance in repetitive assembly processes. They found that HRC can lead to improved performance and reduced cognitive load. Subsequently, they also analysed the evolution of psychophysical signals over time, showing how cobots can be a useful support in the human operators' learning process [9]. Finally, Barravecchia et al. [21] developed a novel approach to manage human-robot symbiosis in collaborative assembly tasks.

Given the critical role of human factors in HRC, it is crucial to consider potential causes that may influence them in manufacturing processes. An important aspect that impacts on the effort required to complete an assembly process may be its "complexity" [10]. The assessment of assembly complexity is a widely addressed topic in manufacturing. Hinckley [22] led the way by proposing a complexity factor based on assembly time, emphasising the need to minimise complexity to improve cost-effectiveness and quality. Shibata [23] and Su et al. [24] explored this further by linking assembly complexity to design and process-based factors, using time as an indirect measure of complexity. Afterwards, Genta et al. [25] proved the existence of a relationship between product complexity and product defects. Alkan [26], inspired by Sinha's generalised product complexity model [27], introduced a method using standard assembly times and DFA theory. This model was later adapted by Verna et al. [28] to predict product defects. Another diffused approach applies the principles of information theory to the assessment of assembly complexity. These methods assume that complexity arises from uncertainty in the assembly process. ElMaraghy and Urbanic [29,30] introduced the Manufacturing Complexity Assessment Tool (MCAT) model, which relates manufacturing complexity to the handling of information. Later, this concept has been extended by Zhu et al. [31] to deal with complexity arising from product variety. Ameri et al. [32] integrated information theory and graph theory to assess the complexity of product design. Samy and El-Maraghy [33] further refined MCAT model to assess product assembly complexity [33,34]. Wang and Hu [35] introduced a complexity measure that accounts for operator selection and fatigue, which was subsequently used to optimise mixed-model assembly systems. Zeltzer et al. [36] and Sun and Fan [37] introduced entropy-based measures that account for task duration variability. More recently, Liu et al. [38] developed an informational entropy measure to optimise the balance of assembly lines in conditions of demand uncertainty. In literature, there is a bunch of methods that provide a subjective assessment of complexity, the so-called "perceived complexity". Specifically, Mattsson et al. [39,40] identified 5 main causes of perceived complexity in assembly processes, i.e., product variants, layout, job content, tools and information). Similarly Falck et al. [41] proposed 16 basic complexity criteria for assembly tasks.

As these studies showed, the complexity of assembly tasks can vary widely, from simple, repetitive actions to complex procedures that require high levels of cognitive engagement and precision. This variability in task complexity can have a primary effect on physiological measures of cognitive effort in workers collaborating with robots. In a pioneering work, Verna et al. [42] explored the effects of perceived complexity on human performances within a collaborative assembly processes. In this work, Verna et al. [42] showed the existence of a relationship between assembly complexity and performance measures, such as product defects and completion times, in collaborative assembly. Specifically they found out that worse performances occurred in more complex products. Similarly, other studies showed the presence of a correlation between perceived assembly complexity and the subjective workload experienced by operators. [43]. Pollak et al. [44] analysed the differences in human stress between manual and autonomous assembly. Similarly, Fournier et al. [45] analysed the relationship between cognitive load and human-robot collaboration. Finally, Zakeri et al. [46] investigated the effects of task complexity and speed on human stress. Unlike previous studies, the novelty of this work lies in the comprehensive assessment of human cognitive load induced by assembly complexity, using several objective physiological metrics. This fills a gap in the existing manufacturing-related literature, providing a quantifiable understanding of the effect of assembly complexity from a physiological point of view.

3. Overview of the experimental approach

This study is designed to investigate the impact on humans' cognitive effort of both product assembly complexity and human-robot collaboration by physiological measurement (see Fig. 1). To this aim, an experimental approach was adopted. Participants were involved in performing a series of product assembly of different complexity, both manually (i.e., "manual modality) and with the support of cobot (i.e., "collaborative modality"). The focus was on measuring and analysing key physiological indicators such as skin conductance, heart rate variability and eye-tracking metrics to assess cognitive load. To clarify the differences between cognitive load and stress, it is important to understand both concepts in relation to human performance and well-being. Cognitive load refers to the amount of mental effort expended in working memory, while stress, on the other hand, can be described as the body's response to any demand or threat. [47–49]. Specifically:

- Cognitive load is more specifically concerned with the mental effort required to process information and perform tasks. High cognitive load can lead to difficulties in processing information, but, at the same time, appropriately managed cognitive load can enhance learning and performance through the efficient use of available cognitive resources [47].
- Stress is a broad concept that encompasses emotional, physical and psychological responses to perceived threats or challenges. It may be

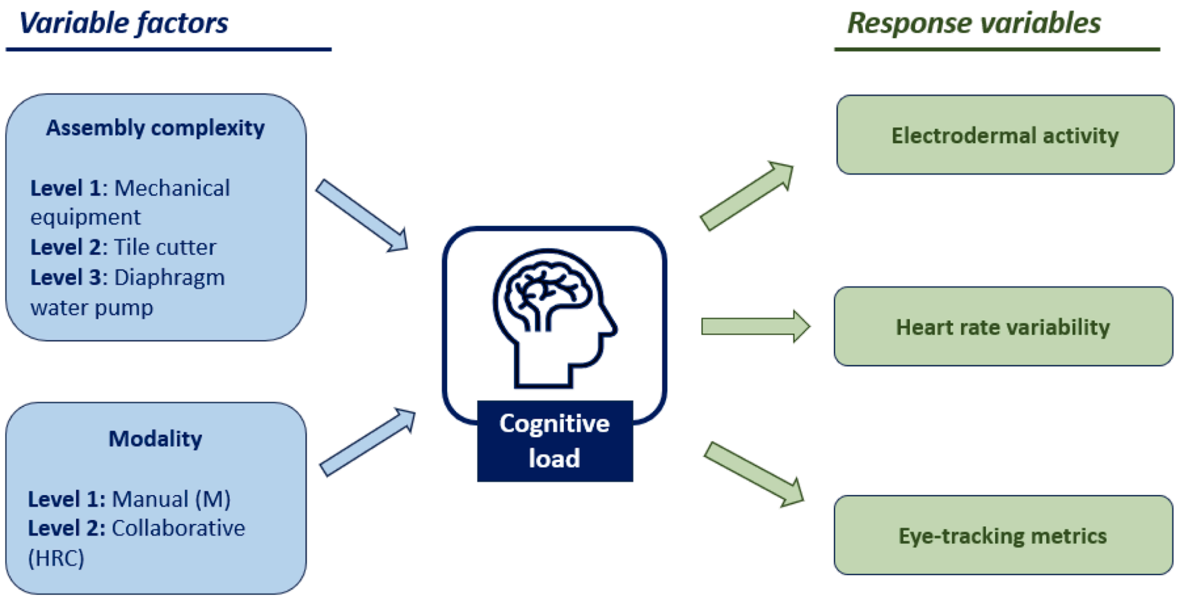

Fig. 1. A conceptual map of the experiment where "assembly complexity" and "modality" are the two variable factors that impact on human cognitive load assessed via electrodermal activity, heart rate variability and eye-tracking metrics.

caused by excessive mental arousal and activates the body's stress response systems, leading to various physiological and psychological changes aimed at coping with the stressor [49].

The relationship between cognitive load and stress is complex. Increased cognitive load can lead to heightened arousal and potentially stress if it exceeds an individual's ability to cope effectively. However, not all increases in cognitive load lead to stress. The outcome depends on factors such as the individual's perception of the task, their coping strategies and their ability to manage the cognitive demand, etc. [47–49]. In this experiment, given that assembly complexity is associated with the cognitive effort required in performing an assembly process [10], the relationship between physiological metrics associated to cognitive load and assembly complexity were tested. The relationship with stress cannot be a-priori identified because an increase in cognitive load may increase stress, but not necessarily [50]. For the sake of clarity, the terms 'cognitive load' or 'cognitive effort' will be used in discussing the results of this experiment. These terms more accurately describe the effort associated with increasing task complexity as opposed to stress which is a broader concept [10]. Details on the experimental settings and procedure follow in the next subsections. This approach aims to bridge the gap in current research by providing insights on how assembly complexity and cobot assistance influence human physiological responses of cognitive effort in a manufacturing context.

The hypotheses that will be tested through the experimental campaign are:

- Hypothesis 1: Increasing assembly complexity leads to higher cognitive load. Products composed of several parts, of different types and with many different alternatives of assembly sequences may require more cognitive effort for the operator.
- Hypothesis 2: The support of cobots relieves human cognitive load. Ideally, the presence of a cobot should support the human operator resulting in lower human effort. The cobot's support can be (i) both mental, for example, when giving the operator the parts in the exact assembly order, it punctuates the assembly sequence; (ii) and physical, since it performs the tasks that would be repetitive and strenuous for humans.

In this study, while recognising the multiple influences on operator cognitive load, we focused on assembly complexity and collaboration modality as the primary factors.

3.1. Assembly processes

In this experiment, participants were asked to assemble the three different products shown in Fig. 2. (i.e., a mechanical equipment, a tile cutter and a diaphragm water pump). Details of components and elementary task necessary to assemble these three products are respectively provided in appendix A.

With the same product, the assembly process was performed in two different modalities: manual and collaborative. In total each participant performed 6 different configurations of assembly process, named respectively 1HRC, 1M, 2HRC, 2M, 3HRC, 3M (see Table 1).

In manual modality participants performed all the tasks of the assembly process, while in collaborative modality some tasks were allocated to the cobot and others to human, according to Appendix A. Fig. 3 shows the assembly work-area at the Mind4Lab laboratory of Politecnico di Torino, where the experiment took place. The work-area was equipped with two collaborative robots UR3. For this experiment only the robot on the right hand-side was used. The work-area consisted of:

- a table on which both humans and cobots can operate simultaneously.
- physical supports to facilitate assembly operations of the tile cutter.
- a component feeding tray where all the parts were arranged. Both cobot and humans can pick parts from the component feeding tray.

3.2. Assessment of product assembly complexity

The three products considered are made of different quantity and type of parts leading to different "complexities". Greater assembly complexity should lead to greater effort required on behalf of the operator [10]. To assess product complexity, the Samy and ElMaraghy's method [33] was implemented. In detail, this method considers the diversity and amount of its components and connectors, as well as their geometric characteristics. They introduced a product assembly complexity index, denoted as $C_{product}$, which is calculated using the formula:

$$C_{product} = \left[\frac{n_p}{N_p} + CI_{product}\right]\left[\log_2(N_p+1)\right] + \left[\frac{n_s}{N_s}\right]\left[\log_2(N_s+1)\right] \quad (1)$$

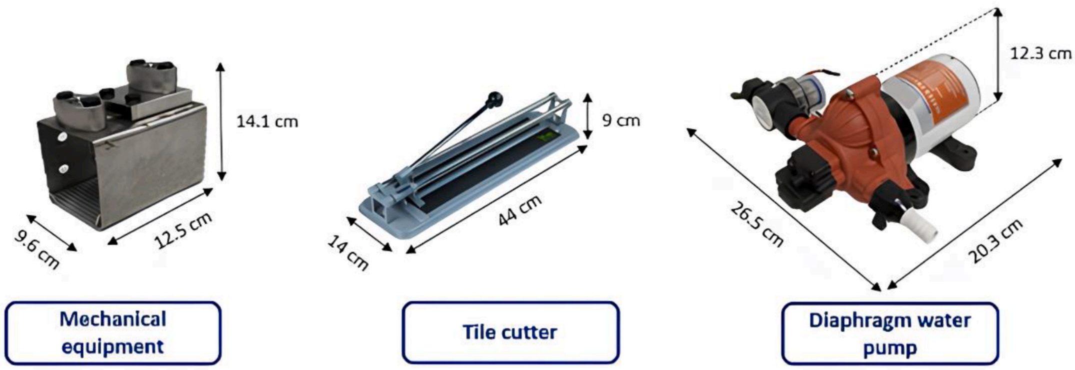

Fig. 2. The three reference products considered in the study.

Table 1
Configuration IDs and their respective descriptions.

Configuration ID	Description
1HRC	Collaborative assembly of the mechanical equipment
1M	Manual assembly of the mechanical equipment
2HRC	Collaborative assembly of the tile cutter
2M	Manual assembly of the tile cutter
3HRC	Collaborative assembly of the diaphragm water pump
3M	Manual assembly of the diaphragm water pump

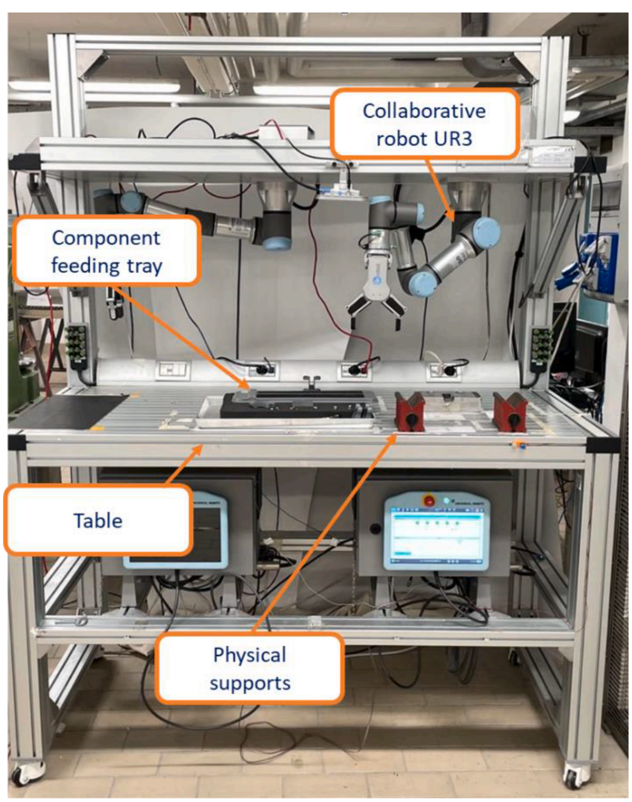

Fig. 3. The collaborative assembly work-area (Mind4Lab laboratory, Politecnico di Torino).

Where:

- n_p is the number of unique parts while N_p is the related total
- n_s is the number of unique fasteners while N_s is the related total
- $CI_{product}$ ranging from 0 to 1, is a complexity index that considers the geometric and dimensional characteristics of the parts. This index can be determined using the difficulty factors of manual handling and joining from Design for Assembly, as outlined in Samy and ElMaraghy's [33].

The implementation of this method to the three products considered in this study leads to the results reported in Table 2. According to the adopted complexity model, the least complex product is the mechanical equipment, while the most complex one is the diaphragm water pump.

The Samy and ElMaraghy's method provides a numerical scale for assessing product assembly complexity. However, it does not distinguish between manual and collaborative assembly processes for the same product. In a collaborative process, humans and robots interact using different technologies, means of communication, and supports which can also affect the effort required of the human operator to complete the process. However, traditional assembly complexity methods do not take into account the complexity of this interaction.

3.3. Experimental procedure

The experiment involved 18 participants. Participants (aged between 20 and 25 years old) were students of management and production engineering and claimed to have no prior experience with cobots. This choice was deliberately made to avoid bias derived from prior experience in assembly tasks. This approach ensures that the observed effects on physiological responses are primarily due to the experimental conditions rather than individual expertise or familiarity with the tasks. In this experiment two factors were considered: assembly complexity and modality. Modality has two levels: manual or collaborative. Assembly complexity has three levels: mechanical equipment, tile cutter and diaphragm water pump which can be ordered according to Samy and ElMaraghy's method [33] from the least to the most complex. For each product and modality, 6 repetitions (including two initial trials) were planned. Fig. 4 shows a flowchart of the experimental procedure for a single participant.

After a brief introduction, participants were shown the randomly

Table 2
Complexity values of the three reference products computed using Samy and ElMaraghy's method.

Configuration	*Samy and ElMaraghy H.*					
	N_p	n_p	N_s	n_s	$CI_{product}$	Complexity value ($C_{product}$)
Mechanical equipment	4	3	6	2	0.668	4.22
Tile cutter	10	8	5	3	0.682	6.67
Diaphragm water pump	13	12	13	4	0.693	7.33

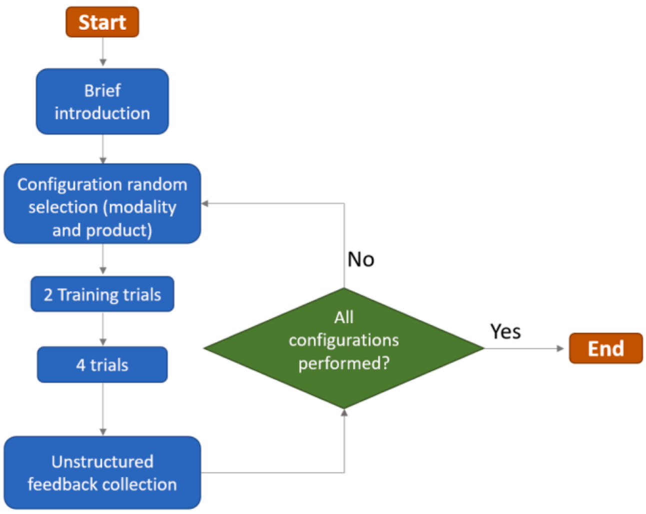

Fig. 4. Flowchart of the experimental procedure per single participant.

selected configuration to be performed. After two training trials, participants performed 4 repetitions of the same assembly process. During the experiment, participants were equipped with two non-invasive biosensors to collect psychophysiological parameters, i.e., Empatica E4 wristband and Tobii Pro Glasses 3 (see Fig. 5). After performing each single configurations, unstructured qualitative feedback on the experiment were collected. According to Charness et al. [51] this experiment involving each participant performing all six configurations can be defined as a "within-subject design" according to which all participants are subjected to all experimental conditions.

3.4. Data collected

During the experiments, several physiological signals were collected to obtain a global understanding of participants' cognitive load. Each signal provides unique insights, and their collective analysis provides a more accurate and deep understanding of physiological responses to different stimuli. The main variables collected were:

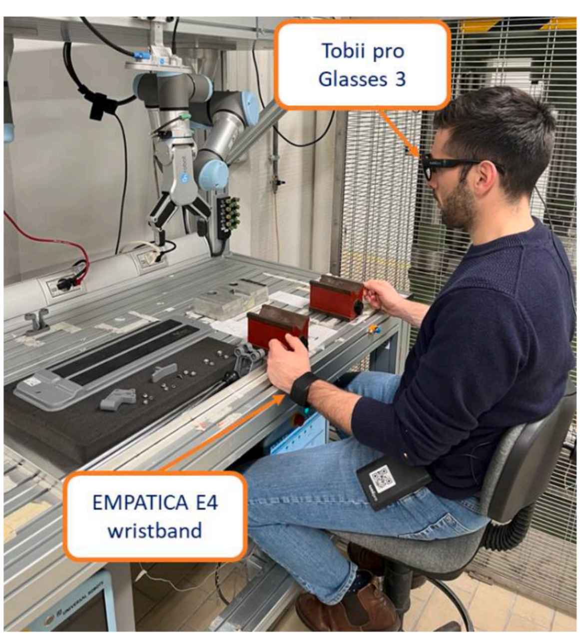

Fig. 5. The setting of a collaborative assembly process for the tile cutter with the operator wearing the Empatica E4 and the Tobii Pro Glasses 3.

- Electro-Dermal Activity and Heart Rate Variability using the non invasive biosensor Empatica E4. The device acquires two types of physiological data: electrodermal activity (EDA) at a frequency of 4 Hz and heart rate data by photoplethysmography (PPG) at 64 Hz. "Ledalab," a MATLAB-based tool, was also used to process the EDA data.
- Eye-tracking data using Tobii Pro Glasses 3 with a 100 Hz sampling rate. The data were processed using Tobii Pro Lab software. For each participant and configuration, time ranges of interest (i.e., the trials of the assembly process) were identified manually by inserting a timestamp indicating the start of the trial and the respective end. Thus, for each eye-tracking metric, the Tobii Pro Lab software computes an overall value for the time of interest considered.

The following subparagraphs provide a short explanation of the data collected.

3.4.1. Electro-dermal activity (EDA)

EDA measures the electrical conductance of the skin, which refers to the activity of the skin sweat glands. This moisture level is controlled by the sympathetic nervous system and is directly related to emotional arousal, regardless of whether the emotion is positive or negative. Increased EDA is typically associated with increased emotional arousal, stress, or cognitive load [52]. Using continuous decomposition analysis (CDA), the EDA signal was separated into phasic and tonic activities. Tonic activity is best identified by observing changes in skin conductance level (SCL), which represent sustained fluctuations in EDA not directly related to external stimuli. Phasic activity, on the other hand, refers to short-lived alterations in EDA that occur in response to an identifiable, externally applied stimulus. These changes are known as skin conductance responses (SCRs), which are essentially shifts in amplitude from the baseline SCL to the highest point of the response, and are discernible through examination of the phasic activity component. [53]. In this work two metrics for EDA were analysed: Average SCL computed as the average value of the SCL signal within a single trial and the average SCR calculated as the average value of SCR for each trial.

3.4.2. Heart rate variability (HRV)

Heart rate variability (HRV) is another critical measure, offering insight into the analysis of cognitive effort. HRV refers to the variation in time intervals between consecutive heartbeats and is an indicator of the body's adaptability to cognitive effort required that may lead to increased stress. Higher variability indicates healthy balance and good adaptive capacity of the autonomic nervous system, while lower variability suggests fatigue or overexertion. By analyzing HRV, we can assess how individuals respond physiologically to stress and how they recover from it. Recent studies also showed that the effects of mental workload can also be described by non-linear metrics related to cardiac activity [54,55]. However, in this work two common HRV metrics will be analysed: RMSSD and SDNN. RMSSD (i.e., Root Mean Square of Successive Differences) is defined as:

$$RMSSD = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N-1}(NN_{i+1} - NN_i)^2} \quad (2)$$

Where N is the number of systolic peaks in the considered time window and NN_i indicates the time interval between the systolic peak i and i+1.

Another common measure for heart rate variability is the so-called "SDNN" (i.e., "Standard Deviation of NN intervals"). The SDNN index reflects all cyclic components responsible for variability over the recording period, so it represents total variability. A higher SDNN value indicates greater heart rate variability, which is often interpreted as a sign of a healthy heart that can respond adaptively to changing environmental and physiological conditions. In contrast, a lower SDNN value suggests less variability and may indicate greater effort and fatigue [56,57].

3.4.3. Eye-tracking measurements

In the last years, many studies proved the efficacy of using eye-tracking data to describe cognitive load [58]. In this experiment, the metrics collected were:

- Fixation number: This metric counts the number of fixations, usually within a specific area of interest (AOI). Fixations can be defined as the periods of time during which the eyes are relatively still, and the central foveal gaze is maintained, allowing the eye-tracking device to collect information about what is being looked at. In Tobii Pro Lab, a fixation is a sequence of raw gaze points where the estimated velocity is below the velocity threshold set in the I-VT gaze filter (e.g., 30°/s) [59]. The amount of fixations is considered indicative of the allocation of visual attention, where a greater number of fixations might suggest less efficient search or greater visual effort, and thus a higher mental workload [58].
- Fixation duration [ms]: This refers to the time spent focusing on a single position. It is computed as the time interval between the first gaze point and the last gaze point belonging to a fixation. A longer duration could indicate more cognitive effort in extracting information from specific point or greater attractiveness of the target [58]. However, in the literature there's still debate to understand how fixation duration and number of fixation can be correlated with cognitive effort [58,60].
- Saccade number: This is the number of saccades within a given area of interest. Saccades are the type of eye movement used to move the fovea rapidly from one point to another. In other words, a saccade event is an involuntary movement of both eyes at the same time that shift the fixation point. In Tobii Pro Lab, a saccade is a sequence of raw gaze points where the associated velocity is above the aforementioned velocity threshold. A higher number of saccades is often related to greater visual effort and thus greater mental workload [58,59].
- Saccade duration [ms]: Measures the time from the beginning to the end of a saccade event, indicating the transition from one fixation to another. A longer duration suggests less processing and more visual search activity [58,59,61].
- Saccade amplitude [°]: Quantifies the degrees of visual arc movement from one fixation centroid to the next. As mental workload increases, saccade amplitude generally decreases [58,59,61].
- Saccade peak velocity [°/s]: This is the speed of the saccade, usually measured by its highest velocity, i.e., the peak velocity. The average peak saccade velocity decreases as mental workload increases [58,59,61].
- Pupil size [mm]: Measures the diameter or area of the pupil. A larger pupil size is often associated with greater mental demand [58].

3.4.4. Final unstructured feedback from participants

At the end of the experiment, each participant underwent an unstructured interview. The question asked were the followings:

- General impressions and feedback
- What was in your opinion the least and the most complex product to assemble?
- How did your experience of assembling the products change when working with the cobot compared to working without it?
- Did the presence of the cobot change your stress level or your approach to the assembly tasks?
- For each of the three products, how did the cobot supported you?
- Did the usefulness of the cobot change with the complexity of the product?

4. Experimental results

A detailed investigation of various physiological parameters described in Section 3.4. was conducted, emphasizing the importance of a holistic approach to obtain reliable results. The experimental database consisted of 648 observations: 18 participants performed 6 trials for each of the 6 configurations (36 trials overall). For each metric, the individual values were standardised to account for personal differences by calculating z-scores using the formula:

$$z_{ij} = \frac{x_{ij} - \overline{x}_j}{s_j} \tag{3}$$

where z_{ij} is the z-score i for participant j, x_{ij} the observation i for participant j, $\overline{x}_j$ the sample mean for participant j, and s_j the sample standard deviation for participant j, with $i = 1, \ldots, 6$ and $j = 1, \ldots, 18$. This standardization process is crucial to mitigate individual variability, ensuring that data from different subjects can be directly compared. This process adjusts for personal baseline levels and individual response magnitudes and allows to discern patterns and changes attributed to experimental conditions rather than inherent physiological differences. As an example, Table 3 shows the calculation of the average SCR values scaled on the participant. The mean $\overline{x}_j$ and the standard deviation s_j are unique for each participant, i.e., calculated on the observations of the entire 6 trials.

Furthermore, to account for potential variability introduced by task duration, physiological metrics averages were calculated for each trial to ensure comparability across different assembly complexities and modalities. For the same reason, the number of fixations was divided by the duration of the trial. The statistical analysis was performed using R and involved the followings step:

- Testing the distribution of each physiological metrics using the Shapiro Wilk-test [62];
- Comparing differences between the six configurations using a paired t-test if the normality distribution cannot be rejected, or the Wilcoxon signed-rank test [63] if the normality hypothesis is rejected. For each physiological signal, the normality assumption was always rejected at least in one of the six configurations to compare. Hence, only Wilcoxon signed-rank tests were performed. Furthermore, this non parametric-test is less influenced by the presence of outliers, since it is based on the comparison of median values.

The two initial training trials were neglected in the following analyses. It was also decided to keep all the outlier observations that were not due to measurement or signal acquisition errors. During the four-hour duration of the experiment, a remarkable range of extreme physiological responses were observed. These can be primarily attributed to the dynamic and prolonged nature of the experimental environment, in which the participant may be exposed to several instantaneous stimuli. Factors such as individual perception of assembly process difficulty, the onset of fatigue over time, or environmental influences may contribute to these pronounced responses. Although standardisation techniques and baseline adjustments were used to reduce inter-individual variability, the outliers predominantly reflect the influence of unavoidable contextual factors inherent in long-term experiments. Therefore, these observations were included as integral part of the dataset to ensure that the results were representative of real biological responses in different scenarios.

4.1. Electro-dermal activity

The analysed metrics related to EDA were: Average SCL and Average SCR. Higher levels of Average SCL and Average SCR are often indicative of increased effort as they reflect increased arousal of the sympathetic nervous system in response to stressors [53,64,65]. Therefore, with

Table 3
Example of a z-score standardisation for the average SCR [microSiemens μS] values of a participant of the experiment.

Participant	Configuration	Trial	Average SCR (x_{ij}) $[\mu S]$	$\bar{x}_j$ $[\mu S]$	$s_j [\mu S]$	Average SCR Scaled (z_{ij})
1	1HRC	1	0.1897	0.1948	0.1555	-0.0323
1	1HRC	2	0.1690	0.1948	0.1555	-0.1659
1	1HRC	3	0.0888	0.1948	0.1555	-0.6813
1	1HRC	4	0.0800	0.1948	0.1555	-0.7379
1	1HRC	5	0.0558	0.1948	0.1555	-0.8933
1	1HRC	6	0.0477	0.1948	0.1555	-0.9457

reference to the hypothesis reported in section 3, it should occur that:

- Hypothesis 1: as product complexity increases, Average SCL and Average SCR should increase.
- Hypothesis 2: in collaborative modalities, Average SCL and Average SCR should be lower than in manual counterparts.

4.1.1. Average SCR

The boxplot presented in Fig. 6 shows the z-scores of the average skin conductance response (SCR) for the six different configurations of assembly processes considered.

From the boxplot, it can be seen that, at the same product, average SCR values for manual configurations generally show a higher median z-score than their human-robot collaboration counterparts. This suggests a trend whereby manual tasks might be associated with higher effort. The boxplot also reveals a slight upward trend in average SCR as product complexity increases from the mechanical equipment to the diaphragm water pump, accordingly to the complexity ranking provided with the Samy and ElMaraghy's method. To test statistical differences between the six configuration the paired Wilcoxon signed rank test was implemented. The results are shown in Table 4. These results provide initial evidence that the complexity of products may influence psycho-physiological responses, as measure by SCR. However, for the same product the support of cobot is not particularly evident.

4.1.2. Average SCL

The boxplot in Fig. 7 represents the z-scores of mean skin conductance level (SCL). It shows a general trend of increasing median SCL values from simpler to more complex products in manual (M) configurations. This trend is consistent with the hypothesis that greater assembly complexity is associated with increased effort levels. For the same product, there is a visible increase in median z scores. It is worth noting that when the product is more complex, the support of the cobot is evident, as suggested by the lower median values of HRC modality for both the tile cutter and the diaphragm water pump with respect to their manual modality configuration. The p-value table (see Table 5) provides

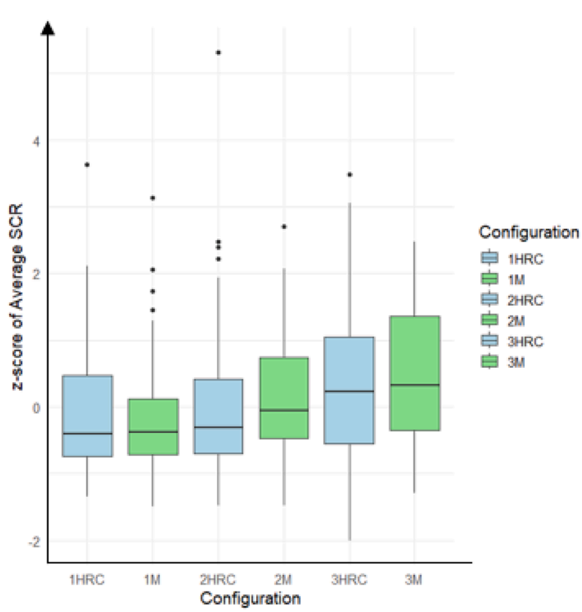

Fig. 6. Boxplot of Average SCR z-scores across the six different configurations considered in the experiment.

Table 4
Results of the pairwise Wilcoxon signed rank test for the z-score median of the average SCR (results in bold are statistically significant, i.e., p-value < 5 %).

p.adjust	1HRC	1M	2HRC	2M	3HRC	3M
1HRC	-					
1M	1.0	-				
2HRC	1.0	1.0	-			
2M	0.3265	0.1829	0.8264	-		
3HRC	0.1574	0.0577	0.4182	1	-	
3M	**0.0076**	**0.0041**	**0.0237**	0.3030	1.0	-

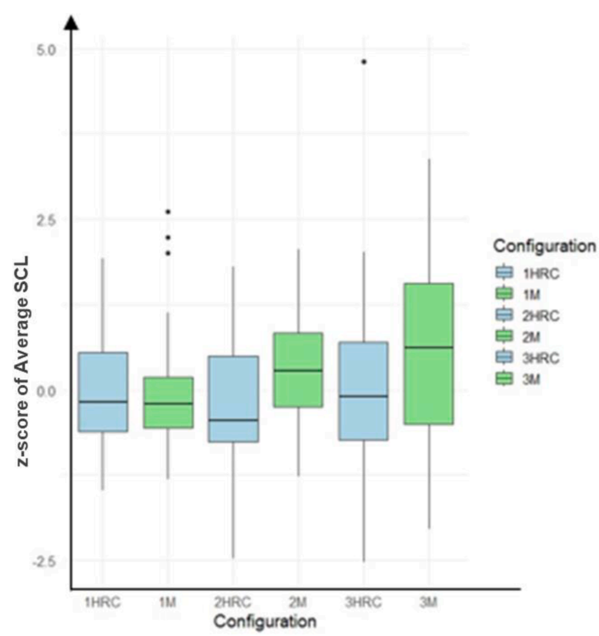

Fig. 7. Boxplot of Average SCL across the six different analysed configurations.

Table 5
Results of the pairwise Wilcoxon signed rank test for the z-score median of Average SCL (results in bold are statistically significant, i.e., p-value < 5 %).

p.adjust	1HRC	1M	2HRC	2M	3HRC	3M
1HRC	-					
1M	1.0	-				
2HRC	1.0	1.0	-			
2M	0.316	**0.019**	**0.005**	-		
3HRC	1	1	1	1	-	
3M	0.160	**0.030**	**0.018**	1	0.27	-

statistical support for some of the observed differences in the boxplot. In particular, there are significant differences between the 2HRC and 2M configurations, as well as between 2M and 1M, supporting the idea that both assembly complexity and the presence of the cobot may affect cognitive effort.

4.2. Heart Rate variability

Regarding heart rate variability, RMSSD and SDNN were analysed. Under condition of increased effort, both RMSSD and SDNN typically decrease, indicating a dominance of sympathetic nervous system activity over parasympathetic activity [57]. With regard to the central hypotheses of this study, it should be proved that:

- Hypothesis 1: as product complexity increases, RMSSD and SDNN should decrease.
- Hypothesis 2: in collaborative modality, RMSSD and SDNN should be higher.

4.2.1. RMSSD metric

The boxplot representing RMSSD z-scores across the six configurations (see Fig. 8) suggests a weak relationship between assembly complexity, modality, and psychophysiological response. Therefore, the median RMSSD z-scores remains broadly consistent across configurations.

The distributions of RMSSD z-scores in the manual and human-robot collaboration (HRC) modality for the three levels of assembly complexity show some overlap with no clear trends. Hence, the RMSSD metric provided no significant evidence to support the hypothesis of this work, suggesting that this particular measure of heart rate variability may not be sensitive to the specific types of cognitive load induced by the assembly complexity and cobot interaction in this study settings. Furthermore, also the results of the Wilcoxon signed-rank test did not show any statistical difference.

4.2.2. SDNN metric

The results of SDNN for the six configurations are shown in Fig. 9 and Table 6. It can be seen that moving from less complex tasks (1HRC, 1M) to more complex tasks (3HRC, 3M), there is no consistent increase or decrease in median z-scores. However, it can be seen that for the mechanical component and the tile cutter, the median values of the collaborative modality are higher than their manual counterparts, thus indicating a lower cognitive load experienced by the operator and confirming the supporting role of the cobot. However, the lack of a uniform trend across all configurations implies that the relationship between the studied variables and the SDNN is unclear and may be influenced by a number of factors. This variability in the data reinforces the importance of a comprehensive approach when using physiological measures such as the SDNN to assess cognitive load.

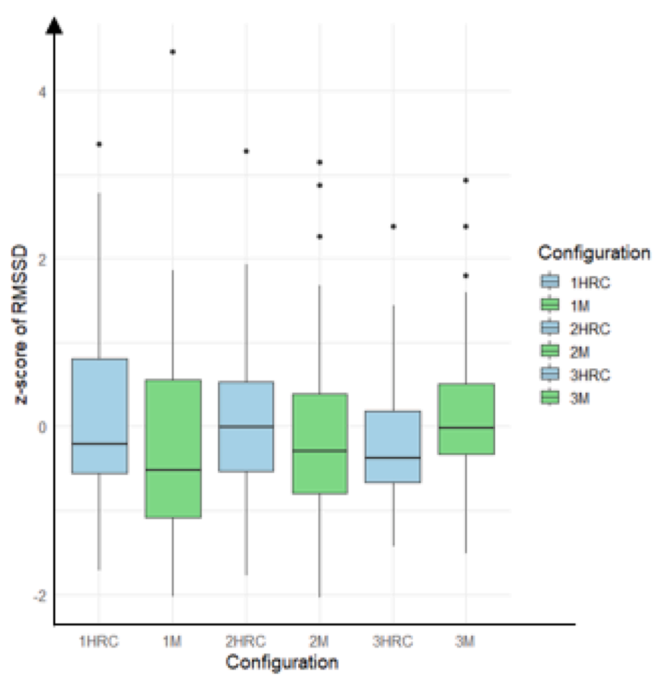

Fig. 8. Boxplot of RMSSD across the six different configurations.

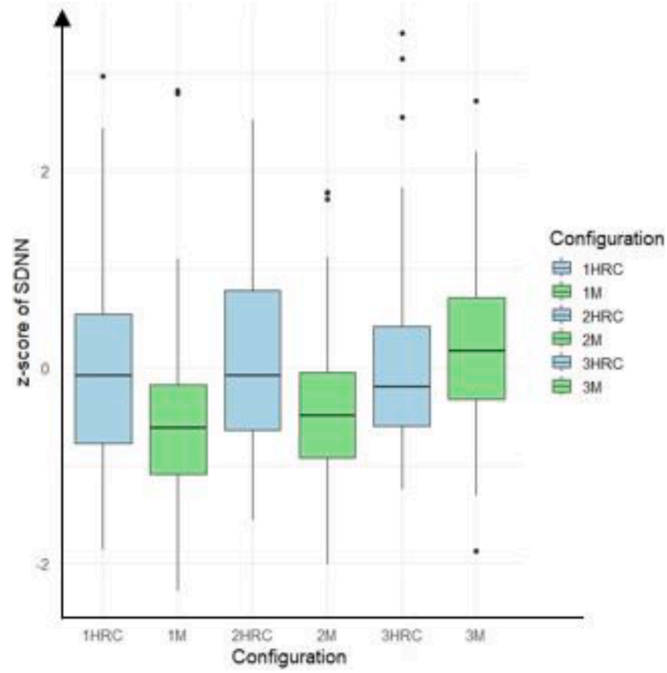

Fig. 9. Boxplot of SDNN across the six different configurations.

Table 6
Results of the pairwise Wilcoxon signed rank test for z-score median of SDNN (results in bold are statistically significant, i.e., p-value < 5 %).

p.adjust	1HRC	1M	2HRC	2M	3HRC	3M
1HRC	-					
1M	0.685	-				
2HRC	1.0	0.073	-			
2M	1.0	0.281	0.647	-		
3HRC	1.0	0.196	1.0	0.166	-	
3M	1.0	**0.029**	1.0000	**0.018**	1.0000	-

4.3. Eye-tracking metrics

The eye-tracking metrics considered were the ratio between number of fixations and trial duration; the average duration of fixations; the average pupil diameter; the average peak velocity of saccades; and the average amplitude of saccades. In eye-tracking analysis, as cognitive effort increase, the first three metrics should increase while the average amplitude of saccades and the average peak velocity of saccades are expected to show an inverse trend [58]. Hence, the hypotheses that will be tested in this work are:

- Hypothesis 1: as product complexity increases, the ratio between number of fixations and trial duration, the average duration of fixations and the average pupil diameter should increase. On the contrary, the average amplitude of saccades and the average peak velocity should decrease.

- Hypothesis 2: in collaborative modalities the ratio between number of fixations and trial duration, the average duration of fixations and the average pupil diameter should be lower than in manual ones (vice versa for the average amplitude of saccades and the average peak velocity).

4.3.1. Number of fixations/trial duration

A greater number of fixations can be related to a greater workload perceived. Obviously, not all the trials had the same duration: assembling a mechanical equipment took less than assembling the diaphragm water pump. To take this into account the number of fixations was divided by the duration of the trial. The boxplot for the ratio number of fixations – duration is shown in Fig. 10.

The boxplot represents the standardized number of fixations per unit time among six task configurations. Although the median z-scores of fixation frequency per unit time do not show a clear increase consistent with greater product complexity, or a decrease for HRC modalities compared to manual tasks. It is plausible that standardization by trial duration may have normalized the data, smoothing out the expected differences. Furthermore, the variability within each configuration, reflected by the range of interquartile distances, suggests that there is a diversity of individual experiences and responses to assembly complexity and modality. The Wilcoxon signed-rank test only showed two statistically significant results: 1M-1HRC (p-value=0.021) and 3M-3HRC (p-value=0.044) (Table 7).

Table 7
Results of the pairwise Wilcoxon signed rank test for z-score median number of fixations/duration (results in bold are statistically significant, i.e., p-value < 5 %).

p.adjust	1HRC	1M	2HRC	2M	3HRC	3M
1HRC	-					
1M	**0.021**	-				
2HRC	1.0	1.0	-			
2M	1.0	1.0	1.0	-		
3HRC	1.0	0.09	1.0	1.0	-	
3M	0.34	1.0	0.100	1.0	**0.044**	-

4.3.2. Average duration of fixations

The average duration of fixation is related to the workload experienced by the participants. Specifically, it increases when the workload increases. The results of this experiment are shown in Fig. 11 and Table 8. Moving from tasks involving the least complex product (1HRC, 1M) to the most complex (3HRC, 3M), there is a gradual increase in median z scores for both manual and human-robot collaboration (HRC) modalities. Although not all pairwise comparisons between configurations show statistical significance, the overall trend within each modality supports the idea that product complexity positively correlates with cognitive load. This trend is consistent with the hypothesis that more complex tasks require more cognitive effort, which is reflected in longer fixation duration. Complex products, indeed, often involve longer and more joining processes with tight tolerances and smaller components involved. In these situations, the need for meticulous attention to specific points results in longer fixations. However, also a different interpretation to these results can be proposed, using the flow theory. Flow is a subjective state that results from deep engagement in an activity. In this case, longer duration of fixations may reflect intense concentration which does not necessarily leads to stress conditions [66, 67]. Again, the existence of different interpretations for the same eye-tracking metrics highlights the importance of analysing multiple physiological signals in human factors studies.

4.3.3. Average amplitude of saccades

The average amplitude of saccades represents an indicator of cognitive load. Specifically, lower values of the amplitude of saccades indicate higher effort. The results for this case study are shown in Fig. 12 and Table 9.

From the boxplot, we can observe that for each level of product complexity, human-robot collaboration (HRC) configurations tend to

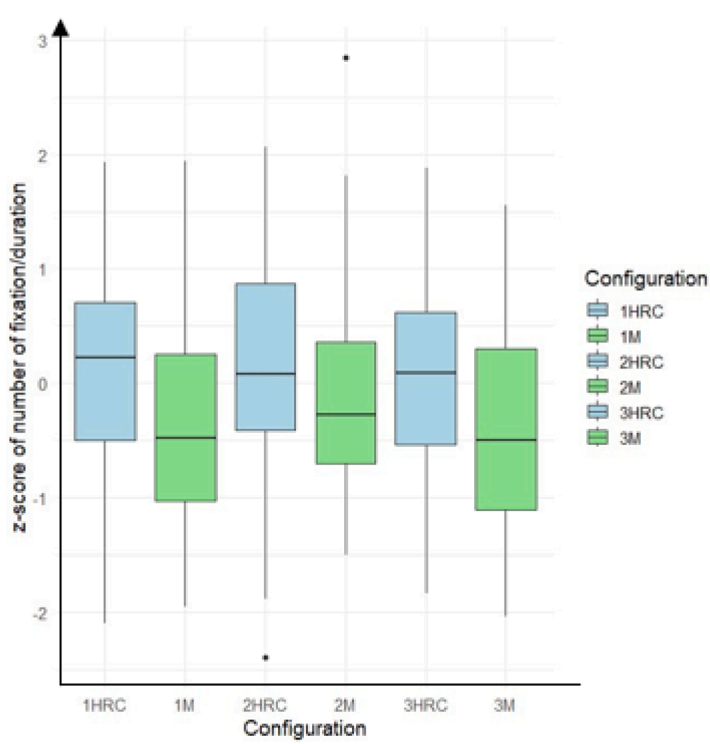

Fig. 10. Boxplot of the ration between number of fixations and trial duration across the six different configurations.

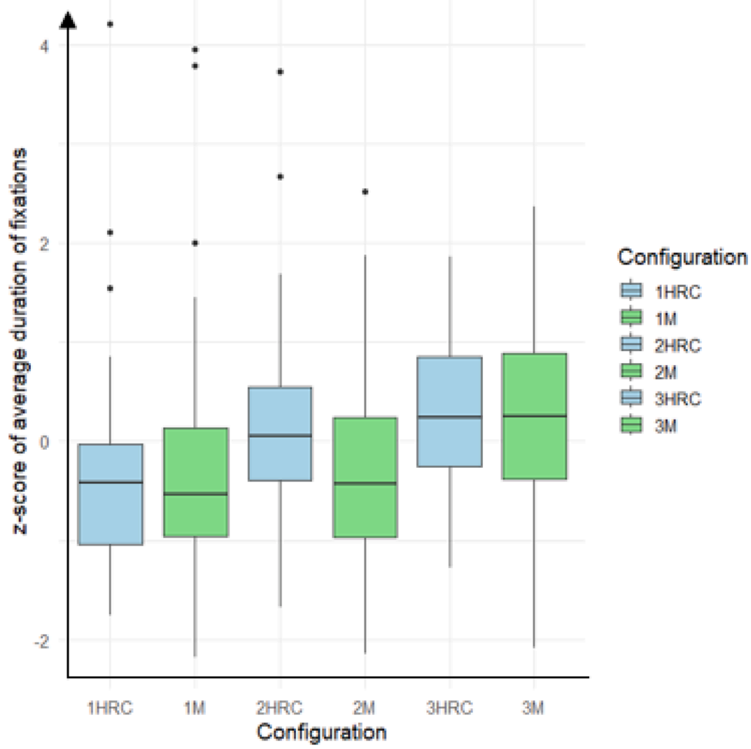

Fig. 11. Boxplot of average duration of fixations across the six different configurations.

Table 8
Results of the pairwise Wilcoxon signed rank test for z-score median of the average duration of fixations (results in bold are statistically significant, i.e., p-value < 5 %).

p.adjust	1HRC	1M	2HRC	2M	3HRC	3M
1HRC	-					
1M	1.0	-				
2HRC	**< 0.001**	**0.00101**	-			
2M	1.0	1.0	0.36040	-		
3HRC	**< 0.001**	**0.00427**	0.92561	**0.00210**	-	
3M	**0.00412**	**0.00195**	1.0000	**0.01094**	1.0000	-

have higher median z-saccade amplitude scores than manual configurations. A slight decreasing trend can also be seen while moving from the mechanical equipment to the diaphragm water pump. This trend suggests that participants experienced lower levels of cognitive effort when working with cobots. The results of Wilcoxon signed rank test show some statistical significances, confirming the fact that both assembly complexity and the presence of cobot may affect human workload. Generally, a slight trend can be seen between the mechanical equipment and the diaphragm water pump, indicating the presence of a relationship between the average amplitude of saccades and the assembly complexity. In conclusion, the data from this boxplot support our hypothesis by showing that the average amplitude of saccades is greater in HRC tasks than in manual tasks, indicating a reduction in cognitive load levels when cobots are involved in the assembly process. This result is significant because it highlights the potential benefits of integrating cobots into industrial workflows, not only to maintain efficiency but also to improve the psycho-physiological working conditions of human operators.

4.3.4. Average peak velocity of saccades

The boxplot of z-scores for average peak velocity of saccades (see Fig. 13 and Table 10) do not show a significant or consistent trend across task configurations. The absence of a clear pattern in the data may suggest that the average peak saccade velocity does not vary systematically with assembly complexity or with the introduction of cobot assistance. The lack of significant differences between most configurations, particularly between different complexities within the same modality or between manual and cobot-assisted tasks, indicates that the relationship with assembly complexity is not straightforward. However, for the same product the average peak velocity is always higher in collaborative modality than in manual one. This suggests a reduced cognitive load required when participants worked with cobots. This result highlights the critical need for a comprehensive approach in the study of physiological signals as indicators of cognitive workload.

4.3.5. Average pupil diameter

The average pupil diameter can be interpreted as an indicator of cognitive load, where larger diameters suggest greater workload. From the results provided in Fig. 14 and Table 11, it can be observed a slight trend among configurations. In the first four configurations (1HRC, 1M, 2HRC, 2M), there seems to be a gradual increase in median z scores. This trend could be indicative of a relationship between assembly complexity and cognitive load, as configurations associated with a higher level of complexity (2M and 2HRC) show an increase in pupil diameter compared to those with lower complexity (1HRC, 1M). Furthermore, when comparing collaborative tasks with their manual counterparts (1HRC vs. 1M and 2HRC vs. 2M), the median z-scores for collaborative tasks are actually lower. This observation is in line with the hypothesis that cobot assistance can lighten the cognitive load of human workers. However, it can be seen an inversion of trend for the last two configurations (i.e., 3HRC and 3M). The reversal trend for the last two boxplot configurations could also reflect subjective perception of complexity. From the unstructured feedback collected during the experiment, a subset of participants (5 out of 18) found the diaphragm water pump assembly process easier and less mentally demanding than the tile cutter assembly. With regard to flow theory, Lu et al. [66] showed pupil diameter is higher in those task where people reported more flow. In this case, the overall pupil diameter seems higher in tasks “2HRC” and “2M”

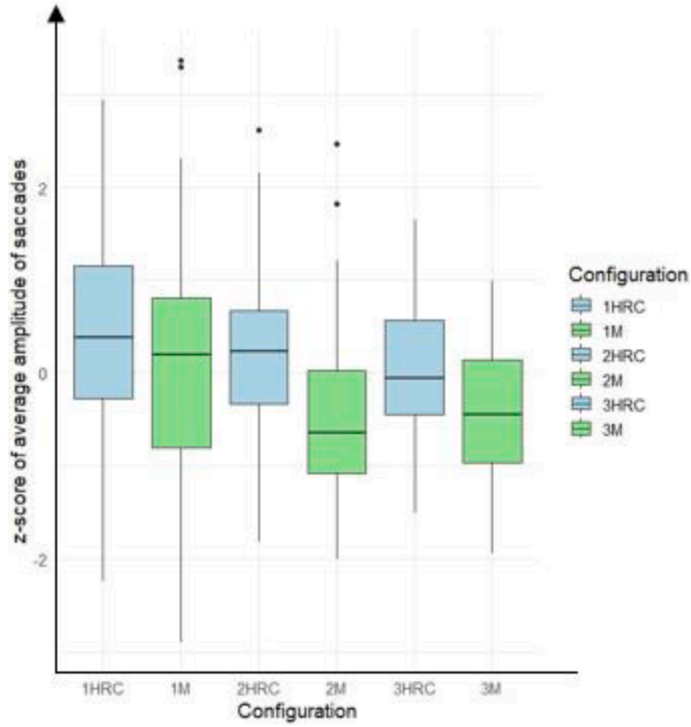

Fig. 12. Boxplot of the ration between average amplitude of saccades across the six different configurations.

Table 9
Results of the pairwise Wilcoxon signed rank test for z-score median of the average amplitude of saccades (results in bold are statistically significant, i.e., p-value < 5 %).

p.adjust	1HRC	1M	2HRC	2M	3HRC	3M
1HRC	-					
1M	1.0	-				
2HRC	1.0	1.0	-			
2M	**< 0.001**	**0.03470**	**0.00163**	-		
3HRC	0.08743	1.0	1.0	**0.00250**	-	
3M	**< 0.001**	0.43	**0.0025**	1.0	**0.05405**	-

that some participants definitely found more engaging. These subjective experiences further underscore the importance of adopting a holistic approach when dealing with physiological signals.

4.4. Unstructured feedback results

During the experiment, unstructured feedback was collected from participants to complement quantitative findings. Initially, all participants viewed the cobot as non-supportive, slowing down processes for mechanical equipment. However, its value was recognised as the complexity of the assembly increased, highlighting its usefulness in complex scenarios. Specifically, 12 out of 18 participants highlighted the cobot's usefulness in assembling the tile cutter's cutting mechanism and maintaining the pump's engine block, where it held parts in place to facilitate manual screwing processes. Furthermore, in terms of product complexity, 13 participants agreed with the complexity order obtained by Samy and ElMaraghy's method and claimed to feel more stressed and scared of making errors when assembling the diaphragm water pump. A notable learning effect was reported by four participants, who noted that the tasks became significantly easier after the fifth trial, suggesting rapid adaptation to the cobot's functionality and the assembly process over time. Finally, no participant claimed to feel uncomfortable of working closely with a robot. Overall, this feedback highlights the potentiality of collaborative environments particularly in complex assembly tasks.

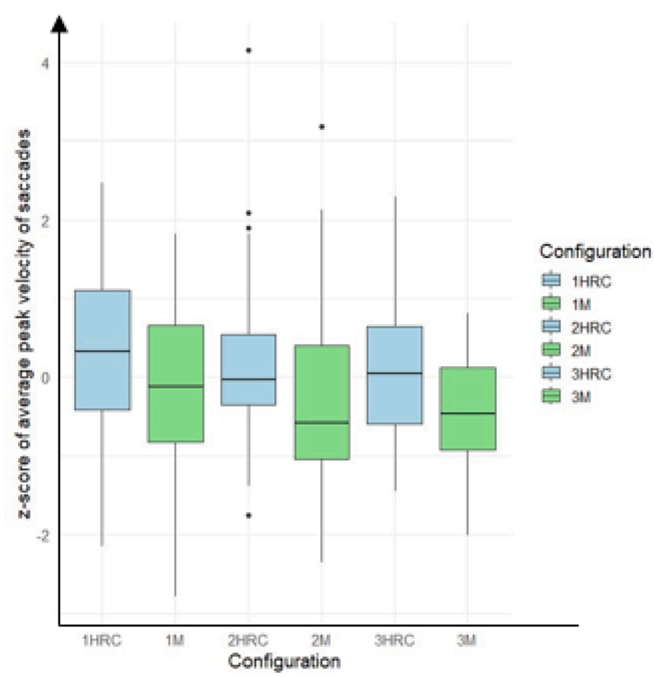

Fig. 13. Boxplot of average peak velocity of saccades across the six different configurations.

5. Discussion

In this experiment, the study of physiological signals has consistently highlighted the difficulties of analysing and interpreting human responses in industrial settings. As shown in Table 12, physiological metrics provide evidence of the multiple nature of human cognitive load under different conditions. Although cobots were introduced with the goal of reducing workers' cognitive load, the data suggest that their

Table 10
Results of the pairwise Wilcoxon signed rank test for z-score median of the average peak velocity of saccades (results in bold are statistically significant, i.e., p-value < 5 %).

p.adjust	1HRC	1M	2HRC	2M	3HRC	3M
1HRC	-					
1M	0.6766	-				
2HRC	1.0	1.0	-			
2M	**0.0224**	1.0	**0.0466**	-		
3HRC	0.3853	1.0	1.0	**0.0373**	-	
3M	**< 0.001**	0.43	**0.0026**	1.0	**0.0025**	-

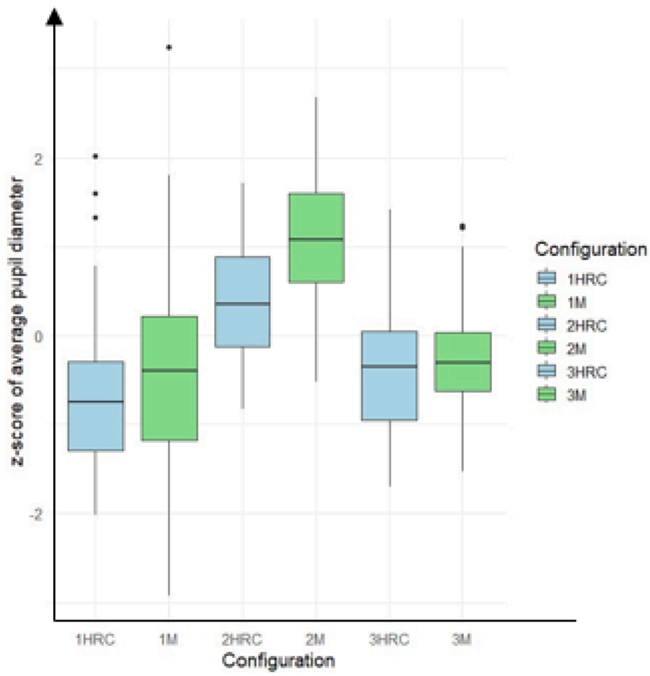

Fig. 14. Boxplot of average pupil diameter across the six different configurations.

Table 11
Results of the pairwise Wilcoxon signed rank test for the z-score median of the average pupil diameter (results in bold are statistically significant, i.e., p-value < 5 %).

p.adjust	1HRC	1M	2HRC	2M	3HRC	3M
1HRC	-					
1M	0.456	-				
2HRC	**< 0.001**	**0.002**	-			
2M	**< 0.001**	**< 0.001**	**< 0.001**	-		
3HRC	0.224	1.0	**< 0.001**	**< 0.001**	-	
3M	**0.042**	0.456	**< 0.001**	**< 0.001**	1.000	-

influence is not homogeneous across all physiological measures. At the same time, an increase in assembly complexity appears to correspond to an increase in human cognitive load, as evidenced by several key physiological indicators. Furthermore, the differences in the analysis of individual metrics highlight the potential for external factors such as the work environment, individual differences between workers and the specific nature of the cobot-task interaction to influence physiological responses. It is evident that no single metric can capture the breadth of human responses. In light of these findings, a holistic approach, capable of capturing a wide variety of responses, is essential to provide an accurate picture of worker's well-being. Such an approach would not only help to optimise human-cobot interaction, but also as a guiding principle to design a work environment that prioritises human well-being alongside operational efficiency.

Table 12
Summary of the main results obtained through the analysis of physiological signals.

Physiological metric	Relationship with cognitive load	Relationship with cobot support	Relationship with product complexity
Average SCR	Positive correlation	Generally lower median z-scores in collaborative modality with respect to manual counterparts	General upward trend
Average SCL	Positive correlation	Generally low median z-scores in collaborative modality with respect to manual counterparts	Upward trend
RMSSD	Negative correlation	Not clear trend	Not clear trend
SDNN	Negative correlation	Not clear trend. Higher median z-scores for 1HRC and 2HRC with respect to manual counterparts	Not clear trend
Number of fixations/ trial duration	Positive correlation	Not clear trend. High variability of data	Not clear trend. High variability of data
Average duration of fixations	Positive correlation	Generally lower median z-scores in collaborative modality with respect to manual counterparts	Upward trend
Average amplitude of saccades	Negative correlation	Generally higher median z-scores in collaborative modality with respect to manual counterparts	Downward trend
Average peak velocity of saccades	Negative correlation	Not clear trend	Not clear trend
Average pupil diameter	Positive correlation	Generally lower median z-scores in collaborative modality	Not clear trend.

Fig. 15 shows a correlation analysis between the physiological signals collected. The correlation values were computed using the Pearson correlation method. These values range between -1 and 1, where 0 represents no correlation, +1 the positive correlation and -1 the negative correlation [68]. The presence of stars in Fig. 15 indicates the statistical significance of these correlations.

From the analysis performed it is worth noting:

- Average SCR vs Average SCL (0.74): The robust positive correlation between these measures suggests that they may indeed capture a similar aspect of the human response. The high degree of correlation indicates that both measures may be reliable indicators of cognitive load, providing comparable information on sympathetic activity.
- Average SCL vs Average amplitude of saccades (-0.21): Although the negative correlation between these variables is statistically significant, the correlation value is low. This observation is in line with theoretical models according to which SCL is positively associated with cognitive effort, whereas average amplitude of saccades tends to decrease.
- SDNN vs RMSSD (0.77): The strong positive correlation between these heart rate variability metrics suggest that they are likely to reflect a similar physiological response to cognitive load, due to changes in product complexity and modality. Given their high correlation, both measures can be used to monitor cardiac activity.
- Number of fixations/ trial duration vs average duration of fixations (-0.32): Despite the theoretical expectation that these metrics would be both positively correlated with cognitive effort, the observed

Fig. 15. Correlation analysis results between physiological signals where: (***) means that p-value < 0.001; (**) p-value <0.01 and (*) p-value<0.05.

negative correlation, and its statistical significance, suggests an inverse relationship in the context of this study. This is due to the high variability of the data collected and it may reflect a complex interplay of cognitive processes during task performance that deserves further investigation.

- Average duration of fixations vs Average peak velocity of saccades (0.24): Again, the positive correlation between these metrics do not supports the theory that as cognitive load increases the duration of fixations and the average peak velocity show opposite behaviours.
- Average amplitude of saccades vs Average peak velocity of saccades (0.81): These two metrics are positively correlated. As one would theoretically expect, they show similar trend as the cognitive load varies. It should also be noted that, as expected, average amplitude of saccades negatively correlates with EDA metrics (average amplitude of saccades vs Average SCR: -0.16; average amplitude of saccades vs Average SCL:-0.21).

Therefore, the metrics that yielded the most valuable results, especially in terms of alignment with theoretical expectations, were those derived from electrodermal activity (EDA) and saccades. The patterns observed through these metrics provide evidence of the interaction between cognitive load and assembly complexity. These results are in line with those obtained by Zakeri et al. [46] who highlighted the significant effects of task complexity and cobot speed on mental load, complementing the hypothesis that increasing assembly complexity correlates with increased cognitive load. Similarly, also Verna et al. [42] showed that increasing product complexity leads to greater values of Skin Conductance Response, thus leading to higher stress level. Although not as significant as Pollak et al. [44], who showed how the presence of cobots affects human stress, this study also showed a slight reduction in cognitive load in the collaborative modality for some metrics. Differently, Fournier et al. [45] showed that working with cobots do not affect perceived cognitive load, which aligns with part of the findings of this work, since cobot support was not uniformly beneficial across all physiological metrics and it varies both with task complexity and with the specific metric considered, indicating a more complex interplay. All of this underlines once again the importance of taking a holistic approach to the analysis of physiological signals.

6. Conclusions

The empirical investigation outlined in this study provides a preliminary insight into the intricate relationship between assembly complexity, cobot support, and human physiological responses. Despite the heterogeneity of the data, some trends were identified:

- Both Electrodermal Activity metrics (i.e., Average SCR and Average SCL) show a general upward trend as assembly complexity increases and a less effort in collaborative modality than in manual one.
- Heart variability data (i.e., RMSSD and SDNN) did not show any particular trend.
- In terms of eye-tracking metrics, the most significant were: average duration of fixation and average amplitude of saccades, which showed a relationship with assembly complexity and modality. Also, average pupil diameter showed a strong relationship with assembly complexity although only for the tile cutter and the mechanical equipment. Finally, the average peak velocity of saccades slightly underlined the supportive role of cobots in assembly processes.

Another key result of the experiment underscored the potential of using physiological responses as indicators of cognitive load in industrial scenarios. It is also evident, though, that no single physiological metric can individually describe cognitive workload and therefore a comprehensive approach remains essential. In light of the findings of this study, one main potential practical implications emerge for improving human well-being in Industry 5.0 environments. In this context, cobots can be programmed to adapt their behaviour based on real-time physiological data, thereby tailoring their operations to the individual stress profiles and cognitive loads of human workers. This adaptive approach could include adjustments to the cobot's behaviour based on metrics such as heart rate variability, skin conductance, and eye-tracking data. For example, if increased cognitive load is detected by a pre-defined dashboard of physiological metrics, the cobot could reduce its work speed to reduce the worker's cognitive load. This approach is in line with the principles of Industry 5.0, which emphasises personalised, human-centred manufacturing environments. It is also worth underlining that individual differences were noted in physiological metrics. Individual differences in physiological responses, such as heart rate variability, electrodermal and eye activity, suggest that workers respond uniquely to similar stressors and cognitive demands. This variability highlights the limitations of generic models in accurately predicting human responses and suggests the need for more personalised models. To account for these individual differences, machine learning algorithms able to learn from a variety of physiological data collected over time from individual workers should be implemented. By integrating these personalised models into the cobot's operating system, cobots could adapt their behaviour in real time to the worker's current state.

Nonetheless, limitations of the study must be acknowledged to contextualize the results. The variability of physiological responses across the six different configurations highlights the impossibility to capture human's physiological state using a single metric. While some trends suggest an increase in cognitive load with assembly complexity, the lack of a consistent pattern across all measures indicates that other unconsidered variables may influence these physiological signals. Another limitation of the current research is the sample size, which, while adequate for initial exploration, may not fully represent a larger population. A larger sample size could provide more robustness to these results. Furthermore, the potential influence of individual differences, such as personal stress tolerance, were not controlled for in this study.

Future research should expand the analysis of this study, using larger datasets and different production environments to validate and generalize the results. In this context, experienced workers may be included in order to analyse how prior expertise affect cognitive effort levels when assembly complexity and modality vary. Furthermore, in the light of the introduction of new supporting technologies, such as cobots, in assembly processes, it is necessary to extend the concept of product complexity to new paradigm of the complexity of collaborative assemblies, where the complexity of human-robot interaction becomes another crucial aspect.

CRediT authorship contribution statement

Matteo Capponi: Writing – original draft, Methodology, Data curation, Conceptualization. **Riccardo Gervasi:** Writing – original draft, Methodology, Data curation, Conceptualization. **Luca Mastrogiacomo:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization. **Fiorenzo Franceschini:** Writing – review & editing, Supervision, Methodology, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Luca Mastrogiacomo reports financial support and administrative support were provided by Politecnico di Torino, Department of Management and Production Engineering. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgments

This study was carried out within the «Empathically enhanced robot for the collaboration with HUmans in MANufacturing (E.HU.MAN)» project – funded by the Ministero dell'Università e della Ricerca – within the PRIN 2022 program (D.D.104 - 02/02/2022). This manuscript reflects only the authors' views and opinions, and the Ministry cannot be considered responsible for them.

Appendix A. Product characteristics, assembly process and HRC task allocation of the (a) mechanical equipment, (b) the tile cutter and (c) the diaphragm water pump (adapted from [43])

(a)

Product	*Product characteristics*			*Assembly process*	*HRC Task allocation*	
	Parts and fasteners	Code	Quantities	Elementary task (same in manual and HRC)	Human	Cobot
Mechanical equipment	Base	Base	1	Pick and place BASE		X
	Elliptical flange	EF1/EF2	2	Pick and place EF1		X
	Square flange	SF	1	Screwing EF1 with Base	X	
	Bolt type 1	B1	4	Pick and place SF		X
	Bolt type 2	B2	2	Screwing SF with Base	X	
	Nuts type 1	N1	6	Pick and place EF2		X
				Screwing EF2 with Base	X	
				Pick the final product and place out of the assembly area		X

(b)

Product	*Product characteristics*			*Assembly process*	*HRC Task allocation*	
	Parts and fasteners	Code	Quantities	Elementary task (same in manual and HRC)	Human	Cobot
Tile cutter 	Base	Base	1	Pick and place Base		X
	Lateral support	C1a/C1b	2	Pick and place C1a and C1b on Base	X	
	Joint component	C2	1	Preliminary screwing C1a and C1b on Base	X	
	Cutting component	C3	1	Placing the subassembly (Base+C1a+C1b) out of the assembly area		X
	Blade	L1	1	Pick and place C2		X
	Tile blocker	C4	1	Pick and place C3 in C2	X	
	Rail rod	P1a/P1b	2	Screwing C3 and C2	X	
	Handle	P2	1	Pick and place L1	X	
	Bolt type 1	B1	2	Screwing L1 and C3	X	
	Bolt type 2	B2	1	Pick and place C4 in C3	X	
	Bolt type 3	B3	2	Screwing C4 and C3	X	
	Nuts type 1	N1	2	Placing the subassembly (C2+C3+C4+L1) out of the assembly area		X
	Nuts type 2	N2	1	Pick and place subassembly (Base+C1a+C1b) back in the assembly area		X
	Nuts type 3	N3	2	Insert sub-assembly (C2+C3+C4+L1) in both P1a/P1b	X	
				Insert P1a/P1b in C1a/C1b	X	
				Final screwing C1a/C1b on Base	X	
				Pick and place P2	X	
				Screwing P2	X	
				Pick the final product and place out of the assembly area		X

(c)

Product	*Product characteristics*			*Assembly process*	*HRC Task allocation*	
	Parts and fasteners	Code	Quantities	Elementary task (same in manual and HRC)	Human	Cobot
Diaphragm water pump 	Engine block	EB	1	Pick and place RF	X	
	Rubber feet	RF	1	Pick and place EB		X
	Ring	R	1	Screwing EB with RF	X	
	Flange 1	F1	1	Pick and place F1		X
	Flange 2	F2	1	Pick and place F2		X
	Diaphragm	D1	1	Insert F1 in F2	X	
	Cover with valves	CV	1	Pick and place D1 on sub-assembly F1+F2	X	
	Cover	C	1	Screwing D1, F1 and insert CV on D1	X	
	Pressure switch	PS	1	Pick and place C		X
	Pressure switch diaphragm	D2	1	Screwing C and F2	X	
	Filter	FIL	1	Insert R on EB	X	
	Flow adapter	AF1/AF2	2	Insert and screwing sub-assembly pump head on EB (joining F1-EB)	X	
	Screws type 1	V1	2	Pick and place D2 and PS on C	X	
	Screws type 2	V2	6	Screwing PS and C	X	
	Screws type 3	V3	3	Pick and place FIL	X	
	Screws type 4	V4	2	Screwing FIL	X	
				Pick and place AF1 and AF2	X	
				Screwing AF1 and AF2	X	
				Pick the final product and place out of the assembly area		X

References

[1] P.K.R. Maddikunta, Q.V. Pham, P. B, N. Deepa, K. Dev, T.R. Gadekallu, R. Ruby, M. Liyanage, Industry 5.0: a survey on enabling technologies and potential applications, J. Ind. Inf. Integr. 26 (2022) 100257, https://doi.org/10.1016/j.jii.2021.100257.

[2] A. Bauer, D. Wollherr, M. Buss, Human–robot collaboration: a survey, Int. J. Humanoid Robot. 05 (2008) 47–66, https://doi.org/10.1142/S0219843608001303.

[3] R.R. Galin, R.V. Meshcheryakov, Human-robot interaction efficiency and human-robot collaboration, in: A.G.(A C.di) Kravets (Ed.), Robotics: Industry 4.0 Issues & New Intelligent Control Paradigms, Springer International Publishing, Cham, 2020, pp. 55–63.

[4] A.K. Inkulu, M.V.A.R. Bahubalendruni, A. Dara, S. K, Challenges and opportunities in human robot collaboration context of Industry 4.0 - a state of the art review, Ind. Robot Int. J. Robot. Res. Appl. 49 (2021) 226–239, https://doi.org/10.1108/IR-04-2021-0077.

[5] R. Gervasi, L. Mastrogiacomo, F. Franceschini, A conceptual framework to evaluate human-robot collaboration, Int. J. Adv. Manuf. Technol. 108 (2020) 841–865, https://doi.org/10.1007/s00170-020-05363-1.

[6] F. Dehais, E.A. Sisbot, R. Alami, M. Causse, Physiological and subjective evaluation of a human–robot object hand-over task, Appl. Ergon. 42 (2011) 785–791, https://doi.org/10.1016/j.apergo.2010.12.005.

[7] B. Kühnlenz, M. Erhart, M. Kainert, Z.Q. Wang, J. Wilm, K. Kühnlenz, Impact of trajectory profiles on user stress in close human-robot interaction, Autom. 66 (2018) 483–491, https://doi.org/10.1515/auto-2018-0004.

[8] R. Gervasi, M. Capponi, L. Mastrogiacomo, F. Franceschini, Manual assembly and Human–Robot Collaboration in repetitive assembly processes: a structured comparison based on human-centered performances, Int. J. Adv. Manuf. Technol. 126 (2023) 1213–1231, https://doi.org/10.1007/s00170-023-11197-4.

[9] R. Gervasi, M. Capponi, L. Mastrogiacomo, F. Franceschini, Analyzing psychophysical state and cognitive performance in human-robot collaboration for repetitive assembly processes, Prod. Eng. 18 (2024) 19–33, https://doi.org/10.1007/s11740-023-01230-6.

[10] W. ElMaraghy, H. ElMaraghy, T. Tomiyama, L. Monostori, Complexity in engineering design and manufacturing, CIRP Ann. 61 (2012) 793–814, https://doi.org/10.1016/j.cirp.2012.05.001.

[11] B. Alkan, D.A. Vera, M. Ahmad, B. Ahmad, R. Harrison, Complexity in manufacturing systems and its measures: a literature review, Eur. J. Ind. Eng. 12 (2018) 116–150, https://doi.org/10.1504/EJIE.2018.089883.

[12] S. Hjorth, D. Chrysostomou, Human–robot collaboration in industrial environments: A literature review on non-destructive disassembly, Robot. Comput.-Integr. Manuf. 73 (2022) 102208, https://doi.org/10.1016/j.rcim.2021.102208.

[13] A. Reiman, J. Kaivo-oja, E. Parviainen, E.P. Takala, T. Lauraeus, Human factors and ergonomics in manufacturing in the industry 4.0 context – A scoping review, Technol. Soc. 65 (2021) 101572, https://doi.org/10.1016/j.techsoc.2021.101572.

[14] W. Patrick Neumann, J. Dul, Human factors: spanning the gap between OM and HRM, Int. J. Oper. Prod. Manag. 30 (2010) 923–950, https://doi.org/10.1108/01443571011075056.
[15] R.K. Khamaisi, A. Brunzini, F. Grandi, M. Peruzzini, M. Pellicciari, UX assessment strategy to identify potential stressful conditions for workers, Robot. Comput.-Integr. Manuf. 78 (2022) 102403, https://doi.org/10.1016/j.rcim.2022.102403.
[16] M.A. Goodrich, A.C. Schultz, Human–robot interaction: a survey, Found. Trends® Human–Computer Interact. 1 (2008) 203–275, https://doi.org/10.1561/1100000005.
[17] A. Khalid, P. Kirisci, Z. Ghrairi, J. Pannek, K.D. Thoben, Towards implementing safety and security concepts for human-robot collaboration in the context of industry 4.0, in: 39th International MATADOR Conference on Advanced Manufacturing, 2017.
[18] A. Colim, C. Faria, J. Cunha, J. Oliveira, N. Sousa, L.A. Rocha, Physical ergonomic improvement and safe design of an assembly workstation through collaborative robotics, Safety 7 (2021) 14, https://doi.org/10.3390/safety7010014.
[19] N. Shirakura, R. Takase, N. Yamanobe, Y. Domae, T. Ogata, Time pressure based human workload and productivity compatible system for human-robot collaboration, in: 2022 IEEE 18th International Conference on Automation Science and Engineering (CASE), 2022, pp. 659–666.
[20] L. Lu, Z. Xie, H. Wang, B. Su, X.X. Edward P Fitts, Assessing workers' mental stress in hand-over activities during human-robot collaboration, Proc. Hum. Factors Ergon. Soc. Annu. Meet. 66 (2022) 537–541, https://doi.org/10.1177/1071181322661194.
[21] F. Barravecchia, M. Bartolomei, L. Mastrogiacomo, F. Franceschini, Redefining human–robot symbiosis: a bio-inspired approach to collaborative assembly, Int. J. Adv. Manuf. Technol. 128 (2023) 2043–2058, https://doi.org/10.1007/s00170-023-11920-1.
[22] C.M. Hinckley, Ph.D. Dissertation, Stanford University, USA, 1994.
[23] H. Shibata, Ph.D. Dissertation, Stanford University, USA, 2002.
[24] Q. Su, L. Liu, D.E. Whitney, A systematic study of the prediction model for operator-induced assembly defects based on assembly complexity factors, IEEE Trans. Syst. Man Cybern. - Part Syst. Hum. 40 (2010) 107–120, https://doi.org/10.1109/TSMCA.2009.2033030.
[25] G. Genta, M. Galetto, F. Franceschini, Product complexity and design of inspection strategies for assembly manufacturing processes, Int. J. Prod. Res. 56 (2018) 4056–4066, https://doi.org/10.1080/00207543.2018.1430907.
[26] B. Alkan, An experimental investigation on the relationship between perceived assembly complexity and product design complexity, Int. J. Interact. Des. Manuf. IJIDeM. 13 (2019) 1145–1157, https://doi.org/10.1007/s12008-019-00556-9.
[27] K. Sinha, Ph.D. Dissertation, Massachusetts Institute of Technology, USA, 2014.
[28] E. Verna, G. Genta, M. Galetto, F. Franceschini, Defect prediction for assembled products: a novel model based on the structural complexity paradigm, Int. J. Adv. Manuf. Technol. 120 (2022) 3405–3426, https://doi.org/10.1007/s00170-022-08942-6.
[29] W.H. ElMaraghy, R.J. Urbanic, Modelling of manufacturing systems complexity, CIRP Ann. 52 (2003) 363–366, https://doi.org/10.1016/S0007-8506(07)60602-7.
[30] W.H. ElMaraghy, R.J. Urbanic, Assessment of manufacturing operational complexity, CIRP Ann. 53 (2004) 401–406, https://doi.org/10.1016/S0007-8506(07)60726-4.
[31] X. Zhu, S.J. Hu, Y. Koren, S.P. Marin, Modeling of manufacturing complexity in mixed-model assembly lines, J. Manuf. Sci. Eng. 130 (2008), https://doi.org/10.1115/1.2953076.
[32] F. Ameri, J. Summers, G. Mocko, M. Porter, Engineering design complexity: an investigation of methods and measures, Res. Eng. Des. 19 (2008) 161–179, https://doi.org/10.1007/s00163-008-0053-2.
[33] S.N. Samy, H. ElMaraghy, A model for measuring products assembly complexity, Int. J. Comput. Integr. Manuf. 23 (2010) 1015–1027, https://doi.org/10.1080/0951192X.2010.511652.
[34] M. Capponi, L. Mastrogiacomo, F. Franceschini, General remarks on the entropy-inspired MCAT (manufacturing complexity assessment tool) model to assess product assembly complexity, Prod. Eng. 17 (2023) 815–827, https://doi.org/10.1007/s11740-023-01212-8.
[35] H. Wang, S.J. Hu, Manufacturing complexity in assembly systems with hybrid configurations and its impact on throughput, CIRP Ann. 59 (2010) 53–56, https://doi.org/10.1016/j.cirp.2010.03.007.
[36] L. Zeltzer, E.H. Aghezzaf, V. Limère, Workload balancing and manufacturing complexity levelling in mixed-model assembly lines, Int. J. Prod. Res. 55 (2017) 2829–2844, https://doi.org/10.1080/00207543.2016.1213452.
[37] H. Sun, S. Fan, Car sequencing for mixed-model assembly lines with consideration of changeover complexity, J. Manuf. Syst. 46 (2018) 93–102, https://doi.org/10.1016/j.jmsy.2017.11.009.
[38] X. Liu, X. Yang, M. Lei, Optimisation of mixed-model assembly line balancing problem under uncertain demand, J. Manuf. Syst. 59 (2021) 214–227, https://doi.org/10.1016/j.jmsy.2021.02.019.
[39] S. Mattsson, M. Karlsson, P. Gullander, H. Van Landeghem, L. Zeltzer, V. Limère, E. H. Aghezzaf, Å. Fasth, J. Stahre, Comparing quantifiable methods to measure complexity in assembly, Int. J. Manuf. Res. 9 (2014) 112–130, https://doi.org/10.1504/IJMR.2014.059602.
[40] S. Mattsson, M. Tarrar, Å. Fast-Berglund, Perceived production complexity – understanding more than parts of a system, Int. J. Prod. Res. 54 (2016) 6008–6016, https://doi.org/10.1080/00207543.2016.1154210.
[41] A.C. Falck, R. Örtengren, M. Rosenqvist, R. Söderberg, Basic complexity criteria and their impact on manual assembly quality in actual production, Int. J. Ind. Ergon. 58 (2017) 117–128, https://doi.org/10.1016/j.ergon.2016.12.001.
[42] E. Verna, S. Puttero, G. Genta, M. Galetto, Exploring the effects of perceived complexity criteria on performance measures of human–robot collaborative assembly, J. Manuf. Sci. Eng. 145 (2023), https://doi.org/10.1115/1.4063232.
[43] M. Capponi, R. Gervasi, L. Mastrogiacomo, F. Franceschini, Assessing perceived assembly complexity in Human-Robot collaboration processes: a proposal based on Thurstone's Law of Comparative Judgement, Int. J. Prod. Res. (2023). Article in press.
[44] A. Pollak, M. Paliga, M.M. Pulopulos, B. Kozusznik, M.W. Kozusznik, Stress in manual and autonomous modes of collaboration with a cobot, Comput. Hum. Behav. 112 (2020) 106469, https://doi.org/10.1016/j.chb.2020.106469.
[45] É. Fournier, D. Kilgus, A. Landry, B. Hmedan, D. Pellier, H. Fiorino, C. Jeoffrion, The impacts of human-cobot collaboration on perceived cognitive load and usability during an industrial task: an exploratory experiment, IISE Trans. Occup. Ergon. Hum. Factors 10 (2022) 83–90, https://doi.org/10.1080/24725838.2022.2072021.
[46] Z. Zakeri, A. Arif, A. Omurtag, P. Breedon, A. Khalid, Multimodal assessment of cognitive workload using neural, subjective and behavioural measures in smart factory settings, Sensors 23 (2023) 8926, https://doi.org/10.3390/s23218926.
[47] J. Sweller, P. Ayres, S. Kalyuga, Cognitive Load Theory, Springer, New York, NY, USA, 2011.
[48] S. Rubio, E. Díaz, J. Martín, J.M. Puente, Evaluation of subjective mental workload: a comparison of SWAT, NASA-TLX, and workload profile methods, Appl. Psychol. 53 (2004) 61–86, https://doi.org/10.1111/j.1464-0597.2004.00161.x.
[49] S. Hopko, J. Wang, R. Mehta, Human factors considerations and metrics in shared space human-robot collaboration: a systematic review, Front. Robot. AI 9 (2022), https://doi.org/10.3389/frobt.2022.799522.
[50] F. Paas, A. Renkl, J. Sweller, Cognitive load theory and instructional design: recent developments, Educ. Psychol. 38 (2003) 1–4, https://doi.org/10.1207/S15326985EP3801_1.
[51] G. Charness, U. Gneezy, M.A. Kuhn, Experimental methods: Between-subject and within-subject design, J. Econ. Behav. Organ. 81 (2012) 1–8, https://doi.org/10.1016/j.jebo.2011.08.009.
[52] O.N. Rahma, A.P. Putra, A. Rahmatillah, Y.S.K.A. Putri, N.D. Fajriaty, K. Ain, R. Chai, Electrodermal activity for measuring cognitive and emotional stress level, J. Med. Signals Sens. 12 (2022) 155–162, https://doi.org/10.4103/jmss.JMSS_78_20.
[53] M. Benedek, C. Kaernbach, A continuous measure of phasic electrodermal activity, J. Neurosci. Methods 190 (2010) 80–91, https://doi.org/10.1016/j.jneumeth.2010.04.028.
[54] S. Delliaux, A. Delaforge, J.C. Deharo, G. Chaumet, Mental workload alters heart rate variability, lowering non-linear dynamics, Front. Physiol. 10 (2019), https://doi.org/10.3389/fphys.2019.00565.
[55] C. Spellenberg, P. Heusser, A. Büssing, A. Savelsbergh, D. Cysarz, Binary symbolic dynamics analysis to detect stress-associated changes of nonstationary heart rate variability, Sci. Rep. 10 (2020) 15440, https://doi.org/10.1038/s41598-020-72034-2.
[56] M.S. Young, K.A. Brookhuis, C.D. Wickens, P.A. Hancock, State of science: mental workload in ergonomics, Ergonomics 58 (2015) 1–17, https://doi.org/10.1080/00140139.2014.956151.
[57] H.G. Kim, E.J. Cheon, D.S. Bai, Y.H. Lee, B.H. Koo, Stress and heart rate variability: a meta-analysis and review of the literature, Psychiatry Investig. 15 (2018) 235–245, https://doi.org/10.30773/pi.2017.08.17.
[58] J. William, R. Murugesh, Potential eye tracking metrics and indicators to measure cognitive load in human-computer interaction research, J. Sci. Res. 64 (2020) 168–175, https://doi.org/10.37398/JSR.2020.640137.
[59] Tobii, https://connect.tobii.com.
[60] J.C. Liu, K.A. Li, S.L. Yeh, S.Y. Chien, Assessing perceptual load and cognitive load by fixation-related information of eye movements, Sensors 22 (2022) 1187, https://doi.org/10.3390/s22031187.
[61] L.L. Di Stasi, A. Antolí, J.J. Cañas, Main sequence: An index for detecting mental workload variation in complex tasks, Appl. Ergon. 42 (2011) 807–813, https://doi.org/10.1016/j.apergo.2011.01.003.
[62] S.S. Shapiro, M.B. Wilk, An analysis of variance test for normality (complete samples)†, Biometrika 52 (1965) 591–611, https://doi.org/10.1093/biomet/52.3-4.591.
[63] F. Wilcoxon, Individual comparisons by ranking methods, Biom. Bull. 1 (1945) 80–83, https://doi.org/10.2307/3001968.
[64] Society for Psychophysiological Research Ad Hoc Committee on Electrodermal Measures, Publication recommendations for electrodermal measurements, Psychophysiology 49 (2012) 1017–1034, https://doi.org/10.1111/j.1469-8986.2012.01384.x.
[65] Y. Liu, S. Du, Psychological stress level detection based on electrodermal activity, Behav. Brain Res. 341 (2018) 50–53, https://doi.org/10.1016/j.bbr.2017.12.021.
[66] H. Lu, D. van der Linden, A.B. Bakker, Changes in pupil dilation and P300 amplitude indicate the possible involvement of the locus coeruleus-norepinephrine (LC-NE) system in psychological flow, Sci. Rep. 13 (2023) 1908, https://doi.org/10.1038/s41598-023-28781-z.
[67] D.J. Harris, S.J. Vine, M.R. Wilson, Flow and quiet eye: the role of attentional control in flow experience, Cogn. Process. 18 (2017) 343–347, https://doi.org/10.1007/s10339-017-0794-9.
[68] D. Montgomery, G. Runger, N. Hubele, Engineering Statistics, John Wiley & Sons Inc., New York, 2010.