International Journal of Social Robotics (2025) 17:211–256
https://doi.org/10.1007/s12369-025-01224-7

Evaluating the Effect of Speed and Acceleration on Human Factors during an Assembly Task in Human–Robot Interaction (HRI)

Ainhoa Apraiz[1] · Ganix Lasa[1] · Maitane Mazmela[1] · Nestor Arana-Arexolaleiba[2] · Antonio Serrano Muñoz[2] · Íñigo Elguea[2,3] · Amaia Etxabe[1]

Accepted: 24 January 2025 / Published online: 19 February 2025
© The Author(s) 2025

Abstract

In the new industrial contexts, the workers' well-being is the central pillar. Therefore, research on methods and technics to improve the workers' user experience in a human–robot collaborative environment is necessary. While the effects of kinematic variables, such as speed and acceleration, on human safety have been extensively studied, their impact on human perception has not been fully explored. This study investigates the effects of the robot's speed and acceleration on humans. An experimental research approach was used, where 20 participants (10 women and 10 men) performed an assembly task collaborating with a robot. An experiment was defined with two procedures, and the participants were evenly distributed: in the first experiment, the participants started by performing the task at a slow robot speed and then performed the same task at a faster speed. In the second experiment, the other half followed the opposite procedure. Key Performance Indicators (KPIs), physiological values (via EEG and EDA), and perceptual values (using the standardised UEQ-S questionnaire) were collected. The results showed that the robot's speed and acceleration impact the task's completion time and participants' emotional responses. Our results lead to a new concept, "HRI speed bell", which indicates that it is necessary to investigate the optimal speed and acceleration to ensure good trust and perceived safety. Furthermore, the task sequence also influences participants' expectations and performance. Finally, the results are examined according to gender perspective.

Keywords Human–robot Interaction (HRI) · User Experience (UX) · Human Factors · Assembly · Physiological monitorisation

1 Introduction

Robots are increasingly implemented in industry [1–5]. The recent paradigm in advanced industrial systems emphasizes prioritizing the well-being of workers within the production process [6]. Human–Robot Interaction (HRI) aims to combine the strengths of the robots with unique human skills by allowing safe and profitable task sharing in a fenceless and common workspace [7]. This enables close and safe interaction between humans and robots, leveraging the benefits of both sides. Optimized communication between human operators and robots can improve results in a robust collaboration framework.

Humans' role will remain important in a collaborative context. The relationship between humans and robots results in a robust collaboration framework that positively impacts flexibility and productivity. The user must perceive the robot as an ally so that they leverage the strengths of both. For robots to become allies in the day-to-day lives of operators, they need to provide positive and fit-for-purpose experiences through smooth and satisfying interactions [8–11]. Therefore, analysing the user experience (UX) in these environments is relevant [12]. ISO 9241–210 [13] defines UX as "a person's perceptions and responses resulting from the use or anticipated use of a product, system, or service". This statement includes user emotions, beliefs, preferences, perceptions, physical and psychological responses, behaviours,

✉ Ainhoa Apraiz
aapraizi@mondragon.edu

1 Faculty of Engineering, Mechanical and Industrial Production, Mondragon Unibertsitatea, Loramendi 4, 20500 Mondragon, Gipuzkoa, Spain

2 Robotics and Automation Electronics and Computer Science Department, Mondragon Unibertsitatea, Loramendi 4, Mondragon, 20500 Arrasate, Spain

3 Electrotécnica Alavesa S.L. Research & Development Department, 1010 Vitoria-Gasteiz, Spain

and achievements that occur before, during, and after use [13]. Similarly, human perception encompasses a multifaceted array of elements, comprising emotions, beliefs, preferences, and psychological responses, all of which wield profound influence over the efficacy of collaborative interactions between humans and robots. When users have a positive perception of robots and engage with them in a similar manner as they would with human counterparts, it enhances the overall user experience and leads to favorable outcomes in collaborative efforts. Therefore, understanding and examining human perception within the realm of HRI is of utmost importance, as it serves as a linchpin for the design and implementation of robots that seamlessly integrate into human environments while augmenting productivity, safety, and holistic well-being.

Previous studies have provided different viewpoints on HRI. This study is in line with the framework proposed by Schmidtler et al. [14], which classifies HRI into three primary categories:

- Human–Robot Coexistence, where both humans and robots operate simultaneously within a shared workspace.
- Human–Robot Cooperation, indicating that both humans and robots work together in the same environment to achieve a shared objective.
- Human–Robot Collaboration (HRC), which necessitates not only cooperation but also direct physical interaction between two entities.

In this line, the research community on social robots has recently recognised the role of UX design in ensuring positive interaction [12]. A negative experience can prevent humans from taking advantage of the robots' features or reject interaction with them. Only the accurate study of the UX from the initial design phase can guarantee an acceptable and pleasant interaction, capable of improving the interaction's efficiency and effectiveness [14]. Considering how the operator interacts with the robot and understanding its UX is difficult, especially in the industrial context. A structured human-centred approach like those based on UX techniques can be a solution. A theoretical framework of HRI from the UX perspective was created by Prati et al. [12], but it had shortcomings in the evaluation phase. To date, no validated model to assess UX in HRI was identified [15].

The robot's kinematic variables, such as speed and acceleration, play an essential role in how humans perceive a robot [16–20]. While there are existing research studies that examine the influence of a robot's kinematic variables on human factors [21–24], to the best of the authors' knowledge, there is a notable absence of comprehensive analyses that incorporate emotional factors objectively, such as through the use of Electroencephalography (EEG). By examining factors such as communication, task allocation, and Reliability between humans and robots, researchers can identify potential challenges and opportunities for improving collaboration and enhancing performance.

The rest of the paper is organised as follows. Section 2 exposes a brief description of related works. Section 3 presents the methodology followed to assess the UX in an HRI assembly workspace. Section 4 shows the results obtained in the case study. Section 5 provides an outline of the main headings to consider for future research, providing points for discussion and reflection. Lastly, Sect. 6 concludes with a summary of the knowledge gained.

2 Literature Review

2.1 Human Factors in HRI

When any human being interacts with an object, a machine, or a system, two types of responses are generated: (i) behavioural and cognitive, and (ii) information and meanings flow from the user to the system in different ways [25, 26]. According to Peruzzini et al. [27], such responses automatically occur when a task is accomplished, regardless of its type and nature. Indeed, any time a human acts, his/her body and his/her brain generate behaviours and cognitive feedback, which affect his/her cognitive and physical workload [27]. When collaborating with a robot, the human response in the workplace could be more delicate, since it is usually characterised by the lengthy execution of tasks, repetitive actions merged with stressful conditions, and the lack of knowledge of the robot's intentionality.

Regarding the human factors that are most likely to influence workers' acceptance in collaborative contexts, Meissner et al. [28] conducted an exploratory study. They identified primary factors that include (i) perceived risk, (ii) perceived benefits, (iii) negative emotions, and (iv) positive emotions. In addition, they identified several secondary factors: object-related factors, subject-related factors, and context-related factors. All these factors influence the attitude towards acceptance of the system and the fact that they are closely related factors. Therefore, they also influence each other.

The study by Hopko et al. [29] systematically reviewed the human factors literature in shared-space HRC, the metrics used to measure human factors, and the implications of human factors on HRC. They found that trust, cognitive workload, and anxiety are the most frequently examined human factors in shared-space HRC, primarily assessed through subjective questionnaires, although alternative methods such as bioinstrumentation, objective behavioural analyses, and mathematical models were also utilized. Hopko et al. [29] also concluded that most studies treated human factors as dependent variables influenced by robot factors or environmental conditions. The authors also indicated that

demographic factors, such as age and gender, were found to be insufficiently studied within the context of shared space, with over half of the reviewed studies omitting demographic participant information.

According to Yan et al. [30], one of the major affecting factors is the comfort of humans in HRC, which is usually less emphasized in COBOT development; however, it is critical to the user acceptance during HRC. They classified human comfort factors under HRC scenarios can be classified in four categories: (i) ergonomic factors, (ii) robot motion-based factors, (iii) anthropomorphism, and (iv) robot sociability factors.

The research conducted by Bröhl et al. [31] delves into the acceptance of HRC based on Technological Acceptance Models (TAMs). TAMs provide foundational frameworks for comprehending how individuals perceive and adopt emerging technologies, such as robots, into their daily interactions. Within the field of HRI, TAMs play a crucial role in facilitating a nuanced comprehension of user needs and expectations from a human-centred perspective. Evaluating user acceptance and human factors with robots in industrial settings can yield insights for enhancing the implementation of robotic systems. In this context, our study is grounded in the principles of human-centered design, aiming to explore the acceptance of robotic systems from a user-centric approach. Through the integration of diverse methods and metrics, our research seeks to evaluate the impact of kinematic variables on user perception and overall user experience during human–robot interactions.

2.2 Methods and Metrics for Assessing Human Factors in HRI

The assessment of human factors impacting HRI acceptance in industrial settings involves a wide range of methods and measures, ranging from the subjective to the objective dimension of interaction. The appropriate choice and combination of these approaches depends on the specific research objectives and the desired depth of human factors analysis in this context.

ANEMONE methodology is a phased and iterative procedure that assesses the effectiveness of HRI by focusing on understanding and predicting a robot's intention and actions [11]. It provides insight into why something works or does not work in a particular context and guides the identification of UX issues in HRI. Steinfeld et al. [32] presented measures and metrics for task-oriented HRI and standard metrics for overall HRI performance. Schillaci et al. [27] aimed to identify the physical and behavioural characteristics that ensure a good UX in HRI and to measure the UX when interacting with a robot. To do so, they adopted two techniques: 1) UX measured by means of qualitative and quantitative questionnaires, and 2) proxemics, estimated from videos recorded during interactions. Prati et al. [12] proposed a UX-centric approach to investigate human–robot dialogue, with an iterative process that places the user persona at the centre of the design. However, this method has limitations by not providing clear guidelines for conducting structured evaluations.

Coronado et al. [33] provided an overview of the factors, metrics, and measures employed in the robotics community to assess both the performance and quality of human-related well-being in HRI contexts. In their analysis, they highlight one relevant point: most tools aimed at assessing humans-related factors require observational techniques. This underscores the need to incorporate monitoring into these processes to achieve more accurate assessments, thus minimizing any potential bias arising from assessor subjectivity and human error. Therefore, in our study, we aim to incorporate physiological monitoring as well, as it represents an intriguing approach to objectively evaluate human factors without retrospective bias. This aligns with the findings in a previous study [15, 34], where we underscored the importance of incorporating monitoring techniques into assessments of human-related factors to achieve a more accurate evaluations and minimize potential retrospective bias arising.

[15] describes an HRI evaluation metrics and measurements classification in terms of performance, posture, robot-related factors, and emotions-related factors. They concluded that the most frequently used evaluation technique is task execution time. The questionnaires to evaluate UX in HRI are a particularly efficient method of application and analysis and are commonly used for user-driven assessments of the system's quality and usability [35]. Nevertheless, the questionnaires are prone to retrospective bias. To reduce the bias, in recent years, there has been an increasing focus on the use of physiological devices [24, 36–38]. The combination of performance indicators, questionnaires and physiological devices makes it possible to carry out a more human-centred evaluation.

2.2.1 Physiological Measurements

In the ongoing quest to comprehend the intricate interaction between humans and technology, the utilization of psychophysiological devices has emerged as a fundamental tool in assessing UX and technological acceptance. According to Neumann and Westbury [38], psychophysiological measures are more objective than self-report measures like questionnaires. The opportunities presented by physiological signals are increasing due to advancements in sensor technology and signal processing [39]. Psychophysiological devices operate at the intersection of psychology and physiology, primarily aiming to measure and quantify the responses the human body undergoes in reaction to various stimuli. These devices

can encompass a variety of parameters, ranging from heart rate and skin conductance to EEG and other relevant markers.

One of the primary attributes distinguishing psychophysiological devices is their ability to provide objective measurements of emotional and cognitive responses. Unlike assessments solely reliant on self-reports, which can be susceptible to biases and individual variations, psychophysiological devices offer a direct window into the physiological responses of the user. Moreover, the real-time nature of psychophysiological measurements adds substantial value to the assessment, allowing the capture of these responses as they occur, providing a detailed understanding of real-time emotional and cognitive changes.

The use of psychophysiological devices not only enriches data collection but also opens the possibility of capturing subtle, non-conscious responses. Users may not be able to accurately verbalize their emotional and cognitive states during interaction with technology. Physiological indicators, such as changes in brain electrical activity or skin conductance, can reveal internal aspects that would otherwise remain concealed, offering a more comprehensive perspective of the user's experience.

In the context of HRI in industrial environments, psychophysiological devices emerge as an emerging tool for assessing UX and robot acceptance by workers. They enable the real-time examination of how the presence and behaviour of robots influence the subjective experience of workers.

Cardiac activity reflects the response of the autonomic nervous system to external or internal physical or emotional stimuli, such as stressful situations [40]. It is typically measured using electrocardiography (ECG), which records the electrical activity of the heart through electrodes applied in standard positions or via wearable devices like smartwatches. Within this context, the key parameter obtained through cardiac activity evaluation is Heart Rate Variability (HRV), encompassing both heart rate (beats per minute) and oscillations in the time interval between consecutive R-waves in the ECG signal [41]. Additionally, other measurements such as heart rate (HR), individual oxygen consumption (VO_2), and maximal oxygen consumption (VO_{2max}) are common.

Electrodermal activity (EDA), also known as Galvanic Skin Response (GSR) or skin conductance (SC), reflects the changing electrical properties of the skin, resulting from fluctuations in sweat production by the human body. The EDA signal can be divided into tonic and phasic components. The tonic component, known as Skin Conductance Level (SCL), pertains to long-term variations in the EDA signal, stemming from individual psychophysiological characteristics. On the other hand, the phasic component of the signal, termed Skin Conductance Response (SCR), represents the skin conductance's reaction to short-term stimuli and is considered a valuable indicator of experienced stress levels [42]. Bitbrain [43] has established a connection between EDA/GSR and the user's Activation and Impact [44–48], highlighting its relevance in assessing UX and emotional responses. Therefore, incorporating EDA sensor into the current study using the Bitbrain Ring device [49], offers the advantage of capturing real-time data on the user's Activation and Impact during the interaction with the robot.

Respiration is particularly influenced by emotional states and stress events and has been recognized as more sensitive than other physiological measures [50]. Peruzzini et al. [27, 51] calculated respiratory frequency, measuring breaths per minute (RPM), in combination with skin temperature (ST) measurement, conducting a comprehensive analysis of the operator's health conditions in the workplace to evaluate stress and fatigue.

Typically, brain activity is recorded using EEG, which is a technique employed for recording the electrical activity generated by the brain. Data is collected by applying electrodes to the scalp, usually following standard positioning configurations [40]. Each electrode measures the voltage produced by neuronal activity in the specific brain region it is placed. Bitbrain [43] has established a connection between brain activity and various emotions, such as emotional Valence [52–54], Memorisation [55–58], Mental Workload [59–62], and Engagement [63–65]. This correlation underscores the utility of EEG in discerning the cognitive and affective responses of individuals during HRI. Integration of the Diadem EEG device by Bitbrain [43] into this study offers an opportunity to explore in-depth the cognitive processes.

Electromyography (EMG) and nerve conduction studies are tests that assess the electrical activity present in muscles and nerves. These tests analyse how nerves emit electrical signals to generate specific responses in muscles. These muscular responses, in turn, generate detectable signals. In the case of electromyography, it investigates the electrical signals emitted by muscles both at rest and during action. On the other hand, nerve conduction studies measure the speed and efficiency with which electrical signals travel through nerves in the body. While this modality is primarily used in device testing, ergonomics, and human factors testing, when connected to facial muscles, it can also serve as another means of measuring emotions.

In previous studies [34], a comprehensive review examined the integration of psychophysiological monitoring in UX studies. Despite none of the identified case studies in that review being from the context of HRI, it was concluded that the interest in incorporating psychophysiological measures in UX tests is growing, with EEG being the most utilized in UX studies. Although none of the identified case studies within that review were situated in the realm of HRI, it was evident that there is a growing inclination towards incorporating psychophysiological measures into UX assessments, with EEG emerging as the most prevalent modality in UX research. Nevertheless, within the domain of HRI UX studies,

as highlighted in [15], EEG remains conspicuously absent. Consequently, the combination of devices such as EDA, the most utilized among physiological monitoring devices in HRI studies [15], along with EEG, the dominant modality in UX studies albeit underutilized during robot interactions, presents an intriguing avenue for exploration.

2.2.2 User Perception Measurements

Questionnaires are widely used tools for assessing user perception, as well as the quality and usability of software programs [35, 66]. They enable efficient quantitative measurement of product characteristics and are often measured using Likert scales or semantic differential pairs [67]. Questionnaires are particularly effective for application and analysis, meeting the evaluation requirements proposed by Väänänen-Vainio-Mattila et al. [68].

The evaluation of UX in the context of HRI presents particular challenges, and so far, no specific questionnaires designed exclusively for this area have been identified. This reflects a gap in the literature and available tools for measuring and understanding the quality of UX in interactions with robots. Due to its interdisciplinary nature, HRI often incorporates and adapts tools and approaches from related disciplines such as Human–Computer Interaction (HCI) and ergonomics. In the analysis of experimental studies, questionnaires like NASA-TLX [69] are included, which, while not specifically designed to evaluate interaction with robots, focus on aspects related to the perceived workload of users. However, it is important to note that NASA-TLX [69] may not fully capture the overall UX. This questionnaire tends to focus more on workload and efficiency aspects, leaving aside more hedonic or emotional elements related to user satisfaction and emotional experience when interacting with robots. This means that while NASA-TLX [69] can provide valuable information about the Mental Workload aspect, it may not be sufficient to evaluate the overall emotional perception of users regarding the robot.

More comprehensive UX questionnaires like the UEQ questionnaire [70] have not yet been applied in HRI. This gap underscores the need to develop and apply specific assessment tools to understand the quality of interactions more accurately and comprehensively between humans and robots. Instead of creating entirely new questionnaires, adapting and extending existing questionnaires like the UEQ [70] to address specific dimensions of HRI could be considered.

Although industrial processes are increasingly automated, human performance remains the primary driver of product quality and factory productivity. In this context, understanding how workers interact with production systems and experience the factory environment is essential for properly modelling human interaction and optimizing processes [71].

The SE-HRI questionnaire [72] evaluates users' self-efficacy perception in the human–robot relationship. In this regard, the questionnaire's applicability in the context of industrial HRC could be observed.

The human factors assessed in the identified experimental studies can also be evaluated using UX questionnaires. In this sense, trust could be measured using SUPR-Q [73] or UEQ + [70], satisfaction could be measured, for example, using the USE questionnaire [74] or UMUX [75]. Learnability, in addition to the SUS questionnaire [76], could also be measured using QUIS [77], SUMI [78], PUTQ [79], WAMMI [80], USE [74], UEQ [35], AltUsability [81] and UEQ + [70].

Furthermore, in the identified experimental studies in Table 1, other questionnaires have been used. For example, the Godspeed questionnaire [82], has been used on two occasions, measuring perceived anthropomorphism, animation, sympathy, perceived intelligence, and safety of robots. The SUS questionnaire [76], has also been used once, measuring usability and learnability. The NASA-TLX questionnaire [69], which assesses Mental Workload, has been used three times (Table 1). Additionally, the experimental studies include non-standardized questionnaires that measure constructs such as trust, satisfaction, perceived safety and comfort, physical and mental stress, and perceived risk.

In the context of HRI studies, questionnaires serve as valuable instruments for assessing UX, to gather perceptual data from the user. However, despite the absence of questionnaires specifically tailored for UX in industrial HRI settings [15], UX questionnaires present an opportune avenue for understanding user perceptions within the framework of Industry 4.0 [66]. Among these, integrating the UEQ-S questionnaire [83] alongside established tools in HRI research holds promise for enhancing our comprehension of user perceptions.

The rationale behind selecting the UEQ-S questionnaire lies in its comprehensive nature, as identified in a prior study [67], which positions the UEQ as one of the most thorough questionnaires for evaluating UX. However, a fundamental requirement, as outlined by [68], is the necessity for a swift assessment method. Consequently, there is a prevalent inclination towards utilizing shorter questionnaire versions [67]. This trend is particularly pertinent in contexts where detailed information is deemed unnecessary, or when users are required to complete the same questionnaire multiple times for comparative analysis across various study objects [67].

Shorter questionnaire iterations, typically comprising 2 to 8 items, are favoured for their expedited completion time. Yet, it is essential to acknowledge that these condensed versions may sacrifice nuanced aspects found in their full-length counterparts. For instance, while the original UEQ [70]

Table 1 Comparative analysis of studies adopting an experimental approach to analyse human factors in industrial HRI (nd = no data)

Ref	Year	Authors	Interaction level	Robot	Measurements and metrics related to human factors			Measurements and metrics related to performance	Sample size				Main contribution
					Evaluation of Perception	Psychophysiological Evaluation			Total	Women	Men	Other	
						Type of sensor	Specific device						
[90]	2013	Schillaci et al	Collaboration	Humanoid robot (Nao version 3.3 from Aldebaran)	Godspeed questionnaire	–	–	–	28	8	20	0	Identification of the physical and behavioral characteristics that should be highlighted to ensure a good UX in the HRI
[91]	2013	Daniel et al	Coexistence	Robotic arm (Nachi MC70)	Open-ended questions about trust in automation	–	–	Task execution time	16	4	12	0	Improve the HRI by implementing a user interface
					Interview with questions adapted from WAI Site Usability Testing Questions about impressions of the current user interface and operation of the robotic cell			Number of interactions performed					
					Flow Conditions Questionnaire			Ratio of touch screen and keyboard interactions					

Table 1 (continued)

Ref	Year	Authors	Interaction level	Robot	Measurements and metrics related to human factors			Measurements and metrics related to performance	Sample size				Main contribution
					Evaluation of Perception	Psychophysiological Evaluation			Total	Women	Men	Other	
						Type of sensor	Specific device						
[16]	2015	Lasota & Shah	Collaboration	Robotic arm (ABB IRB-120)	Questionnaire (Satisfaction with the robot as a companion; and perceived safety and comfort)	–	–	Task execution time Simultaneous movement Average separation distance Robot idle time Person idle time	20	7	13	0	Analysis of human response to different levels of robot adaptation in collaborative tasks
[92]	2017	Danielsson et al	Collaboration	Robotic arm (UR3 from Universal Robots)	SUS Questionnaire General Interest Questionnaire	–	–	–	25 27 25 21	7 7 7 4	17 19 18 14	1 1 0 1	Evaluation of the effectiveness of AR instructions for collaborative human–robot assembly
[93]	2018	Aromaa et al	Coexistence	Robot system composed of a 4WD rover, a Lynx AL5D robotic arm, and a pan-tilt camera	NASA-TLX [68]	–	–	–	10	2	8	0	Identification of the physical and behavioral characteristics essential to ensure optimal UX in HRI through the use of DHM in the evaluation of prototype gestural interfaces

Table 1 (continued)

Ref	Year	Authors	Interaction level	Robot	Measurements and metrics related to human factors			Measurements and metrics related to performance	Sample size				Main contribution
					Evaluation of Perception	Psychophysiological Evaluation			Total	Women	Men	Other	
						Type of sensor	Specific device						
									9	4	5	0	
[94]	2018	Tang & Webb	Collaboration	Robotic arm (UR5 from Universal Robots)	SUS Questionnaire	–	–	–	8	4	4	0	Evaluation to address ergonomic issues in the design and comparison of the UX of this system with a conventional learning controller in the study
[95]	2019	Kidal et al	Collaboration	Dual-arm robot (Yumi IRB 14000 by ABB)	Interview	–	–	–	1	0	1	0	Qualitative pilot evaluation in which a cognitively impaired assembly worker and multidisciplinary experts interacted with a collaborative robot in routine assembly tasks
[84]	2020	Almeida et al	Coexistence	nd	IBM Computer Usability Satisfaction Questionnaire	–	–	Task execution time	25	2	23	0	Analysis of the transfer of skills and experience to remote locations through teleoperation systems

Table 1 (continued)

Ref	Year	Authors	Interaction level	Robot	Measurements and metrics related to human factors			Measurements and metrics related to performance	Sample size				Main contribution
					Evaluation of Perception	Psychophysiological Evaluation			Total	Women	Men	Other	
						Type of sensor	Specific device						
					Questionnaire based on the presence questions of Slater et al. [96] y Usoh et al. [97]			Errors Accuracy of path-following					
[98]	2020	Hietanen et al	Collaboration	Robotic arm (UR5 from Universal Robot)	Self-generated questionnaire to assess UX composed of the sub-constructs safety, information processing, ergonomics, autonomy, competence and affinity	–	–	Task execution time Robot idle time	20	nd	nd	nd	Model based on depth sensors for workspace monitoring and an interactive AR UI for safe HRC in industrial manufacturing
[99]	2020	Pantano et al	Collaboration	Robotic arm (Kuka iiwa™ R820 14)	NASA-TLX [68]	–	–	-	19	4	15	0	Evaluation of the workload perceived by the users when interacting with the robot and its interface
[100]	2020	Beschi et al	Collaboration	Robotic arm (UR10 from Universal Robots)	Self-generated questionnaire on perceived danger-safety	–	–	Task execution time	32	4	28	0	Analysis of motion planning parameters in the HRC

Table 1 (continued)

Ref	Year	Authors	Interaction level	Robot	Measurements and metrics related to human factors			Measurements and metrics related to performance	Sample size				Main contribution
					Evaluation of Perception	Psychophysiological Evaluation			Total	Women	Men	Other	
						Type of sensor	Specific device						
								Average separation distance Ratio of the time required to complete the task with and without the robot					
[101]	2021	Colim et al	Collaboration	Robotic arm (UR10e from Universal Robots)	Questionnaire to assess workers' perceptions of robots in industry and associated ergonomic improvements	–	–	Task execution time	4	nd	nd	nd	Assessment of the musculoskeletal risk associated with the pre-assembly position before and after the implementation of HRC control, as well as the analysis of indicators of well-being and acceptance of robotics, based on workers' perceptions

Table 1 (continued)

Ref	Year	Authors	Interaction level	Robot	Measurements and metrics related to human factors			Measurements and metrics related to performance	Sample size				Main contribution
					Evaluation of Perception	Psychophysiological Evaluation			Total	Women	Men	Other	
						Type of sensor	Specific device						
								Variability in production times Production rate					
[23]	2021	Joosse et al	Coexistence	Humanoid robot (Nomad Scout)	Godspeed questionnaire	–	–	–	85 80	35 35	50 45	0 0	Analysing human responses to personal space violations committed by robots compared to those of humans and determining the effects of robot characteristics (sound, size, and speed) on acceptance

Table 1 (continued)

Ref	Year	Authors	Interaction level	Robot	Measurements and metrics related to human factors			Measurements and metrics related to performance	Sample size				Main contribution
					Evaluation of Perception	Psychophysiological Evaluation			Total	Women	Men	Other	
						Type of sensor	Specific device						
[24]	2022	Gervasi et al	Collaboration	Robotic arm (UR3e from Universal Robots)	NARS Questionnaire	Skin Conductance Responce (SCR)	Empatica E4 biosensor wristband	Task execution time	42	12	30	0	Experimental evaluation on the relationships between HRC speed, distance and control, perceived interaction quality, affective state, and stress
					Quality of Interaction Questionnaire	HRV		Robot speed					
					SAM			Distance between person and robot					
[102]	2022	Akalin et al	Collaboration	Humanoid robot (Pepper from Aldebaran)	Questionnaire on comfort based on the study by Kim et al. [103]	EDA		–	27	17	10	0	Relationship between factors such as comfort, predictability, sense of control and trust, and perceived safety in HRI using subjective and objective measures
					Questionnaire on perceived safety [104]								
					Questionnaire on sense of control adapted from Strube and Werner [105]	HRV							
					Trust Perception Scale-HRI questionnaire [106]								

Table 1 (continued)

Ref	Year	Authors	Interaction level	Robot	Measurements and metrics related to human factors			Measurements and metrics related to performance	Sample size				Main contribution
					Evaluation of Perception	Psychophysiological Evaluation			Total	Women	Men	Other	
						Type of sensor	Specific device						
					Post-performance questionnaire by Weiss et al. [107] SAM	IBI							
[108]	2023	Gervasi et al	Collaboration	nd	NASA-TLX [68] SAM	SCR HRV	Empatica E4 biosensor wristband	–	36	19	17	0	Experimental evaluation that emulates shift work and investigates the differences between manual repetitive assembly processes and repetitive assembly processes
[85]	2023	Gervasi et al	Collaboration	Robotic arm (UR3e from Universal Robots)	–	SCR HRV	Empatica E4 biosensor wristband	Number and type of process failures	12	6	6	0	To explore the evolution of the psycho-physical state (focusing on stress, cognitive load, and fatigue) and performance of operators during continuous interaction with a collaborative robot in a repetitive assembly process

offers detailed insights into dimensions such as learnability, efficiency, controllability, attractiveness, and originality, the abbreviated UEQ-S [70] delineates between pragmatic and hedonic quality [56]. Nonetheless, empirical evidence suggests that the UEQ-S demonstrates predictive validity concerning pragmatic and hedonic quality when compared to the full version.

2.2.3 Analysis of the Experimental Studies to Measure Human Factors in HRI

Table 1 presents a comparative analysis of studies employing an experimental approach to assess human factors in HRI. A total of 17 studies conducted over the past decade, spanning from 2013 to 2023, are examined. These studies encompass a broad spectrum of research conducted throughout the last ten years, allowing us to observe both the evolution and trends in the field of HRI during this period. The heterogeneity observed among these diverse studies suggests a need for greater standardization and the exploration of more comprehensive approaches to evaluate HRI in industrial settings.

The inclusion of performance indicators serves to provide valuable insights into the efficiency and effectiveness of HRI. As can be seen in Table 1, 8 out of the 17 studies incorporated assessments related to performance. The most commonly used performance metric in these studies is the "task completion time", also according to a previous study [84]. Additionally, tracking errors during HRI provides crucial information about the quality and accuracy of task execution. Errors can encompass a range of issues, including misinterpretations of instructions, incorrect actions taken by the robot, or user errors. While error collection was only conducted in the study by Almeida et al.[85] among those reviewed in Table 1, the analysis of errors provides insights into areas requiring improvement, facilitating the refinement of the process to enhance performance and UX.

Regarding physiological measurements, it is notable that four of the studies have incorporated such evaluations into their methodologies. What is particularly striking is that these studies are notably recent, with publication dates in the years 2022 and 2023, underscoring how technological advancements and the development of more sophisticated wearables are enabling the integration of physiological monitoring in research. However, it is important to highlight that the recorded physiological measures primarily focus on ECG and EDA. In fact, three out of the four studies utilize the Empatica E4 biosensor wristband device. While effective for collecting physiological data, Gervasi et al. [86], indicated some limitations: (1) the accuracy of the measurements obtained from the Empatica E4 may be influenced by factors such as motion artifacts or skin contact issues, which could affect the reliability of the data; (2) the battery life of the device limited the duration of the data collection; (3) while it captures EDA and HR data, it does not provide a comprehensive range of physiological parameters for a detailed analysis of the operator's physiological state.

Thus far, no study has been identified that has utilized EEG in conjunction with these physiological measurements to assess human factors in experimental HRI studies in industrial settings. Nevertheless, EEG offers direct insights into cognitive processes and emotional responses during the interaction. According to a systematic review on the use of psychophysiological devices in UX studies [34], EEG emerged as the most commonly employed tool. The extensive presence of EEG in various studies indicates its growing application in UX testing and its significant potential for extracting emotional insights, despite its complex application protocols and interpretation. Nevertheless, it appears that its utilization in UX studies within domains such as HRI has yet to be fully realized. This may be attributed to its potentially more intricate implementation in multifaceted environments like HRI, prompting current HRI studies to favor less invasive and more straightforward devices. Integrating EEG can greatly enhance our understanding of UX in HRI studies, as it gives us direct insights into cognitive processes and emotional responses during interactions, complementing information from EDA.

Out of the 17 identified studies, 15 of them incorporate at least one questionnaire as part of their evaluation methods. Among the variety of questionnaires used, the NASA-TLX [69] stands out, being employed in 3 of the studies. This frequent choice of the NASA-TLX [69] suggests that it is a well-recognized and widely validated tool for measuring perceived workload and other aspects related to HRI. However, it's important to note that the NASA-TLX [69] does not address hedonic aspects of the interaction experience with a robot, underscoring the need to identify or develop a more comprehensive questionnaire that allows for the assessment of users' perceptions.

Regarding the sample size used in the case studies, Table 1 shows that it ranges from 4 to 85. The wide variability in sample size among the case studies primarily depends on the methodology employed in each of them. According to a previous study [34], physiological monitoring is a relatively emerging technology in usability and UX studies, thus an ideal estimation for the sample size in such studies is not readily available. However, several studies [87–89] suggest that the commonly known "magic number 5" may suffice to uncover approximately 80% of usability issues in technology acceptance. Nevertheless, employing small sample sizes may result in significant variability in test outcomes that cannot be adequately mitigated [90].

As for gender representation, the studies reviewed indicate a significant underrepresentation of women. Only 33% of the total sample across the studies are women, while 62% are men. Although more recent studies are starting to use

gender-balanced samples, the obtained results are often not disaggregated by gender, resulting in a lack of a gender perspective in the research. Regardless of whether there are actual gender differences or not, it is crucial to explore and acknowledge the potential gender disparities in the findings. This is further underscored by the results obtained by Hopko et al. [29] and Apraiz et al. [15].

2.3 Kinematic Variables and Human Factors in HRI

The robot's kinematic variables influence human perception and, therefore, his/her experience and performance. Human operators have mental strain when industrial robots move, even though they would get accustomed to the robots [18]. The speed and acceleration of movements have been considered in collaborative robots mainly for safety reasons. This causes researchers and developers to disregard the operator's emotional responses when specifying the robot's kinematic variables [16].

Kulic and Croft [110] investigated the potential impact of an industrial robot's movements on the physiological state of humans. They utilized measures such as HR and EDA to assess this effect. The study revealed that the robot's proximity and speed had a significant influence on human physiological responses. Unlike their work, the present study endeavours to delve not only into physiological responses such as HR and EDA, but also into EEG. As identified in Sect. 2.2.3, there are no existing HRI studies in industrial settings that incorporate this measure. Furthermore, this study aims to concurrently investigate perceptual measurements and performance during HRI.

Arai et al. [18] found that the speed of a robot moving toward a human operator should be less than 500 mm/s. However, their focus was primarily on mental strain, with less consideration given to more emotional factors. Similarly, Koppenborg et al. [17] studied the effects of movement speed and path predictability on human operators using a virtual reality set-up. They found that faster movements resulted in higher ratings for task load and anxiety. In this case, the authors centered their evaluation on mental workload, risk cognition, and risk perception through the NASA-TLX [69], leaving aside emotional aspects. They also collected data on mean Interbit Intervals (IBI).

Kühnlenz et al. [111] explored the impact of different trajectory patterns executed by an industrial robot on human subjects. They assessed these effects by measuring HRV and EDA in the participants. Their results revealed a rise in both average HRV and average skin resistance when participants were exposed to the trapezoidal trajectory pattern, indicating a reduction in mental stress. However, their study did not take into account more emotional aspects. Unlike their approach, our research aims to incorporate these emotional factors.

Lasota and Shah [16] and Beschi et al. [101] also explored the effects of motion. Nevertheless, they did not explore with an integral human-centred approach considering cognitive and perceptual aspects. The study by Rubagotti et al. [22] identified the main factors related to the robot's motion and characteristics that determine perceived safety. Among other things, they concluded that the lower the robots speed is, the safer it is perceived. In contrast to their work, our study expands the scope beyond just safety considerations. Our research aims to explore the impact of kinematic variables, speed and acceleration, on a broader range of human factors, including physiological and perceptual aspects.

Based on the analysis of the current literature, the HRI within the manufacturing sector has yet to fully incorporate human perception into its design and implementation [12]. Notably, the effect of kinematic variables, such as speed and acceleration, on human perception during HRI still needs to be explored from a brain activity and gender perspective approach. Although motion levels have been widely studied in the literature, to date, no study has been identified that explores its effects from a cognitive and experiential approach.

The objective of affective robotics is to alleviate the cognitive burden on users when they are faced with tasks that exceed their mental capacities. This is achieved by adapting the robot's behaviour and incorporating an appropriate level of autonomy [112]. Affective robotics holds the potential to provide an intuitive interface, thus easing the cognitive demands placed on users as the robot dynamically adjusts to their circumstances. However, prior to implementing affective robotics, it is essential to gain insight into individuals' emotional responses to various robot behaviours. In this context, the introduction of EEG together with an EDA as a tool to assess human's perceptions in response to distinct kinematic variables of the robot presents an intriguing and previously unexplored opportunity. It is imperative to initiate the exploration of users' reactions before employing affective robotics, specifically investigating the impact of robot kinematic variables on users' physiological and perceptual responses. Furthermore, it is noteworthy that existing literature has generally overlooked the gender perspective in human–robot interactions, warranting additional investigation in this regard.

Given the growing importance of HRI, it is imperative to understand how kinematic variables affect this interaction. A thorough investigation of these factors would allow the identification of improvements in the interaction design and could enhance the effectiveness and efficiency of HRI. Additionally, a more comprehensive understanding of cognitive and experiential variables would provide valuable insights into the development of user-centred design strategies for robots, which could ultimately increase their adoption and use in a variety of contexts. In light of these considerations,

the execution of an experiment to evaluate the effects of speed and acceleration on human perception during a robot assembly task, while taking into account gender perspectives, is warranted. Such an endeavour would contribute to a more comprehensive understanding of the kinematics of HRI and could inform the development of more effective and efficient interaction.

3 Research Methodology

The experiment aims to find out how kinematic variables (speed and acceleration) influence human factors. We used a multi-method and human-centred approach, considering data of diverse nature: performance values, physiological values, and perceptual values. This research also relies on observation.

3.1 Research Questions and Hypothesis

This paper aims to provide a study on how different kinematic variables of a robotic system affect the UX, determining the acceptance of the robot by the user. The research questions that will be investigated in this paper are the following:

Research Question 1 (RQ1) In the context of industrial HRI assembly task, what is the impact of different robot kinematic variables (i.e., movement speed and acceleration) on human factors?

To provide a more refined focus for the RQ1, the following hypotheses have been formulated:

Hypothesis 1 (H1) Within the context of industrial HRI assembly task higher values of robot kinematic variables (i.e., speed and acceleration) will lead to an increased cognitive workload.

Hypothesis 2 (H2) Within the context of industrial HRI assembly task higher values of robot kinematic variables (i.e., speed and acceleration), will decrease the user's emotional responses.

Research Question 2 (RQ") Do these kinematic variables have a different impact on men and women?

To address this question, H1 and H2 will be examined, considering the gender perspective. In essence, an equitable sample will be employed, and the collected data will be disaggregated based on gender for further analysis.

To answer these RQs, a collaborative assembly task was implemented where 20 participants (10 men and 10 women) worked together with a KUKA iiwa collaborative robot. We observed how well the participants performed, checked their emotional state, and measured their physical responses like Electrodermal Activity (EDA), Heart Rate Variability (HRV), and brain activity (by means of an EEG).

3.2 Experimental indicators

As can be seen in Fig. 1, to evaluate cognitive workload (H1), we will employ physiological indicators derived from EEG, including Memorisation, Mental Workload, and Engagement. For the assessment of emotional responses (H2), physiological indicators such as Valence as detected by EEG, will be considered. Additionally, post-interaction perception of emotional responses holds significance. Also, Activation, and Impact, as detected by an EDA/GSR. Additionally, post-interaction perception of emotional responses holds significance. To this end, perceptual indicators including Pragmatic Quality, Hedonic Quality, Reliability, Controllability, and Perceived Usefulness will be evaluated through a questionnaire.

We integrated an approach to capturing a person's perceptions at different stages, i.e., before, during, and after an interaction. We included different measurement methods at different moments of the interaction—quantitative and qualitative, objective and subjective—to guarantee the interpretation of the data and the understanding of the entire flow of interaction. In this sense, we used 1) Key Performance Indicators (task execution time and errors), which are quantitative and objective measures; 2) physiological values, which are quantitative and objective; and 3) perceptual values, which are quantitative and subjective. This approach was also supported by observing participants' reactions, which is a qualitative and subjective measure.

3.2.1 Key Performance Indicators (KPIs)

Performance refers to how participants perform their tasks. Users' capabilities shaped it, and it is the consequence of the human factors displayed in the system. These pragmatic and objective indicators provide the model with an objective view of task performance. In this case, we measured two indicators:

- *Task execution time***,** from the task's assignment to completion.
- *Errors***.** Considering as errors all those actions that were not foreseen in the sequence.

3.2.2 Physiological Indicators

Among the various physiological tools commonly utilized in UX studies, EEG and GSR/EDA have emerged as robust choices due to their well-established relationships with emotional and affective states [113]. EEG offers valuable insights into cognitive and emotional processes by detecting electrical brain activity, while GSR/EDA provides information about emotional arousal through changes in skin conductance.

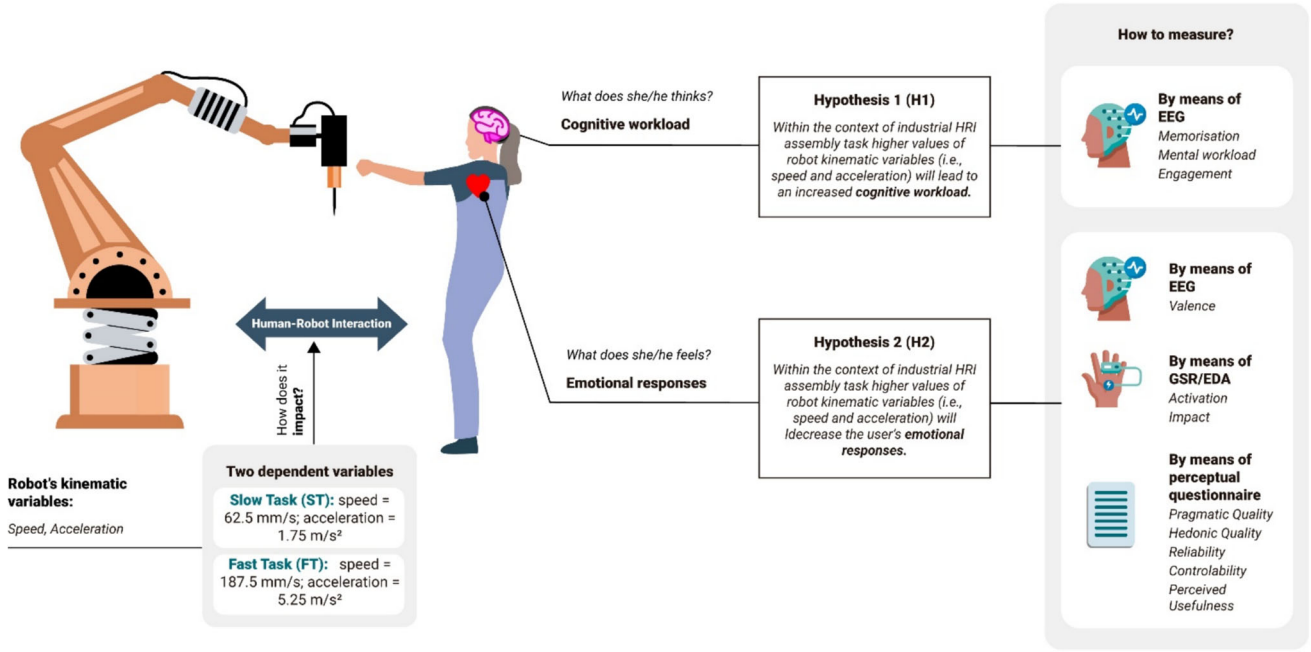


Fig. 1 Outline of the experimental design

These physiological measures have repeatedly demonstrated their ability to capture nuanced aspects of the user's emotional and cognitive experience during interactions with technology.

Furthermore, it is noteworthy that, despite their established relevance in UX research, there is a notable gap in the existing literature regarding their application in the context of HRI within industrial tasks. Specifically, no prior HRI studies utilizing EEG to monitor users' emotional responses in industrial settings have been identified in the literature.

In accordance with the formulated Hypothesis 1, which posits that, "*Within the context of industrial HRC assembly task higher values of robot kinematic variables (i.e., speed and acceleration) will lead to an increase in Memorisation, Mental Workload, Engagement, Impact, and Activation, while concurrently decreasing emotional Valence.*", we selected devices capable of measuring these aspects. Therefore, we used two devices:

- *EEG***.** We employed the Diadem of Bitbrain [43], a dry-sensor designed for real-world applications requiring high user comfort, agile set-up, and outstanding signal quality. Diadem has been developed with 12 dry sensors located in specific brain areas (pre-frontal, frontal, parietal, and occipital) to estimate emotional states (Fig. 2). The indicators collected through it are 1) Valence, 2) Memorisation, 3) Mental Workload, and 4) Engagement.
- *GSR/EDA*. We used the Ring of Bitbrain [49]. It is a mobile and robust biosignal measurement device for monitoring skin conductance (electrodermal sensor—EDA/galvanic skin response—GSR) and cardiovascular activity. It integrates sensors positioned on the first and second phalanges of the fingers, which are optimal measurement points for accurate data collection (Fig. 3). Also, it incorporates a three-axis solid-state accelerometer (ACC) to mitigate motion artifacts. The indicators collected through Ring are 1) Activation and 2) Impact.

3.2.3 Perceptual Indicators and Observation

User perception values were collected through the questionnaire UEQ-S [83]. It contains two subscales: Pragmatic Quality (PQ) and Hedonic Quality (HQ). It also reflects the overall UX value. Moreover, in the HRI context, perceived safety, trust and Perceived Usefulness are relevant human factors [33, 114, 115]. As the UEQ-S questionnaire does not include them, we have added three items to assess them also in a 7-point Likert scale:

- *Reliability* Unreliable–Reliable.
- *Controllability***:** Uncontrollable–Controllable.
- *Perceived Usefulness* Not valuable–Valuable.

The concepts of Reliability and Controllability are important measures to assess the user's perception of the robot. Reliability refers to the consistency and dependability of the robot, while Controllability refers to the ease of operation and the ability of the user to control the robot. These factors play a crucial role in determining the level of Trust and Perceived

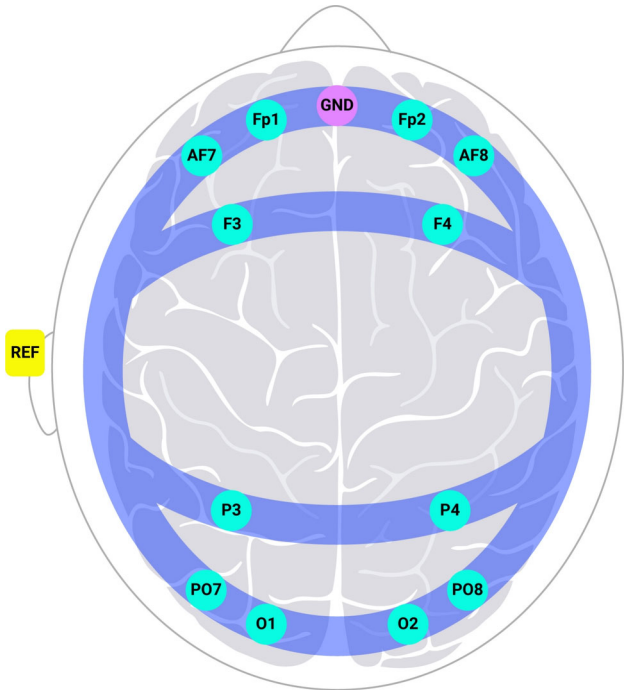


Fig. 2 Electrode placement in the EEG. Adapted from Bitbrain [43]

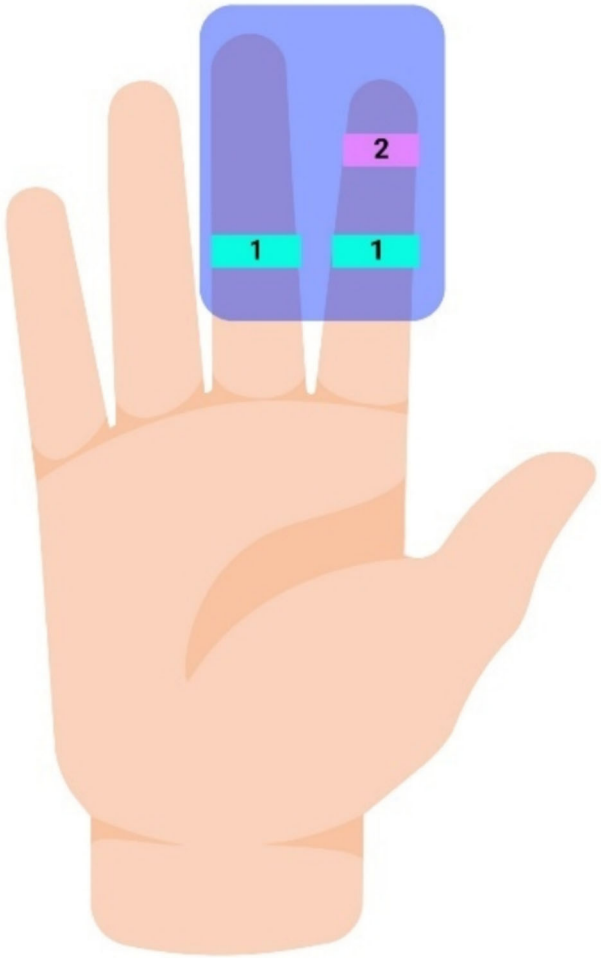


Fig. 3 Sensor placement in the Ring. Adapted from Bitbrain [43]

Safety in the interaction [103]. High scores for Reliability and Controllability indicate that users trust the robot to perform consistently and can easily control it, leading to a sense of safety. Conversely, low scores for these concepts can lead to a lack of trust and decreased perception of safety, which can impact the overall experience. On the other hand, Perceived Usefulness refers to the degree to which an individual believes that using the robot will enhance their job performance or increase their productivity. It reflects the degree to which the participant finds the robot helpful and adds value to their experience. It is a crucial factor in determining the overall satisfaction of the participant and their willingness to continue using the robot.

In addition to the questionnaire, the users' interaction was observed during the execution of tasks. This observation was done by another person, not by the test moderator. The person dictating the tasks should focus only on directing the user. The observer focuses on picking up problems that arise during the test and the user's reactions to these situations, and on the comments the user did while executing the task and after. That is, participants' verbal and non-verbal reactions were observed.

3.2.4 Data Processing

Table 2 summarises the response and independent variables considered in the study. Physiological data were collected by Sennslab [116], a multimodal data recording software. Then, we processed the data by Sennsmetrics [117], an extension of Sennslab that allows us to analyse the physiological data.

3.3 Experimental Procedure

In the experimental procedure, three phases are differentiated (Fig. 4): 1) pre-task execution phase, 2) task execution phase, and 3) post-task execution phase.

3.3.1 Pre-task Execution Phase

To begin with, participants had to sign the consent form, where they were informed about how the test would be carried out, and the data that would be requested and collected (image and sound recordings; and biometric monitoring using electroencephalogram -EEG-, electrodermal activity -EDA-).

Participants then filled in a socio-demographic questionnaire asking for information such as age, gender, role at university, and whether they had previous experience with

Table 2 Summary of the variables involved in the study

Type of variable	Variable category	Variable name	Hypothesis to assess	Description
Independent variables	Configuration factor	Kinematic variables		Speed and acceleration (low, high)
Response variables	Performance indicator	Task execution time		From the moment the task is assigned until its completion
		Errors		All those actions that were not foreseen in the sequence
	Perceptual indicators	Pragmatic Quality (PQ)	Emotional responses (H2)	Perceived ability to support the achievement of "do-goals" [118]
		Hedonic Quality (HQ)	Emotional responses (H2)	Perceived ability to support the achievement of "be-goals" [118]
		Reliability	Emotional responses (H2)	Perceived confidence towards the robot
		Controllability	Emotional responses (H2)	Perceived ability to control the robot
		Perceived Usefulness	Emotional responses (H2)	The degree of belief that using it will enhance their performance or increase their productivity
	Physiological response	Valence	Emotional responses (H2)	The degree of attraction experienced [52–54]
		Memorisation	Cognitive workload (H1)	The intensity of cognitive processes related to forming future memories [55–58]
		Mental Workload	Cognitive workload (H1)	The use of cognitive resources to carry out a task [59–62]
		Engagement	Cognitive workload (H1)	The degree of involvement or connection between the participant and the task [63–65]
		Activation	Emotional responses (H2)	Basal level of physiological Activation [44–48]
		Impact	Emotional responses (H2)	The number and intensity of point changes in the emotional state [44–48]

robots. The participants were also instructed on how to interact with the robot using voice commands.

3.3.2 During the Task Execution Phase

Participants were fitted with physiological devices (EEG and EDA) to start with the task. As the SennsLab software [119] indicates, the first step was calibrating the devices. This validated calibration protocol is provided in the software itself [120].

Once physiological devices were calibrated, two procedures were differentiated:

- *In the Slow-Fast (SF) procedure*,[1] the participants (1–10) first completed the task in a slow robot motion mode, i.e., the speed of the robot was slow (62,5 mm/s) and the acceleration smooth (1.75 m/s^2). Once they finished the task; they filled in the perceptual questionnaire (see subsection 3.4.3). Then, they repeated the same task in a fast motion mode, i.e., with the high robot's speed (187,5 mm/s) and high acceleration (5.25 m/s^2). Then, they filled in the perceptual questionnaire again.
- *Fast-Slow (FS) procedure*.[2] In the Fast-Slow (FS) procedure, the participants (11–20) first performed the task in the fast-motion mode. Then, they filled in the perceptual indicators' questionnaire. After filling in the questionnaire, they performed the task in slow-motion mode. Then they filled in the perceptual indicators' questionnaire.

3.3.3 Post-task Execution Phase

After the interaction, a free conversation was held with the participants. They were asked questions about how they felt

[1] Slow-Fast (SF) procedure.

[2] Fast-Slow (FS) procedure.

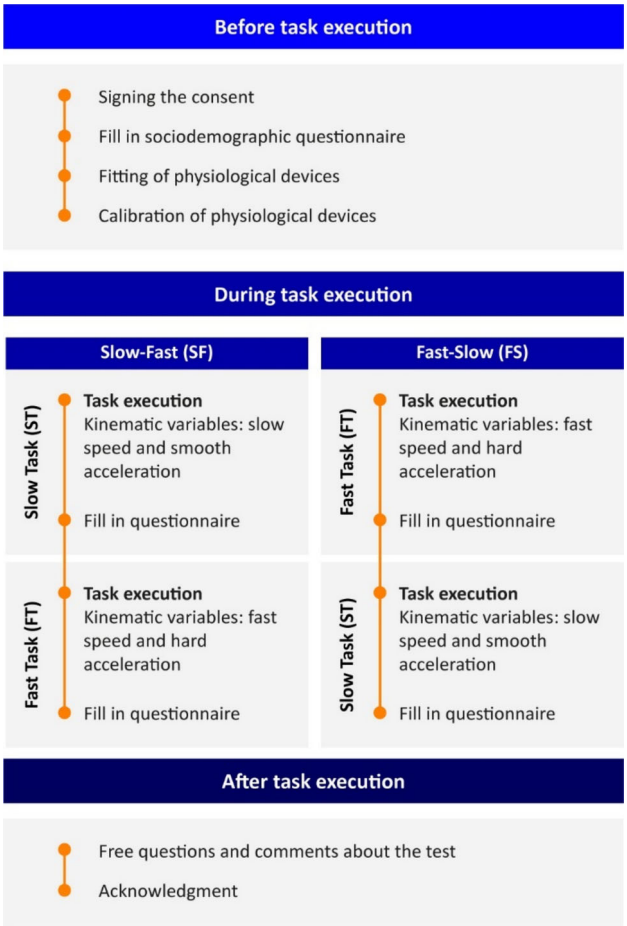


Fig. 4 The sequence of the experimental procedure throughout the case study

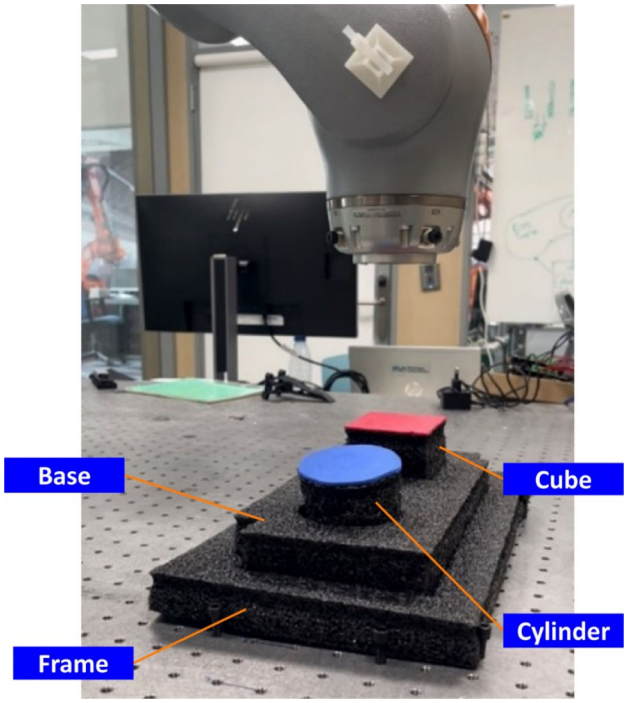


Fig. 5 Robot task and environment setup

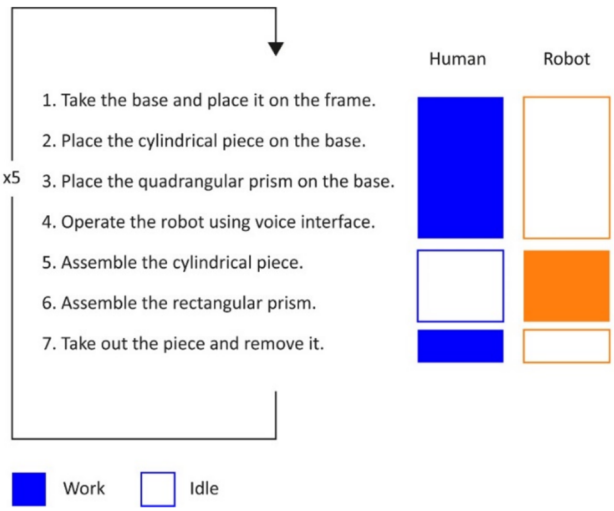


Fig. 6 The sequence of one assembly during the task

and what they expected. The goal was for participants to feel free to discuss their emotions.

3.4 Task Environment and Robot

The execution of tests was carried out in a laboratory setting (Fig. 5), being a controlled environment, it allows to focus on the variables to be studied, i.e., the kinematic variables in human perception. The KUKA LBR iiwa 14 R820 [121] collaborative robot was used [122]. It has seven degrees of freedom and a torque sensor mounted in each joint.

3.5 Task

We designed a collaborative assembly task. The participant's duty was to pick & place the three parts (cylinder, cube, and a base) to be assembled. The participants had to be careful to place the base on a frame. Bearing in mind that the base has two holes, one with each geometrical shape, then the cylinder and cube has to be located in each corresponding position (Fig. 3). After, the participants, using a voice interface, would command the robot to push on the two pieces down to assemble them with the base. Once assembled, the participant had to remove the base (already assembled with the two pieces) from the frame. The process had to be repeated up to five times. Figure 6 shows the sequence of activities during the task. What makes this task collaborative is that the participants, utilizing a voice interface, would then instruct the robot to apply downward pressure on the two components, effectively assembling them with the base. Once the assembly was complete, the participants had to skilfully remove

the base, now securely joined with the two pieces, from the frame. This process was repeated up to five times, providing a clear demonstration of human–robot collaboration.

As the aim of the study is to analyse how kinematic variables influence perception, two robot behaviours were programmed: Slow Task[3] (ST, speed = 62,5 mm/s; acceleration = 1.75 m/s^2 and Fast Task[4] (FT, speed = 187,5 mm/s; acceleration = 5.25 m/s^2). The speed and acceleration of the robot were three times higher in FT than in ST. The selection of two distinct robot behaviours, ST and FT, was driven with the aim of comprehensively assessing the influence on kinematic variables on human factors to be studied. ST and FT represent two ends of the spectrum in terms of speed and acceleration, allowing for a broader exploration of how these variables impact the UX. ST was deliberately set at a slower pace to simulate tasks that require precision and careful control, which are common in industrial settings. FT, on the other hand, was configured with significantly higher speed and acceleration o represent scenarios where rapid task completion is essential.

To mitigate the possibility of order effects that might introduce bias into the results, a counterbalancing strategy was employed. This approach resulted in the presentation of two procedures: the Slow-Fast (SF) procedure and the Fast-Slow (FS) Procedure, as elaborated in Sect. 3.3.2. Half of the participants began the experiment with ST, while the other half commenced with FT. This systematic approach ensures that the sequence of task exposure does not confound the findings and enables a more robust analysis of the impact of kinematic variables on user perception.

During the tasks, participants remained seated in a position in front of the robot (Fig. 7). In addition, as they were monitored with biometric devices, EEG and EDA (see subsection 3.4.2), they could not move the non-dominant hand.

3.6 Sample

Twenty healthy volunteers participated in the experiment (ten men and ten women) evenly distributed in both procedures. The SF procedure involved ten people aged between 21 and 41 years (average = 28.7, std = 6,7), five university students, and five university staff members. Four of them had previous experience with a domestic robot, five with industrial robots, and one had never interacted with a robot before. And the FS procedure involved ten people aged between 24 and 50 years (average = 30,5 std = 7,8). Two were university students, and eight were university staff members. Two participants had previously interacted with a domestic robot, four with an industrial robot, and four had never interacted with a robot.

[3] Slow Task (ST).

[4] Fast Task (FT).

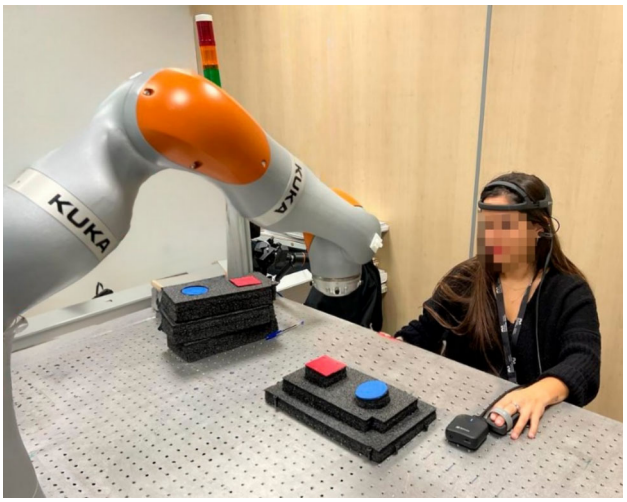


Fig. 7 Participant executing experimental study

The dominant hand of all participants was the right hand; therefore, all participants wore the Ring on the left.

4 Results

Data analysis was structured according to the proposed research methodology, so the results are grouped into three sections: 1) the key performance indicators, 2) the physiological indicators, and 3) the perceptual indicators resulting from the questionnaire. The presented dataset is stored at Mendeley data [123].

4.1 Key Performance Indicators (KPIs)

KPIs were analysed by measuring task execution time and errors. In ST, the robot's speed was 62,5 mm/s and the time needed to execute the task without the person's help was 2 min and 28 s. In FT, the robot's speed was 187,5 mm/s, and the time needed to execute the task without considering the time needed by the person was 1 min and 20 s.

Analysing the absolute values without considering the order in which the tasks were performed, the results show that participants took less time to complete FT (Fig. 8). These results are logical because, as the robot performed faster, the participants had a better chance of finishing the task sooner.

If the results are analysed considering the order of the tasks, in procedure SF, all participants took longer to complete ST than the participants in procedure FS. Similarly, the time to complete FT was shorter for participants in procedure SF than those in procedure FS. Therefore, it is observed that participants who performed the tasks in the opposite order of procedures (SF and FS) took a longer time to complete the tasks and made more errors in the first task than in the

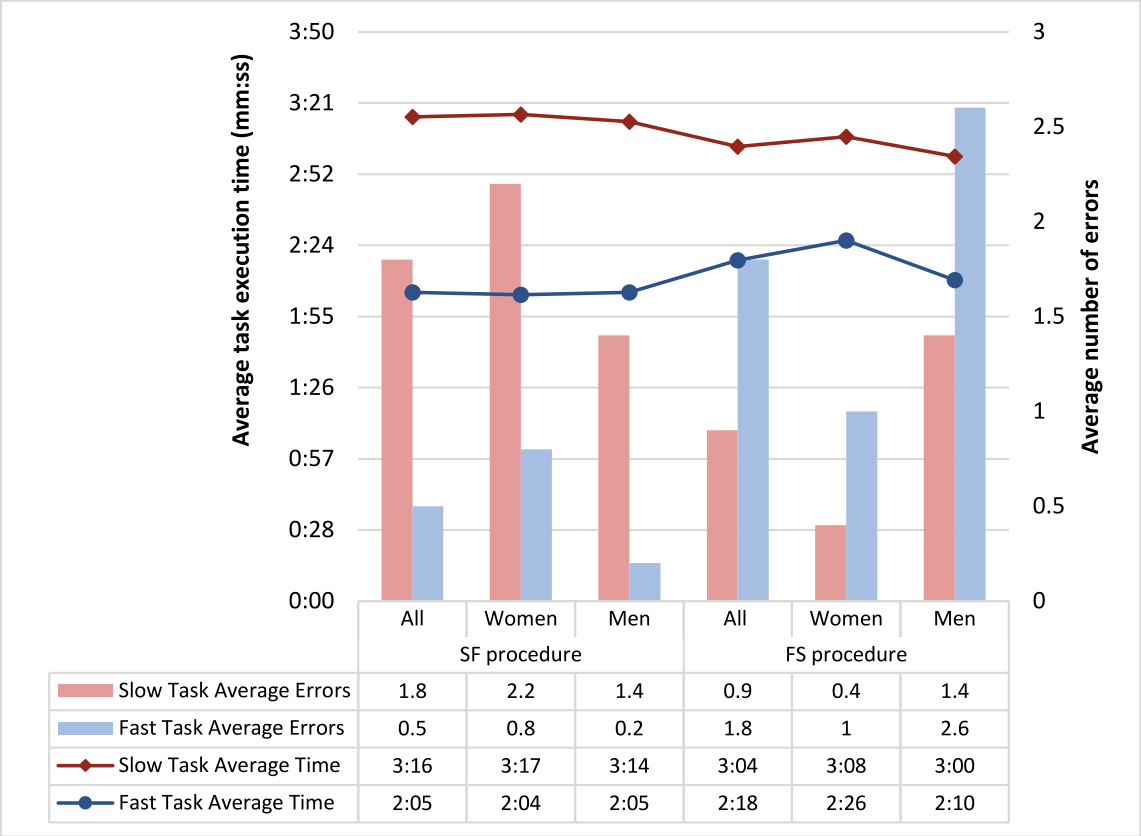


Fig. 8 General comparison of average Execution Time and Errors

Table 3 T-test results for assessing performance indicator differences between tasks in each procedure considering $\alpha = 0{,}05$. Statistically significant values are highlighted in bold

SF Procedure		FS Procedure	
Time	Errors	Time	Errors
6,65E-11	**3,85E-03**	**2,20E-08**	**1,40E-01**

second task. This indicates that the order in which the tasks are performed can influence the results and suggest that participants may have to become more familiar with the task to lead to better performance.

On average, participants in the SF procedure took 48 s more to execute ST than the expected time. Participants in the FS procedure also took 36 s more on average to execute ST than the expected time. As for FT, participants in the SF procedure took 45 s more than expected on average, and those in the FS procedure took 58 s more than expected.

Table 3 presents the results of the T-test to determine differences between tasks in both procedures.

As for the errors, the order in which the tasks were carried out influenced the performance. Participants in the SF procedure made more errors in ST than in FT, and the participants in procedure FS made more errors in FT than in ST.

Therefore, the order factor alters the results during the execution of the tasks. The second task was completed in both procedures with better results than the first one. The number of errors was also reduced in the task performed in the second place. However, the time needed to complete the task is faster in the FT. In this sense, the order factor has a significant impact on learnability, as it can greatly influence the speed and efficiency with which users are able to learn and effectively use the system.

Table 4 presents the results of the T-test to determine differences in the task disaggregated by gender. it is observed that for women, in both procedures, the difference in task execution time is statistically significant. Conversely, for men in the SF procedure, the difference is not statistically significant, but it is in the FS procedure. Regarding errors, in the SF procedure, the difference is considered significant for both men and women, while in the FS procedure, neither men nor women show statistically significant differences.

Table 4 T-test results for assessing performance indicator differences between tasks in each procedure considering $\alpha = 0{,}05$ and disaggregating by gender. Statistically significant values are highlighted in bold

	SF Procedure		FS Procedure	
	Time	Errors	Time	Errors
Women	**0,0160**	**0,0371**	**0,0000270**	0,4045
Men	0,0001	**0,0509**	**0,0001**	0,1347

4.2 Physiological Indicators

This section shows and analyses the results obtained with the physiological devices. Then, the results obtained using EEG in the SF Procedure are shown (Fig. 5) and analysed.

- *Engagement* A higher level of Engagement was observed in participants in the FT (M = 47.81; SD = 57.40) compared to the ST (M = 33.81; SD = 4.53). However, this difference is not statistically significant (Table 5).
- *Memorisation* Slight lower Memorisation was observed in participants in the ST (M = 30.40; SD = 6.00) compared to the FT (M = 31.03; SD = 5.05). However, this difference is not statistically significant (Table 5).
- *Valence* Emotional Valence was recorded as more negative in the ST (M = -3.04; SD = 7.62) compared to the FT (M = -1.89; SD = 9.76). However, this difference is not statistically significant (Table 5).
- *Mental Workload* Mental workload was observed to be lower in the ST (M = 29.58; SD = 3.10) compared to the FT (M = 37.21; SD = 19.59). However, this difference is not statistically significant (Table 5).
- *Activation* It is observed that Activation was higher in the ST (M = 18.75; SD = 16.15) than in the FT (M = -0.74; SD = 27.59). However, this difference is not statistically significant (Table 5).
- *Impact* It is observed that the Impact was lower in the ST (M = 34.94; SD = 1.68) compared to the FT (M = 36.43; SD = 2.55). However, this difference is not statistically significant (Table 5).

However, the differences depicted in Fig. 9 are not statistically significant (Table 5).

In order to address the second research question, we have disaggregated the data obtained in the SF procedure by gender (Fig. 10).

- *Engagement* In the case of women, it is observed that the level of Engagement has been slightly higher in the ST (M = 32.45; SD = 5.50) than in the FT (M = 29.44; SD = 5.72). However, for men, the level of Engagement has been significantly lower in the ST (M = 35.16; SD = 3.36) than in the FT (M = 66.18; SD = 80.85). Nonetheless, these differences are not statistically significant (Table 6).
- *Memorisation* Among women, Memorisation appears to be slightly lower in the ST (M = 30.27; SD = 7.28) than in the FT (M = 31.96; SD = 6.06). Conversely, for men, Memorisation is slightly higher in the ST (M = 30.53; SD = 5.43) than in the FT (M = 30.11; SD = 4.31). However, these differences are not statistically significant (Table 6).
- *Valence* In the case of women, emotional Valence is better in the ST (M = -1.07; SD = 9.84) than in the FT (M = -3.31; SD = 2.79). However, in the case of men, emotional Valence is lower in the ST (M = -5.01; SD = 4.29) than in the FT (M = -0.47; SD = 14.19). Nonetheless, these differences are not statistically significant (Table 6).
- *Mental Workload* In the ST, it is observed that Mental Workload is lower for both men and women. However, the difference is more pronounced in men, where mental load is higher in the FT compared to the ST. The data shows that for women, in the ST (M = 28.77; SD = 2.07) and in the FT (M = 29.89; SD = 3.95), while for men, in the ST (M = 30.39; SD = 3.96) and in the FT (M = 44.52; SD = 26.72). However, these differences are not statistically significant (Table 6).
- *Activation* Both men and women show higher Activation in the ST (Women: M = 24.32; SD = 18.88; Men: M = 13.17; SD = 12.37) than in the FT (Women: M = 16.63; SD = 18.62; Men: M = -18.11; SD = 24.74). In women, this difference is not statistically significant, but in men, it is considered statistically significant (Table 6).
- *Impact* Both men and women show less Impact in the ST (Women: M = 35.14; SD = 1.93; Men: M = 34.74; SD = 1.58) than in the FT (Women: M = 35.77; SD = 1.16; Men: M = 37.09; SD = 3.49). However, these differences are not statistically significant (Table 6).

Therefore, as observed in Table 6, the only significant difference is in the case of men in the Activation indicators, as they have obtained lower Activation values in the FT.

Below, the results obtained through EEG and EDA in the FS Procedure (Fig. 11) are presented and analyzed:

- *Engagement* A lower level of Engagement was observed in participants in the FT (M = 32.84; SD = 4.67) compared to the ST (M = 36.04; SD = 7.71). However, this difference is not statistically significant (Table 7).
- *Memorisation* Participants were found to have a higher level of Memorisation in the FT (M = 32.12; SD = 4.69)

Table 5 T-Test results to determine significant differences between tasks in the SF procedure considering $\alpha = 0.05$. Statistically significant values are highlighted in bold

Engagement	Memorisation	Valence	Mental workload	Activation	Impact
0,4613	0,8011	0,7722	0,2532	0,0737	0,1420

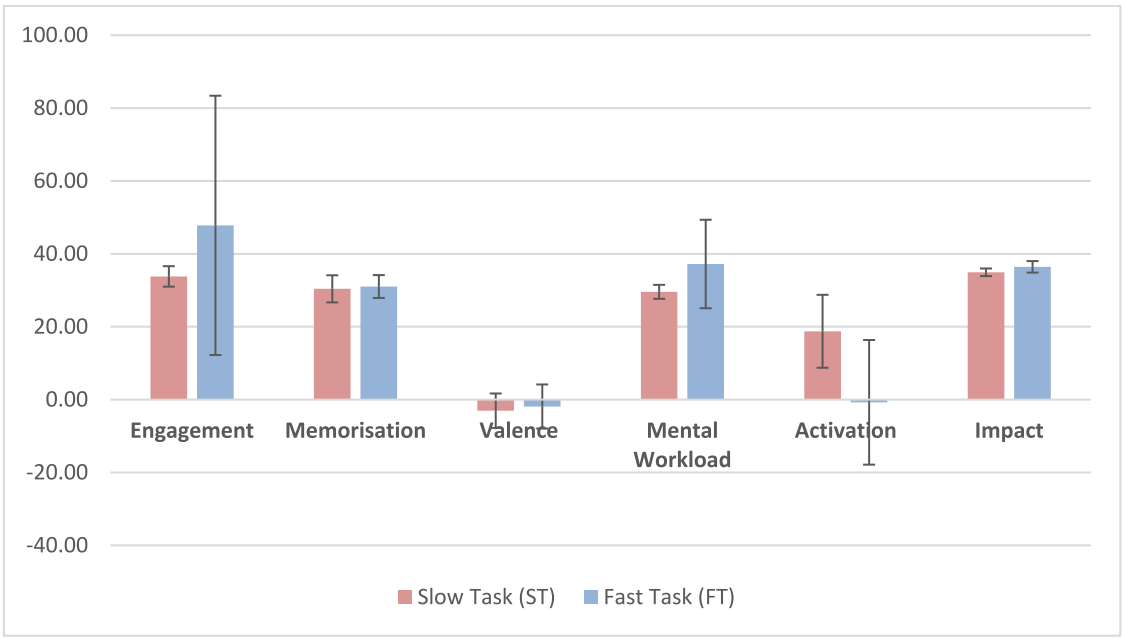


Fig. 9 Results of Physiological Indicators in the SF Procedure

compared to the ST (M = 23.74; SD = 4.93). This difference is considered statistically significant (Table 7).

- *Valence* Emotional Valence was recorded as more negative in the FT (M = -3.08; SD = 6.19) in contrast to the ST (M = 5.00; SD = 19.63). However, this difference is not statistically significant (Table 7).
- *Mental Workload* It was observed that Mental Workload was similar in the FT (M = 30.35; SD = 3.68) compared to the TL task (M = 30.53; SD = 3.97). Indeed, this difference is not statistically significant (Table 7).
- *Activation* It is noted that Activation has been less negative in the ST (M = -0.71; SD = 34.29) than in the FT (M = -9.79; SD = 31.84). However, this difference is not statistically significant (Table 7).
- *Impact* It is observed that the Impact has been lower in the FT (M = 35.06; SD = 1.71) than in the ST (M = 35.50; SD = 4.17). However, this difference is not statistically significant (Table 7).

However, it is important to highlight that, according to the results in Table 7, only the difference in the Memorisation indicators is statistically significant between the tasks.

Breaking down the data from the FS procedure by gender has yielded the following results (Fig. 12):

- *Engagement* In both men and women, a lower level of Engagement is observed in the ST (Women: M = 33.75; SD = 4.17; Men: M = 31.93; SD = 5.45) compared to the FT (Women: M = 35.04; SD = 9.54; Men: M = 37.04; SD = 6.35). However, these differences are not considered statistically significant (Table 8).
- *Memorisation* In both men and women, a higher level of Memorisation is observed in the ST (Women: M = 30.35; SD = 3.40; Men: M = 33.89; SD = 5.49) compared to the FT (Women: M = 25.24; SD = 5.69; Men: M = 22.24; SD = 4.10). In women, the difference is not considered statistically significant, but it is for men (Table 8).
- *Valence* In both men and women, a lower emotional Valence is observed in the ST (Women: M = -2.90; SD = 4.15; Men: M = -3.27; SD = 8.29) compared to the FT (Women: M = 7.92; SD = 28.64; Men: M = 2.07; SD = 5.03). However, these differences are not considered statistically significant (Table 8).
- *Mental Workload* In the case of women, a higher Mental Workload is observed in the ST (M = 29.61; SD = 3.45) compared to the FT (M = 28.67; SD = 2.29). However, in the case of men, a lower Mental Workload is observed in the ST (M = 31.10; SD = 4.14) compared to the FT (M

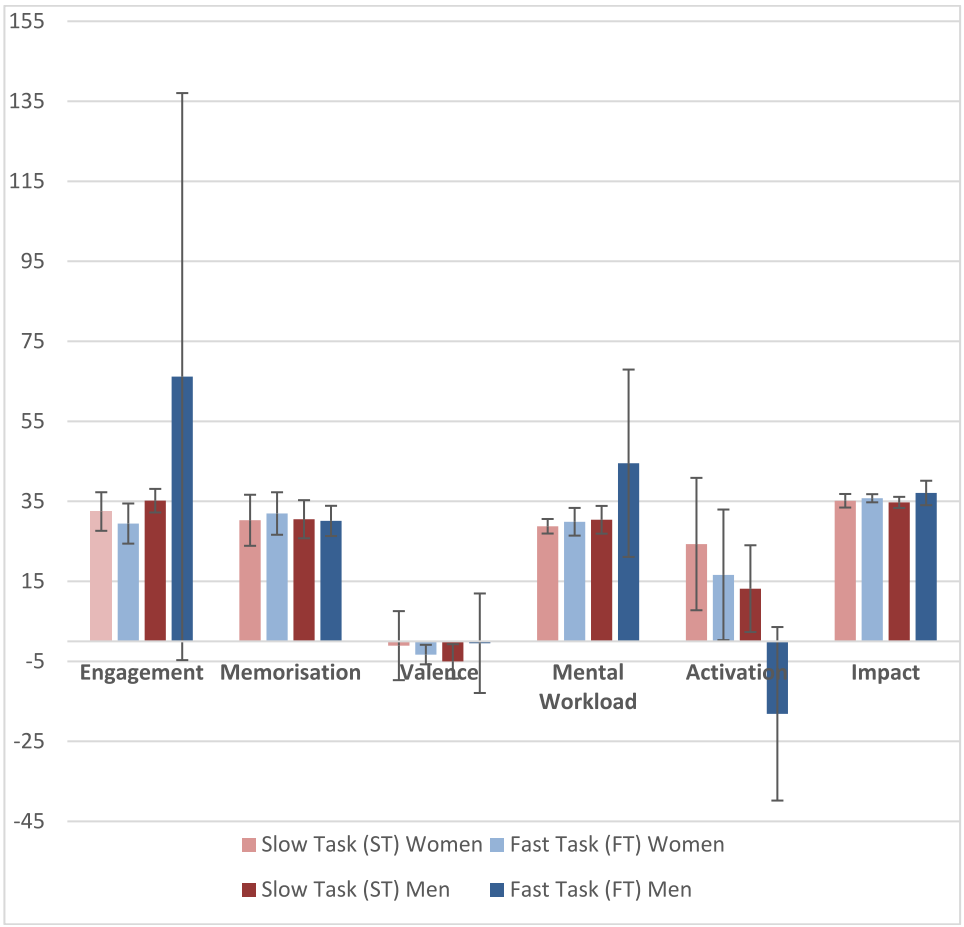


Fig. 10 Results of Physiological Indicators in the Slow-Fast (SF) procedure disaggregated by gender

Table 6 T-Test results to determine significant differences between tasks in the SF procedure disaggregated by gender, considering $\alpha = 0.05$. Statistically significant values are highlighted in bold

	Engagement	Memorisation	Valence	Mental workload	Activation	Impact
Women	0,4214	0,6981	0,6474	0,5934	0,5347	0,5540
Men	0,4396	0,8953	0,5294	0,3043	**0,0455**	0,2218

= 32.38; SD = 4.64). However, these differences are not considered statistically significant (Table 8).

- *Activation* In the case of women, a less negative Activation is observed in the ST (M = -13.05; SD = 32.48) compared to the FT (M = -23.09; SD = 42.27). For men, a higher Activation is observed in the ST (M = 11.64; SD = 34.78) compared to the FT (M = 3.50; SD = 7.21). However, these differences are not considered statistically significant (Table 8).
- *Impact* In the case of women, a similar Impact is observed in both the ST (M = 34.56; SD = 1.88) and the FT (M = 34.38; SD = 1.68). For men, a lower Impact is observed in the ST (M = 35.56; SD = 5.75) compared to the FT (M = 36.63; SD = 5.75). Nonetheless, these differences are not considered statistically significant (Table 8).

However, it is relevant to note that, according to the results in Table 6, there is only a significant difference between tasks in the indicators of Memorisation in men.

Below are the results of the physiological measurements without considering the processing; in other words, we have combined both procedures (Fig. 13):

- *Engagement* A lower level of Engagement was observed in participants in the ST (M = 34.92; SD = 6.26) compared to the FT (M = 40.33; SD = 40.37). However, this difference is not statistically significant (Table 9).
- *Memorisation* Participants were found to have a higher level of Memorisation in the FT (M = 31.58; SD = 4.78) compared to the ST (M = 27.07; SD = 6.34). This difference is considered statistically significant (Table 9).

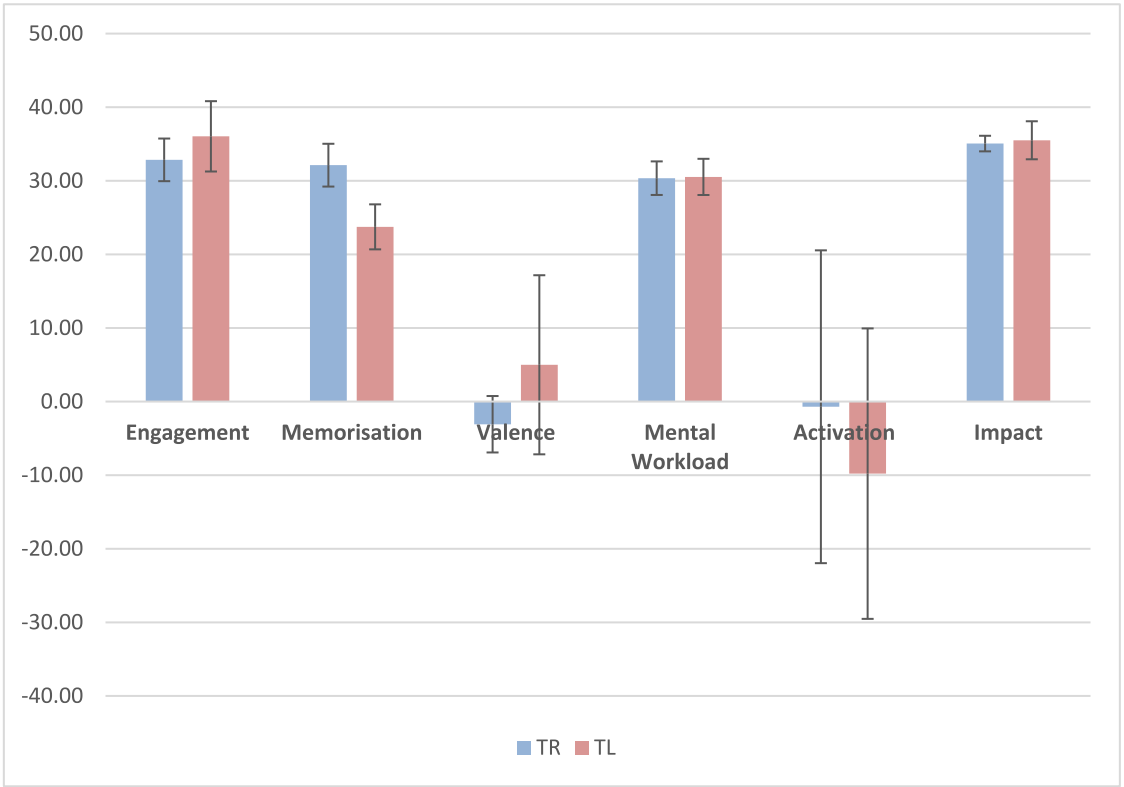


Fig. 11 Results of Physiological Indicators in the FS Procedure

Table 7 T-Test results to determine significant differences between tasks in physiological responses in the FS procedure, considering $\alpha = 0.05$. Statistically significant values are highlighted in bold

Engagement	Memorisation	Valence	Mental workload	Activation	Impact
0,2801	0,0011	0,2408	0,9208	0,5468	0,7587

- *Valence* Emotional Valence was recorded as more negative in the FT (M = -2.48; SD = 7.97) in contrast to the ST (M = 0.98; SD = 15.07). However, this difference is not statistically significant (Table 9).
- *Mental Workload* It was observed that Mental Workload was higher in the FT (M = 33.78; SD = 14.16) compared to the ST (M = 30.05; SD = 3.50). However, this difference is not statistically significant (Table 9).
- *Activation* It is noted that Activation has been more positive in the ST (M = 4.48; SD = 28,60) than in the FT (M = -0.72; SD = 30.29). However, this difference is not statistically significant (Table 9).
- *Impact* It is observed that the Impact has been similar in the FT (M = 35.74; SD = 2.24) and ST (M = 35.22; SD = 3.11). Indeed, this difference is not statistically significant (Table 9).

Following this, we present the results obtained by merging both procedures while also considering gender disaggregation (Fig. 14):

- *Engagement* In the case of women, a higher level of Engagement is observed in the ST (M = 33.74; SD = 7.46) than in the FT (M = 31.60; SD = 5.24). Contrary, in the case of men, a higher level of Engagement is observed in the FT (M = 49.06; SD = 56.97) than in the ST (M = 36.10; SD = 4.87). However, these differences are not considered statistically significant (Table 10).
- *Memorisation* In both men and women, a higher level of Memorisation is observed in the FT (Women: M = 31.15; SD = 4.71; Men: M = 32.00; SD = 5.06) compared to the ST (Women: M = 27.75; SD = 6.66; Men: M = 26.38; SD = 6.30). In women, the difference is not considered statistically significant, but it is for men (Table 10).
- *Valence* In both men and women, a higher emotional Valence is observed in the ST (Women: M = 3.42; SD

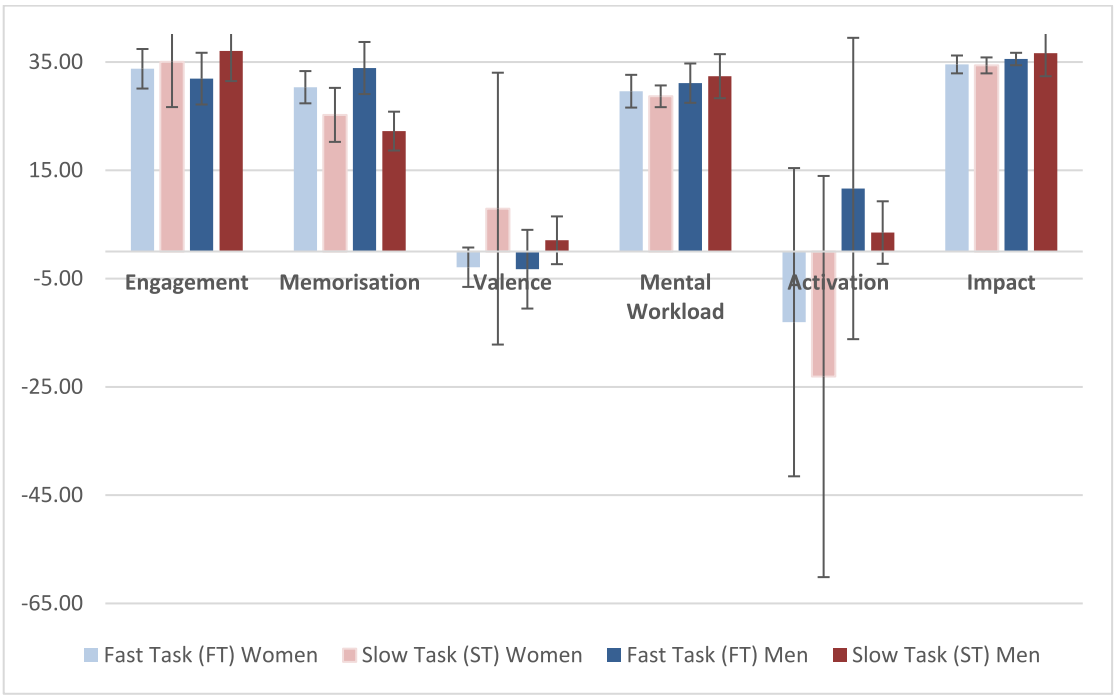


Fig. 12 Results of Physiological Indicators in the FS Procedure disaggregated by gender

Table 8 T-Test results to determine significant differences between tasks in the SF procedure disaggregated by gender, considering $\alpha = 0.05$. Statistically significant values are highlighted in bold

	Engagement	Memorisation	Valence	Mental workload	Activation	Impact
Women	0,7929	0,1313	0,4485	0,6275	0,6855	0,8774
Men	0,2100	**0,0060**	0,2605	0,6568	0,6334	0,7046

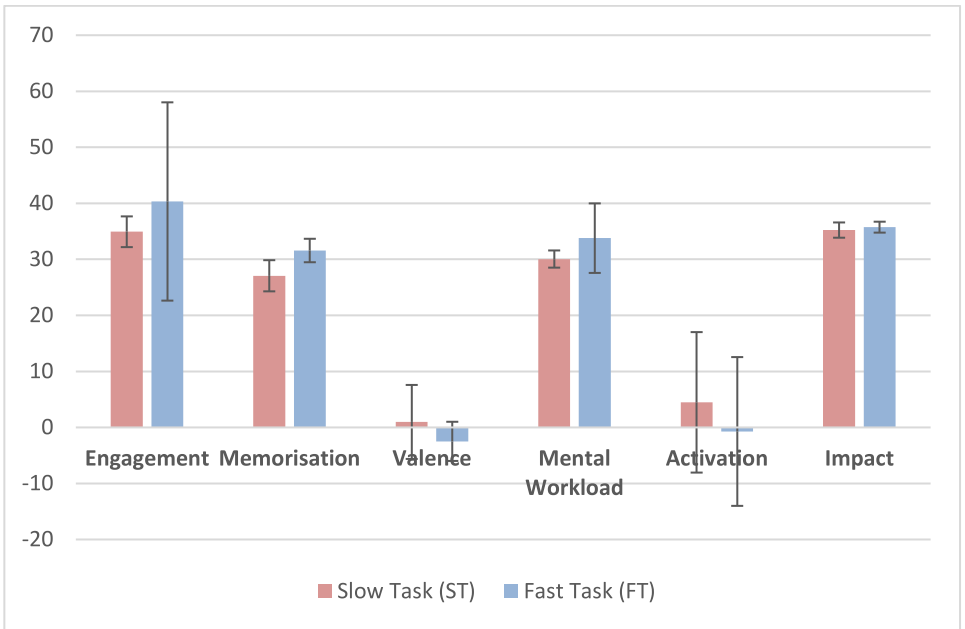


Fig. 13 Results of Physiological Indicators by merging both procedures

Table 9 T-Test results to determine significant differences between tasks in physiological responses merging both procedures, considering $\alpha = 0.05$. Statistically significant values are highlighted in bold

Engagement	Memorisation	Valence	Mental workload	Activation	Impact
0,5608	0,0157	0,3710	0,2656	0,5800	0,5452

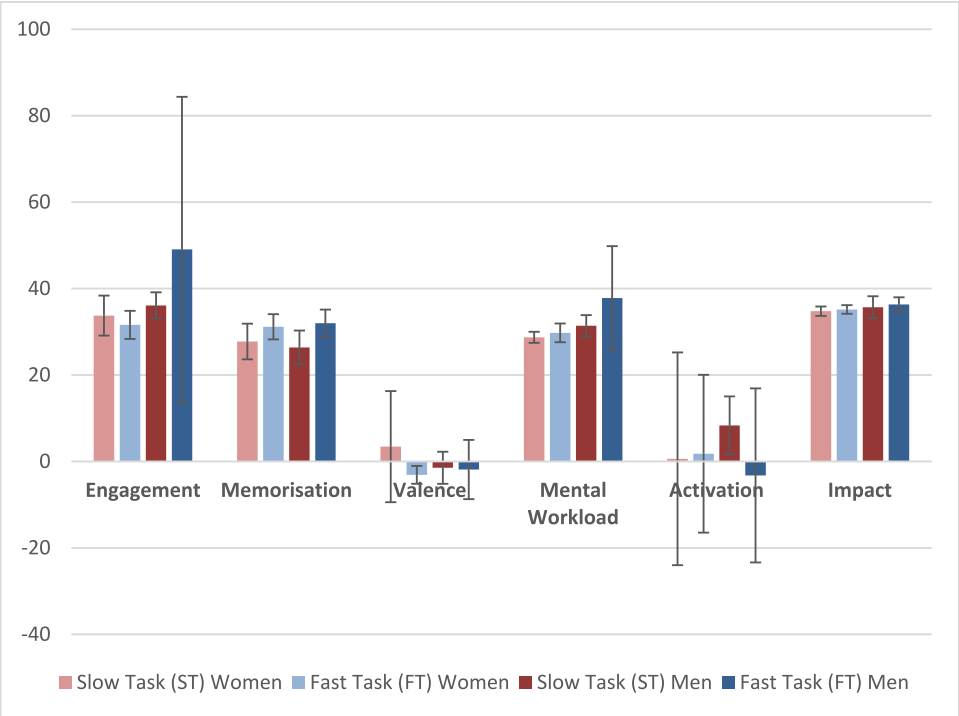


Fig. 14 Results of Physiological Indicators by merging both procedures and disaggregating by gender

= 20.74; Men: M = -1.46; SD = 5.99) compared to the FT (Women: M = -3.10; SD = 3.34; Men: M = -1.86; SD = 11.06). However, these differences are not considered statistically significant (Table 10).

- *Mental Workload* In both men and women, a higher Mental Workload is observed in FT (Women: M = 29.75; SD = 3.50; Men: M = 37.81; SD = 19.36) than in ST (Women: M = 28.72; SD = 2.06; Men: M = 31.38; SD = 3.99). However, these differences are not considered statistically significant (Table 10).
- *Activation* In the case of women, a higher Activation is observed in the FT (M = 1.79; SD = 29.46) compared to the ST (M = 0.61; SD = 39.71). Contrary, for men, a higher Activation is observed in the ST (M = 8.34; SD = 10.82) compared to the FT (M = -3.23; SD = 32.49). However, these differences are not considered statistically significant (Table 10).
- *Impact* In the case of women, a similar Impact is observed in the FT (M = 35.16; SD = 1.61) than in the ST (M = 34.76; SD = 1.61). For men, a lower Impact is observed in the ST (M = 35.68; SD = 4.10) compared to the FT (M = 36.32; SD = 2.67). Nonetheless, these differences are not considered statistically significant (Table 10).

4.3 Perceptual Indicators and Observation

In this section, a descriptive analysis of the perceptual indicators is conducted. For an overall view of the differences between the various tasks in each aspect evaluated in the questionnaire, Figs. 52 and 54 are presented, corresponding to the SF and FS procedures, respectively. Furthermore, for a more detailed understanding, specific values are detailed in Table 64.

Figure 15. displays the results obtained in the SF procedure. Additionally, as can be observed in Table 7, the results of the PQ are identified as a significant difference.

- *PQ* A lower level of PQ was observed in participants in the ST (M = 5.45; SD = 0.71) compared to the FT (M = 6.075; SD = 0.54). This difference is considered statistically significant (Table 11).

Table 10 T-Test results to determine significant differences between tasks in physiological responses merging both procedures and disaggregating by gender, considering $\alpha = 0.05$

	Engagement	Memorisation	Valence	Mental workload	Activation	Impact
Women	0,47	0,21	0,35	0,43	0,94	0,60
Men	0,49	0,04	0,92	0,33	0,31	0,68

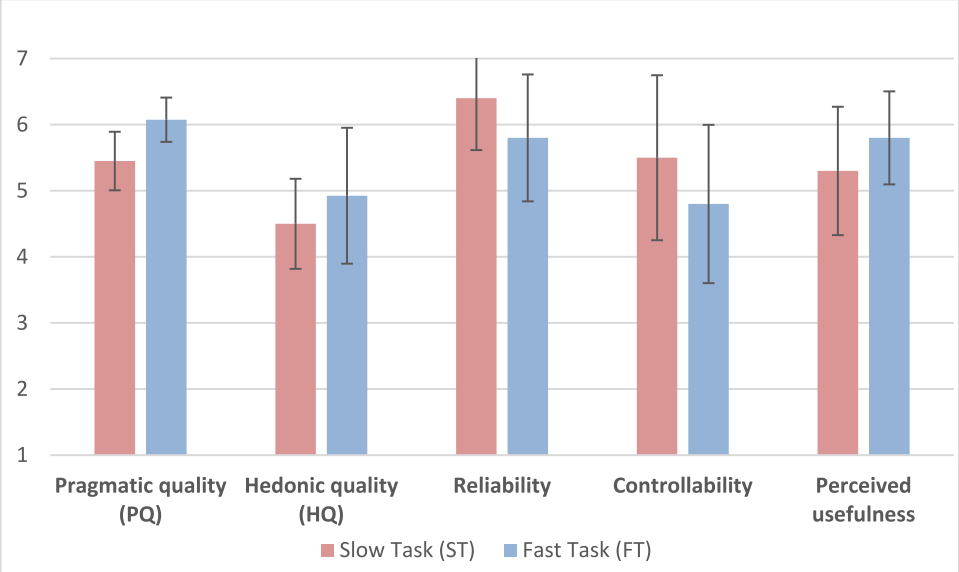


Fig. 15 Comparative results of the questionnaire for the SF procedure

- *HQ* A lower level of HQ was observed in participants in the ST (M = 4.5; SD = 1.09) compared to the FT (M = 4.925; SD = 1.66). However, this difference is not considered statistically significant (Table 11).
- *Reliability* A higher level of Reliability was observed in the ST (M = 6.4; SD = 1.26) compared to the FT (M = 5.8; SD = 1.55). However, this difference is not considered statistically significant (Table 11).
- *Controllability* A higher level of Controllability was observed in the ST (M = 5.5; SD = 2.01) compared to the FT (M = 4.8; SD = 1.93). However, this difference is not considered statistically significant (Table 11).
- *Perceived Usefulness* A lower level of Perceived Usefulness was observed in the ST (M = 5.3; SD = 1.57) compared to the FT (M = 5.8; SD = 1.35). However, this difference is not considered statistically significant (Table 11).

Furthermore, Fig. 16 presents the results of the SF procedure disaggregated by gender.

- *PQ* Both men and women show lower ratings in PQ for the ST (Women: M = 5.3; SD = 0.99; Men: M = 5.6; SD = 0.33) compared to the FT (Women: M = 6.1; SD = 0.28; Men: M = 6.05; SD = 0.76). However, these differences are not considered statistically significant (Table 12).
- *HQ* Both men and women exhibit lower ratings in HQ for the ST (Women: M = 4.4; SD = 0.96; Men: M = 4.6; SD = 1.33) compared to the FT (Women: M = 5.15; SD = 1.22; Men: M = 4.7; SD = 2.14). However, these differences are not considered statistically significant (Table 12).
- *Reliability* In the case of women, a higher rating is observed for Reliability in the ST (M = 7; SD = 0) compared to the FT (M = 5.6; SD = 2.07). Conversely, in men, a lower rating for Reliability is observed in the ST (M = 5.8; SD = 1.64) compared to the FT (M = 6; SD = 1). However, these differences are not considered statistically significant (Table 12).
- *Controllability* Both men and women show a higher level of Controllability for the ST (Women: M = 5.6; SD = 2.6; Men: M = 5.4; SD = 1.51) compared to the FT (Women: M = 4.4; SD = 2.3; Men: M = 5.2; SD = 1.64). However, this difference is not considered statistically significant (Table 12).
- *Perceived Usefulness* Both men and women exhibit a lower rating in the Perceived Usefulness for the ST (Women: M = 4.6; SD = 1.94; Men: M = 6; SD = 0.7) compared to the FT (Women: M = 5; SD = 1; Men: M = 6.6; SD = 0.55). However, these differences are not considered statistically significant (Table 12).

In addition, Fig. 17 presents the results obtained in the FS procedure, which are described below:

Table 11 T-Test results to determine significant differences between tasks in perceptual responses in the FS procedure, considering $\alpha = 0.05$. Statistically significant values are highlighted in bold

PQ	HQ	Reliability	Controllability	Perceived usefulness
0,0417	0,5093	0,3558	0,4380	0,4256

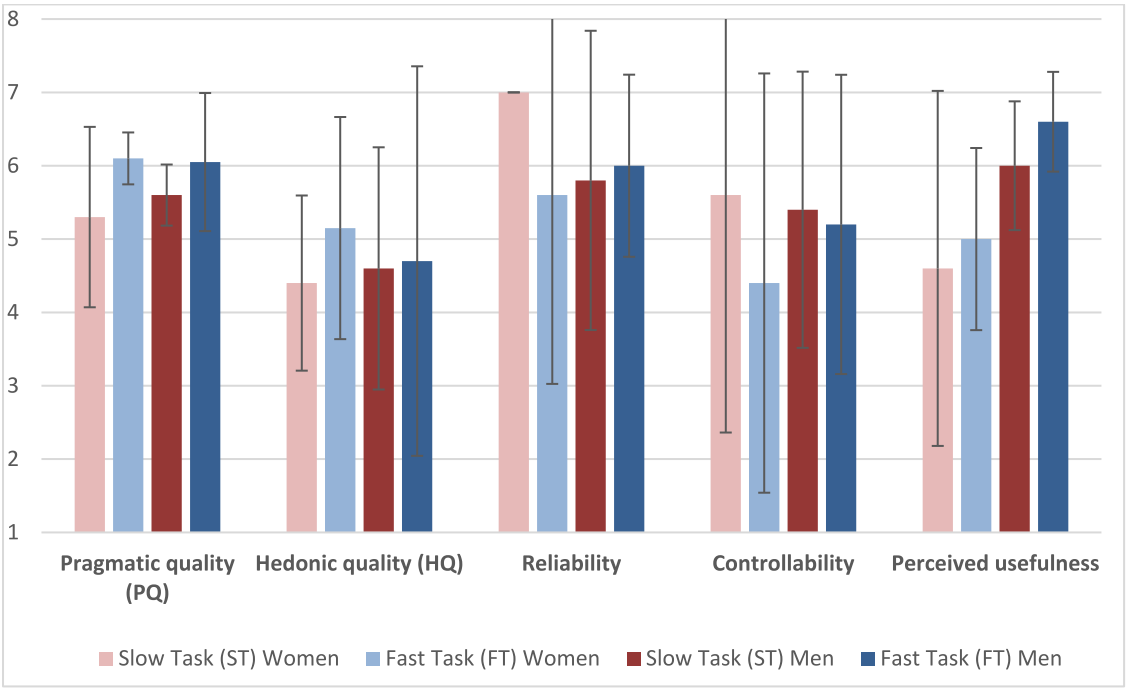


Fig. 16 Results of perceptual indicators in the SF Procedure disaggregated by gender

Table 12 T-Test results to determine significant differences in the perceptual indicators between tasks in the SF procedure disaggregated by gender, considering $\alpha = 0.05$. Statistically significant values are highlighted in bold

	PQ	HQ	Reliability	Controllability	Perceived usefulness
Women	0,1132	0,8616	0,2056	0,7901	0,9860
Men	0,0860	0,4902	0,2825	0,2062	0,0753

- *PQ* A higher level of PQ was observed in the participants in the FT (M = 5.97; SD = 0.86) compared to the ST (M = 4.35; SD = 0.65). This difference is considered statistically significant (Table 13).
- *HQ* A higher level of HQ was observed in the participants in the FT (M = 4.72; SD = 1.72) compared to the ST (M = 3.4; SD = 1.28). However, this difference is not considered statistically significant (Table 13).
- *Reliability* A lower level of Reliability was observed in the FT (M = 5.9; SD = 1.45) compared to the ST (M = 5.8; SD = 0.79). However, this difference is not considered statistically significant (Table 13).
- *Controllability* A higher level of Controllability was observed in the ST (M = 6.1; SD = 0.74) compared to the FT (M = 6.4; SD = 0.84). However, this difference is not considered statistically significant (Table 13).
- *Perceived Usefulness* A higher level of Perceived Usefulness was observed in the FT (M = 5.6; SD = 1.78) compared to the ST (M = 3.3; SD = 1.7). This difference is statistically significant (Table 13).

Furthermore, Fig. 18 presents the results of the FS procedure broken down by gender, which are described as follows:

- *PQ* Both in men and women, a higher rating is observed in PQ for the FT (Women: M = 6.25; SD = 0.35; Men: M = 5.7; SD = 1.16) compared to the ST (Women: M = 4.2; SD = 0.74; Men: M = 4.5; SD = 0.59). In the

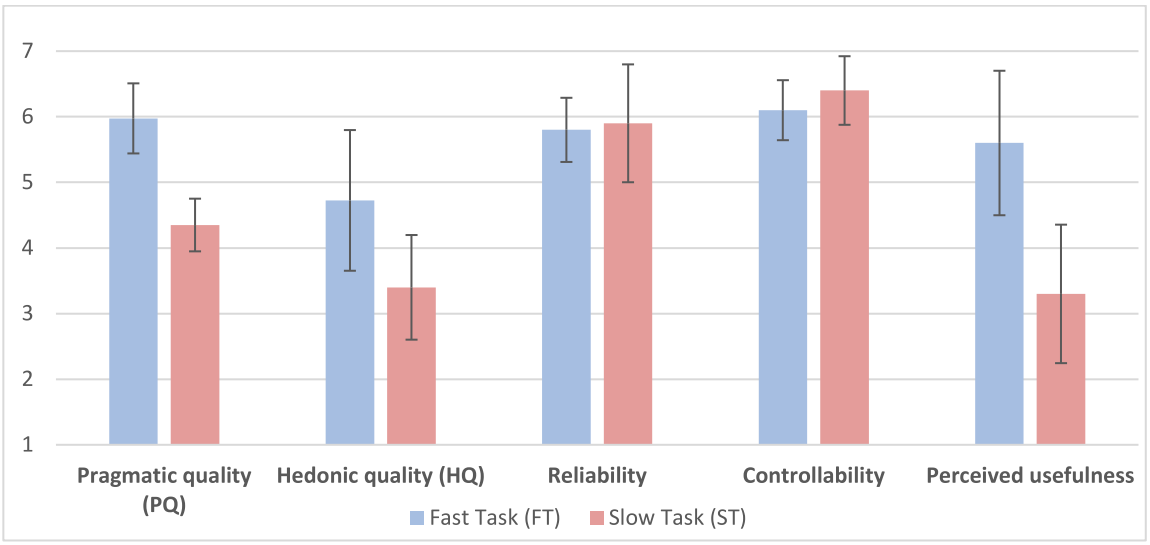
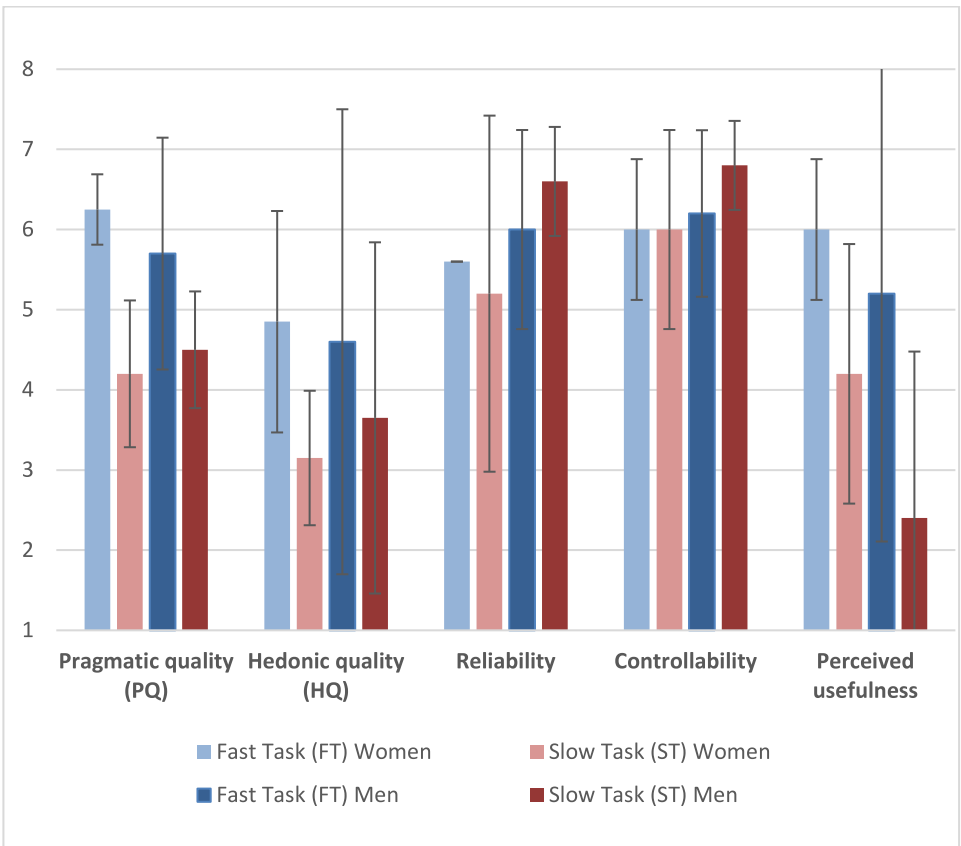


Fig. 17 Results of perceptual indicators in the FS Procedure

Table 13 T-Test results to determine significant differences in the perceptual indicators between tasks in the FS procedure, considering $\alpha = 0.05$. Statistically significant values are highlighted in bold

PQ	HQ	Reliability	Controllability	Perceived usefulness
0,0002	0,0691	0,8508	0,4085	**0,0085**

Fig. 18 Results of perceptual indicators in the FS Procedure disaggregated by gender



case of women, the difference is considered statistically significant, but not for men (Table 14).

- *HQ* Both in men and women, a lower rating is observed in HQ for the ST (Women: M = 4.85; SD = 1.11; Men: M = 4.6; SD = 2.34) compared to the FT (Women: M = 3.15; SD = 0.68; Men: M = 3.65; SD = 1.76). In the case of women, the difference is considered statistically significant, but not for men (Table 14).
- *Reliability* In the case of women, a higher rating is observed in Reliability for the FT (M = 5.6; SD = 0.55) compared to the ST (M = 5.2; SD = 1.79). Conversely, in men, a lower rating of Reliability is observed in the ST (M = 6.6; SD = 0.55) compared to the FT (M = 6; SD = 1). However, these differences are not considered statistically significant (Table 14).
- *Controllability* In women, the same rating is observed in Controllability for the FT (M = 6; SD = 0.7) as in the ST (M = 6; SD = 1). In men, a lower perception of Controllability is observed in the FT (M = 6.2; SD = 0.84) compared to the ST (M = 6.8; SD = 0.45). These differences are not considered statistically significant (Table 14).
- *Perceived Usefulness* Both in men and women, a lower rating is observed in the Perceived Usefulness of the ST (Women: M = 4.2; SD = 1.3; Men: M = 2.4; SD = 1.6) compared to the FT (Women: M = 6; SD = 0.7; Men: M = 5.2; SD = 2.5). In the case of women, the difference is considered statistically significant, although not for men (Table 14).

Below are the results of the perceptual without considering the procedure; in other words, we have merged the results of both procedures (Fig. 19):

- *PQ* A higher level of PQ was observed in the participants in the FT (M = 6.03; SD = 0.70) compared to the ST (M = 4.90; SD = 0.87). This difference is considered statistically significant (Table 15).
- *HQ* A higher level of HQ was observed in the participants in the FT (M = 4.83; SD = 1.65) compared to the ST (M = 3.95; SD = 1.29). However, this difference is not considered statistically significant (Table 15).
- *Reliability* A lower level of Reliability was observed in the FT (M = 5.80; SD = 1.20) compared to the ST (M = 6.15; SD = 1.35). However, this difference is not considered statistically significant (Table 15).
- *Controllability* A higher level of Controllability was observed in the ST (M = 5.95; SD = 1.57) compared to the FT (M = 5.45; SD = 1.57). However, this difference is not considered statistically significant (Table 15).
- *Perceived Usefulness* A higher level of Perceived Usefulness was observed in the FT (M = 5.70; SD = 1.45) compared to the ST (M = 4.30; SD = 1.89). This difference is statistically significant (Table 15).

Following this, we present the results obtained by merging both procedures while also considering gender disaggregation (Fig. 20):

- *PQ* Both in men and women, a higher rating is observed in PQ for the FT (Women: M = 6.175; SD = 0.31; Men: M = 5.875; SD = 0.94) compared to the ST (Women: M = 4.75; SD = 1.00; Men: M = 5.05; SD = 0.73). These differences are considered statistically significant (Table 16).
- *HQ* Both in men and women, a lower rating is observed in HQ for the ST (Women: M = 3.775; SD = 1.02; Men: M = 4.125; SD = 1.55) compared to the FT (Women: M = 5,00; SD = 1.11; Men: M = 4.65; SD = 2.11). In the case of women, the difference is considered statistically significant, but not for men (Table 16).
- *Reliability* Both women and men, rated higher Reliability for the ST (Women: M = 6.1; SD = 1.52; Men: M = 6.2; SD = 1.23) compared to the FT (Women: M = 5.6; SD = 1.43; Men: M = 6.00; SD = 0.94). However, these differences are not considered statistically significant (Table 16).
- *Controllability* Both women and men, rated higher Controllability for the ST (Women: M = 5.8; SD = 1.87; Men: M = 6.1; SD = 1.29) compared to the FT (Women: M = 5.2; SD = 1.871; Men: M = 5.7; SD = 1.31). However, these differences are not considered statistically significant (Table 16).
- *Perceived Usefulness* Both in men and women, a lower rating is observed in the Perceived Usefulness of the ST (Women: M = 4.4; SD = 1.58; Men: M = 4.2; SD = 2.25) compared to the FT (Women: M = 5.5; SD = 0.97; Men: M = 5.9; SD = 1.85). In the case of women, the difference is considered statistically significant, although not for men (Table 16).

In addition to the questionnaire, the participants also shared some comments, and with the observation, these insights can be summarised:

> *"I was impatient and didn't wait for the robot to finish its motion before proceeding with the next piece. The robot was moving at a slow pace during the initial stages of the task, and I was able to complete the task quicker on my own. If I were a worker in a real-life scenario, I would have instructed the robot to move faster to save time. Waiting for the robot to complete its motion would slow me down."*
>
> *"As I interacted with the robot, I noticed my confidence in its behaviour increased over time. This made me feel more comfortable in taking actions such as reaching in before the robot was finished or sending commands faster. The more time I spent working with the robot, the*

Table 14 T-Test results to determine significant differences in the perceptual indicators between tasks in the FS procedure disaggregated by gender, considering $\alpha = 0.05$. Statistically significant values are highlighted in bold

	PQ	HQ	Reliability	Controllability	Perceived usefulness
Women	**0,0016**	**0,0239**	0,6538	1,0000	**0,0340**
Men	0,0860	0,4902	0,2825	0,2062	0,0753

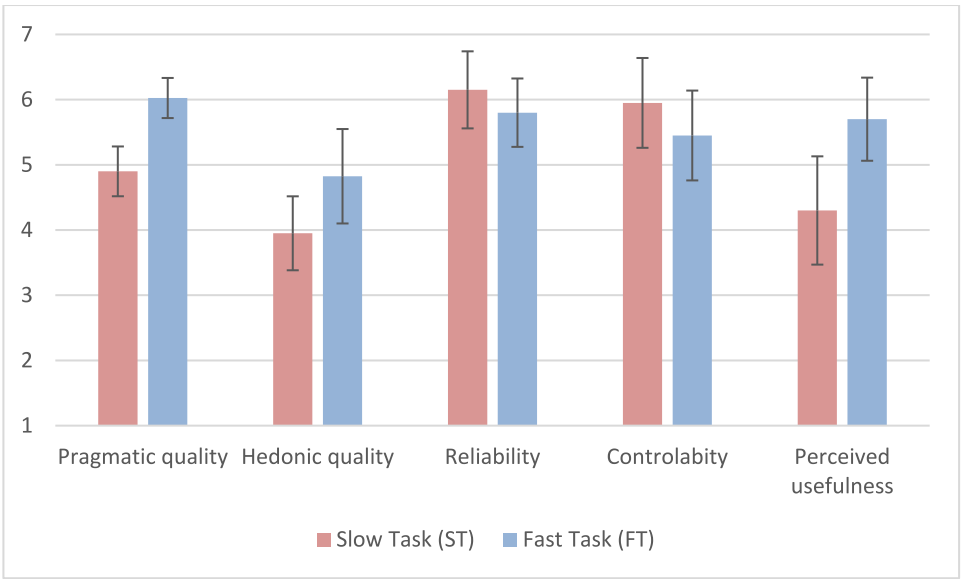


Fig. 19 Results of Perceptual Indicators by merging both procedures

Table 15 T-Test results to determine significant differences between tasks in perceptual responses merging both procedures, considering $\alpha = 0.05$. Statistically significant values are highlighted in bold

Pragmatic quality	Hedonic quality	Reliability	Controlabity	Perceived usefulness
0,00007	0,0705	0,3908	0,3209	**0,0128**

more I was able to anticipate its movements and understand its behaviour. This confidence boosts ultimately led to a better experience and improved performance in our collaboration."

"I was frustrated with the slow speed of the robot at first. But when it picked up speed, I felt more confident in its ability to complete the task accurately. The faster pace made the robot seem more intelligent and in control, while the slow speed made it appear unsure and less competent."

"Initially, I was wary of activating the robot as I wasn't sure of its movements. I took precautions and kept my distance, not wanting to risk getting hit. But as I became more familiar with its behaviour, I grew more confident and started to activate it without hesitation. The more I worked with it, the more natural the interaction felt."

Based on the comments provided, it appears that participants' experiences with the robot were influenced by several factors, including kinematic variables, trust, learnability, and interaction time. Some participants felt frustrated with the slow movements of the robot, finding it to be less precise and intelligent. On the other hand, when the robot's speed increased, the participants felt more confident in its ability to complete tasks accurately. Participants also reported that their confidence in the robot's behaviour increased with interaction time, which led to a more comfortable and natural interaction. As a result, participants were more willing to take actions such as reaching in before the robot was finished or sending commands faster. These findings suggest that the kinematic variables of the robot, as well as the participant's trust and interaction time, play a role in shaping their experience with the robot.

5 Discussion

In this section, the results of the study on the effect of the kinematic variables (speed and acceleration) on human factors

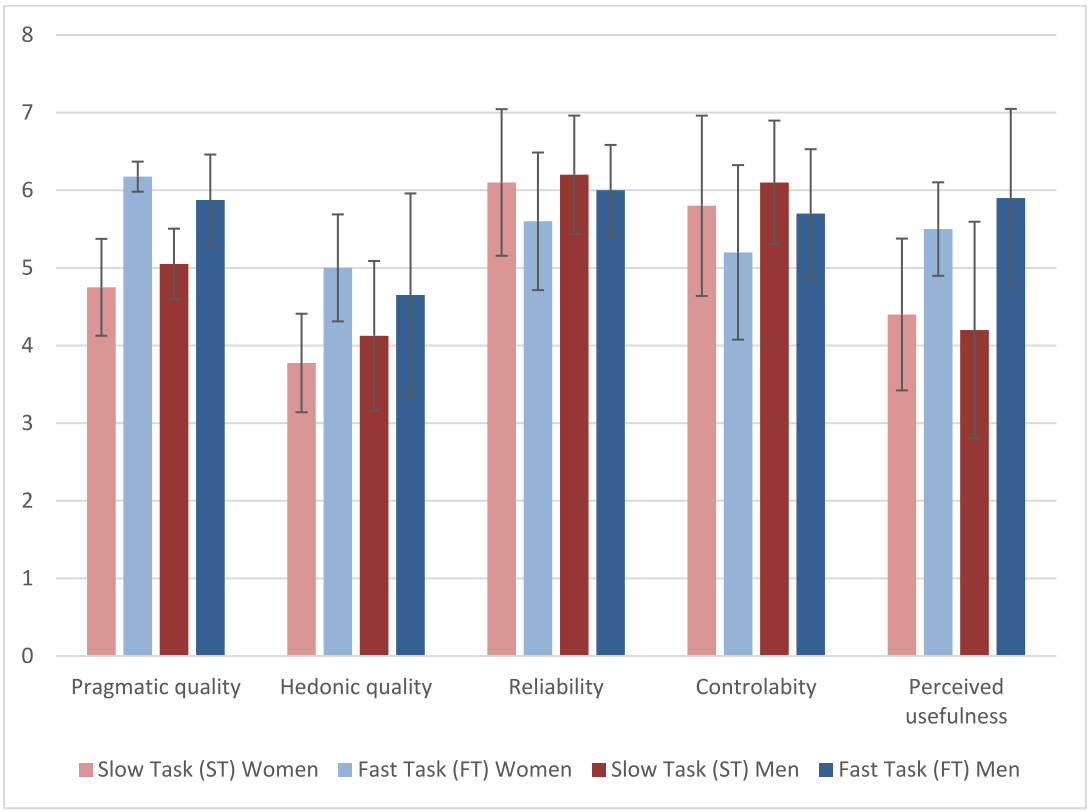


Fig. 20 Results of Perceptual Indicators by merging both procedures and disaggregating by gender

Table 16 T-Test results to determine significant differences between tasks in perceptual responses merging both procedures and disaggregating by gender, considering $\alpha = 0.05$. Statistically significant values are highlighted in bold

	Pragmatic quality	Hedonic quality	Reliability	Controlability	Perceived usefulnes
Women	**0,0014**	**0,0196**	0,4591	0,4762	0,0801
Men	**0,0436**	0,5355	0,6882	0,5042	0,0823

are interpreted and discussed. The study aimed to investigate how the robot's speed and acceleration changes during a task affect human responses and emotions. We find results in 1) KPIs, 2) Physiological indicators, and 3) Perceptual indicators and observation. The discussion will provide a comprehensive overview of the results, providing insights into the effects of kinematic variables on HRI and highlighting areas that require further investigation. Additionally, the limitations of the study will also be discussed.

5.1 Results for RQ 1: In Industrial HRC Assembly Task, What is the Impact of Different Robot Kinematic Variables (i.e., Movement Speed and Acceleration) on Human Factors?

Table 17 provides an overview of the significant differences identified throughout the case study.

Regarding performance, it is evident that the FT leads to shorter task execution times. In terms of errors, it is observed that due to task learnability, the task performed second results in fewer errors.

As for performance indicators, a significant difference is only observed in Memorisation within the Fast-Slow (FS) procedure, with FT leading to higher levels of Memorisation. No significant differences are observed in Engagement, Valence, Mental Workload, Activation, and Impact.

In the realm of perceptual indicators, a significant difference is noted in PQ in both procedures, indicating that higher kinematic variables result in a better assessment of PQ. However, no significant differences are observed in HQ, Reliability, and Controllability in either procedure. In the case of Perceived Usefulness in the FS procedure, a significant difference is observed, with higher kinematic variables correlating with a higher assessment of Perceived Usefulness.

Table 17 Summary of the differences identified by the T-test in the indicators collected (performance, physio-logical and perceptual) in both procedures

		Slow-fast (SF) procedure	Fast-slow (FS) procedure	Merging procedures
Performance	Time	Higher kinematic variables ↑ resulted in shorter task execution time✓	Higher kinematic variables ↑ resulted in shorter task execution time✓	Higher kinematic variables ↑ resulted in shorter task execution time✓
	Errors	Higher kinematic variables↑ resulted in fewer errors✓	Lower kinematic variables↓ resulted in fewer errors✓	No
Physiological indicators	Engagement	No	No	No
	Memorisation	No	Higher kinematic variables↑ resulted in a higher level of Memorisation ×	No
	Valence	No	No	No
	Mental Workload	No	No	No
	Activation	No	No	No
	Impact	No	No	No
Perceptual indicators	Pragmatic quality (PQ)	Higher kinematic variables↑ resulted in higher PQ✓	Higher kinematic variables↑ resulted in higher PQ✓	Higher kinematic variables↑ resulted in higher PQ✓
	Hedonic quality (HQ)	No	No	No
	Reliability	No	No	No
	Controllability	No	No	No
	Perceived Usefulness	No	Higher kinematic variables↑ resulted in higher Perceived Usefulness✓	Higher kinematic variables↑ resulted in higher Perceived Usefulness✓

In this context, H1 stands disproven, as no substantial impact on Engagement, Memorisation, and Mental Workload.

Similarly, with regards to H2, it is also refuted, contradicting initial assumptions. Higher kinematic variables yield higher ratings in terms of PQ and Perceived Usefulness. Conversely, in the case of Valence, Activation, Impact, HQ, Reliability, and Controllability, no statistically significant disparities are observed.

5.2 Results for RQ 2: Do these Kinematic Variables have a Different Impact on Men and Women?

We have integrated gender perspective in our research to avoid biases, where sometimes the realities, experiences, and expectations of a group of people (considering men as a reference) are constructed as the norm, resulting in partial and non-universal results. Our results confirm differences between genders when interacting with a robot in terms of task completion time, Valence, Memorisation, Mental Workload, Engagement, and Perception. Throughout the research, T-tests were conducted with data disaggregated by gender. The outcomes of these tests are summarized in Table 18.

Concerning performance indicators, for both men and women in both procedures, the FT consistently resulted in shorter task execution times. Regarding errors, it was observed that, due to task learnability, the task performed second resulted in fewer errors.

Regarding physiological indicators, no significant differences were observed in Engagement for either gender in any of the procedures. In terms of Memorisation, a significant difference was observed only in the case of men in the FS procedure, where FT resulted in a higher level of Memorisation. However, no significant differences were observed in participants of the SF procedure or in women in the FS procedure. Emotional Valence and Mental Workload also showed no significant differences among any of the genders or in any of the procedures. While examining Activation in a general context revealed no significant differences, it is worth noting that higher kinematic variables led to lower Activation values in men in the SF procedure. However, no such differences were observed in women in the SF procedure or in participants in the FS procedure. Lastly, no significant differences were observed in Impact among any of the genders or in any of the procedures.

Regarding perceptual indicators, a significant difference in PQ was observed only in women in the FS procedure.

Table 18 Summary of the differences identified by the T-test in the indicators collected (performance, physio-logical and perceptual) in both procedures disaggregated by gender

		Slow-fast (SF) Procedure		Fast-Slow (FS) Procedure		Merging both procedures	
		Women	Men	Women	Men	Women	Men
Performance	Time	Higher kinematic variables ↑ resulted in shorter task execution time✓	Higher kinematic variables ↑ resulted in shorter task execution time✓	Higher kinematic variables↑ resulted in shorter task execution time✓	Higher kinematic variables↑ resulted in shorter task execution time✓	Higher kinematic variables↑ resulted in shorter task execution time✓	Higher kinematic variables↑ resulted in shorter task execution time✓
	Errors	Higher kinematic variables↑ resulted in fewer errors✓	Higher kinematic variables↑ resulted in fewer errors✓	Lower kinematic variables↓ resulted in fewer errors✓	Lower kinematic variables↓ resulted in fewer errors✓	No	No
Physiological indicators	Engagement	No	No	No	No	No	No
	Memorisation	No	No	No	Higher kinematic variables↑ resulted in a higher level of Memorisation ×	No	No
	Valence	No	No	No	No	No	No
	Mental Workload	No	No	No	No	No	No
	Activation	No	Higher kinematic variables↑ resulted in lower Activation ×	No	No	No	No
	Impact	No	No	No	No	No	No
Perceptual indicators	Pragmatic quality (PQ)	No	No	Higher kinematic variables↑ resulted in higher PQ✓	No	Higher kinematic variables↑ resulted in higher PQ✓	Higher kinematic variables↑ resulted in higher PQ
	Hedonic quality (HQ)	No	No	Higher kinematic variables↑ resulted in higher HQ✓	No	Higher kinematic variables↑ resulted in higher HQ✓	No
	Reliability	No	No	No	No	No	No
	Controllability	No	No	No	No	No	No
	Perceived Usefulness	No	No	Higher kinematic variables↑ resulted in higher Perceived Useful-ness✓	No	No	No

This result is intriguing because when not disaggregating the results by gender in both procedures, the T-test result was statistically significant. As for HQ, a significant difference was also observed only in women in the FS procedure, where FT resulted in higher scores in HQ. No statistically significant differences were found in Reliability and Controllability. In terms of Perceived Usefulness, significant differences were observed only in women in the FS procedure, where FT resulted in higher scores in Perceived Usefulness.

During the analysis of research results, separate T-tests were conducted for each gender, without a direct intergender comparison. In Table 19, we present the T-test results, which compare outcomes between both genders in each task of every procedure.

Notably, a significant gender-based difference emerged in the time needed to complete the FT within the FS procedure. In this instance, women averaged 2 min and 26 s, while men finished the task in 2 min and 10 s.

Furthermore, another distinctive gender difference surfaced in Activation levels during the FT within the SF procedure. Women displayed an average Activation level of 16.63, while men averaged -18.11.

Significantly, divergent gender perspectives were evident in terms of Perceived Usefulness during the FT in the SF procedure. Women rated it at an average of 5, while men scored an average of 6.6.

Finally, another noteworthy discrepancy appeared when merging both procedures in relation to errors during the ST, with women committing fewer errors.

5.3 The Influence of Kinematic Variables on UX

During the execution of the tasks, the participants' verbal and non-verbal reactions were observed. In addition, the comments made by the participants after completing the tasks showed that the speed and acceleration of the robot influenced human's confidence. These results are based on the perceptual values and what has been observed and analysed from the reactions.

The kinematic variables exert influence in such a manner that an excessively low robot speed instils mistrust in individuals, giving rise to thoughts and doubts regarding the robot's precision and "intelligence". However, results obtained from the Reliability and Controllability items suggest that higher kinematic variables lead to a suboptimal experience, albeit not necessarily translating into reduced perceived trust and safety. In this regard, the impact of kinematic variables on perceived trust and safety occasionally appears contradictory. While excessively low speed triggers distrust in the robot's accuracy and intelligence, results from the Reliability and Controllability items suggest that higher kinematic variables result in a poorer experience. This inconsistency underscores the need for ongoing research to comprehend the optimal balance. According to Beschi et al. [101], an operator's emotions are more affected by the duration of a trajectory than by its speed and acceleration.

Conversely, Rubagotti et al. [22] found that the lower the robot's speed, the safer it is perceived. Thus, high speed and acceleration can decrease UX because it can lead to a lack of control and predictability. Too-high kinematic variables hinder people from anticipating robot movement and reacting accordingly. This leads to feeling unsafe and lacking trust, which can negatively impact UX. Additionally, high speed and acceleration can lead to safety concerns. A too-high speed would also generate distrust in humans, bringing up emotions such as anxiety and risk cognition, thus considerably reducing perceived safety and trust [17].

To visually represent this intricate relationship between the robot's kinematic variables and the perceived safety and Reliability by users, the concept of the "HRI Speed Bell" is introduced (Fig. 21). This graphical representation adopts the form of a Gaussian bell curve, where the x-axis reflects the robot's kinematic variables, while the y-axis denotes the UX. The conception of this bell curve is based on the one hand on the research by Rubagotti et al. [22], who indicated that both excessive speed and acceleration of the robot were perceived as unsafe and unreliable by users. We interpreted it as a Gaussian bell because we believe that it better captures human behaviours in response to varying kinematic variables. On the other hand, the results of the present case study suggest that excessively low speed and acceleration also evoke distrust in participants. Therefore, this Gaussian bell curve representation underscores that both excessively low speed and acceleration, as well as excessively high kinematic configurations, lead to a decrease in perceived safety and Reliability.

The choice of representing this relationship using a Gaussian bell curve is justified due to its suitability in describing natural and social phenomena characterized by statistical distributions. The Gaussian bell curve is a continuous representation that captures variability in participant's preferences and perceptions regarding the robot's kinematic variables. This makes it a more appropriate choice than a stepped or pyramid-shaped representation, as the latter would imply abrupt transitions in perception as the variables change. Given that relationships between variables in the context of HRI are often complex and nonlinear, the Gaussian bell curve offers a nuanced representation illustrating how different levels of kinematic variables can gradually influence users' perceptions of safety and Reliability.

It is pertinent to highlight that individual preference variability is a fundamental aspect. What one user considers "safe and reliable" in terms of speed and acceleration may significantly differ from another user's perception. The "HRI Speed

Table 19 T-test results to determine if there are significant differences between genders in each task of each procedure, considering $\alpha = 0.05$. Significant differences are highlighted in bold

		Slow-Fast (SF) Procedure		Fast-Slow (FS) Procedure		Merging procedures	
		Slow task (ST)	Fast task (FT)	Fast task (FT)	Slow task (ST)	Slow task (ST)	Fast task (FT)
Performance	Time	0.3143	0.8182	**0.0281**	0.2245	0.1662	0.1208
	Errors	0.3749	1.0000	0.1635	0.0651	**0.0394**	0.2188
Physiological indicators	Engagement	0.3804	0.3677	0.5706	0.7069	0.4160	0.3593
	Memorisation	0.9494	0.5951	0.2612	0.3699	0.6430	0.7029
	Valence	0.4549	0.6817	0.9324	0.6750	0.4894	0.7416
	Mental workload	0.4487	0.2898	0.5547	0.1612	0.0947	0.2255
	Activation	0.3064	**0.0386**	0.2795	0.2340	0.5656	0.7216
	Impact	0.7241	0.4601	0.3868	0.4409	0.5249	0.2583
Perceptual indicators	Pragmatic quality (PQ)	0.5500	0.8955	0.3611	0.4977	0.4572	0.3610
	Hedonic quality (HQ)	0.7928	0.6962	0.8364	0.5790	0.5608	0.6502
	Reliability	0.1778	0.7116	0.4617	0.1583	0.8736	0.4712
	Controllability	0.8867	0.5465	0.6941	0.1577	0.6820	0.4926
	Perceived usefulness	0.1911	**0.0193**	0.5225	0.0965	0.8209	0.5554

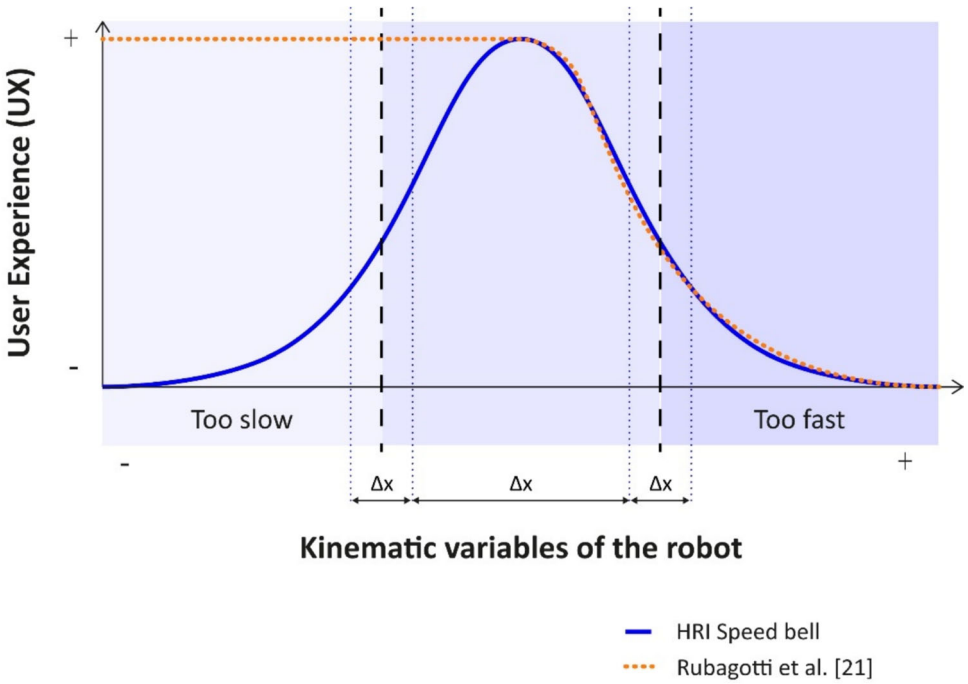


Fig. 21 "HRI speed bell" concept

Bell" reflects this diversity of individual perspectives regarding kinematic variables. Each incremental Δx on the graph can represent a unique perception in each user. This concept underscores the inherent complexity in understanding and adapting the robot's kinematic variables to the individual preferences and requirements of users.

Furthermore, this graphical representation suggests that finding the optimal balance in the robot's kinematic variables constitutes a complex and highly individualized challenge. Future research could focus on developing more precise and efficient methods to determine these optimal configurations for each user. This might involve collecting data on individual preferences and dynamically adapting the robot's kinematic variables during interaction. This personalized approach has significant implications, as it could pave the way for the

development of affective robotics, where the robot's kinematic variables are personalized to each user's preferences and needs.

5.4 Learnability Paradox

In both the SF and FS procedures, participants generally performed better when the robot moved at a faster speed and higher acceleration. However, the order in which tasks were performed appeared to influence participants' expectations of the robot and learnability. In the SF procedure, participants improved their experience by first performing ST and then FT, which led to a higher-rated Perceived Usefulness of FT. In contrast, the participants in the FS procedure had a worse experience because they first performed FT, which led to higher expectations. When they then performed ST, the experience worsened considerably. The order in which tasks were performed in the SF and FS procedures had a significant impact on the performance. The results showed that when participants first performed the ST and then the FT, they rated the Perceived Usefulness of FT higher, indicating a better experience and improved learnability. On the other hand, when participants first performed the FT and then the ST, their expectations were higher, leading to a worse experience and decreased learnability. This suggests that the order in which tasks were performed greatly influenced the learnability of the participants and their ability to effectively use the robot.

The results also suggest that the order in which the tasks were performed may have influenced Mental Workload and Engagement. In the FS procedure, participants had a higher Mental Workload in FT and a lower Mental Workload in ST. Similarly, participants in the FS procedure had a higher Engagement in FT than in ST.

The results of the study suggest that there is a trade-off between learnability and experience. As participants spend more time with the robot, their performance improves, and learnability increases. However, the study also reveals that longer interaction time leads to a decline in participants' Activation and Impact, creating a "learnability paradox" in Fig. 22. This paradox highlights the need for further research to find the optimal balance between performance and experience in HRC.

5.5 Limitations

The main limitations of the current experimental study refer to the following main points:

- *Bias from the test configuration itself*. The proposed assembly task has been simplified to focus on how kinematic variables affect human's perception. In such a way, no participant could have had any problems or difficulties performing the assembly. However, this may have biased the experiment, as the task had been seen as too easy to perform, so participants perceived that it could be easier without the robot.
- *Task simplicity***.** While the study focused on the impact of robot speed variations in a controlled and simplified assembly task, it is important to acknowledge that the proposed task may not fully capture the complexity of real-world HRC scenarios. The task's simplicity may have limited the generalizability of the findings to more intricate collaborative settings. However, this task was designed to establish a foundational understanding of the effects of robot kinematic variables, as it allowed to isolate and examine these specific variables in a controlled environment.
- *Sample size***.** In [87–89] claimed that 5 participants are sufficient to identify 80% of usability problems. However, small sample sizes can lead to high variability in test results that cannot be fully adjusted [90]. As concluded in [34], there is no widely validated sample size for UX tests with physiological monitoring. In fact, according to a systematic review on the use of physiological monitoring in UX studies [34], the median sample size in studies utilizing EEG stands at 19.4 participants. Furthermore, half of the EEG case studies feature sample sizes ranging from 9.5 to 24.5 participants. Therefore, the sample size in this experiment could be too small to obtain the possible correlations. A larger sample size would be needed to ensure that the results are statistically significant and can be generalised to a broader population.
- *The applicability of the proposed set-up*. The devices proposed in this paper for the evaluation of physiological factors are, applicable in a laboratory setting but not in an actual context. The devices themselves have several limitations, such as the fact that the participant must be seated, that they do not allow the user to behave naturally as he/she would in a real context, and that the participant can only use the dominant hand. According to Peruzzini et al. [51] improvements are necessary to move into a real industrial environment; physiological devices should be introduced as less intrusive.
- *Acceleration profile*: In the context of this study, the robot's acceleration was maintained at a constant rate. However, it is important to acknowledge that this approach may not entirely encapsulate the characteristics of 'human-like' movements. Typically, such movements are associated with an acceleration profile that commences from zero, escalates, and then reverts back to zero, all the while ensuring that the jerk over the trajectory is minimized. This simplification, necessitated by the constraints of our experimental setup, may have exerted an influence on the outcomes of the study.

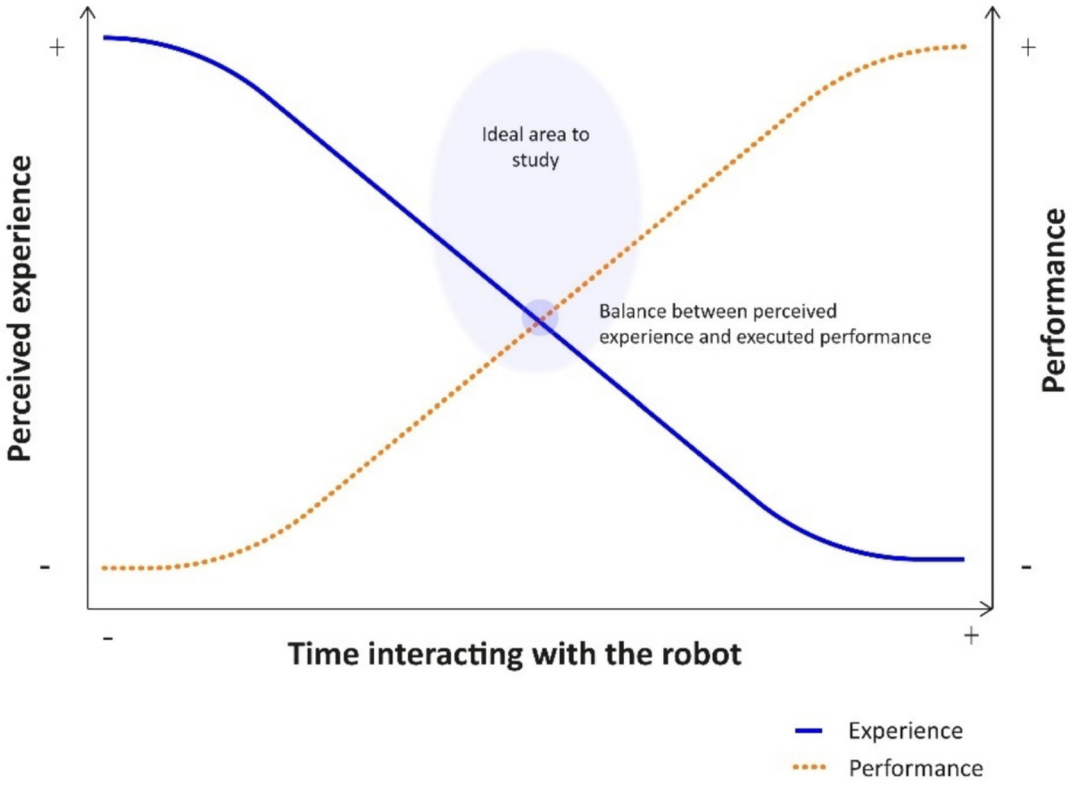


Fig. 22 Learnability paradox

Despite those limitations, the research is valuable since it demonstrated the feasibility of the approach to assess the UX of the operators in industrial settings. Promising results push the research to continue and overcome the current limitations.

6 Conclusions and Future Directions

In an increasingly collaborative work environment, the need for research on human factors requirements is evident. This paper presents an analysis from the UX and human-centred design perspective to see how kinematic variables influence human factors. The study involved two procedures in which participants performed a task at different speeds and acceleration levels.

The protocol analysis was based on a set of metrics into three investigation areas (KPIs, physiological, and perceptual indicators) to investigate the UX thoroughly from a human-centred approach. The set-up merges the traditional UX evaluation methods and biometric measuring devices (EEG and EDA). The analysis protocol has proved to be correct, and the lack of correlations between the different variables indicates that all measures are necessary. However, the UEQ-S was not the most suitable questionnaire to better understand perceptual values. The UEQ-S has some disadvantages, such as the loss of detailed information [83], as it does not measure the UX fully as it does the full UEQ. However, we also note that the questions are far from human–robot reality. Thus, there is a need to identify a more suitable questionnaire for industrial robotic environments. Additionally, it is important to consider that the three items related to Reliability, Controllability, and Perceived Usefulness are self-generated, which implies that these measures were developed specifically for this study and may not have been previously validated in other research.

The results of this study provide valuable insights into how the kinematic variables of a robot, such as speed and acceleration, influence the human factors. By means of KPIs, it was observed that the faster the robot's movement, the sooner the task was completed. Therefore, in all cases, participants were quicker in the FT. Nevertheless, the order of tasks influences results in a task performance study. Participants who did the tasks in reverse order took longer to complete the task and made more errors in the first task than in the second task. The second task was completed better with fewer errors. This suggests that the learnability of the interaction improves performance.

Both H1 and H2 have been invalidated. Concerning H1, the anticipated influence on Engagement, Memorisation or Mental Workload is not substantiated. Similarly, H2 has been debunked, contradicting the initial suppositions. Increased kinematic variables result in higher ratings for PQ and Perceived Usefulness. However, in the context of Valence,

Activation, Impact, HQ, Reliability, and Controllability, no statistically significant distinctions are apparent.

The experimental outcomes underscore that excessively high speeds have a detrimental impact on the UX during interactions with the robot, while overly slow speeds also yield negative effects. Consequently, we introduce the concept of the HRI speed bell, illustrating that both excessively high and excessively low robot speeds and accelerations are associated with poorer overall user experiences. Therefore, there is a clear need to explore and determine the ideal speed and acceleration settings to ensure a positive UX.

Additionally, the order in which tasks are performed influences participants' expectations of the robot and emotional responses. This is evident in the differences in scores between both procedures, where the order of tasks appears to impact the outcomes. It is essential to consider the influence of learnability when designing HRIs, as it may influence UX and performance. By understanding the effect of task order in HRI, it is possible to optimise the design of robots to meet operators' needs and expectations better and increase their well-being in the new industrial era. The study has revealed a "learnability paradox" in HRI, where longer interaction times improve participants' performance and learnability, but at the same time led to decline in UX. These results highlight the importance of considering the effect of learnability when designing HRI, as it has an impact on UX and performance. By understanding the influence of task order, it is possible to optimise the design of robots to better meet operators' needs and expectations and enhance their well-being in the new industrial era.

The study incorporates a gender perspective into the research to avoid biases that often arise from considering men as the default reference group. This approach provides a more inclusive and comprehensive understanding of how different genders interact with robots in industrial settings. The research findings confirm the existence of gender-based differences in task Execution Time, Activation, and Perceived Usefulness.

The study reveals that, regardless of gender, the higher kinematic variables consistently lead to shorter task execution times in both procedures. It is observed that the task performed second tends to result in fewer errors, indicating that learnability plays a role in error reduction. Memorisation showed a significant difference among men in the FS procedure, with higher kinematic variables leading to higher Memorisation levels. However, this difference was not evident in women in the FS procedure or in participants of the SF procedure. While the general analysis of Activation did not yield significant gender differences, an interesting observation was made in men in the SF procedure, where higher kinematic variables led to lower Activation levels. This trend was not observed in women in the SF procedure or participants in the FS procedure. Engagement, Mental Workload, and Valence exhibited no significant differences in any of the procedures in any of the genders, implying that these factors remain relatively consistent regardless of gender.

A significant gender-based difference in PQ was noted, primarily among women in the FS procedure, where higher kinematic variables resulted in higher PQ scores. HQ displayed a significant difference in women in the FS procedure, with higher kinematic variables resulting in higher HQ scores. No significant differences were found in Reliability and Controllability. Perceived Usefulness exhibited significant gender-related differences in women in the FS procedure, with higher kinematic variables leading to higher scores in Perceived Usefulness.

Overall, our results suggest that several factors influence how humans interact with robots, including the robot's speed and acceleration, the order in which tasks are performed, and the individual's background and knowledge. Future research could aim to replicate the study with a larger sample size to increase the generalizability of the results. It would also be valuable to extend the study to different populations or contexts.

6.1 Future Directions

In contemplating the future trajectory of this study, several promising avenues emerge for further exploration and refinement. Firstly, it is imperative to build upon the foundational insights gleaned from this exploratory study by conducting a larger-scale investigation with a more extensive sample size. This expanded study would provide greater statistical power, enabling more confident discernment of the observed effects' statistical significance.

Furthermore, the findings of this study have highlighted the importance of kinematic variables in the realm of industrial HRI. The identification of the "HRI speed bell" depicted in Fig. 21 accentuates the critical necessity of striking an optimal equilibrium among these variables. Deviations from this balance, whether excessive or deficient, can profoundly impact UX during interactions with robots. Therefore, a more profound inquiry into how these kinematic variables can be precisely defined and finely tuned to optimise HRI in industrial settings is warranted.

Moreover, given the potential variance in task complexity and cognitive demands across different tasks and specific contextual situations, a comprehensive exploration of the variability of these variables is imperative. Future research endeavours may thus pivot towards the development of models and algorithms capable of dynamically regulating these kinematic variables in real-time, contingent upon user requirements and preferences.

In this vein, this study could serve as a precursor to the burgeoning field of cognitive robotics [124, 125] and affective robotics [126, 127]. Cognitive robotics seeks to imbue

robots with cognitive abilities akin to those of humans, enabling them to perceive, reason, learn, and interact with their environment in a manner that mirrors human cognition. By exploring the intricate interplay between kinematic variables and UX, this study lays the groundwork for future research aimed at developing robots that possess advanced cognitive capabilities to adapt to the appropriate kinematic variables based on the individual with whom they are interacting.

On the other hand, the robot's acceleration in this study was constant, which may not fully replicate 'human-like' movements that typically have a variable acceleration profile with minimal jerk. Future research could include such 'humanised' acceleration profiles on the robot. This could lead to more natural and intuitive collaborations, offering a more realistic representation of human–robot interactions. This approach could potentially enhance the effectiveness of human–robot collaborations and contribute significantly to the field. Indeed, the robot's acceleration in this study was constant, which may not fully replicate 'human-like' movements that typically have a variable acceleration profile with minimal jerk. Future research could include such 'humanised' acceleration profiles on the robot. This could lead to more natural and intuitive collaborations, offering a more realistic representation of human–robot interactions. This approach could potentially enhance the effectiveness of human–robot collaborations and contribute significantly to the field.

Another compelling avenue for future investigation lies in extending the applicability of the insights garnered from this study to a broader spectrum of industrial contexts and specific tasks. While this study has provided valuable insights into HRI within an assembly environment, it is essential to acknowledge the potential divergence of challenges and dynamics across different tasks and industries. Conducting expansive research encompassing diverse industrial sectors, such as logistics, electronics, food industry, or agri-food, holds promise for enriching our understanding of HRI across varied application domains.

Furthermore, an essential aspect for future inquiry pertains to the structural dynamics of human–robot teams. Exploring how human operators interact with multiple robots in collaborative work settings and discerning the influence of leadership on task perception and efficiency are critical areas for further investigation. Addressing these questions through case studies that simulate realistic interaction scenarios between humans and multiple robots in diverse roles and responsibilities will be instrumental in advancing our understanding of HRI dynamics.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12369-025-01224-7.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature.

Data availability The datasets presented in this study can be found in online repositories. The name of the repository and accession number can be found below: https://doi.org/10.17632/yskpsxc9sk.1

Declarations

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Informed consent All the procedures in this study were conducted in accordance with the Mondragon Unibertsitatea-approved protocols. Written informed consent was obtained from the participants for their anonymized information to be published in this article.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Marvel JA, Bagchi S, Zimmerman M, Antonishek B (2020) Towards effective interface designs for collaborative HRI in manufacturing: metrics and measures. ACM Trans Hum-Robot Interact (THRI) 9(4):1–55. https://doi.org/10.1145/3385009
2. Karabegović I, Turmanidze R, and Dašić P, (2020) "Global trend of implementation of industrial robots relating to industry 4.0," in *Design, Simulation, Manufacturing: The Innovation Exchange*, Springer, pp. 147–155
3. Karabegović I, Turmanidze R, Dašić P (2022) Structural network for the implementation of 'industry 4.0' in production processes. Industry 4.0 7(1):3–6
4. Kopacek P (2013) Development trends in robotics. Elektrotechnik und Informationstechnik 130(2):42–47. https://doi.org/10.1007/S00502-013-0129-1/FIGURES/4
5. Bröhl C, Nelles J, Brandl C, Mertens A, Nitsch V (2019) Human–robot collaboration acceptance model: development and comparison for Germany, Japan, China and the USA. Int J Soc Robot 11(5):709–726. https://doi.org/10.1007/s12369-019-00593-0
6. D.-G. for R. and I. European Commission, "Industry 5.0 - Towards a sustainable, human-centric and resilient European industry," European Commission, 2021. Accessed: Sep. 22, 2022. [Online]. Available: https://op.europa.eu/en/publication-detail/-/publication/468a892a-5097-11eb-b59f-01aa75ed71a1/
7. Gualtieri L, Rauch E, Vidoni R, Matt DT (2020) Safety, ergonomics and efficiency in human-robot collaborative assembly: design guidelines and requirements. Procedia CIRP 91:367–372. https://doi.org/10.1016/J.PROCIR.2020.02.188
8. Chen M, Nikolaidis S, Soh H, Hsu D, Srinivasa S (2020) Trust-aware decision making for human-robot collaboration: Model

learning and planning. ACM Trans Human-Robot Interact (THRI) 9(2):1–23. https://doi.org/10.1145/3359616
9. Boden M et al (2017) Principles of robotics: regulating robots in the real world. Conn Sci 29(2):124–129. https://doi.org/10.1080/09540091.2016.1271400
10. Kahn PH Jr et al (2007) What is a Human?: toward psychological benchmarks in the field of human–robot interaction. Interact Stud 8(3):363–390. https://doi.org/10.1075/is.8.3.04kah
11. Lindblom J, Alenljung B (2020) The ANEMONE: theoretical foundations for UX Evaluation of action and intention recognition in human-robot interaction. Sensors 20(15):4284. https://doi.org/10.3390/s20154284
12. Prati E, Peruzzini M, Pellicciari M, Raffaeli R (2021) How to include user experience in the design of human-robot interaction. Robot Comput Integr Manuf 68:102072. https://doi.org/10.1016/J.RCIM.2020.102072
13. ISO 9241–210, "Ergonomics of human-system interaction—Part 210: Human-centred design for interactive systems," *International Standard*, vol. 2.Udgave. 2019.
14. Schmidtler J, Knott V, Hölzel C, Bengler K (2015) Human centered assistance applications for the working environment of the future. Occup Ergon 12(3):83–95. https://doi.org/10.3233/OER-150226
15. Apraiz A, Lasa G, Mazmela M (2023) Evaluation of user experience in human-robot interaction: a systematic literature review. Int J Soc Robot. https://doi.org/10.1007/S12369-022-00957-Z/TABLES/13
16. Lasota PA, Shah JA (2015) Analyzing the effects of human-aware motion planning on close-proximity human-robot collaboration. Hum Factors J Hum Factors Ergonomics Soc 57(1):21–33. https://doi.org/10.1177/0018720814565188
17. Koppenborg M, Nickel P, Naber B, Lungfiel A, Huelke M (2017) Effects of movement speed and predictability in human–robot collaboration. Hum Factors Ergon Manuf 27(4):197–209. https://doi.org/10.1002/hfm.20703
18. Arai T, Kato R, Fujita M (2010) Assessment of operator stress induced by robot collaboration in assembly. CIRP Ann 59(1):5–8. https://doi.org/10.1016/J.CIRP.2010.03.043
19. Dehais F, Sisbot EA, Alami R, Causse M (2011) Physiological and subjective evaluation of a human–robot object hand-over task. Appl Ergon 42(6):785–791. https://doi.org/10.1016/J.APERGO.2010.12.005
20. Dragan AD, Bauman S, Forlizzi J, and Srinivasa SS, (2015) "Effects of robot motion on human-robot collaboration," In *International Conference on Human-Robot Interaction*, pp. 51–58. https://doi.org/10.1145/2696454.2696473.
21. Koppenborg M, Nickel P, Naber B, Lungfiel A, Huelke M (2017) Effects of movement speed and predictability in human–robot collaboration. Hum Factors Ergonomics Manufact Service Ind 27(4):197–209. https://doi.org/10.1002/HFM.20703
22. Rubagotti M, Tusseyeva I, Baltabayeva S, Summers D, Sandygulova A (2021) Perceived safety in physical human–robot interaction—A survey. Robot Autonomous Syst 151:104047. https://doi.org/10.1016/j.robot.2022.104047
23. Joosse M, Lohse M, van Berkel N, Sardar A, Evers V (2021) Making appearances: how robots should approach people. ACM Trans Hum-Robot Interact 10(1):1–24. https://doi.org/10.1145/3385121
24. Gervasi R, Aliev K, Mastrogiacomo L, Franceschini F (2022) User experience and physiological response in human-robot collaboration: a preliminary investigation. J Intell Robot Syst 106(2):1–30. https://doi.org/10.1007/S10846-022-01744-8
25. M. S. Sanders and E. J. McCormick, (1998) "Human factors in engineering and design," Industrial Robot: An International Journal.
26. D. Norman, The design of everyday things: revised and expanded edition. Basic books, 2013.
27. Peruzzini M, Grandi F, Pellicciari M (2018) How to analyse the workers' experience in integrated product-process design. J Ind Inf Integr 12:31–46. https://doi.org/10.1016/j.jii.2018.06.002
28. Meissner A, Trübswetter A, Conti-Kufner AS, Schmidtler J (2020) Friend or Foe? understanding assembly workers' acceptance of human-robot collaboration. ACM Trans Hum-Robot Interact 10(1):1–30. https://doi.org/10.1145/3399433
29. Hopko S, Wang J, Mehta R (2022) Human factors considerations and metrics in shared space human-robot collaboration: a systematic review. Front Robot AI 9:799522. https://doi.org/10.3389/FROBT.2022.799522/BIBTEX
30. Yan Y, Jia Y (2022) A review on human comfort factors, measurements, and improvements in human–robot collaboration. Sensors 22(19):7431. https://doi.org/10.3390/s22197431
31. Bröhl C, Nelles J, Brandl C, Mertens A, Nitsch V (2019) Human–robot collaboration acceptance model: development and comparison for Germany, Japan, China and the USA. Int J Soc Robot 11(5):709–726. https://doi.org/10.1007/s12369-019-00593-0
32. A. Steinfeld *et al.*, "Common metrics for human-robot interaction," in *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*, 2006, pp. 33–40.
33. Coronado E, Kiyokawa T, Ricardez GAG, Ramirez-Alpizar IG, Venture G, Yamanobe N (2022) Evaluating quality in human-robot interaction: a systematic search and classification of performance and human-centered factors, measures and metrics towards an industry 5.0. J Manuf Syst 63:392–410. https://doi.org/10.1016/J.JMSY.2022.04.007
34. Apraiz A, Lasa G, Mazmela M (2021) Evaluating user experience with physiological monitoring: a systematic literature review. DYNA New Technol 8(1):21. https://doi.org/10.6036/NT10072
35. B. Laugwitz, T. Held, and M. Schrepp, "Construction and evaluation of a user experience questionnaire," In: *Symposium of the Austrian HCI and usability engineering group*, Springer, 2008, pp. 63–76.
36. Charles RL, Nixon J (2019) Measuring mental workload using physiological measures: a systematic review. Appl Ergon 74:221–232. https://doi.org/10.1016/J.APERGO.2018.08.028
37. Argyle EM, Marinescu A, Wilson ML, Lawson G, Sharples S (2021) Physiological indicators of task demand, fatigue, and cognition in future digital manufacturing environments. Int J Hum Comput Stud 145:102522. https://doi.org/10.1016/J.IJHCS.2020.102522
38. Neumann DL and Westbury HR, "The psychophysiological measurement of empathy," *Psychology of empathy*, pp. 119–142, 2011.
39. Oliveira IIP and Guimarães NM, (2013) "Practical neurophysiological analysis of readability as a usability dimension," In: International conference on human factors in computing and informatics, Springer, pp. 194–211.
40. Blandino G (2023) How to measure stress in smart and intelligent manufacturing systems: a systematic review. Systems 11:167. https://doi.org/10.3390/SYSTEMS11040167
41. Castaldo R, Melillo P, Bracale U, Caserta M, Triassi M, Pecchia L (2015) Acute mental stress assessment via short term HRV analysis in healthy adults: a systematic review with meta-analysis. Biomed Signal Process Control 18:370–377. https://doi.org/10.1016/J.BSPC.2015.02.012
42. Setz C, Arnrich B, Schumm J, La Marca R, Tröster G, Ehlert U (2010) Discriminating stress from cognitive load using a wearable eda device. IEEE Trans Inf Technol Biomed 14(2):410–417. https://doi.org/10.1109/TITB.2009.2036164
43. Bitbrain Technologies, "Diadem | EEG móvil de sensores secos | Bitbrain." Accessed: Oct. 01, 2021. [Online].

Available: https://www.bitbrain.com/es/productos-neurotecnologia/dry-eeg/diadem
44. W. Boucsein, *Electrodermal activity*. Springer Science & Business Media, 2012.
45. Benedek M, Kaernbach C (2010) Decomposition of skin conductance data by means of nonnegative deconvolution. Psychophysiology 47(4):647–658
46. Benedek M, Kaernbach C (2010) A continuous measure of phasic electrodermal activity. J Neurosci Methods 190(1):80–91. https://doi.org/10.1016/J.JNEUMETH.2010.04.028
47. Kreibig SD (2010) Autonomic nervous system activity in emotion: a review. Biol Psychol 84(3):394–421. https://doi.org/10.1016/J.BIOPSYCHO.2010.03.010
48. Golland Y, Keissar K, Levit-Binnun N (2014) Studying the dynamics of autonomic activity during emotional experience. Psychophysiology 51(11):1101–1111. https://doi.org/10.1111/PSYP.12261
49. Bitbrain Technologies, "Dispositivo GSR, BVP móvil para investigación | Bitbrain." Accessed: Apr. 11, 2022. [Online]. Available: https://www.bitbrain.com/es/productos-neurotecnologia/amplificador-de-biosenales/ring
50. Nicolò A, Massaroni C, Schena E, Sacchetti M (2020) The importance of respiratory rate monitoring: from healthcare to sport and exercise. Sensors 20(21):6396. https://doi.org/10.3390/S20216396
51. Peruzzini M, Grandi F, Pellicciari M (2020) Exploring the potential of operator 4.0 interface and monitoring. Comput Ind Eng 139:105600. https://doi.org/10.1016/j.cie.2018.12.047
52. Davidson RJ (1992) Anterior cerebral asymmetry and the nature of emotion. Brain Cogn 20(1):125–151. https://doi.org/10.1016/0278-2626(92)90065-T
53. Allen JJB, Coan JA, Nazarian M (2004) Issues and assumptions on the road from raw signals to metrics of frontal EEG asymmetry in emotion. Biol Psychol 67(1–2):183–218. https://doi.org/10.1016/J.BIOPSYCHO.2004.03.007
54. Harmon-Jones E, Gable PA, Peterson CK (2010) The role of asymmetric frontal cortical activity in emotion-related phenomena: a review and update. Biol Psychol 84(3):451–462. https://doi.org/10.1016/J.BIOPSYCHO.2009.08.010
55. Sederberg PB, Kahana MJ, Howard MW, Donner EJ, Madsen JR (2003) Theta and gamma oscillations during encoding predict subsequent recall. J Neurosci 23(34):10809–10814
56. Long NM, Burke JF, Kahana MJ (2014) Subsequent memory effect in intracranial and scalp EEG. Neuroimage 84:488–494. https://doi.org/10.1016/J.NEUROIMAGE.2013.08.052
57. Klimesch W, Doppelmayr M, Russegger H, Pachinger T (1996) Theta band power in the human scalp EEG and the encoding of new information. NeuroReport 7:1235–1240
58. Klimesch W, Doppelmayr M, Schimke H, Ripper B (1997) Theta synchronization and alpha desynchronization in a memory task. Psychophysiology 34(2):169–176
59. Brouwer A-M, Hogervorst MA, Van Erp JBF, Heffelaar T, Zimmerman PH, Oostenveld R (2012) Estimating workload using EEG spectral power and ERPs in the n-back task. J Neural Eng 9(4):045008
60. M. Stikic *et al.*, (2014) "Modeling temporal sequences of cognitive state changes based on a combination of EEG-engagement, EEG-workload, and heart rate metrics," https://doi.org/10.3389/fnins.2014.00342.
61. Klimesch W (1999) EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. Brain Res Rev 29(2–3):169–195. https://doi.org/10.1016/S0165-0173(98)00056-3
62. Gevins A et al (1998) Monitoring working memory load during computer-based tasks with EEG pattern recognition methods. Hum Factors 40(1):79–91
63. Freeman FG, Mikulka PJ, Prinzel LJ, Scerbo MW (1999) Evaluation of an adaptive automation system using three EEG indices with a visual tracking task. Biol Psychol 50(1):61–76. https://doi.org/10.1016/S0301-0511(99)00002-2
64. Mikulka PJ, Scerbo MW, Freeman FG (2002) Effects of a biocybernetic system on vigilance performance. Hum Factors 44(4):654–664
65. Pope AT, Bogart EH, Bartolome DS (1995) Biocybernetic system evaluates indices of operator engagement in automated task. Biol Psychol 40(1–2):187–195. https://doi.org/10.1016/0301-0511(95)05116-3
66. Apraiz A, Lasa G, Mazmela M (2021) Aplicación de cuestionarios para evaluación de la Experiencia de Usuario en entornos industriales. Dyna (Medellin) 96(3):235–235. https://doi.org/10.6036/10025
67. Apraiz A, Lasa G (2020) Questionnaires for the evaluation of user experience and usability in industrial environments: a bibliographic review. DYNA New Technol. https://doi.org/10.6036/nt9845
68. Väänänen-Vainio-Mattila K, Roto V, and Hassenzahl M, (2008) "Towards practical user experience evaluation methods," *Meaningful measures: Valid useful user experience measurement (VUUM)*, pp. 19–22
69. Hart SG, Staveland LE (1988) Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. Adv Psychol. https://doi.org/10.1016/S0166-4115(08)62386-9
70. Schrepp M, Thomaschewski J (2019) Design and validation of a framework for the creation of user experience questionnaires. Int J Int Multimedia Artif Intell. https://doi.org/10.9781/ijimai.2019.06.006
71. Peruzzini M, Grandi F, Pellicciari M (2017) Benchmarking of tools for user experience analysis in industry 4.0. Procedia Manuf 11:806–813. https://doi.org/10.1016/J.PROMFG.2017.07.182
72. von der Pütten AR, Bock N (2018) Development and validation of the self-efficacy in human-robot-interaction scale (SE-HRI). ACM Trans Human-Robot Interaction (THRI). https://doi.org/10.1145/3139352
73. Sauro J (2015) SUPR-Q: a comprehensive measure of the quality of the website user experience. J Usabil Stud 10(2):68–86
74. Lund AM (2001) Measuring usability with the use questionnaire12. Usability interface 8(2):3–6
75. Finstad K (2010) The usability metric for user experience. Interact Comput 22(5):323–327. https://doi.org/10.1016/j.intcom.2010.04.004
76. Brooke J (1996) SUS-A quick and dirty usability scale. Usabil Evaluat Ind 189(194):4–7
77. Chin JP, Diehl VA, and Norman KL, "Development of an instrument measuring user satisfaction of the human-computer interface," in *Proceedings of the SIGCHI conference on Human factors in computing systems*, 1988, pp. 213–218.
78. Kirakowski J, Corbett M (1993) SUMI: the software usability measurement inventory. Br J Edu Technol 24(3):210–212
79. Lin HX, Choong Y-Y, Salvendy G (1997) A proposed index of usability: a method for comparing the relative usability of different software systems. Behav inf technol 16(4–5):267–277
80. Kirakowski J, Cierlik B (1998) Measuring the usability of web sites. Proce Hum Factors Ergonomics Soc Ann Meet 42(4):424–428. https://doi.org/10.1177/154193129804200405
81. Lewis JR, Utesch BS, Maher DE (2015) Measuring perceived usability: the SUS, UMUX-LITE, and AltUsability. Int J Hum Comput Interact 31(8):496–505
82. Bartneck C, Kulić D, Croft E, Zoghbi S (2008) Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. Int J Soc Robot 1(1):71–81. https://doi.org/10.1007/s12369-008-0001-3

83. Schrepp M, Hinderks A, Thomaschewski J (2017) Design and evaluation of a short version of the user experience questionnaire (UEQ-S). Int J Interact Multimedia Artif Intell 4(6):103–108
84. Apraiz A, Lasa G, Mazmela M (2023) Evaluation of user experience in human-robot interaction: a systematic literature review. Int J Soc Robot. https://doi.org/10.1007/s12369-022-00957-z
85. Almeida L, Menezes P, Dias J (2020) Interface transparency issues in teleoperation. Appl Sci 10(18):6232. https://doi.org/10.3390/APP10186232
86. Gervasi R, Capponi M, Mastrogiacomo L, Franceschini F (2023) Analyzing psychophysical state and cognitive performance in human-robot collaboration for repetitive assembly processes. Prod Eng Res Devel 2023:1–15. https://doi.org/10.1007/S11740-023-01230-6
87. Virzi RA (1992) Refining the test phase of usability evaluation: how many subjects is enough? Hum Factors J Hum Factors Ergonomics Soc 34(4):457–468. https://doi.org/10.1177/001872089203400407
88. Lewis JR (1994) Sample sizes for usability studies: additional considerations. Hum Factors 36(2):368–378. https://doi.org/10.1177/001872089403600215
89. J. Nielsen, "Why you only need to test with 5 users." Useit. com Alertbox, 2000.
90. Cazañas A, de Miguel AS, Parra E (2017) Estimating sample size for usability testing. Enfoque UTE 8(1):172–185. https://doi.org/10.29019/ENFOQUEUTE.V8N1.126
91. Schillaci G, Bodiroža S, Hafner VV (2013) Evaluating the effect of saliency detection and attention manipulation in human-robot interaction. Int J Soc Robot 5(1):139–152. https://doi.org/10.1007/S12369-012-0174-7
92. Daniel B, Thomessen T, Korondi P (2013) Simplified human-robot interaction: modeling and evaluation. Model Identif Control Norwegian Res Bull 34(4):199–211. https://doi.org/10.4173/MIC.2013.4.4
93. Danielsson O, Syberfeldt A, Brewster R, Wang L (2017) Assessing instructions in augmented reality for human-robot collaborative assembly by using demonstrators. Procedia CIRP 63:89–94. https://doi.org/10.1016/J.PROCIR.2017.02.038
94. Aromaa S, Frangakis N, Tedone D, Viitaniemi J, Aaltonen I (2018) digital human models in human factors and ergonomics evaluation of gesture interfaces. Proc ACM Hum Comput Interact. https://doi.org/10.1145/3229088
95. Tang G, Webb P (2018) The design and evaluation of an ergonomic contactless gesture control system for industrial robots. J Robot 2018:1–10. https://doi.org/10.1155/2018/9791286
96. Kildal J, Martín M, Ipiña I, Maurtua I (2019) Empowering assembly workers with cognitive disabilities by working with collaborative robots: a study to capture design requirements. Procedia CIRP 81:797–802. https://doi.org/10.1016/J.PROCIR.2019.03.202
97. Slater M, Usoh M, Steed A (1994) Depth of presence in virtual environments. Presence: Teleoperators Virtual Environ 3(2):130–144. https://doi.org/10.1162/pres.1994.3.2.130
98. Usoh M, Catena E, Arman S, Slater M (2000) Using presence questionnaires in reality. Presence Teleoperators Virtual Environ 9(5):497–503. https://doi.org/10.1162/105474600566989
99. Hietanen A, Pieters R, Lanz M, Latokartano J, Kämäräinen JK (2020) AR-based interaction for human-robot collaborative manufacturing. Robot Comput Integr Manuf 63:101891. https://doi.org/10.1016/J.RCIM.2019.101891
100. Pantano M, Regulin D, Lutz B, Lee D (2020) A human-cyber-physical system approach to lean automation using an industrie 4.0 reference architecture. Procedia Manuf 51:1082–1090. https://doi.org/10.1016/J.PROMFG.2020.10.152
101. Beschi M, Faroni M, Copot C, Pedrocchi N (2020) How motion planning affects human factors in human-robot collaboration. IFAC-PapersOnLine 53(5):744–749. https://doi.org/10.1016/J.IFACOL.2021.04.167
102. Colim A et al (2021) Lean manufacturing and ergonomics integration: defining productivity and wellbeing indicators in a human-robot workstation. Sustainability 13:13. https://doi.org/10.3390/SU13041931
103. Akalin N, Kristoffersson A, Loutfi A (2022) Do you feel safe with your robot? Factors influencing perceived safety in human-robot interaction based on subjective and objective measures. Int J Hum Comput Stud 158:102744. https://doi.org/10.1016/j.ijhcs.2021.102744
104. Kim Y, Mutlu B (2014) How social distance shapes human–robot interaction. Int J Hum Comput Stud 72(12):783–795
105. Akalin N, Kristoffersson A, and Loutfi A, (2019) "Evaluating the sense of safety and security in human–robot interaction with older people," *Social robots: Technological, societal and ethical aspects of human-robot interaction*, pp. 237–264,.
106. Strube MJ, Werner C (1984) Personal space claims as a function of interpersonal threat: The mediating role of need for control. J Nonverbal Behav 8:195–209
107. Schaefer KE, "Measuring trust in human robot interactions: development of the 'trust perception scale-HRI,'" in *Robust intelligence and trust in autonomous systems*, Springer, 2016, pp. 191–218.
108. Weiss A, Bernhaupt R, Tscheligi M, and Yoshida E, "Addressing user experience and societal impact in a user study with a humanoid robot," In *AISB2009: proceedings of the symposium on new frontiers in human-robot interaction (Edinburgh, 8–9 April 2009), SSAISB*, 2009, pp. 150–157.
109. Gervasi R, Capponi M, Mastrogiacomo L, Franceschini F (2023) Manual assembly and Human-Robot Collaboration in repetitive assembly processes: a structured comparison based on human-centered performances. Int J Adv Manuf Technol. https://doi.org/10.1007/S00170-023-11197-4
110. Kulic D, Croft EA (2007) Affective state estimation for human–robot interaction. IEEE Trans Rob 23(5):991–1000
111. Kühnlenz B, Erhart M, Kainert M, Wang ZQ, Wilm J, Kühnlenz K (2018) Impact of trajectory profiles on user stress in close human-robot interaction. Automatisierungstechnik 66(6):483–491. https://doi.org/10.1515/AUTO-2018-0004/MACHINEREADABLECITATION/RIS
112. Villani V, Pini F, Leali F, Secchi C (2018) Survey on human–robot collaboration in industrial settings: safety, intuitive interfaces and applications. Mechatronics 55:248–266. https://doi.org/10.1016/J.MECHATRONICS.2018.02.009
113. Liu Y, Ritchie JM, Lim T, Kosmadoudi Z, Sivanathan A, Sung RCW (2014) A fuzzy psycho-physiological approach to enable the understanding of an engineer's affect status during CAD activities. CAD Comput Aided Des 54:19–38. https://doi.org/10.1016/j.cad.2013.10.007
114. Hancock PA, Billings DR, Schaefer KE, Chen JYC, de Visser EJ, Parasuraman R (2011) A meta-analysis of factors affecting trust in human-robot interaction. Hum Factors 53(5):517–527. https://doi.org/10.1177/0018720811417254
115. Onnasch L, Roesler E (2021) A taxonomy to structure and analyze human–robot interaction. Int J Soc Robot 13(4):833–849. https://doi.org/10.1007/s12369-020-00666-5
116. "Human Behaviour Research | Software de sincronización biométrica | Bitbrain." Accessed: Oct. 01, 2021. [Online]. Available: https://www.bitbrain.com/es/productos-neurotecnologia/software/sennslab
117. Bitbrain Technologies, "SennsMetrics | Software para el Análisis de BioMétricas | Bitbrain." Accessed: Oct. 01, 2021. [Online]. Available: https://www.bitbrain.com/es/productos-neurotecnologia/software/sennsmetrics
118. Hassenzahl M (2007) The hedonic/pragmatic model of user experience. Towards a UX manifesto 10:2007

119. Bitbrain Technologies, "Human Behaviour Research | Software de sincronización biométrica | Bitbrain." Accessed: Feb. 23, 2021. [Online]. Available: https://www.bitbrain.com/es/productos-neurotecnologia/software/sennslab
120. I. Aliagas Ocaña, (2022) "Análisis neuropsicofisiológico de la eficacia del emplazamiento de producto en videojuegos," Universidad Complutense de Madrid, Madrid.
121. KUKA, "KUKA LBR iiwa 14 R820 Data sheet."
122. "LBR iiwa | KUKA AG." Accessed: Nov. 02, 2021. [Online]. Available: https://www.kuka.com/en-de/products/robot-systems/industrial-robots/lbr-iiwa
123. A. Apraiz *et al.*, (2024) "Impact of robotic kinematic variables on user experience: dataset on performance, physiological response, and user perception in human-robot interaction during an assembly task," *Mendeley Data*.
124. Soori M, Arezoo B, Dastres R (2023) Artificial intelligence, machine learning and deep learning in advanced robotics, a review. Cognit Robot 3:54–70. https://doi.org/10.1016/J.COGR.2023.04.001
125. Han L, Yang S, Tyroller Q (2021) Cognitive human-robot-collaboration in assembly: a framework for cognitive interaction planning and subject study. Procedia Manuf 55:24–31. https://doi.org/10.1016/J.PROMFG.2021.10.005
126. Landi CT, Villani V, Ferraguti F, Sabattini L, Secchi C, Fantuzzi C (2018) Relieving operators' workload: towards affective robotics in industrial scenarios. Mechatronics 54:144–154. https://doi.org/10.1016/J.MECHATRONICS.2018.07.012
127. Y. Bar-Cohen, (2016) "Affective robotics: human motion and behavioral inspiration for safe cooperation between humans and humanoid assistive robots," in *Biomimetics*, CRC Press, pp. 481–496. https://doi.org/10.1201/B11230-20.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Dra. Ainhoa Apraiz is a researcher and lecturer specialising in Industrial Design Innovation at Mondragon Unibertsitatea (MGEP). She completed her undergraduate studies in Industrial Design Engineering in 2018 and obtained a Master's degree in Strategic Design of Products and Services in 2020. She received her PhD from Mondragon Unibertsitatea in 2023, where she presented her thesis "New Model for Evaluating Human-Robot Interaction in Industrial Environments: ITPX-ROBOTS" and was awarded Cum Laude. Since 2024, Ainhoa has been teaching and conducting research at MGEP, focusing on the interaction and perception of technologies and user experience. She has participated in various collaborative projects with companies and several European projects under the Horizon Europe framework. She is the author of several scientific articles on interaction design, technological acceptance, and user experience.

Dr. Ganix Lasa is a researcher and lecturer in the area of Innovation in Industrial Design at MGEP and is currently the coordinator of the Design Innovation Center (DBZ) research group. Earned a PhD from Mondragon Unibertsitatea in 2015 after defending the doctoral thesis titled Eyeface: A New Tool for Evaluating Conceptual Ideas for Products and Services Based on User Experience. Continuing his research career, he completed a postdoctoral stay at the Centre for Interaction Design at Edinburgh Napier University in Scotland (United Kingdom). His research activity focuses on the interaction and perception of products, with a particular emphasis on aspects such as intuitiveness, acceptance, and usability. He has led and participated in projects in collaboration with companies as well as in several European projects. He is the author of more than 20 scientific articles on user-centered design, interaction design and human factors.

Dra. Maitane Mazmela is a Lecturer and Researcher at the Polytechnic School of Mondragon Unibertsitatea (MU). She earned her Bachelor's degree in Industrial Design Engineering in 2010, followed by a Master's degree in Strategic Design of Products and Associated Services in 2012. Her academic journey culminated in a PhD with Cum Laude distinction, awarded by Mondragon Unibertsitatea in 2020. Since 2012, she has held the role of Digital Strategy and User Experience Lead at several design agencies, leading projects for a range of institutions and companies. She has been teaching at MU since 2016, specializing in Industrial Design and Product Development Engineering, as well as in the Master's program in Strategic Design of Products and Services. In addition to her research and teaching activities, Maitane has supervised PhD research in human-robot interaction (2023). Her research focuses on technology acceptance and human-robot interaction, areas in which she has contributed to numerous privately funded projects and European initiatives. She has also authored multiple scientific publications on these topics. In recognition of her expertise, she has been a member of the scientific committee for the International Congress on Human-Computer Interaction since 2019.

Dr. Nestor Arana-Arexolaleiba holds a Ph.D. degree in automatic systems from INSAT-CNRS-LAAS (Robotics and AI Research Group) Toulouse, France (2002). Since 2002, he has been actively involved with the Robotics and Automation Research and Intelligent Systems for Industrial Systems groups at Mondragon University. Additionally, he served as a Guest Researcher at Aalborg University, Denmark, during 2018–2019 and 2020–2022. His research interests span safe machine learning for remanufacturing, human-robot collaboration, and autonomous vehicles.

Dr. Antonio Serrano Muñoz obtained his B.S. degree in Automatic Control and Computing from the Technological University of Havana "José Antonio Echeverría" (CUJAE), Cuba, in 2015, and his M.S. degree in Computer Science from the University of Havana, Cuba, in 2018. In 2023, he received his Ph.D. in Engineering from Mondragon Unibertsitatea, Spain.

Dr. Íñigo Elguea received the B.S. degree in biomedical engineering from Mondragon University, Spain, in 2018, and the M.S. degree in biomedical engineering from the Polytechnic University of Valencia, Spain, in 2019. In 2024, he obtained the Ph.D. in collaborative robotics and reinforcement learning.

Amaia Etxabe holds a Bachelor's degree in Industrial Design Engineering and Product Development (2021) and a Master's degree in Strategic Product and Service Design (2023), both from Mondragon Unibertsitatea. Her main research interest lies in Inclusive Design, which has shaped her Dual Education pathway and Master's Thesis, focusing on the creation and definition of inclusive design tools that integrate digital competencies. Currently, she works in the Industrial Design Innovation area at DBZ (Diseinu Berrikuntza Zentroa) within the Faculty of Engineering at Mondragon Unibertsitatea. She is also conducting doctoral research on design tools for the development of accessible emerging technologies.