POLITECNICO DI TORINO
Repository ISTITUZIONALE

Does size matter? Exploring the effect of cobot size on user experience in human–robot collaboration

Original
Does size matter? Exploring the effect of cobot size on user experience in human–robot collaboration / Gervasi, R., Capponi, M., Mastrogiacomo, L., Franceschini, F.. - In: INTERNATIONAL JOURNAL, ADVANCED MANUFACTURING TECHNOLOGY. - ISSN 0268-3768. - STAMPA. - 133:11-12(2024), pp. 5777-5791. [10.1007/s00170-024-14060-2]

Availability:
This version is available at: 11583/2991562 since: 2024-08-06T13:20:12Z

Publisher:
Springer

Published
DOI:10.1007/s00170-024-14060-2

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

26 November 2026

The International Journal of Advanced Manufacturing Technology (2024) 133:5777–5791
https://doi.org/10.1007/s00170-024-14060-2

ORIGINAL ARTICLE

Does size matter? Exploring the effect of cobot size on user experience in human–robot collaboration

Riccardo Gervasi[1] · Matteo Capponi[1] · Luca Mastrogiacomo[1] · Fiorenzo Franceschini[1]

Received: 23 March 2024 / Accepted: 22 June 2024 / Published online: 9 July 2024
© The Author(s) 2024

Abstract
In the vision of Industry 5.0, collaborative robots (or cobots) play a central supporting role in various industries, especially manufacturing. Close interaction with cobots requires special attention to user experience to fully exploit the benefits of this paradigm. Consequently, understanding the impact of a cobot's physical size on user experience becomes critical to optimizing human–robot collaboration (HRC). This research aims to investigate the relationship between cobot size (UR3e – small cobot vs. UR10e – large cobot) and user experience in HRC contexts, in conjunction with other factors (i.e., cobot movement speed and product assembly complexity). Through a series of controlled experiments involving 32 participants, user experience data were obtained by collecting physiological measures (i.e., electro-dermal activity, heart activity, eye-tracking metrics) and subjective responses with questionnaires (i.e., perceived workload, interaction quality, and affective state). Results showed that the large cobot was generally perceived to be safer, more natural, efficient, fluid, and trustworthy. With the large cobot, there was a decrease in dominance; however, it was offset by the learning effect. Perceived workload was mainly influenced by product complexity. No clear difference in terms of mental strain emerged from the physiological data comparing the cobot sizes. In addition, the interaction term between cobot size and cobot movement speed never emerged as significant. The results of this research can offer practical insights to improve the effectiveness and acceptance of cobots during the implementation phase.

Keywords Human–robot collaboration · Industry 5.0 · Cobot size · User experience · Physiological measures · Mental strain

1 Introduction

In the era of Industry 5.0, the integration of collaborative robots, commonly known as cobots, is revolutionizing the landscape of various industries, particularly in manufacturing [1, 2]. These kinds of robots are designed to work alongside humans, emphasizing a collaborative approach to enhance overall efficiency and productivity [3]. As cobots are central to the vision of Industry 5.0, it becomes imperative to study the implications of human–robot collaboration (HRC) to fully exploit the potential benefits of this paradigm shift [4].

A critical aspect demanding attention in HRC is the user experience during close interaction with cobots [5]. For optimal collaboration, it is essential to comprehend how certain factors, such as the physical size of cobots, influence the user's perception during the interaction [3].

In one of the first works aimed to explore this aspect, Rahimi and Waldemar Karwowski [6] found out that the selection by the operator of safe speed for a traditional industrial robot depended on the size of the robot, leading to the conclusion that larger robots produce the perception of a more dangerous motion in space. Karwowski et al. [7] investigated the effects of exposure to a simulated accident, involving a robot hitting a mannequin placed inside the work envelope by exploring also industrial robot size. Results,

✉ Fiorenzo Franceschini
fiorenzo.franceschini@polito.it

Riccardo Gervasi
riccardo.gervasi@polito.it

Matteo Capponi
matteo.capponi@polito.it

Luca Mastrogiacomo
luca.mastrogiacomo@polito.it

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

surprisingly, showed that industrial workers felt safer in the presence of an arm of the larger robot than the smaller one. Hirori and Ito [8] explored the influence of size, in terms of robot height (small, medium, and high), of a mobile robot on the anxiety or threat felt by a human to be caused by a robot approaching him or her and determining the appropriate human–robot distance. They observed a tendency for the human–robot distance to increase along with the size of the robot, but surprisingly, the one that resulted in less anxiety was the robot with the medium height.

From the literature, it is not entirely clear how the size of the robot affects the user experience, especially in HRC. This research aims to address this crucial aspect by investigating the relationship between cobot size – specifically, comparing a small cobot (UR3e) with a large one (UR10e) – and user experience in HRC contexts in terms of both subjective and physiological responses. To provide a comprehensive understanding, this exploration is conducted in conjunction with other factors, namely cobot movement speed and product assembly complexity. Two main research questions are addressed: (i) *How does the cobot's physical size affect user experience?* (ii) *Do the joint effects of the cobot size with movement speed or product complexity affect user experience?*

The study employs a methodology incorporating controlled experiments with 32 participants. Data collection encompasses both physiological measures, including electro-dermal activity, heart activity, and eye-tracking metrics, as well as subjective responses obtained through questionnaires addressing perceived workload, interaction quality, and affective state. Through this multifaceted approach, the dynamics that modulate user experience in the context of cobots of different sizes are explored. The main novel elements of this study reside in (i) the comparison of two morphologically identical cobots that differ only in size, effectively eliminating possible other confounding elements that can affect human perception; (ii) the joint analysis of multiple physiological signals to provide a broader view on the cognitive aspects involved in the HRC.

The significance of this research extends beyond conceptual considerations, offering practical insights that can significantly impact the implementation of cobots in real-world scenarios. By understanding how users perceive and interact with cobots of different sizes, we aim to provide valuable guidance for designers, engineers, and industry stakeholders seeking to optimize HRC for enhanced effectiveness and acceptance. The subsequent sections present detailed findings and analyses derived from our experiments, shedding light on the intricate interplay between cobot size, movement speed, and product assembly complexity in shaping the user experience.

The paper is structured as follows. The Section 2 presents the methodology used to explore the research questions, describing the experimental setup, collected data, and data processing. The Section 3 illustrates the results obtained by analyzing the data collected through the experiment. In the Section 4, a discussion of the implications of the cobot size and movement speed on user experience is presented. Finally, future work and conclusions are presented in the Section 5.

2 Methodology

In this section, the implemented methodology to explore the cobot size effect is explained.

2.1 Experimental setup

In the "Mind4Lab" (Manufacturing Industry 4.0 Laboratory) at the "Politecnico di Torino," collaborative assembly tasks were implemented in an HRC setting. To explore the effects of cobot size (*Size*) on user experience, two collaborative workstations were used: one with the UR3e cobot (*Small*) and the other with the UR10e cobot (*Large*) (Fig. 1). These cobots were chosen for their similarity in terms of esthetic features, differing only in physical size, and both had a setup hanging above the operator (i.e., overhead) (Fig. 2). The UR3e weights 11.2 kg, has a reach of 0.5m, and a payload of 3 kg, whereas the UR10e weights 33.5 kg, has a reach of 1.3 m, and a payload of 12.5 kg. Two types of products with different assembly complexity were considered (*Product*). A mechanical component (Fig. 3a), consisting of a base, two oval flanges, and one square flange, was considered a product of low assembly complexity (*Low*), whereas a tile cutter was considered a more complex product (*High*) (Fig. 3b).

The operations for assembling the mechanical component can be divided into the following phases:

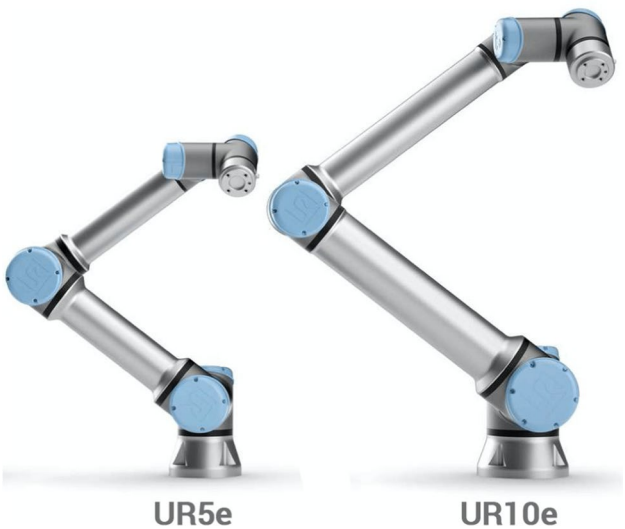


Fig. 1 Size comparison between cobots UR3e and UR10e [9]

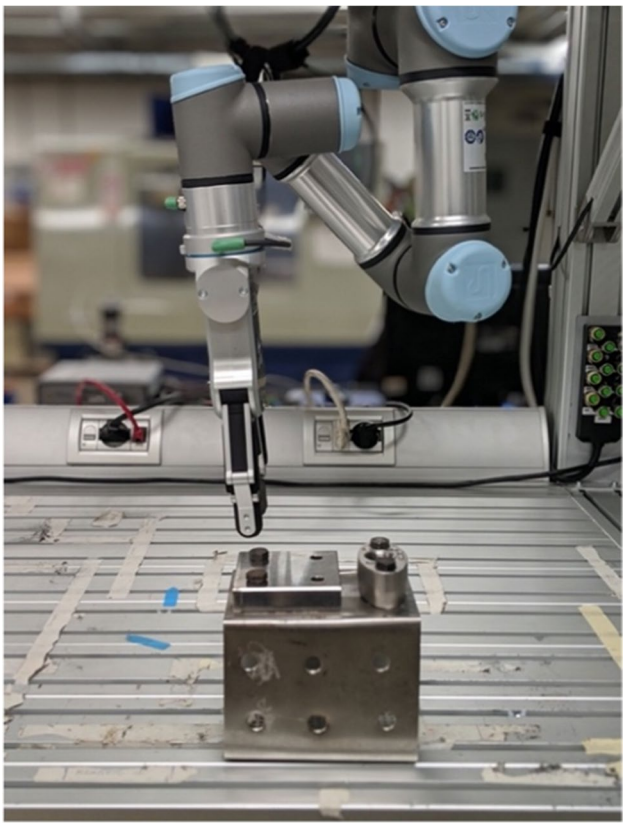


Fig. 2 Collaborative workstation with overhead UR3e setup

1. *Positioning the base*. The cobot picks the base and brings it closer to the operator.
2. *Assembling the square flange*. The cobot picks the square flange and places it in the correct position on the base. The operator takes the screws, inserts them into the holes, and tightens them.
3. *Assembling the first oval flange*. The cobot takes the oval flange and places it correctly on the base of the square flange. The operator takes two screws, inserts them into the holes, and tightens them.
4. *Assembling the second oval flange*. The cobot takes the remaining oval flange and places it correctly on the base. The operator takes the final two screws, inserts them into the holes, and tightens them. The cobot takes the assembled mechanical component and places it in another work area.

The operations for assembling the tile cutter can be broken down into four macro-phases:

1. *Assembly of the side supports.* The cobot brings the base of the tile cutter closer to the operator, who assembles the side supports for the rail rods. When the operation is finished, the cobot sets the assembled component aside.
2. *Assembly of the cutting mechanism*. The cobot takes the main component of the cutting mechanism and holds it in an ergonomic position in front of the operator. The operator positions and assembles the round blade and the remaining two components of the mechanism. When the operations are finished, the cobot releases the cutting mechanism close to the operator.
3. *Assembling the base with the cutting mechanism.* The cobot brings the base with assembled side supports back to the operator; the operator picks up the cutting mechanism, inserts the rail rods through the appropriate slots of the cutting mechanism, and joins the rail rods to the base through the side supports.
4. *Completion of the tile cutter*. The operator takes the handle and inserts it into the threaded slot of the cutting mechanism. When finished, the cobot takes the completed tile cutter and sets it aside.

Two levels of the robot's joint speed (*Speed*) were implemented: 90°/s (*Normal*) and 270°/s (*High*). These values represent the maximum speed that all the robot's joints could reach.

A within-subject experimental design was implemented in this study to examine the effects of two fixed factors (i.e., cobot size and cobot movement speed) on

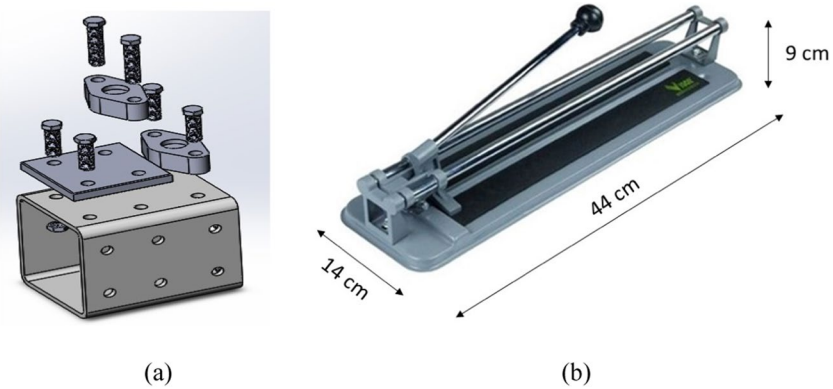


Fig. 3 The mechanical component (**a**) and the tile cutter (**b**)

perceived interaction quality, affective state, workload, and physiological response. A between-subject approach was used to explore the effect of product complexity in order to reduce the duration of the experimental session per participant (Fig. 4).

Thirty-two participants, with no prior experience with cobots, were recruited from the "Politecnico di Torino" and the surroundings for the study (27.3% females and 72.7% males). After randomly selecting the product to be assembled, each participant performed the assembly task twice in all four possible configurations (i.e., small cobot – normal speed, small cobot – high speed, large cobot – normal speed, and large cobot – high speed) in random order.

At the end of each trial, three different questionnaires were administered. An interaction quality questionnaire (Table 1) composed of seven items was used to collect participants' perceptions on different dimensions related to the interaction with the cobot. The items were evaluated on a seven-point scale (from "strongly disagree" to "strongly agree").

The self-assessment manikin (SAM) [11, 12] was administered to participants to gather their affective state in the different task configurations by evaluating on a nine-point scale three dimensions: valence, arousal, and dominance (Fig. 5). Valence represents the pleasantness relative to a stimulus (for instance, happiness and relaxation are associated with a high valence, whereas anxiety or anger with a low valence). Arousal refers to the intensity of emotion provoked by a stimulus (e.g., fear and anger are usually associated with a high arousal, whereas relaxation and boredom with a low

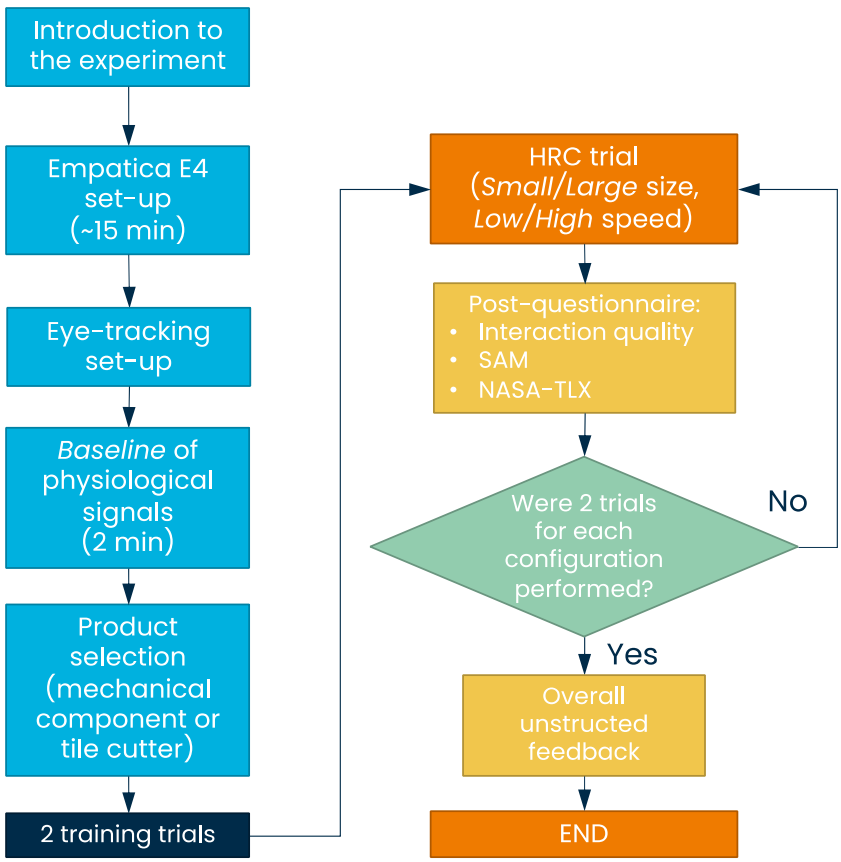


Fig. 4 Flowchart of the experimental process

Table 1 Questionnaire for interaction quality [10]

Item No	Dimension	Questionnaire item
Q1	Robot helpfulness	The robot was helpful in accomplishing the task
Q2	Interaction unsafety	I felt the interaction was not safe
Q3	Interaction naturalness	The collaboration felt natural
Q4	Team efficiency	The robot and I worked efficiently together
Q5	Team fluency	The robot and I worked fluently together
Q6	Discomfort	I felt uncomfortable with the robot
Q7	Robot trustworthiness	The robot was trustworthy

arousal). Dominance represents the degree of control felt relative to a stimulus (e.g., relaxation or anger is usually associated with a high dominance, while fear or anxiety with low dominance).

The commonly used NASA-TLX [13] was implemented to assess operator workload during the collaborative assembly (Fig. 6). NASA-TLX is composed of the following six dimensions:

- *Mental demand*, referring to the amount of mental activity required by the task.
- *Physical demand*, which represents the amount of physical activity required by the task.
- *Temporal demand*, which refers to how much time pressure is perceived due to the task pace.
- *Performance*, concerning with the degree of success and satisfaction with the results obtained in performing the task.
- *Effort*, which represents how hard one has to work (both mentally and physically) to achieve a certain level of performance.
- *Frustration*, referring to the amount of annoyance, frustration, or irritation felt during the task.

Each NASA-TLX dimension is rated on a 0–100 scale with five-point steps (see Fig. 6), and the final workload score is obtained by averaging the dimension ratings.

In addition to subjective evaluations, physiological signals were also collected to deepen the participant's state during the experimental trials. Electro-dermal activity (EDA) data and heart data through photopletismogram (PPG) were obtained, respectively, at 4 Hz and 64 Hz using the non-invasive biosensor Empatica E4 wristband (Fig. 7a). From EDA and PPG signals, stress and arousal indicators were derived for each HRC configuration, as explained in the following sub-section. Eye-tracking data were also collected using the Tobii Pro Glasses 3 at 100Hz (Fig. 7b), providing information on cognitive processes and user experience, and analyzing pupillometry and metrics related to fixations and saccades.

2.2 Data processing and modeling

Table 2 provides a summary of all the dependent and independent variables included in the analysis.

From physiological signals, potential artifacts were identified and removed. By using the MATLAB "Ledalab," EDA signals were decomposed in tonic components and phasic components through continuous decomposition

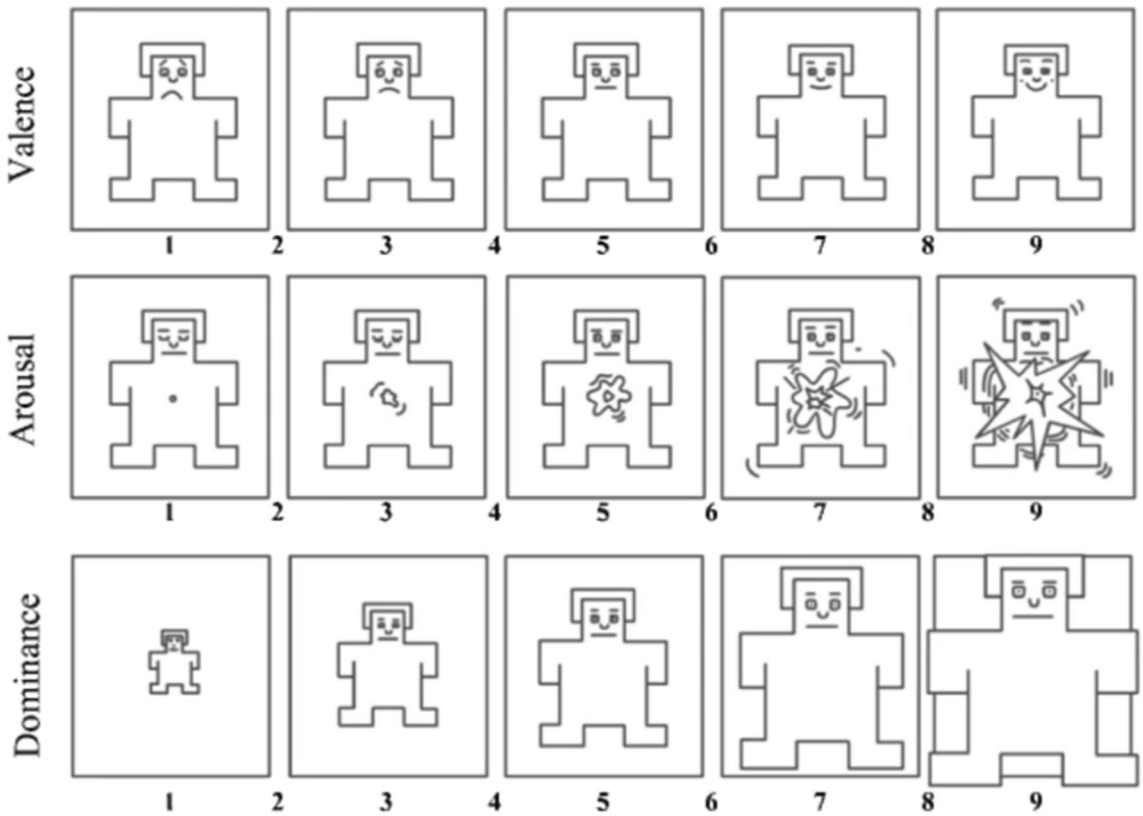


Fig. 5 Self-assessment manikin (SAM) [11]

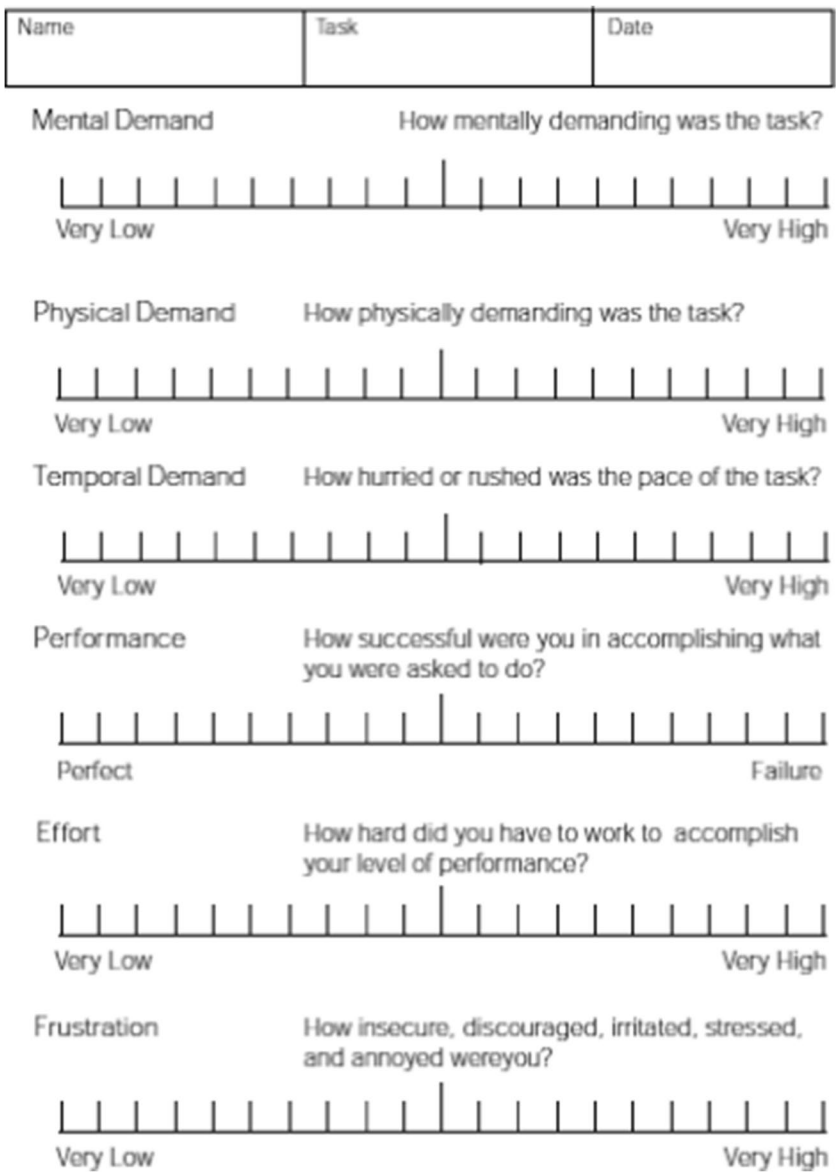


Fig. 6 NASA-TLX questionnaire [13]



(a) *(b)*

Fig. 7 The Empatica E4 [14] (**a**) and the wearable eye-tracking Tobii Pro Glasses 3 [15] (**b**)

analysis (CDA) [16]. The tonic component is characterized by the skin conductance level (SCL), which represents the long-term fluctuations in EDA that are not directly derived from external stimuli. Short-term EDA fluctuations elicited by an external stimulus represent a phasic component. From the phasic component, Skin conductance responses (SCRs), which are amplitude differences between the SCL and response peaks, are detected. In the present study, the average SCR (*EDA – average SCR*) and SCL (*EDA – average SCL*) were calculated for each HRC configuration, representing stress response and mental strain indicators, respectively. Regarding heart data, NN intervals (i.e., time intervals between systolic peaks) were obtained from PPG. As an HRV measure for stress, the standard deviation of NN intervals (SDNN) was included due to its widespread usage [17, 18].

Table 2 Summary of independent and dependent variables

Type of variable	Category	Variable name	Description
Independent variables	*Fixed factors*	*Size*	Cobot physical size (*Small, Large*)
		Speed	Robot movement speed (*Low, High*)
		Product	Product assembly complexity (*Low, High*)
	Covariate	*Exposure*	Number of exposures to the same configuration (*1, 2*)
	Random block	*Participant*	ID of the participant
Dependant variables	*Interaction quality*	*Q1_Helpful*	Assessment of perceived robot helpfulness (seven-point scale)
		Q2_NotSafe	Assessment of perceived interaction unsafety (seven-point scale)
		Q3_Natural	Assessment of perceived interaction naturalness (seven-point scale)
		Q4_Efficient	Assessment of perceived team efficiency (seven-point scale)
		Q5_Fluid	Assessment of perceived team fluency (seven-point scale)
		Q6_Uncomfortable	Assessment of perceived discomfort (seven-point scale)
		Q7_Trustworthy	Assessment of perceived robot trustworthiness (seven-point scale)
	SAM	*Valence*	SAM dimension assessing how positive the emotion is (nine-point scale)
		Arousal	SAM dimension assessing how agitated a person feels (nine-point scale)
		Dominance	SAM dimension assessing how strong is the dominance feeling (nine-point scale)
	NASA-TLX	*Mental demand*	NASA-TLX dimension representing how mentally demanding a task
		Physical demand	NASA-TLX dimension representing how physically demanding a task
		Temporal demand	NASA-TLX dimension representing how temporally demanding a task
		Performance	NASA-TLX dimension representing how well a task is perceived to be performed
		Effort	NASA-TLX dimension representing how much effort is perceived during the task
		Frustration	NASA-TLX dimension representing how many negative feelings are perceived in a task
		Workload	NASA-TLX dimension representing the overall perceived workload score
	Physiological metrics	*Average pupil diameter*	Average dilatation of pupils [mm]
		Average duration of fixations	Mean duration of fixations in the area of interest [ms]
		Average peak velocity of saccades	Mean of peak velocities of saccades [°/ms]
		Average amplitude of saccades	Mean of saccade amplitudes [°]
		EDA – average SCR	Average skin conductance response [µS]
		EDA – average SCL	Average skin conductance level [µS]
		HRV – SDNN	Standard deviation of heart rate NN intervals [ms]

Eye-tracking data were processed using the "Tobii Pro Lab" software, which filtered raw data and extracted the metrics of interest reported in Table 2. Pupillary dilation is influenced by mental and cognitive processes, and an increase in it is often attributable to a stress response or increased use of mental resources [19, 20]. Fixations represent periods of stable gaze during which the eyes remain relatively still and gain detailed visual information. A longer duration of fixations may be attributable to increased interest in a certain area of interest or increased mental effort resulting from fatigue or difficulty in processing information [19, 20]. Saccades are rapid movements of the eyes, shifting the gaze from one point of interest to another. Analysis of saccade amplitude and velocity provides additional insight into mental effort. A decrease in amplitude as well as velocity of saccades is often related to an increase in mental effort [20, 21].

A series of mixed-effect ordinal logistic regression (MOLR) models were implemented in order to investigate the relationship of the fixed factors of the experiment and their interactions with the subjective responses (i.e., interaction quality and SAM dimensions). The MOLR model was chosen for its suitability in (i) modeling dependent variables defined on an ordinal scale and (ii) handling the participant effect as a random block effect [22, 23]. MOLR models are based on cumulative link mixed models (CLMMs) and can be specified in terms of cumulative logits as follows (supposing to have K levels of the ordinal response variable, n observations, and J participants):

$$\begin{aligned}\operatorname{logit}\left(\mathbb{P}\left(Y_{ij} \le k\right)\right) &= \log\left(\frac{\mathbb{P}\left(Y_{ij} \le k\right)}{1-\mathbb{P}\left(Y_{ij} \le k\right)}\right) = \theta_k \\ &\quad - \boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta} - u_j \qquad \text{for } k = 1, \ldots, K-1 \end{aligned} \tag{1}$$

with $i = 1, \ldots, n$ and $j = 1, \ldots, J$. The term $\mathbb{P}\left(Y_{ij} \le k\right)$ is the probability that the observation i of the participant j is associated with a rating below or equal to k. Note that the logit is not defined for $k = K$ since $\mathbb{P}\left(Y_{ik} \le K\right) = 1$. The coefficients θ_k are called threshold parameters and act as model intercepts. Data of the independent variables of observation i of participant j are contained in the vector $\boldsymbol{x}_{ij}^{\mathrm{T}}$, whereas the model parameters are in vector $\boldsymbol{\beta}$. The term u_j represents the random effect for participant j, where $u_j \sim N(0, \sigma^2)$. Note that the negative sign in front of $\boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}$ ensures that positive parameters are associated with increased probability for higher levels as the independent variables increase. The "ordinal" package from the software R was used to fit MOLR models [24].

Since the NASA-TLX dimension scores and selected indicators for physiological response can be considered continuous variables, linear mixed models (LMMs) were implemented to explore the relationship with the considered factors and to handle the participant effect as a random block effect. LMMs can be represented as follows:

$$Y_{ij} = \boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta} + u_j \qquad \text{with } i = 1, \ldots, n \text{ and } j = 1, \ldots, J \tag{2}$$

where $\boldsymbol{x}_{ij}^{\mathrm{T}}$ is the vector containing data of the independent variables for observation i of participant j, $\boldsymbol{\beta}$ the vector for the model parameters, and u_j the random effect for participant j with $u_j \sim N(0, \sigma^2)$. Models were fitted using the "lmerTest" package from the software R.

The formula used for the models is reported below using the Wilkinson notation [25]:

$$\begin{aligned} Y \sim 1 &+ \text{Size} * \text{Speed} * \text{Exposure} + \text{Size} * \text{Product} \\ &+ \text{Speed} * \text{Product} + \text{Exposure} * \text{Product} \\ &+ (1|\text{Participant}) \end{aligned} \tag{3}$$

The significance of the model terms will be analyzed and discussed in the following section for each response variable.

3 Results

In this section, the results of the fitted regression models are described and analyzed. See the Appendix for additional details on the resulting models.

3.1 Perceived interaction quality

The relationships between the experimental factors and the different aspects of the perceived quality of interaction with the robot will be presented. In Table 3, the results of the analysis of deviance (ANODE) for each MOLR model are reported, showing the significance of each term. The large cobot was felt during the interaction slightly safer, more efficient, fluid, and trustworthy. In addition, when the cobot movement speed was high, the cobot was perceived as more helpful, efficient, and fluid and also slightly more uncomfortable. The second experience with the same configuration tended to result in an overall improved interaction quality: The cobot was perceived to be safer, more natural, efficient, fluid, and trustworthy. However, it was also perceived as slightly less helpful and less comfortable when moving slowly. By assembling a more complex product, the cobot was perceived in the second experience to be furthermore helpful, natural, efficient, fluid, and comfortable. In addition, assembling a more complex product with higher movement speed resulted in a slight decrease in perceived efficiency and fluency.

3.2 Perceived workload

Table 4 contains the results of the analysis of variance (ANOVA) for each linear mixed model of the NASA-TLX dimensions, showing the significance of each model term.

Table 3 ANODE table with resulting *p*-values for quality of interaction

Model term	Q1_Helpful	Q2_NotSafe	Q3_Natural	Q4_Efficient	Q5_Fluid	Q6_Uncomfortable	Q7_Trustworthy
Size	0.099	0.012*	0.055	0.039*	0.046*	0.602	0.016*
Speed	0.020*	0.098	0.473	< 0.001***	< 0.001***	0.005**	0.703
Exposure	0.034*	< 0.001***	< 0.001***	< 0.001***	< 0.001***	< 0.001***	0.016*
Product	0.834	0.822	0.660	0.799	0.470	0.745	0.164
Size · Speed	0.891	0.894	0.727	0.651	0.587	0.398	0.311
Size · Exposure	0.485	0.373	0.945	0.776	0.453	0.964	0.922
Speed · Exposure	0.660	0.446	0.545	0.627	0.540	0.098	0.621
Exposure · Product	0.003**	0.147	0.008**	0.002**	0.001**	0.043*	0.541
Size · Product	0.355	0.624	0.602	0.896	0.231	0.658	0.421
Speed · Product	0.080	0.671	0.635	< 0.001***	0.031*	0.300	0.357
Size · Speed · Exposure	0.304	0.632	0.639	0.306	0.965	0.701	0.361

*$0.05 > \boldsymbol{p} \geq 0.01$

**$0.01 > \boldsymbol{p} \geq 0.001$

***$\boldsymbol{p} < 0.001$

In general, it emerges that the perceived workload decreases in the second experience, especially with the more complex product, due to a learning effect. Looking in detail at the various dimensions of workload, there emerges a decrease in mental demand, physical demand, temporal demand, and effort and an increase in performance. Especially with the more complex product, the second experience resulted in a further decrease in perceived mental demand and effort and an increase in performance. Regarding perceived performance, a decrease was observed with the more complex product, especially with even higher speed. Concerning frustration, a slight increase was observed in the second experience, especially when the cobot was moving slower. In fact, a decrease in frustration was observed with higher speed in the second experience.

3.3 Affective state

Table 5 contains the results of the ANODE for each MOLR model for the SAM dimensions, showing the significance of each model term. Initially, the higher speed with a complex product resulted in more negative emotions due to more initial effort on the part of the participant. However, in the second experience, the higher speed was more appreciated once they became familiar with the product assembly. In terms of arousal, an increase was observed with the highest speed and in the second experience. However, there was also a significant decrease in the second experience with the more complex product. The larger cobot implied an initial feeling of less dominance, but this increased during the second experience, especially with the more complex product.

Table 4 ANOVA table with resulting *p*-values for NASA-TLX

Model term	Workload	Mental demand	Physical demand	Temporal demand	Performance	Effort	Frustration
Size	0.458	0.270	0.577	0.753	0.073	0.883	0.901
Speed	0.295	0.146	0.705	0.158	0.837	0.586	0.708
Exposure	< 0.001***	< 0.001***	< 0.001***	< 0.001***	< 0.001***	< 0.001***	0.049*
Product	0.149	0.072	0.908	0.685	0.045*	0.206	0.472
Size · Speed	0.633	0.721	0.717	0.789	0.599	0.806	0.441
Size · Exposure	0.603	0.291	0.700	0.898	0.375	0.833	0.716
Speed · Exposure	0.156	0.527	0.290	0.156	0.505	0.173	0.031*
Exposure · Product	0.012*	0.021 *	0.605	0.623	0.001**	0.002**	0.250
Size · Product	0.649	0.142	0.101	0.292	0.277	0.531	0.932
Speed · Product	0.179	0.921	0.346	0.867	0.026*	0.442	0.070
Size · Speed · Exposure	0.524	0.958	0.614	0.293	0.949	0.112	0.312

*$0.05 > \boldsymbol{p} \geq 0.01$

**$0.01 > \boldsymbol{p} \geq 0.001$

***$\boldsymbol{p} < 0.001$

Table 5 ANODE table with resulting *p*-values for SAM dimensions

Model term	Valence	Arousal	Dominance
Size	0.141	0.730	0.036*
Speed	0.912	0.002**	0.180
Exposure	0.072	< 0.001***	< 0.001***
Product	0.596	0.929	0.864
Size · Speed	0.353	0.679	0.432
Size · Exposure	0.714	0.882	0.404
Speed · Exposure	0.023*	0.682	0.744
Exposure · Product	0.590	< 0.001***	0.022*
Size · Product	0.982	0.549	0.593
Speed · Product	< 0.001***	0.377	0.287
Size · Speed · Exposure	0.192	0.173	0.321

*$0.05 > \boldsymbol{p} \geq 0.01$

**$0.01 > \boldsymbol{p} \geq 0.001$

***$\boldsymbol{p} < 0.001$

3.4 Physiological parameters

Analysis of multiple physiological parameters allows for a broader representation of the mental and cognitive aspects involved during the process. At the same time, finding concordances between multiple physiological signals can provide stronger evidence of the observed phenomena. Table 6 contains the results of the ANOVA for each linear mixed model of the physiological measures considered. Looking at HRV, no particularly significant terms emerged. Focusing on EDA, only a significant reduction in average SCR and SCL was noted in the second experience, mainly due to a learning effect. Regarding eye-tracking metrics, interesting effects emerged in pupillary dilation, peak saccade velocity, and saccade amplitude. The larger cobot and higher speed resulted in an increase in pupillary diameter, likely due to higher user engagement situations. A decrease in pupillary diameter was also observed when the user interacted with the larger cobot while assembling the more complex product. Another decrease emerged in the second experience, denoting a learning effect. Having to assemble a more complex product resulted in a slightly smaller saccade amplitude, indicative of slightly higher cognitive effort. However, in the interaction with the larger cobot and in the second experience, increases in the peak velocity of saccades were observed with the more complex product, indicative of an alleviation of cognitive load.

4 Discussion

The quantitative analysis revealed the influence of the configuration factors, i.e., cobot size, speed, and product complexity, on the various response variables related to user experience, highlighting some interesting relationships.

Table 6 ANOVA table with resulting *p*-values for physiological measures

Model term	Average pupil diameter	Average duration of fixations	Average peak velocity of saccades	Average amplitude of saccades	EDA – average SCR	EDA – average SCL	HRV – SDNN
Size	< 0.001***	0.089	0.304	0.458	0.214	0.084	0.631
Speed	< 0.001***	0.180	0.628	0.648	0.126	0.331	0.638
Exposure	< 0.001***	0.994	0.087	0.161	< 0.001***	0.005**	0.158
Product	0.138	0.791	0.091	0.035*	0.156	0.284	0.152
Size · Speed	0.638	0.517	0.638	0.890	0.752	0.675	0.428
Size · Exposure	0.821	0.257	0.160	0.092	0.494	0.305	0.850
Speed · Exposure	0.133	0.703	0.116	0.090	0.796	0.910	0.361
Exposure · Product	0.186	0.061	0.030*	0.125	0.348	0.499	0.696
Size · Product	< 0.001***	0.471	0.037*	0.072	0.666	0.589	0.683
Speed · Product	0.245	0.862	0.202	0.192	0.580	0.180	0.307
Size · Speed · Exposure	0.718	0.502	0.601	0.334	0.975	0.701	0.883

*$0.05 > \boldsymbol{p} \geq 0.01$

**$0.01 > \boldsymbol{p} \geq 0.001$

***$\boldsymbol{p} < 0.001$

Perceived workload was mainly influenced by product complexity and learning effect, meaning that the cobot size and movement speed did not have a significant influence. Therefore, the workload was influenced mainly by the product to be assembled.

Cobot size and movement speed were most influential in the perceived quality of interaction. When the movement speed was higher, the cobot was perceived as more helpful, efficient, and fluid, especially during the second experience of the user. Surprisingly, the larger cobot generally resulted in better interaction quality, with higher perceived safety, efficiency, fluency, and trustworthiness. Although there was lower initial dominance with it, the learning effect totally compensated for the loss.

From a physiological point of view, observing pupillometry, SCR, and SCL confirmed the learning effect, which resulted in a general decrease in mental strain. From the pupillary dilation, it was also noticed a higher activation by the participants with the large cobot and high speed, which can be interpreted as a higher engagement. Through saccade-related metrics, a slight positive influence of product complexity on cognitive load was noted and how it was also slightly mitigated by the large cobot. No significant effect emerged in terms of HRV.

The participants' unstructured feedback was in line with the results obtained from the questionnaires. A large proportion of participants showed a slight preference toward the larger cobot, perceiving it to be more reliable and stable during movements. This result is in line with the one observed by Karwowski et al. [7], in which industrial workers felt safer in the presence of an arm of a larger robot than a smaller one. Initially, before starting the experiment, some participants had the impression that interacting with the large cobot might result in more anxiety. However, they noticed that during the assembly process, their attention toward the cobot was mainly focused on its end-effector (i.e., the gripper), which was the same for both cobots. As a result, during the interaction, they did not pay special attention to the entire size of the cobot, which suggests that the characteristics of the end-effector may have a greater influence. This aspect needs to be explored further in future work. Another aspect that will be interesting to explore in future work is to compare, in terms of cobot size, the effects of unexpected movements by the cobot on the user experience of the operator.

5 Conclusions

The proposed study aimed to explore the effect of cobot size, in conjunction with robot movement speed and product complexity, on user experience during an HRC assembly process.

Experimental results revealed that the larger cobot was the most preferred during collaborative assembly and did not induce significant stress responses or negative mental strain. The large cobot was perceived as slightly safer, more efficient, fluent, and reliable. Initially, a lower sense of dominance and higher activation in terms of pupillary dilation were found with the large cobot, but with the second exposure to the same configuration, these effects were compensated.

This result can be particularly useful when a company decides to implement collaborative robotics in a production process and needs to choose the cobot model taking into account also human factors. A larger cobot, such as the UR10e, offers the opportunity to work with larger payloads, potentially facilitating reconfigurability and deployment in production processes. Additionally, there was no significant interaction observed between cobot speed and size. This is because, during interaction, the operator's attention was more focused on the cobot's end-effector (which was the same for both) than the entire body of the cobot. This phenomenon suggests that it would be beneficial, in future work, to explore the effect of different end-effectors on user experience in HRC. The results of this study should indeed be interpreted as an initial exploration of the effect of cobot size that needs further investigation of other factors that may intervene during HRC. Future work will focus on strengthening the results obtained by evaluating the effect of the cobot larger than the UR10e and other cobot settings, e.g., with the base resting on the floor or the work surface, on user experience.

Appendix. List of estimated parameters of the models

In this section, the estimated parameters for the fitted models are reported. Tables 7, 8, 9, and 10 contain the estimated parameters of the models for interaction quality dimensions, SAM dimensions, NASA-TLX dimensions, and physiological signals, respectively.

Table 7 MOLR model parameter estimates for each interaction quality dimension

Model term	Model parameter	Coefficient						
		Q1 Helpful	Q2 NotSafe	Q3 Natural	Q4 Efficient	Q5 Fluid	Q6 Uncomfortable	Q7 Trustworthy
Thresholds (intercepts)	θ_1	−8.44	1.24	−8.37	-	-	2.86	−7.36
	θ_2	−6.78	3.18	−6.26	−5.33	−4.03	4.43	−6.25
	θ_3	−5.41	3.83	−4.24	−3.06	−2.70	5.90	−4.64
	θ_4	−3.81	4.82	−2.70	−1.63	−0.90	6.92	−3.76
	θ_5	−0.59	6.14	−0.13	0.78	1.69	7.68	−1.19
	θ_6	2.28	7.26	3.67	3.42	3.84	-	1.12
Size	*Large*	0.20	−0.86	0.32	0.12	1.10	0.11	1.40
Speed	*High*	0.67	0.28	0.13	1.73	1.74	1.74	0.25
Exposure	*2*	−0.76	−1.32	0.43	−0.04	0.67	0.05	1.20
Product	*High*	−0.36	0.59	−1.19	0.37	0.70	0.84	−1.17
Size · Speed	*Large · High*	0.64	−0.19	0.06	0.74	−0.26	0.39	−1.03
Size · Exposure	*Large · 2*	0.93	0.26	0.28	0.65	−0.36	−0.29	−0.45
Speed · Exposure	*High · 2*	0.79	0.24	0.55	0.26	−0.29	−1.57	−0.74
Exposure · Product	*2 · High*	1.63	−1.04	1.40	1.57	1.67	−1.59	−0.33
Size · Product	*Large · High*	−0.51	−0.34	0.27	−0.07	−0.61	−0.33	−0.44
Speed · Product	*High · High*	−0.96	0.30	−0.25	−1.76	−1.11	−0.78	0.50
Size · Speed · Exposure	*Large · High · 2*	−1.12	0.67	−0.49	−1.03	−0.04	0.58	0.98
Random effect	-	St. deviation						
		Q1 Helpful	Q2 NotSafe	Q3 Natural	Q4 Efficient	Q5 Fluid	Q6 Uncomf.	Q7 Trustw.
Participant	-	3.52	2.61	3.10	2.24	2.42	2.54	2.39

Table 8 MOLR model parameter estimates for each SAM dimension

Model term	Model parameter	Coefficient		
		Valence	Arousal	Dominance
Thresholds (intercepts)	θ_1	-	−2.68	-
	θ_2	−8.41	0.18	−7.77
	θ_3	−7.42	2.18	−7.27
	θ_4	−5.54	3.97	−6.34
	θ_5	−2.11	5.23	−5.13
	θ_6	−0.59	6.91	−3.27
	θ_7	1.67	9.61	−0.63
	θ_8	5.29	-	2.34
Size	*Large*	−0.11	0.50	−0.92
Speed	*High*	−0.25	1.29	−0.62
Exposure	*2*	−0.50	0.38	0.37
Product	*High*	1.48	1.61	−0.60
Size · Speed	*Large · High*	1.14	−0.45	0.90
Size · Exposure	*Large · 2*	0.48	−0.76	0.09
Speed · Exposure	*High · 2*	1.82	−0.85	0.68
Exposure · Product	*2 · High*	0.27	−2.36	1.19
Size · Product	*Large · High*	0.01	−0.29	0.28
Speed · Product	*High · High*	−1.95	−0.43	−0.55
Size · Speed · Exposure	*Large · High · 2*	−1.32	1.32	−1.02
Random effect	-	St. deviation		
		Valence	Arousal	Dominance
Participant	-	3.21	2.44	2.90

Table 9 LMM parameter estimates for each NASA-TLX dimension

Model term	Model parameter	Coefficient						
		Workload	Mental demand	Physical demand	Temporal demand	Performance	Effort	Frustration
Intercept	-	18.44	15.26	14.36	22.85	21.86	15.56	11.25
Size	*Large*	−0.95	−0.45	0.54	1.04	−4.51	−1.10	1.38
Speed	*High*	0.95	2.71	−0.21	3.26	−1.80	−0.67	2.51
Exposure	*2*	−3.89	−6.07	−2.88	−7.58	−2.76	−3.78	2.00
Product	*High*	8.83	12.46	−0.32	5.21	12.68	8.65	2.58
Size · speed	*Large · High*	0.31	−0.72	0.98	2.30	−1.26	3.52	−3.57
Size · exposure	*Large · 2*	2.27	2.62	1.01	2.71	2.59	2.65	−1.30
Speed · exposure	*High · 2*	−1.54	−1.37	−0.63	−1.08	−1.64	0.44	−6.37
Exposure · product	*2 · High*	−4.98	−5.46	0.59	−1.44	−8.83	−6.00	−2.30
Size · Product	*Large · High*	−0.89	−3.47	−1.86	−3.09	2.97	−1.20	−0.17
Speed · Product	*High · High*	2.65	0.23	1.07	−0.49	6.13	1.47	3.64
Size · Speed · Exposure	*Large · High · 2*	−2.51	−0.25	−1.14	−6.17	−0.35	−6.10	4.06
Random effect	-	St. deviation						
		Workload	Mental demand	Physical demand	Temporal demand	Performance	Effort	Frustration
Participant	-	13.30	11.68	10.03	17.95	16.64	12.17	11.81

Table 10 LMM parameter estimates for each physiological indicator

Model term	Model parameter	Coefficient						
		Average pupil diameter	Average duration of fixations	Average peak velocity of saccades	Average amplitude of saccades	EDA – average SCR	EDA – average SCL	HRV – SDNN
Intercept	-	3.759	417.84	232.05	7.226	0.0841	1.1494	62.82
Size	*Large*	0.201	−19.04	−3.94	−0.004	−0.0231	−0.1869	−0.50
Speed	*High*	0.129	27.30	7.98	0.464	0.0124	−0.0243	5.21
Exposure	*2*	−0.125	−16.76	−0.96	0.151	−0.0416	−0.2217	−3.70
Product	*High*	0.361	−3.13	−21.73	−1.361	0.1102	0.4953	5.22
Size · Speed	*Large · High*	−0.024	−22.35	−0.25	−0.179	0.0074	0.0059	−7.48
Size · Exposure	*Large · 2*	−0.017	7.82	−9.06	−0.573	0.0169	0.1066	0.33
Speed · Exposure	*High · 2*	−0.053	−17.82	−9.86	−0.575	−0.0075	−0.0824	−8.43
Exposure · Product	*2 · High*	0.038	32.01	10.29	0.333	−0.0243	−0.1121	3.10
Size · Product	*Large · High*	−0.098	12.25	9.88	0.391	−0.0112	−0.0896	3.20
Speed · Product	*High · High*	−0.034	−2.96	−6.03	−0.283	0.0144	0.2231	8.03
Size · Speed · Exposure	*Large · High · 2*	0.021	22.75	4.91	0.417	0.0016	0.1273	2.33
Random effect	-	St. deviation						
		Average pupil diameter	Average duration of fixations	Average peak velocity of saccades	Average amplitude of saccades	Average SCR	Average SCL	SDNN
Participant	-	0.550	173.5	21.58	1.352	0.1873	1.2699	20.61

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

Author contribution All authors contributed to the study's conception and design. Material preparation, data collection, and analysis were performed by R. Gervasi and M. Capponi. The first draft of the manuscript was written by R. Gervasi and M. Capponi under the supervision of L. Mastrogiacomo and F. Franceschini. All authors read and approved the final manuscript.

Funding Open access funding provided by Politecnico di Torino within the CRUI-CARE Agreement. This study was carried out within the "Empathically enhanced robot for the collaboration with HUmans in MANufacturing (E.HU.MAN)" project – funded by the Ministero dell'Università e della Ricerca – within the PRIN 2022 program (D.D.104 – 02/02/2022).

Data availability The datasets generated and analyzed during this study are not currently publicly available.

Declarations

Ethics approval The authors respect the ethical guidelines of the journal.

Consent to participate Informed consent was obtained from all individual participants included in the study.

Consent for publication Not applicable.

Conflicts of interest/Competing interests The authors declare that they have no conflict of interest.

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. Wang L, Gao R, Váncza J, Krüger J, Wang XV, Makris S, Chryssolouris G (2019) Symbiotic human-robot collaborative assembly. CIRP Ann 68:701–726. https://doi.org/10.1016/j.cirp.2019.05.002
2. Maddikunta PKR, Pham Q-V, B P, Deepa N, Dev K, Gadekallu TR, Ruby R, Liyanage M (2022) Industry 5.0: a survey on enabling technologies and potential applications. J Ind Inf Integr 26:100257. https://doi.org/10.1016/j.jii.2021.100257
3. Kopp T, Baumgartner M, Kinkel S (2021) Success factors for introducing industrial human-robot interaction in practice: an empirically driven framework. Int J Adv Manuf Technol 112:685–704. https://doi.org/10.1007/s00170-020-06398-0
4. Gervasi R, Mastrogiacomo L, Franceschini F (2020) A conceptual framework to evaluate human-robot collaboration. Int J Adv Manuf Technol 108:841–865. https://doi.org/10.1007/s00170-020-05363-1
5. 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 126:1213–1231. https://doi.org/10.1007/s00170-023-11197-4
6. Rahimi M, Karwowski W (1990) Human perception of robot safe speed and idle time. Behav Inf Technol 9:381–389. https://doi.org/10.1080/01449299008924252
7. Karwowski W, Rahimi M, Parsaei H, Amarnath BR, Pongpatanasuegsa N (1991) The effect of simulated accident on worker safety behavior around industrial robots. Int J Ind Ergon 7:229–239. https://doi.org/10.1016/0169-8141(91)90006-8
8. Hiroi Y, Ito A (2008) Are bigger robots scary? —The relationship between robot size and psychological threat—. In: 2008 IEEE/ASME International Conference on Advanced Intelligent Mechatronics. Xi'an, China, pp 546–551
9. Universal Robots (2024) Collaborative robotic automation | Cobots from Universal Robots. https://www.universal-robots.com/. Accessed 15 Mar 2024
10. 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 106:36. https://doi.org/10.1007/s10846-022-01744-8
11. Bradley MM, Lang PJ (1994) Measuring emotion: the self-assessment manikin and the semantic differential. J Behav Ther Exp Psychiatry 25:49–59. https://doi.org/10.1016/0005-7916(94)90063-9
12. Lang PJ (1980) Behavioral treatment and bio-behavioral assessment: computer applications. In: Sidowski JB, Johnson JH, Williams TA (eds) Technology in mental health care delivery systems. Ablex, Norwood, pp 119–137
13. Hart SG, Staveland LE (1988) Development of NASA-TLX (task load index): results of empirical and theoretical research. In: Hancock PA, Meshkati N (eds) Advances in Psychology. North-Holland, Amsterdam, pp 139–183
14. Empatica (2022) E4 wristband. https://www.empatica.com/research/e4. Accessed 13 Apr 2022
15. Tobii (2023) Tobii Pro Glasses 3. https://www.tobii.com/products/eye-trackers/wearables/tobii-pro-glasses-3. Accessed 20 Dec 2023
16. Benedek M, Kaernbach C (2010) A continuous measure of phasic electrodermal activity. J Neurosci Methods 190:80–91. https://doi.org/10.1016/j.jneumeth.2010.04.028
17. Kim H-G, Cheon E-J, Bai D-S, Lee YH, Koo B-H (2018) Stress and heart rate variability: a meta-analysis and review of the literature. Psychiatry Investig 15:235–245. https://doi.org/10.30773/pi.2017.08.17
18. Young MS, Brookhuis KA, Wickens CD, Hancock PA (2015) State of science: mental workload in ergonomics. Ergonomics 58:1–17. https://doi.org/10.1080/00140139.2014.956151
19. Sharafi Z, Shaffer T, Sharif B, Guéhéneuc Y-G (2015) Eye-tracking metrics in software engineering. In: 2015 Asia-Pacific Software Engineering Conference (APSEC). New Delhi, India, pp 96–103
20. William J, Murugesh R (2020) Potential eye tracking metrics and indicators to measure cognitive load in human-computer interaction research. J Sci Res 64:168–175. https://doi.org/10.37398/JSR.2020.640137
21. Di Stasi L, Renner R, Staehr P, Helmert J, Velichkovsky B, Cañas J, Catena A, Pannasch S (2010) Saccadic peak velocity sensitivity to variations in mental workload. Aviat Space Environ Med 81:413–417. https://doi.org/10.3357/ASEM.2579.2010
22. Agresti A (2002) Categorical data analysis. John Wiley & Sons, New Jersey
23. Gervasi R, Mastrogiacomo L, Franceschini F (2023) An experimental focus on learning effect and interaction quality in

human–robot collaboration. Prod Eng 17:355–380. https://doi.org/10.1007/s11740-023-01188-5
24. Christensen RHB (2019) Ordinal: regression models for ordinal data. https://CRAN.R-project.org/package=ordinal
25. Wilkinson GN, Rogers CE (1973) Symbolic description of factorial models for analysis of variance. J R Stat Soc Ser C Appl Stat 22:392–399. https://doi.org/10.2307/2346786

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