Assessing Trust in Collaborative Robotics with Different Human-Robot Interfaces

Matteo Menolotto
Tyndall National Institute
University College Cork
Cork, Ireland
matteo.menolotto@tyndall.ie

Dimitrios-Sokratis Komaris
Mechanical, Biomedical and Design Engineering Department
Aston University
Birmingham, UK
s.komaris@aston.ac.uk

Patricia O'Sullivan
Tyndall National Institute
University College Cork
Cork, Ireland
patricia.osullivan@tyndall.ie

Brendan O'Flynn
Tyndall National Institute
University College Cork
Cork, Ireland
brendan.oflynn@tyndall.ie

***Abstract*—Human-robot interfaces (HRIs) serve as the main communication tools for controlling and programming robots in industry 4.0 applications. To be effective, the design of these interfaces should consider not only functional and morphological characteristics, but also factors influencing human interactions, such as trust. A lack of trust is linked to the underutilization or misuse of collaborative robots, leading to ineffective automation implementation and compromised safety. The assessment of human factors is therefore gaining traction in robotics, with the emergence of both objective and subjective methodologies. Nevertheless, the absence of a holistic approach hinders the development of a unified assessment framework. This study introduces a novel assessment methodology that integrates self-reporting questionnaires with human-centric data collected through wearable sensing technologies. The approach aims to offer a comprehensive evaluation of HRIs, considering both perceptual and behavioral dimensions. Empirical testing on three different HRIs substantiates the effectiveness of this methodology. Preliminary results reveal variations in trust levels based on the combination of tasks performed and the specific HRI used for communication with a collaborative robot. These findings not only contribute to advancing our understanding of trust dynamics in human-robot interactions but also lay the groundwork for a more inclusive evaluation framework, enhancing our comprehension of the intricate interplay between humans and robots in the context of smart manufacturing.**

***Index Terms*—human-robot interface, collaborative robotics, trust, IMU, GSR**

I. Introduction

The constant advancement of collaborative robots (cobots), and their implementation in smart manufacturing, is making human machine interaction (HMI) a crucial aspect in many industrial and service sectors. The level of flexibility of the deployment process for cobots, combined with reduced costs and low disruption for the production line, are key factors of such rapid expansion, enabling more and more low- and middle-scale enterprises to implement automation [1].

However, true collaboration in the form of an interaction based on negotiation, intuition and sharing of information is still far from being achieved in collaborative robotics. Ideally, in a cobots application, the intrinsic capabilities of robots (precision, power and repeatability) are combined with the soft skills of the human worker (dexterity, problem solving) to improve specific production factors (e.g. safety, quality and productivity and reduce fatigue [2]) or to introduce automation in previously impossible tasks. To enable this scenario, the mutual exchange of information between human worker and robot has to pass through direct (e.g. verbal, visual) and indirect (e.g. body language biophysical signals) communication. To this end, not only the sensing framework connected to the robot needs to boost its level of autonomy, but also the human-robot interface (HRI) needs to evolve.

How effective a tool/object is in enabling the accomplishment of a specific task depends on a combination of physical and functional characteristics. The geometry, size, weight and material, to name a few, have an impact on the usability of the tool; in a similar way, how simple or intuitive the tool is, or how it reacts to manipulation, also have an impact on its usability. Such complex aggregation of human factors is closely related with the concept of ergonomics defined by ISO26800 [3]. For this reason, HMI is currently a multidisciplinary subject, relating computer science with human studies, such as psychology, sociology and arts [4].

Several attempts have been made to investigate how human factors impact on human-robot collaboration (HRC). Gervasi *et al.* [5] developed a conceptual framework combining several latent dimensions to assess HRC. One of those considers human factors, which account for five different sub-dimensions: workload, trust, robot morphology, physical ergonomics and usability. Here, the HRC is analyzed as an emotional and

This work was carried out with the support of Enterprise Ireland and Science foundation Ireland through the SFI Centre CONFIRM (16/RC/3918), VistaMilk (SFI 16/RC/3835), CONNECT Centre for Future Networks and Communications (13/RC/2077) and the Insight Centre for Data Analytics (SFI/12/RC/2289). This work is co-funded under the European Regional Development Fund.

cognitive process, where each sub-dimension was assessed via self-reported questionnaires. Trust, in particular, has been the focus of several works [6]–[8], as some of the attributes of robots, e.g., different degrees of anthropomorphism and autonomy, introduce a level of uncertainty not common for other automated systems. The lack of trust can result in underuse or misuse of the robot's capabilities, reducing productivity and quality [9]. Moreover, low trust levels impact negatively on the overall usage experience of the operator, causing more anxiety and mental stress. Objective methodologies to assess mental stress and frustration can also be found in the literature. In particular, galvanic skin response (GSR) and human-robot proximity detection [7] have been used to assess trust and emotional arousal. Here, however, the results did not conclude any direct correlation between the test conditions and the change in skin conductivity or human-robot distance. Hesitation, as a lack of movement fluency, was instead assessed via inertial measurement units (IMUs) by evaluating the jerk [10], which is correlated with the tremor and the sudden change of movement execution.

In general, in the abovementioned works, the assessment was focused on the direct interaction between the robot and the human, neglecting to account for the role played by the HMI. However, in the majority of the industrial applications, the robot itself is not directly used as a controlled interface. Rather, keyboards, joypads, touchscreens or tablets are the most common HRIs. Understanding how to effectively design such devices so that they can convey trust, for instance, by providing adequate level of functionality for a kind of cobot to control and perform a task, is therefore highly desirable in the field.

In this work, we combined subjective and objective data to gain a broader understanding on how trust changes by using different HRIs while performing different collaboration tasks with a cobot. Using the taxonomy identified by Semeraro *et al.* [11], two classic collaborative robotics tasks, object handover and collaborative manufacturing, were performed using a cobot and a human subject via three different HRI technologies. A selected set of wearable sensing technologies and questionnaires were used to holistically assess frustration and emotional stress in an attempt to identify factors in the HRI design that affect HRC. The final aim was to develop an assessment framework that is able to objectively quantify trust to offer system designers a valuable tool to identify and optimize the features that affect trust in HMIs for industry 4.0.

II. Methods

A. Participants

A total of 11 subjects (6 male) were recruited for the study from the Tyndall National Institute, with ages ranging from 23 to 53 years (M=31.8, SD=8.1). Before commencing the trial, a comprehensive briefing of the nature of the study was given, followed by the opportunity to confirm consent. We only retained participant information that was relevant to the interpretation of the data: age, sex (M/F) and level of experience with a robot and/or virtual reality/augmented reality (VR/AR) technologies. The study had ethical approval from the university's ethical committee (CREC Review Reference number: ECM 4 (p) 6/7/2021). No information that can directly identify participants was included in the data collection, in accordance with the General Data Protection Regulation 2018 (GDPR) of the EU.

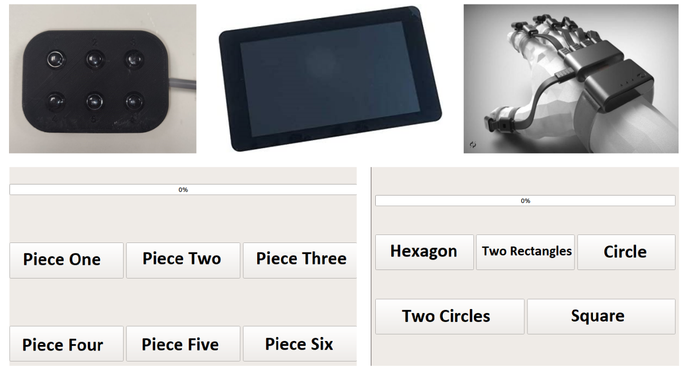

Fig. 1. Upper raw: the three different HRIs tested: from the left, 6-buttons keypad, touchscreen and data glove. Bottom raw: example of the graphical interface of the touchscreen device.

B. Human-Robot Interfaces and Collaborative Robots

Three different HRIs were tested for trust: a button keypad, a touchscreen and a data glove (Fig. 1). The button keypad was designed and 3D printed to allocate 6 buttons ($\varnothing 8\mathrm{mm}$ push button switch) arranged in two rows of three buttons. The touchscreen selected was a Raspberry Pi controlled 7′ capacitive touchscreen that was mounted into a screen frame to facilitate usage and replicate a common tablet-like HRI; a simple graphic user interface was developed for the touchscreen in order to complete the two tasks.

The data glove used is an IMU-based wearable device developed by Tyndall National Institute [12], which mounts two IMUs per finger, one IMU for the palm and one for the wrist. Here only three fingers were used (thumb, index, middle) to maximizing hand dexterity and minimize discomfort. A simple gesture recognition algorithm was created to recognize up to six different gestures. Before commencing the prescribed task with the glove, the participant was asked to follow a simple procedure aimed to train the gesture recognition technique based on six different hand shapes decided by the subject. Later on, the same gestures were used to trigger the robot's movements. The industrial collaborative robot used is a UR16e (Universal Robots, Odense, Denmark), mounting a Robotiq 3-finger adaptive end-effector from (Robotiq, Lévis, Canada). Each HRI was directly connected to the robot via a TCP/IP socket or to the digital input panel of the control box of the robot itself.

C. Wearable Sensors

IMUs and a galvanic skin response (GSR) sensor were used to extract objective data related to the level of stress from the subjects during their interaction with the robot (Fig. 2). Five wearable IMUs form Xsens (MTw Awinda, Movella,

Fig. 2. Wearable sensing technologies used in the trust assessment. From the left: IMU, and GSR sensor.

Enschede, The Netherlands) were placed directly on the upper body: trunk, shoulder, upper arm and forearm of the left and the right arms. The GSR information were extracted using a Shimmer 3GSR (Shimmer, Dublin, Ireland), which includes an IMU, temperature and PPG sensors. In this initial study, only the variations of the GSR were used to infer stress condition of the operator while performing the tasks and questionnaires.

Although both wearable sensors were used to acquire data from the subject, in this preliminary work only the IMUs were used to evaluate jerk during the task. Here, we derived the log dimensionless jerk ($LDLJ$), calculated directly from the IMU's acceleration. As previously described by Melendez-Calderon and Shirota [12], the $LDLJ$ metric is defined as:

$$LDLJ = -ln\left(\frac{t2 - t1}{a_{peak}^2}\int_{t1}^{t2} \dddot{x}(t)^2 + \dddot{y}(t)^2 + \dddot{z}(t)^2 dt\right) \tag{1}$$

where a_{peak} is equal to the magnitude of the peak total acceleration minus the mean total acceleration of the movement, and $t1$ and $t2$ represent the time at beginning and end of the recording, respectively. Such value is commonly reported between -2 to -10, where a number closer to zero corresponds to smoother movements.

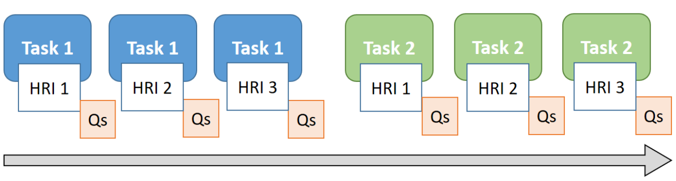

Fig. 3. Block diagram of the experimental protocol. Qs phase represent the administration of the three questionnaires: Charalambous, NASA-TLX and SUS. The order of the task and the HRI were pseudo renadomised.

D. Questionnaires

To subjectively assess the level of trust when interacting with the robot using the three different HRIs, the Charalambous [6] questionnaires were administrated after each iteration. This questionnaire was selected as it is already validated for robot interaction. To take into consideration the effect of the different workload and usability associated with each task and the three HRIs on trust, the NASA-TLX [14] and the System Usability Scale (SUS) [15] questionnaires were also administrated after completion of each step of the assessment protocol (Fig. 3). The items of the SUS and the Trust questionnaires were modified in order for all strongly positive statements to score 1 at the 5-point of the linked scale, neutral statements to score 3, and strongly negative statements to score 5. This was done in accordance with the NASA task load index, where very low demanding tasks score 1 on the linked scale, and very high demanding tasks score 20. Therefore, the best and worst possible total scores for the SUS and the Trust questionnaires were 10 and 50, respectively; similarly for the NASA questionnaire, the best possible score is 6, and the worst is 120. The total scores for each individual test, and the sum of all three questionnaires (with a range from 26 to 220), were compared with Friedman Tests with significance set at $p < 0.05$. Post hoc pairwise comparisons were conducted with Wilcoxon signed-rank tests with a Bonferroni correction applied, resulting in a significance level set at $p < 0.017$.

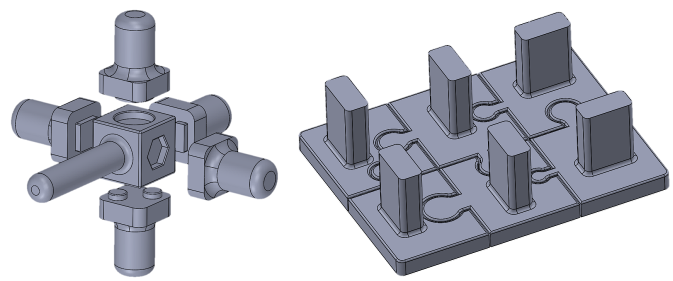

Fig. 4. From the left: 5 pieces + core assembly and 6 pieces jigsaw puzzle parts. Each part was custom designed and 3D printed with black PLA.

E. Trust Assessment Protocol

Each subject performed two different tasks which simulate typical collaborative robotic applications: object handover and collaborative manufacturing [11]. Each task was carried out three times, every time using a different HRI from those under test. Both the order of the task and the HRI used were pseudo randomised to exclude independent factors, such as tiredness or task memorisation.

1) *Object handover (puzzle)*: this task consists of the assembly of a 6-pieces jigsaw puzzle (Fig. 4, right). Each piece was designed to include a grabbing attachment to facilitate the pick and place operation of the cobot's 3-finger adaptive end-effector. The six pieces were placed in front of the operator but outside their reaching volume. When the command was prompted via one of the HRIs, the cobot picked the selected piece and released it in front of the subject at mid-air. The subject, therefore, after giving the command, extended their arm in position, ready to receive the piece, grabbed it, and assembled the puzzle (Fig. 5). The task is concluded when the jigsaw was correctly assembled in full.
2) *Collaborative manufacturing (3D core)*: the second task consists of connecting 5 components to a central core (Fig. 4, left), held by the cobot. The core is formed by a

a) HMI 1: Keypad b)HMI 2: Touchscreen c) HMI 3: Data glove

Fig. 5. Use of the three different HRIs during the puzzle tasks. From the left: the keypad, the touchscreen and the data glove.

cube with 5 sides designed to receive a component via a dedicated socket, and one side connected to a rod for the end-effector to grasp. After prompting the robot, the end-effector rotated the core exposing a different side every time. Each side has a socket with a unique geometry (e.g., circular, triangular, hexagonal), which corresponds to a specific component. The operator grabbed the desired component, assembled it to the core and commanded the robot to assume another orientation via a different HRI. The task is successfully competed when all 5 components were correctly assembled to the core. An overview of the whole operation is shown in Fig. 6.

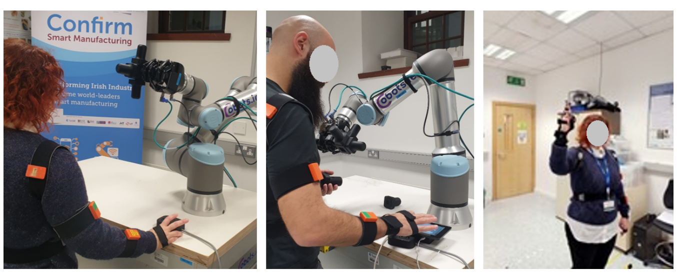

Fig. 6. Use of the three different HRIs during the 3D core tasks. From the left: the keypad, the touchscreen and the data glove.

After the completion of each task, the volunteer was asked to complete the three self-reported questionnaires, for a total of six times

III. Results

Only one participant had previously interacted with a robot, and the average experience level with VR/AR technology was rated at 2.6 out of 10. The median total scores for the button keypad, the data glove and the touchscreen for the puzzle task were equal to 45 (26 to 87), 67 (42 to 107) and 48 (26 to 96), respectively (Fig. 7, left). Similarly, for the 3D core task (Fig. 7, right), the median total scores were 50 (26 to 97), 79 (29 to 157) and 41 (25 to 86). There were statistically significant differences in the total scores for both tasks ($p = 0.033$ for the puzzle task, and $p = 0.001$ for the 3D core task).

There was a statistically significant difference in the perceived effort as measured with the NASA Task Load Index on which type of technology (button, touchpad, glove) was used to manipulate the robotic arm for the puzzle task ($p = 0.018$). Post-hoc analysis with Wilcoxon signed-rank tests returned a statistically significant reduction in the perceived effort when using the button vs when using the glove ($p = 0.007$). There were no statistically significant differences in the scores of the SUS and the Trust questionnaires for the puzzle task. Similarly, for the 3D core test, there were statistically significant differences in the perceived trust in the operation of the robotic arm, for all three questionnaires (Trust, $p = 0.018$; NASA, $p = 0.001$; SUS, $p = 0.001$), as shown in Table I. Post-hoc pairwise comparisons demonstrated that the glove was inferior to both the touchpad and button for the completion of the 3D core task ($p < 0.008$). There were no significant differences between the operation of the robotic arm with the touchpad or the button for both tasks ($p > 0.058$).

TABLE I
Friedman Test Results

Task	**Friedman test (p-values), significance at $p = 0.05$**			
	Q1 (Trust)	***Q2 (NASA TLI)***	***Q 3 (SUS)***	***Sum***
Puzzle	0.913	**0.018**	0.145	**0.033**
3D Core	**0.018**	**0.001**	**0.001**	**0.001**

TABLE II
Wilcoxon Test Results

Task	**Wilcoxon signed-rank pairwise tests (p-values) Significance at $p = 0.017$ after Bonferroni adjustment**				
	Comparison	***Q1 (Trust)***	***Q2 (NASA TLI)***	***Q3 (SUS)***	***Sum***
Puzzle	Touchpad - Button	0.721	0.999	0.796	0.878
	Glove - Button	0.729	**0.007**	0.126	**0.016**
	Glove - Touchpad	0.933	0.061	0.068	0.13
3D Core	Touchpad - Button	0.083	0.066	0.058	0.017
	Glove - Button	0.168	**0.008**	**0.007**	**0.007**
	Glove - Touchpad	0.058	**0.008**	**0.003**	**0.003**

As for the $LDLJ$, higher values were evaluated for the button interface and the touchpad respectively for the puzzle task and the 3D core task (Fig. 8), which highlight a smoother interaction between operator and robot with these two HRIs. Table III summarises the results.

TABLE III
$LDLJ$ Values for the different HRI

HRI	**Task**	
	Puzzle	***3D core***
Button	-7.97±2.36	-7.22±2.94
Touchpad	-8.46±1.25	-4.88±2.11
Glove	-8.96±2.26	-6.16±2.85

IV. Discussions

In our research, we introduced a novel holistic approach, integrating self-reported questionnaires and human-centric data from wearable sensors. This framework facilitates the assessment of HRIs in collaborative robotic applications, with

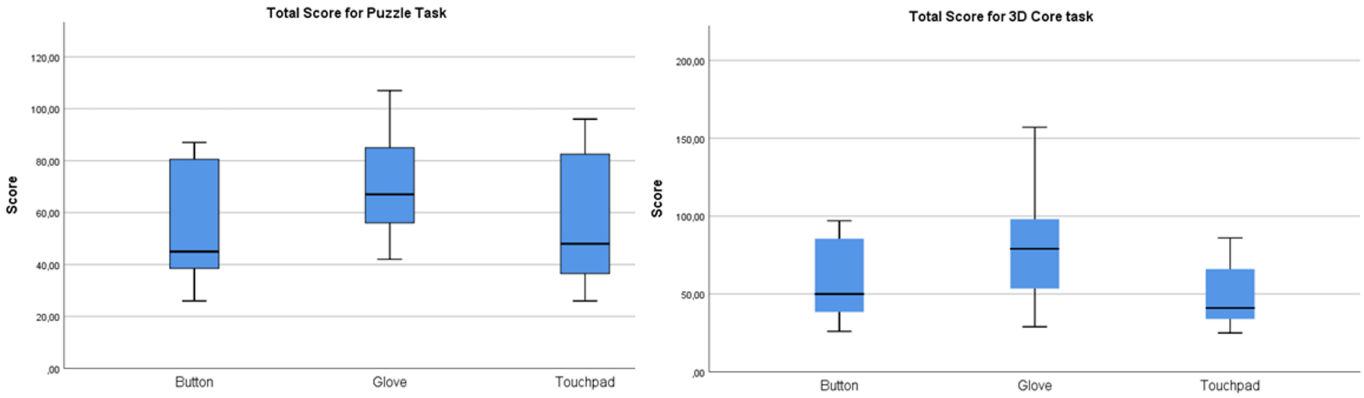

Fig. 7. Boxplot of the total score from all three questionnaires for the puzzle task (left) and the 3D Core task (right).

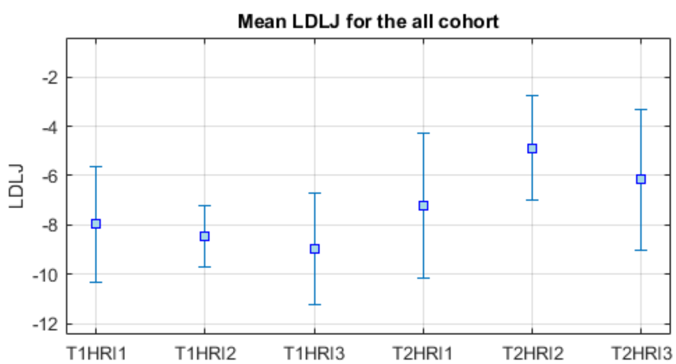

Fig. 8. Different value of $LDLJ$ for T1=puzzle and T2=3D core, using HRI1=buttons, HRI2=touchpad and HRI3=Data glove.

a focus on trust, a critical factor for safe and effective collaboration with cobots. Workload and usability were also evaluated through self-reported questionnaires, recognizing their influence on trust.

Notably, the data glove proved less suitable for collaborative tasks, particularly due to task typology and operational intricacies. Simple interfaces like keypads were found more apt for straightforward operations, such as pick and place, compared to the less reliable and less intuitive nature of gesture-based actions triggered by the glove. Usability tests highlighted the impediment the glove posed to task completion, associated with a steeper learning curve and less intuitive gesture mapping. Button and touchpad HRIs, on the other hand, demonstrated intuitive design, visually correlating puzzle pieces to button numbers and positions. This intuitive mapping enhanced trust, reduced workload, and improved usability scores.

Wearable technologies, including haptic gloves, are gaining prominence in industry and smart manufacturing, serving both as tools for extracting human-centric information and as HRIs, such as in teleoperation [16]. However, a comprehensive consideration of ergonomics is crucial in the design of such gloves to optimize usability. This research contributes valuable insights into the practical implications of different HRIs in collaborative robotics, offering guidance for their effective implementation in Industry 4.0 settings.

V. Limitations and Future Works

In the context of human factors measurement in Industry 4.0 context, our current approach acknowledges certain limitations. While the preliminary results align with comparable studies, the sample size of 11 subjects may be considered small for an in-depth exploration of human factors. Notably, although GSR was part of the sensing framework, it wasn't incorporated into the preliminary results. The data glove, explored as a HRI for its potential in smart manufacturing, revealed ergonomic challenges that demand attention in subsequent iterations. The current haptic data glove's hardware characteristics hindered fine manipulation with the same hand, contributing to stress and frustration in collaborative tasks.

To address these limitations and advance our research, ongoing efforts involve recruiting additional volunteers to achieve a more robust sample size of 30 subjects. Future enhancements include integrating heart rate monitoring through photoplethysmography and video recording to better correlate GSR and Jerk epochs with stress-inducing events. Exploring alternative off-the-shelf data gloves, particularly textile gloves or designs minimizing impact on hand dexterity, will be a focal point in optimizing the HRI for enhanced usability in Industry 4.0 applications.

Acknowledgment

We thank Eoghan Cravengrace for his help with the experimental setup and many insightful discussions.

References

[1] E. Vicentini, “Collaborative robotics: a survey,” Journal of Mechanical Design, vol. 143, no. 4, 2021.

[2] O'sullivan, P., Menolotto, M., O'Flynn, B. and Komaris, D.S., 2022, July. Estimation of Maximum Shoulder and Elbow Joint Torques Based on Demographics and Anthropometrics. In 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) (pp. 4346-4349). IEEE.

[3] ISO 26800:2011: Ergonomics - general approach, principles and concepts, 2011.

[4] D. Hutchison *et al.*, Human Machine Interaction: Research Results of the MMI Program. Springer Berlin Heidelberg, 2009

[5] R. Gervasi, L. Mastrogiacomo, and F. Franceschini, ”A conceptual framework to evaluate human-robot collaboration,” The International Journal of Advanced Manufacturing Technology, vol. 108, no. 3, pp. 841-865, 2020.

[6] G. Charalambous, S. Fletcher, and P. Webb, ”The Development of a Scale to Evaluate Trust in Industrial Human-robot Collaboration,” International Journal of Social Robotics, vol. 8, no. 2, pp. 193-209, 2015.

[7] K. Hald, M. Rehmn, and T. B. Moeslund, ”Human-Robot Trust Assessment Using Motion Tracking & Galvanic Skin Response,” pp. 6282-6287, 2020.

[8] P. A. Hancock, D. R. Billings, K. E. Oleson, J. Y. Chen, E. De Visser, and R. Parasuraman, ”A meta-analysis of factors influencing the development of human-robot trust,” 2011.

[9] R. Parasuraman and V. Riley, ”Humans and Automation: Use, Misuse, Disuse, Abuse,” Human Factors: The Journal of the Human Factors and Ergonomics Society, vol. 39, no. 2, pp. 230-253, 2016.

[10] D. S. Komaris, G. Tarfali, B. O’Flynn, and S. Tedesco, ”Unsupervised IMU-based evaluation of at-home exercise programmes: a feasibility study,” BMC Sports Sci Med Rehabil, vol. 14, no. 1, p. 28, Feb 19 2022.

[11] F. Semeraro, A. Griffiths, and A. Cangelosi, ”Human–robot collaboration and machine learning: A systematic review of recent research,” Robotics and Computer-Integrated Manufacturing, vol. 79, p. 102432, 2022.

[12] M. P. Wilk, J. Torres-Sanchez, S. Tedesco, and B. O’Flynn, ”Wearable Human Computer Interface for Control Within Immersive VAMR Gaming Environments Using Data Glove and Hand Gestures,” pp. 1-9, 2018.

[13] A. Melendez-Calderon, C. Shirota, and S. Balasubramanian, ”Estimating Movement Smoothness From Inertial Measurement Units,” Front Bioeng Biotechnol, vol. 8, p. 558771, 2020.

[14] S. G. Hart, ”NASA-task load index (NASA-TLX); 20 years later,” in Proceedings of the human factors and ergonomics society annual meeting, 2006, vol. 50, no. 9, pp. 904-908: Sage publications Sage CA: Los Angeles, CA.

[15] J. Brooke, ”SUS-A quick and dirty usability scale,” Usability evaluation in industry, vol. 189, no. 194, pp. 4-7, 1996.

[16] M. Menolotto, D. S. Komaris, S. Tedesco, B. O’Flynn, and M. Walsh, ”Motion Capture Technology in Industrial Applications: A Systematic Review,” Sensors (Basel), vol. 20, no. 19, Oct 5 2020.