Journal of Intelligent & Robotic Systems (2026) 112:47
https://doi.org/10.1007/s10846-026-02383-z

REGULAR PAPER

Human-Centered Evaluation of Control and Perception Interface Factors in a Teleoperational Cutting Task

Inês F. Ramos[1] · Keerthi Sagar[2] · Philip Long[3] · Maria Chiara Leva[4] · Ernesto Damiani[1] · Gabriele Gianini[5]

Received: 21 July 2025 / Accepted: 3 March 2026

Abstract

Despite significant advances in teleoperated systems, optimal human performance remains a challenging task due to one key aspect, i.e. *reduced situational awareness*. Active telepresence is one of the major contributing factors affecting operator's situational awareness. Providing the user with multiple perspectives of the remote environment can be very beneficial, although it may also introduce challenges in maintaining effective control. Prior work has largely examined discrete manipulation tasks using single-modality assessments, whereas this study applies a multimodal neuroergonomic approach to a continuous precision-cutting task, providing first insights into control–view alignment and multi-camera perception effects. In this work, we focus on two interaction and interface design factors: (a) *control interaction factor*, where we compare the effects of fixed and camera view-aligned control user interfaces on task-performance; (b) *perception interface factor*: the impact of different visual feedback configurations on operator attention and workload state, and interaction with the interface. A telerobotic system focused on a cutting task is chosen for this evaluation as it demands precision, depth perception, and also continuous feedback. A multimodal bio-sensor network is used to record operator-centric data, i.e Eye-tracking pupil measures, gaze measures and EEG signals, and assess the proposed interfaces using a neuroergonomics framework. User-study results show that using a view-aligned control frame in a multi-camera interface significantly improves cutting quality compared to using a fixed control frame, while the display of multiple camera perspectives in the same window is preferred to a single camera per window, promoting operator visual attention with no significant impact on the mental workload.

Keywords Teleoperation · Human-machine interface · Human-machine interaction · Human-centered design

1 Introduction

Teleoperation systems have become ubiquitous in safety-critical applications involving dangerous tools or environments (space, military, underwater, search and rescue), or performance-critical domains (such as medical surgery). The visibility of the remote workspace is crucial in these systems for safe and successful task completion, typically achieved through video streams from remote cameras. This visibility mirrors the natural coupling between manipulation and

Inês F. Ramos and Keerthi Sagar contributed equally to this work.

✉ Gabriele Gianini
gabriele.gianini@unimib.it

Inês F. Ramos
ines.fernandes@unimi.it

Keerthi Sagar
keerthi.sagar@imr.ie

Philip Long
philip.long@atu.ie

Maria Chiara Leva
mariachiara.leva@tudublin.ie

Ernesto Damiani
ernesto.damiani@unimi.it

1 Department of Computer Science, Università degli Studi di Milano, via Celoria 18, Milan 20133, MI, Italy

2 Robotics and Automation Group, Irish Manufacturing Research Centre, National Science Park, Dublin Road, Mullingar N91 TX80, Co. Westmeath, Ireland

3 Department of Mechanical & Industrial Engineering, Atlantic Technological University, Dublin Road, Galway H91 T8NW, Co. Galway, Ireland

4 School of Food Science and Environmental Health, Technological University Dublin, Grangegorman Lower, Dublin 7 D07 H6K8, Ireland

5 Department of Informatics, Systems and Communication, Università degli Studi di Milano-Bicocca, viale Sarca 336, Milan 20126, MI, Italy

Published online: 26 March 2026

viewpoint in everyday tasks, such as a person crouching to adjust a valve or moving one's head to navigate a cluttered space [1], indicating the operator's innate preference to certain perspectives for optimal task execution.

Although several works on camera perspective assessment have focused on telerobotic tasks, little attention has been paid to telerobotic cutting. Despite the success of teleoperation in diverse fields, it has been seen as an unsuitable modality for manufacturing, mainly because of throughput concerns. We investigate a teleoperational task that replaces an existing manual cutting operation called "skiving", where operator uses a razor-blade to remove heat-shrinks or covering layer material for creating exit ports for catheters, in a clean-room medical device manufacturing context. The existing method presents an ergonomic risk, leads to variation in product quality, and requires significant investment in operator training. Moreover, the cutting operation requires straight and curved cuts on a deformable object, which increases task complexity and limits the viability of full automation. Hence, replacing it with remotely controlled human-in-the-loop telerobotic cutting is advantageous, providing the flexibility and adaptability required for the manipulation of miniature and malleable objects, precise cutting tasks, and low-volume production with high personalization, while maintaining a sterile environment. Considering these characteristics, having sufficient view-points and depth information is crucial for the operator to fully capture the remote environment, while performing collision-free movements.

We propose and evaluate Human Machine Interface (HMI) design solutions based on (a) intuitive control and (b) visual information feedback for situational awareness [2] for this teleoperational cutting task. In this work, experimental trials are conducted using a representative Styrofoam work object for the cutting operation. The Styrofoam workpiece exhibits local deformability characteristics comparable to catheter tubing, thereby enabling the study to focus on the operator's cutting trajectory. As the primary objective of this work is to evaluate operator performance during a teleoperated cutting task and not to replicate the full industrial skiving process used in manufacturing, the use of a Styrofoam surrogate is appropriate. This material provides a practical and controlled means to study cutting behavior while remaining within the scope of the investigation.

The main contributions of this work are:

- HMI solutions based on ergonomic and human-centered principles.
- A multimodal human-centered experimental framework for the evaluation of proposed solutions and for design guidance.
- Detailed analysis of the impact of two teleoperation interfaces and interaction factors (perception and control) on human state, behavior, and performance.

The paper is organized as follows, in Section 2 we describe work related to the design and evaluation of the telerobotic system. The HMI teleoperation designs are proposed subsequently in Section 3 and the framework employed for human-centered evaluation is described in Section 4. The user experiments and results are presented in Sections 5 and 6, respectively. A final discussion of the results and concluding remarks is presented in Section 7 and 8.

2 Related Work

Two major problems exist in remote human-machine interaction: (a) control and (b) awareness of the remote environment. A key factor influencing operator performance in remote environment awareness, as noted in [3], is the field of view (FOV) and the positioning of camera viewpoints. For single camera systems, an adaptive FOV solution can mitigate limited visual information, requiring, for example, the automation of the camera-in-hand robot arm to provide effective viewpoints [4]. Alternatively, multiple fixed cameras can provide a variety of view-points and perspectives, but increase the mental computations required by the operator to integrate information from mismatched camera frames [3]. Fixed frame, in here refers to the camera view whose orientation and position are rigidly attached to the environment rather than the robot or the operator's viewpoint. Because the camera does not move with the robot, the visual information is presented in a static, world-aligned reference frame. In teleoperation, the control frame denotes the coordinate frame in which operator inputs are interpreted and mapped to robot motion. Multiple view-aligned control frames [5] lead to better telemanipulation performance, but without significant differences in user performance between an explicit (button) or implicit (eye-gaze-based) mode of control frame view selection. View-aligned control frames in the literature and in the current teleoperation study refer to control frame aligned along the operator's perspective of the environment, thereby, the operator remotely controlling a robot through a camera feed, the data or commands sent are structured according to how the operator views the environment. Although automatic switching (implicit) is the preferred way to select the active view [5], it can hinder performance if the user utilizes the other views for visual information and not control. In this work, we investigate a three-camera interface approach with explicit selection of view-aligned control, since it is critical that the operator is completely aware and in control of the active control-frame to avoid irreversible cutting errors.

Virtual-(VR), augmented-(AR), mixed-reality (MR), and augmented virtuality (AV) have been used in teleoperation to compensate for degraded video image [3]. They provide additional visual functionalities [3, 6], generate a greater sense of embodiment [7–9], and integrate multiple 2D/3D visual streams in a virtual reality environment, thus enhancing depth perception and spatial awareness [8, 10–12].

Most works use mixed reality for integrating a virtual workspace with 3D depth image information from the real workspace. While being efficient, the quality of these 3D reconstruction techniques are still insufficient for dynamic environments involving small parts, and are typically useful only for large object grasping, manipulation and assembly. In contrast, in tasks requiring real-time quality visual feedback, for instance, cutting tasks, we show that direct camera streams are preferred to a mixed-reality interface with a 3D virtualization of the remote scene.

To evaluate the effectiveness of the proposed user interfaces operators' biosignals can me are measured. Biosignals offer an objective means of assessing cognitive workload, as elevated cognitive demands require greater allocation of mental resources to maintain task performance; this increased demand manifests itself in measurable changes in several bodily systems, including alterations in cardiac activity, variations in electrical brain signals such as EEG patterns, and modifications in oculomotor behavior such as eye movements and pupil dilation. These physiological responses provide valuable insights into the operator's mental state and can be leveraged for real-time monitoring in high-stakes environments [13, 14]. A substantial body of literature addresses this subject [15–17]. Several physiological-based objective indicators have been shown to complement task-based performance metrics and subjective user questionnaires (e.g., NASA Task Load Index (NASA-TLX)) in the design and evaluation of systems from the user/human perspective [18–20]. On the one hand, task-based performance metrics are not applicable for all teleoperation modes or operator tasks, such as in supervisory roles, and on the other hand, subjective assessments are biased by operator preference and difficult to use for the comparison of small interface changes. Objective physiological indicators of operator state are better suited for the analysis of HMI features as performance shaping factors [21, 22].

This work proposes a human-centered evaluation approach for designing HMIs, based on a neuroergonomics framework [23], which addresses the limitations of the neurophysiological construct of mental workload in human performance analysis. The framework has theoretical foundations underlying the identification of important degraded states of human performance, mapped to two neurophysiological dimensions, arousal, and task engagement. The concept of task engagement captures the objective-oriented aspect of consciousness, and arousal captures the levels of disengagement that occur in high or low stimulation states, as shown in [24]. In this work, we use multimodal physiological indicators of arousal and engagement, and compare them with baseline indicators of mental workload, from a consumer-grade eye-tracking device and mobile electroencephalography (EEG) device, for increased robustness against noise, motion artifacts, and subject variability. The results of this work present evidence supporting the use of engagement and arousal constructs to capture the impact of teleoperation HMI design on the operator state, that cannot be fully captured by the typical assessment methods.

More recently, Large Language Models (LLMs) have been employed to further assist operators in decision-making and task planning, through context-awareness and understanding [25]. Our work is essential for identifying the most effective and contextually relevant information about the human operator to be used as feedback in LLM-based adaptive planning. Such systems have been used to autonomously manage robot routine tasks, allowing the operator to focus on monitoring and higher-level decision-making [26], to support and augment human decision-making with fast processing and analysis of complex sensor data [26], and to facilitate communication through automatic speech recognition, providing another channel for feedback and operator intent recognition [27].

User studies have assessed performance and workload in teleoperation, also using Eye-tracking or EEG modalities, by designing tasks to induce varying levels of workload [5, 28, 29]. In contrast, this study brings the user closer to the work domain under real conditions [30], by selecting a case study-based task and the testing of functionally relevant ergonomic design factors. Moreover, the use of multimodal physiological (both Eye-tracking and EEG) indicators of the same psychophysiological constructs allows us to capture a broader range of individual responses under the same conditions.

The needs and constraints of the case-study task are presented in the next section.

3 Experimental Setup

An operator must perform a smooth cutting operation on a styrofoam workpiece using a hot-wire tool mounted on the robot wrist, as in Fig. 1 (Left), by moving the tool along a collision-free spatial trajectory depicted in Fig. 7I. The cutting action must be executed in a single attempt to ensure success; a correctional action is not possible. The high-magnification camera system relays the current state of the remote environment, however visual occlusions can occur due to the three-dimensional nature of the task.

3.1 Teleoperation Cell

The teleoperation cell consists of two linked stations: the operator side and the remote robot station, as shown in Fig. 2. The system uses a KUKA KR4-Agilus robot with a hot wire

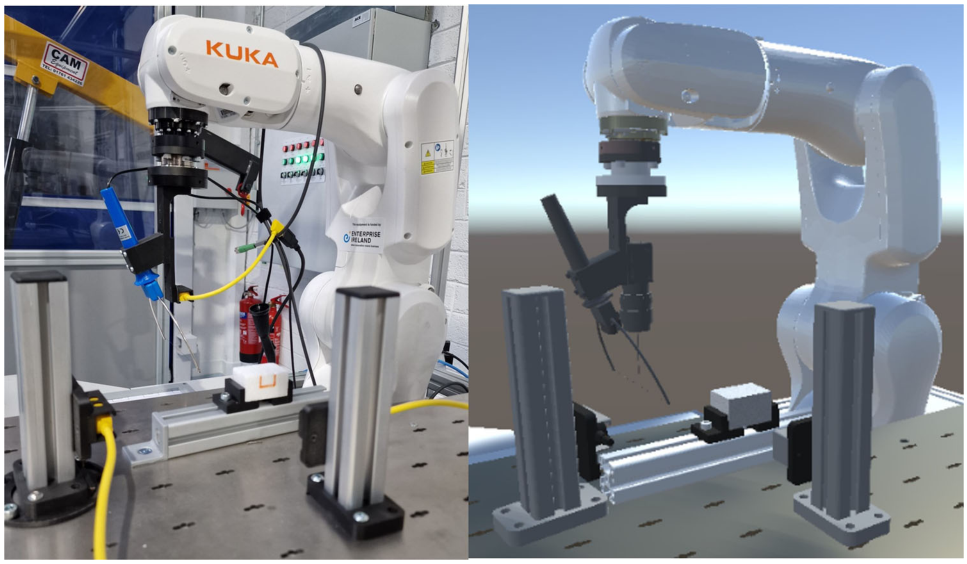

Fig. 1 Experimental setup. (Left) KUKA KR4-Agilus station shown a hot-wire cutting setup. (Right) Real-time digital visualization (user-interface) during execution

tool, controlled via a Phantom Omni haptic device through ROS-Noetic. Three cameras (two fixed, one wrist-mounted) monitor the cutting area. A Unity interface displays video feeds and representative virtual model, connected through the Unity-Robotics-Hub toolbox over an ethernet network. The styrofoam workpiece features a guiding line for trajectory planning.

3.2 Interface Design

Under the assumption that HMI factors can affect the cognitive/mental state of the operator and consequently the teleoperation task performance, we built upon prior human-centered solutions to propose four user interface designs. These HMIs are used to test two formulated hypotheses regarding their impact on the operator.

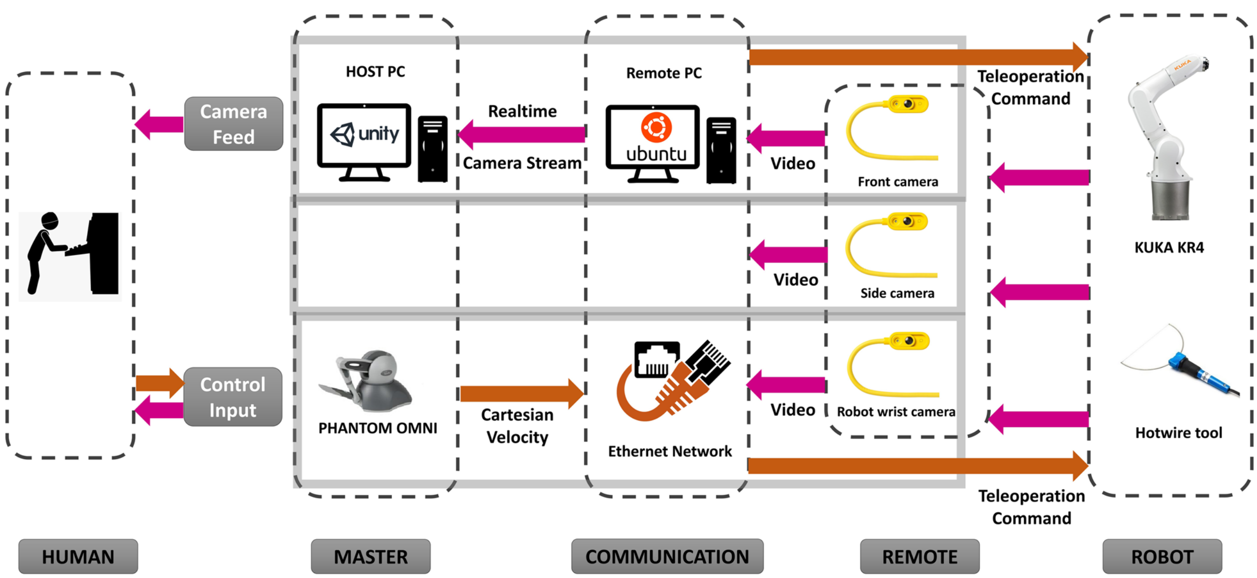

Fig. 2 Overview of the proposed human-in-the-loop telerobotic system architecture and communication structure, connecting the operator station and the remote robot station. The pink arrows represent the data received from the robot station, i.e. real-time video streams from three cameras. The orange arrows represent the Cartesian velocity input data sent from the remote station by the operator using the phantom omni device to control the robot. All data communication is enabled through Ethernet network between the Remote station (Ubuntu-OS) to the operator station (Windows-OS) to ensure low-latency

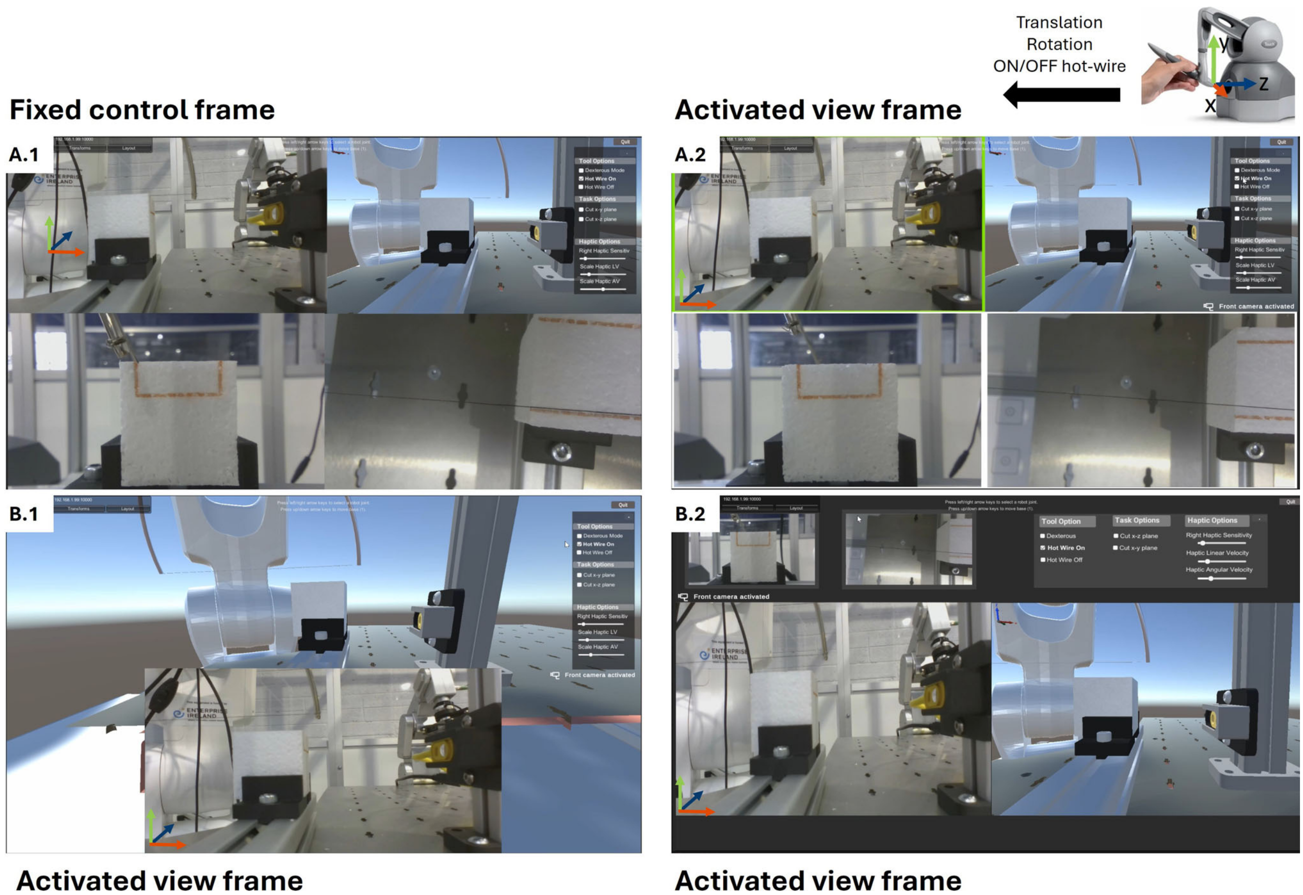

Fig. 3 The HMI designs for teleoperation of a cut task: (**A.1**) "Fixed control HMI" with a four-window split screen with three cameras and 3D environment, using a fixed control frame; (**A.2**) "Explicit view-aligned control HMI" with a green border highlighting the active control frame, switchable by clicking another camera stream; (**B.1**) "Single view focus HMI" with a two-window design showing the 3D environment and a view-aligned control frame, with the active view changeable by clicking the corresponding camera on the 3D model, which highlights it in red to help map the camera's workspace location to the displayed view; (**B.2**) "Multiple view visual hierarchy HMI" with a four-window layout with the active control frame and 3D environment centered, and smaller camera streams at the top left, switchable by clicking a stream

All interfaces' designs shown in Fig. 3, provide three camera streams and a 3D virtual robot environment. The cameras provide real-time workpiece visualization and cutting tool perspective. The virtual 3D environment (VE) is shown for global spatial awareness and depth perception (Fig. 1). A "free-roaming camera" is added to the VE to allow the user to interact with the scene in three dimensions, without being restricted to the fixed camera perspective views.

3.2.1 Experimental Hypotheses

Hypothesis 1 - "Fixed control HMI" (design A.1 in Fig. 3) vs. "Explicit view-aligned control HMI" (design A.2 in Fig. 3): Multiple view-aligned control frames (design A.2) enhance cut task performance and interaction efficiency compared to a fixed control frame (design A.1), in a multi-camera mixed-reality interface [5]. Design A.2 allows users to switch control frames by selecting a view-aligned frame (highlighted in green), requiring the operator to keep the awareness of which is the active control frame.

Hypothesis 2 - "Single view focus HMI" (design B.1 shown in Fig. 3) vs. "Multiple view visual hierarchy HMI" (design B.2 shown in Fig. 3): Displaying a single control-aligned camera view in separate windows (design B.1) improves remote awareness and depth perception with lower mental workload [31], compared to displaying multiple camera streams simultaneously (design B.2), which increases distraction and control confusion [3].

4 Human-Centered Design Evaluation Framework

We use the neuroergonomics framework as a model of human performance for the proposed design evaluation. Two types of biosensor are used to capture the relevant psychophysiological constructs, EEG and Eye-tracking. EEG offers high temporal resolution and low latency measures of neural activity in response to external stimuli or mental processes [32–35]. Eye tracking can be performed with less intrusive sensors (e.g. glasses or sensors attached to a screen) and provide temporal and spatial information on blinks, gaze fixations, and pupil changes, although with much lower temporal resolution compared to EEG [36].

5 User Experiments

5.1 Study Protocol

The within-group experiment was implemented using Tobii Pro Lab version 1.181.37603, separately for each of the hypotheses, assigning each subject to one of the hypotheses. The experiment received ethical approval N° REC-20-52C and all subjects gave their informed consent. Each subject tested the two conditions associated with the hypothesis, assigned in counterbalanced order. Subjects were asked to perform a total of six tests (cut six workpieces) using two designs/conditions, with a baseline recording of 30 seconds, small rest periods in between trials, and a longer rest between conditions. After each trial, the participant assessed their perceived state and performance on the task using self-rating questionnaires (created with Psytoolkit version 3.4.0). The subjects performed a familiarization trial before the recording started. Participants must control the hot wire at the end-effector of the robotic arm to cut a (2cm x 1cm) rectangular profile section on a styrofoam workpiece. The task starts with the end effector located at a fixed home position and ends with the participant clicking on a quit button on the interface, after the cut is performed to the satisfaction of the participant.

The final dataset was collected for 27 subjects (17 males, 10 females) within the ages of 24 and 64 (mean age = 36, std age = 11), with varied reported level of previous experience with teleoperation (#low experience level 1 – 5, #experience level 2 – 8, #experience level 3 – 5, #high experience level 4 - 9). Each hypothesis was assigned 13 different subjects. Sample size estimation was not conducted due to data collection and participant recruitment constraints.

5.2 Bio-Sensor Data Collection and Pre-Processing

The subjects performed the tasks while wearing the mBrainTrain's wireless EEG cap (Smarting mobi, mBrainTrain, Serbia), with a sampling rate of 500Hz, and facing a Tobii Pro Nano Eye-tracking device mounted on the task monitor, sampling eye data at a rate of 60Hz. The EEG and Eye-tracking devices were set up and calibrated at the beginning of the experiment, and the signals were synchronized during collection with the Lab Streaming Layer (LSL).

Before metric computation, both EEG and Eye-tracking require pre-processing, to handle noise and other artifacts. A conservative pre-processing pipeline of EEG recordings is chosen with manual removal of bad components using EEGLAB [37], to avoid biased findings. EEG data are filtered with 40 Hz low-pass filter and 1 Hz high-pass filter, then bad channels are identified visually, removed and interpolated, remaining channels are re-referenced to the average, followed by Artifact Subspace Reconstruction [38] for artifact correction with std dev parameter set to 13, and Independent Component Analysis with the use of the Independent Component Label tool [39] for the identification of non-brain components. The cleaned recordings are segmented for each baseline and trial and the associated features are computed. For pupil size recordings, the data are segmented for each baseline and trial, and consequently filtered to remove invalid samples associated with blink or pupil dilation speed artifacts, by sample imputation with the mean of 20 neighboring samples. The resulting signal for the right and left eye is averaged, followed by the computation of pupil size features. Gaze fixation durations and time samples are obtained directly from the Tobii Pro Lab software. The location and duration of blinks are identified from the coordinated loss of data for both eyes, with a duration between 100 and 400 milliseconds.

To mitigate the variation between trials in physiological signals, the average of the baseline values is subtracted from the trial data to obtain the normalized trial signal. Other metrics associated with behavior, such as blinks and gaze, are not subject to the normalization process.

5.3 Human-Centered Metrics

5.3.1 EEG-Based Features

Under user state metrics, three EEG indexes based on changes in spectral power are calculated, as seen in Fig. 4 I. for one trial. Mental Workload Level (MWL) is estimated by the ratio between the frontal midline theta band power (4–7 Hz) and parietal midline alpha band power (8–12 Hz) [40]. Task engagement index (EI) is monitored by the ratio between theta power, alpha power, and beta band power (13-30 Hz) [32]. The arousal level (AL) can also cause spectral power changes, namely on the F3 and F4 (prefrontal) channels' beta power ratio over alpha power ratio [41]. The trial-based mean and standard deviation features are extracted from the normalized trial values.

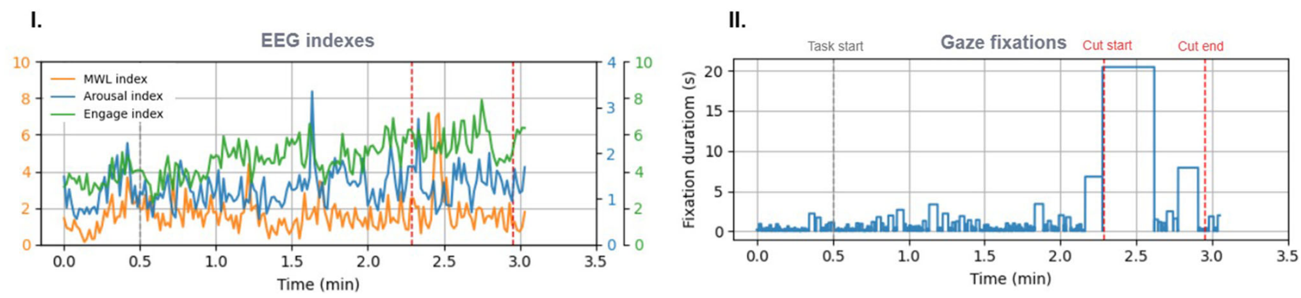

Fig. 4 User state physiological indicators: **I**. EEG-based MWL, Arousal, Engagement indexes, depicting a high peak of MWL in orange and depression of arousal in blue, during the cut stage, and slow increase of engagement in green during the task. **II**. Gaze fixation signals, showing a long fixation aligned with the start of the cut. Grey stripes mark the task start, and red lines indicate the start and end of the cut

5.3.2 Eye-Tracking-Based Features

Under Eye-tracking metrics associated with the internal user state, statistical features are extracted for the normalized pupil diameter signal, gaze fixations as seen in Fig. 4II, and the identified blinks. Under Eye-tracking user behavior metrics, the transition and stationary gaze entropy are computed for the trials, using AOI gaze hit sequence data, considering the set of AOIs defined for the interfaces (S_AOI = front_cam, side_cam, robot_cam, virtual_env, option_menu). The transition entropy provides insights into the pattern of use of the different AOIs in the interface, such that the higher the entropy, the more randomness there is in the gaze transitions, indicating an exploratory visual attention pattern. While the stationary entropy of the gaze can inform about the use of specific AOIs for the task, such that a high value indicates equally distributed visual attention between the AOIs, and a low value indicates a concentration on specific AOIs. The entropy metrics are computed following the equations in [42]. To assess gaze focus on specific interface AOIs, the gaze fixation time ratio on each AOI is computed, visualized in Fig. 6I. through the use of heat maps.

5.3.3 Reported User State Features

Under reported user state metrics, NASA-TLX questionnaire [43] measures are obtained for each trial, on a scale between 1 (Very low) - 10 (Very High), for the perceived mental demand, physical demand, temporal demand, performance, effort, and frustration. The self-assessment Manikin (SAM) questionnaire [44] measures are obtained for the emotional valence on a scale between 1 (Negative/Unhappy) - 5 (Positive/Happy), the emotional arousal on a scale between 1 (Calm/relaxed) - 5 (Excited/vigilant), and the emotional dominance on a scale between 1 (Not in control) - 5 (In control). The Engagement questionnaire (EQ) [45] measures are obtained on a scale between 1 and 7, corresponding to the level of the subject's reported active involvement and level of reported purposeful intent. The average scores obtained across subjects for each design can be visualized in Fig. 5. An additional usability questionnaire was performed at the end of the experiment to collect general preference information between interfaces A and B, used in an exploratory analysis.

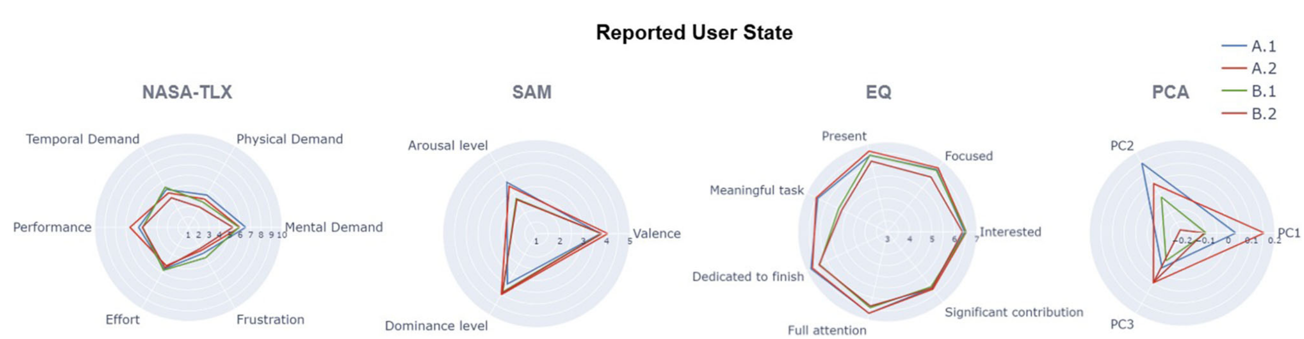

Fig. 5 Reported user state results: averaged NASA-TLX, SAM, and EQ scores across conditions. PCA reveals: (PC1) A.2 with highest task meaningfulness, valence, and dominance, while B.1 and B.2 were lowest; (PC2) A.1 had the highest mental/physical demand, frustration, and arousal, decreasing from A.2 to B.2; (PC3) A.2 and B.2 showed the lowest temporal demand and highest arousal

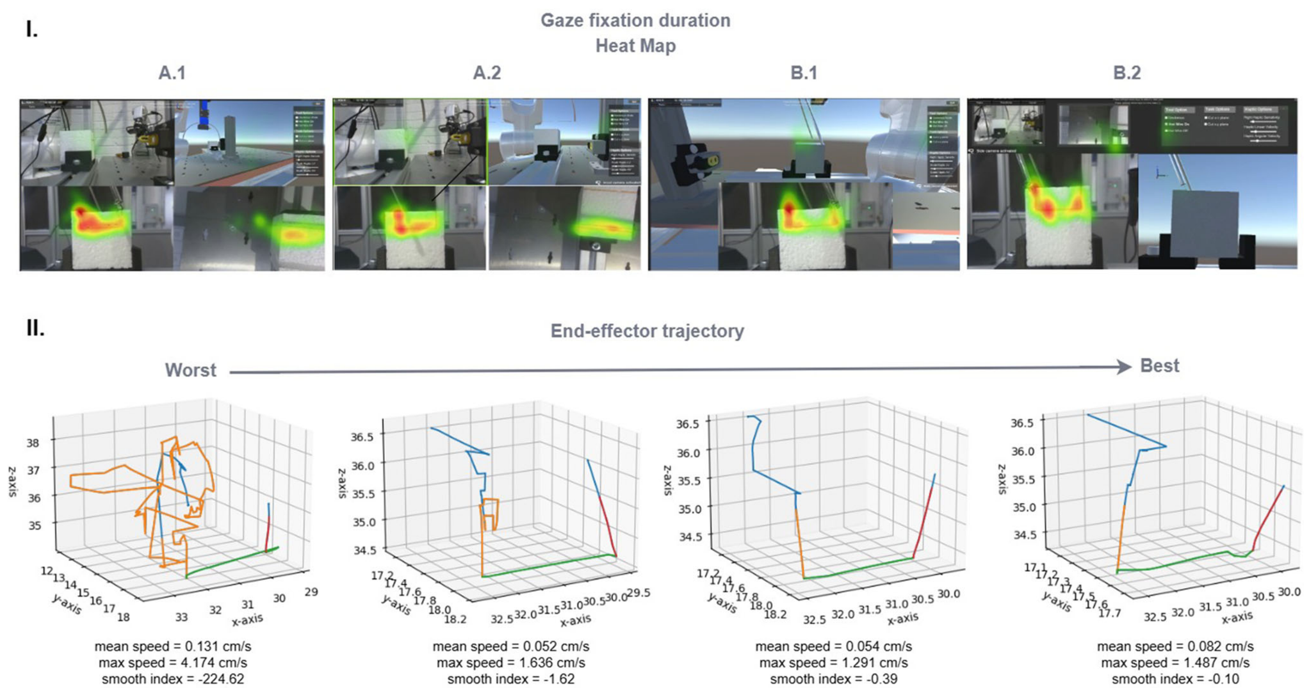

Fig. 6 User behavior data: **I**. Fixation heatmaps reveal limited attention to the virtual environment and the front camera view, affecting the users' situational awareness of the position and location of the end effector, focusing most of their attention on following the cut trajectory on the side and the robot cameras. This can be an indication that the mental demand of the cutting task is too high to maintain the situational awareness through the use of the available interface information. **II**. End-effector trajectories from lowest to highest smoothness index. The two worst trajectories depict an user with difficulty in aligning the hot-wire with the cut line, as shown by the orange lines tracing the end-effector position since its first contact with the workpiece

5.3.4 User Behavior Features

User behavior metrics are retrieved from ROS topics saved as Rosbags. The trial features include the number of clicks for translation and rotation commands, changes and rate of change of the active control frame, mean end-effector velocity, path length, cut-to-total trial time ratio and the active time ratio for each control frame ($S_{cf} = front\_cam, side\_cam, robot\_cam$). Path smoothness S (1) is measured to assess control effectiveness and quality using: the log dimensionless jerk for uncontrolled acceleration variations [46], spectral arclength for speed variations [46], lateral deviation (standard deviation of position values perpendicular to movement) and the number of positional peaks along the movement direction, reflecting bumps and corrective re-cutting.

$$S = \left(\frac{1}{n} \sum_{k=1}^{n} (j_k + sl_k) ld_k p_k \right) \frac{l}{100}, n = 3 \quad (1)$$

where n corresponds to the number of discontinued path segments, j_k corresponds to the log dimensionless jerk of the segment, sl_k to the spectral arclength of the segment, ld_k to the lateral deviation of the segment, p_k to the total peaks in the position coordinates along the segment, and l to the path-length. Values are scaled by path length to penalize extended paths. The robot sample trajectories and the corresponding smoothness index are shown in Fig. 6II.

5.3.5 Task Performance Scores

Under user performance metrics, task time is used as a performance score, computed as the time from when the subject clicked the play button on the interface to the time the subject ended the task. A qualitative cut quality score is attributed to each trial workpiece, on a scale from 1 (worst) to 10 (best), considering the worst cut and the best cut of the experimental samples, as shown in Fig.7I. The distribution of scores can be seen in Fig. 7II.

6 Experimental Results

6.1 Statistical Analysis of the Hypotheses

To test the hypotheses, we perform an "within-group" pairwise comparison for each of the two groups: control factor and perception factor. The non-parametric two-sided

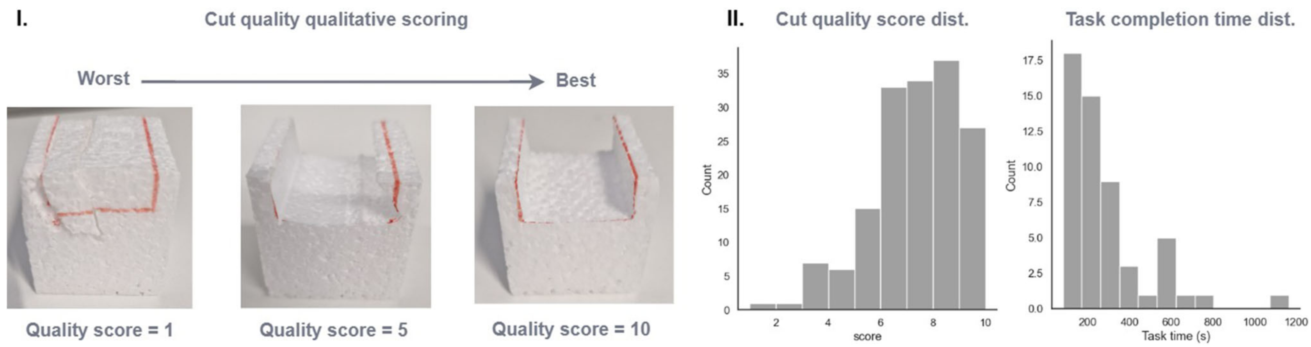

Fig. 7 User performance data: **I**. Cut workpiece examples with quality scores of 1 (worst), 5, and 10 (best). The example for quality score 5 shows the difficulty of the user in accurately controlling the end-effector and following the cut trajectory, creating bumps and slanted cut lines in the workpiece, compared to the example of quality score 10. **II**. Performance metrics show a skewed distribution, with most participants achieving high scores

Wilcoxon signed-rank test with matched pairs is used, as it is suitable for low sample sizes (13 subjects per hypothesis). The effect sizes, calculated as the matched pair rank-biserial correlation for the Wilcoxon sign rank tests, are interpreted as r = 0.1 = small, 0.3 = medium, 0.5 = large, 0.7 = very large. Considering that comparisons are made between the averages of the three trials, it is expected that these individual tests can only identify with high probability true significant differences between conditions (0.8 power, 0.05 α error prob.) that have very large effect sizes (as computed with Gpower [47]). As our goal is not to draw an overall conclusion about the effect of the interface design, but instead explore separately each variable (some that are indicators used in a novel context) related to either task performance, user state or interaction efficiency, we did not correct for multiple comparisons.

6.1.1 Hypothesis 1 - Statistical Analysis of the Control Interaction Factor

We investigated the statistical differences between interfaces A, implementing different control strategies. Table 1 presents significant and non-significant (n.s.) differences, including visible trends.

Physiological and questionnaire metrics assessed internal user states. Although EEG-based indicators did not present

Table 1 Hypothesis 1 - control interaction factor P-values of pair-wise comparison for conditions A.1 and A.2

Metric	Feature (A.1 vs. A.2)	p-value
EEG (User State)	MWL mean (A.2 > A.1)	0.216
	MWL std (A.2 > A.1)	0.110
	Arousal std (A.1 > A.2)	0.057
	Engagement std (A.1 > A.2)	0.127
EYE (User State)	Gaze speed std (A.1 > A.2)	0.006**, r = 0.79
	Blink duration (A.2 > A.1)	**0.042*, r = 0.62**
Self-Report	Mental demand (A.1 > A.2)	0.240
	Perceived performance (A.2 > A.1)	**0.036*, r = 0.73**
Behavior	Clicks (A.1 > A.2)	0.080
	Path length (A.1 > A.2)	0.340
	Cut/Task time ratio (A.1 > A.2)	0.194
Performance	**Cut score (A.2 > A.1)**	**0.039*, r = 0.65**

$**\,0.001 < p \leq 0.01$, $*0.01 < p \leq 0.05$

std - standard deviation, r indicates the effect size

The user state EEG-based indicators did not present significant differences. The EYE-based indicators reveal a significant mean blink duration difference between the conditions, indicating lower workload with A.2, that along with more efficient control (lower number of controller clicks), may have led to the significantly higher user performance significant differences, the differences obtained for other metric groups and the correlation with Eye and self-reported user state metrics observed in Fig. 10, present some support for the EEG user state results. For example, the "Explicit view-aligned control HMI" (A.2) is potentially associated with more stable arousal and engagement levels (lower std, n.s.), indicating consistent physiological activation and task engagement that can have impacted the performance differences observed. Eye-tracking-based differences include significant effects on gaze speed std ($p < 0.01, r = 0.79$) and mean blink duration ($p < 0.05, r = 0.62$), with very high and high effect sizes. Shorter blink durations (46.33 ± 19.57 ms) imply higher workload states, possibly due to visual misalignment with the control frame in A.1. Significant large and very large effects were found on perceived performance (1.12 ± 0.89, $p < 0.05$, $r = 0.73$) and cut quality score (1.06 ± 0.78, $p < 0.05$, $r = 0.65$), both favoring the "Explicit view-aligned control HMI". General behavior trends (n.s.) suggested that A.2 promoted more efficient control (fewer controller clicks) and longer alignment times, positively influencing cut quality.

6.1.2 Hypothesis 2 - Statistical Analysis of the Perception Interaction Factor

Statistical differences between interfaces B, which implement different visual designs, are presented in Table 2, including both significant results and non-significant (n.s.) trends.

The perception interaction factor did not show significant effects on EEG user state indicators, but significantly influenced eye tracking metrics, revealing again an effect on user state. The "Multiple view visual hierarchy HMI" (B.2) resulted in larger pupil diameter (0.21 ± 0.20 mm, $p < 0.01, r = 0.89$) and greater deviation (0.24 ± 0.24 mm, $p < 0.01, r = 0.85$), lower blink rate (0.05 ± 0.04 Hz, $p < 0.01, r = 0.89$), and a n.s. trend for longer blinks, suggesting increased visual attention and lower visual workload than the "Single view focus HMI" (B.1). Gaze patterns also differed significantly. Interface B.2 promoted a balanced use of visual feedback as measured by a higher gaze entropy. In contrast, B.1 showed significantly higher fixation on the side camera (0.12 ± 0.06, $p < 0.05$) and VE (0.09 ± 0.05,

Table 2 *Hypothesis 2* - perception interaction factor P-values for conditions B.1 and B.2

Metric	Feature (B.1 vs. B.2)	p-value
EYE (User State)	**Pupil diameter mean (B.2 > B.1)**	**0.002**, r = 0.89**
	Pupil diameter std (B.2 > B.1)	**0.005**, r = 0.85**
	Gaze speed mean (y) (B.1 > B.2)	0.057
	Gaze speed std (y) (B.1 > B.2)	0.057
	Fixation duration max (B.2 > B.1)	0.244
	Blink duration mean (B.2 > B.1)	0.150
	Blink rate (B.1 > B.2)	**0.002**, r = 0.89**
Self-Report	Mental demand (B.1 > B.2)	0.11
	Temporal demand (B.1 > B.2)	0.12
	Frustration (B.1 > B.2)	**0.025*, r = 0.74**
Behavior	Path length (B.2 > B.1)	0.127
	End-effector velocity mean (B.2 > B.1)	0.057
	Gaze Stationary Entropy (B.2 > B.1)	0.191
	Front cam. fix. ratio (B.2 > B.1)	0.024*, r = 0.76
	Side cam. fix. ratio (B.1 > B.2)	**0.048*, r = 0.62**
	Robot cam. fix. ratio (B.2 > B.1)	0.010**, r = 0.78
	Virtual env. fix. ratio (B.1 > B.2)	**0.001***, r = 0.93**
Performance	Task time (B.1 > B.2)	0.216

$**0.001 < p \leq 0.01$, $*0.01 < p \leq 0.05$

std - standard deviation, r indicates the effect size

The user state EYE-based indicators suggest B.2 interface is associated with lower visual workload and increased attention compared to interface B.1. To understand the differences in interface use, we refer to the significant differences in interface AOI fixation ratio, observing a more balanced gaze fixation spatial distribution with interface B.2 and increased fixation on the VE with interface B.1, as expected. This interface use difference was associated with a reported user state difference, specifically higher frustration with B.1, but no significant difference was observed in the user performance measures

$p < 0.001$), likely due to its manual camera switching procedure, leading to greater reported frustration (1.38 ± 1.06, $p < 0.05$) and perceived demand (n.s.). Though the "Single view focus HMI" aimed to reduce visual load and improve awareness by requiring interaction with the VE for camera selection, user feedback indicated it was cumbersome and disrupted performance. Despite these differences in user state and interface use, no significant differences in performance were observed, although B.1 showed a n.s. trend of longer task times, possibly due to lower mean end-effector velocities (associated with lower control efficiency).

6.2 Interface Design Evaluation

To evaluate the interface designs across groups and participants, a correlation matrix between metric groups was used to identify features best suited for assessing human performance, including variations from experimental conditions and individual differences (e.g., skill, experience). Features with significant correlations ($p < 0.05$) and Pearson's $r \geq 0.1$ were prioritized for interface comparisons. The sensitivity to actual and perceived performance was estimated using effect sizes (Pearson's r), with thresholds: small ($|r| = 0.1$), medium ($|r| = 0.3$), and large ($|r| = 0.5$).

A Principal Component Analysis (PCA) was performed on normalized questionnaire data to reduce dimensionality and capture meaningful reported user states. The first component (PC1) reflected task preference, combining positive contributions from perceived task meaningfulness, valence, dominance, and negative from frustration. The second (PC2) reflected higher mental and physical workload, arousal, and frustration, while the third (PC3) captured perceived arousal positively and temporal demand negatively. PC2 and PC3 suggest higher activation/vigilance may link to higher task demand or lower perceived temporal demand, as the task had no time limit.

The results in Fig. 10 indicate that few individual features exhibit a strong linear correlation with task performance, specifically EEG metrics. However, EEG features along with reported experience level, did demonstrate a medium-strength correlation with perceived task demand from the questionnaires, suggesting that individual physiological features alone may be insufficient for performance prediction due to complex and potentially non-linear relationships. Task behavioral metrics showed strong correlations with performance but remain highly task-dependent, with limited applicability in other scenarios. Reducing extreme cognitive demands and enhancing user comfort may better help prevent performance degradation.

A comparison of all four interfaces using physiological features may be influenced by subject-wise variations, as normalization across subjects for all conditions was not feasible. However, significant correlations were still observed. The larger deviations in EEG arousal ($r = 0.35$) and engagement index ($r = 0.21$) were positively correlated with task time, suggesting that in short tasks such as the present (mean task time = 4.72 minutes, as observed in Fig. 7II), the difficulty in maintaining activation and engagement can reduce efficiency. The "Explicit view-aligned control HMI" (A.2) showed a lower trial-averaged EEG arousal standard deviation as shown in Fig. 8, panel I), while the "Fixed control HMI" (A.1) showed higher EEG engagement and arousal standard deviation as seen in panel I. and II. of Fig. 8.

Eye tracking features showed small to medium correlations with task quality (pupil diameter: $r = 0.26$, gaze fixation deviation: $r = 0.25$) and task efficiency (pupil diameter: $r = -0.36$, gaze fixation deviation: $r = -0.27$), along with inverse correlations between blink rate and performance measures (cut quality: $r = -0.22$, task time: $r = 0.29$). The larger pupil diameter and lower blink rates suggest higher task engagement and visual attention, which contribute to improved efficiency and quality. Fixation duration varied across task stages: the alignment stage involved shorter fixations as all three cameras were used for end-effector alignment, while the cutting stage required longer fixations on the side camera for line tracking, as seen in Fig. 4II. Consistent long fixations may indicate perseveration and reduced situational awareness. Both interfaces A.2 and B.2 exhibited generally higher mean pupil diameter and lower blink rates (Fig. 8III, IV).

The perceived/reported user state features demonstrated a significant correlation with task completion time but did not capture variations in cut quality. This suggests participants may have associated longer task duration with higher task demand rather than cut accuracy or that quality variations were difficult to perceive. As shown in the PCA graph in Fig. 5, the "Explicit view-aligned control HMI" was considered the preferred interface. However, both A.1 and A.2 were associated with higher perceived mental and physical demand, frustration, and arousal than the "Multiple view visual hierarchy HMI".

Control efficiency indicators, including the number of controller clicks, mean end-effector velocity, and path length, correlated with task time ($r = 0.52$, $r = -0.41$, $r = 0.56$, respectively), while trajectory smoothness correlated more with control quality and skill ($r = 0.37$ with cut quality and $r = 0.22$ with experience level). The use of the front camera as the control frame correlated with longer task times ($r = 0.26$), suggesting lower control usability compared to other views. Still, the "Multiple view visual hierarchy HMI" (B.2) promoted a more balanced use of control frames potentially associated with achieving generally

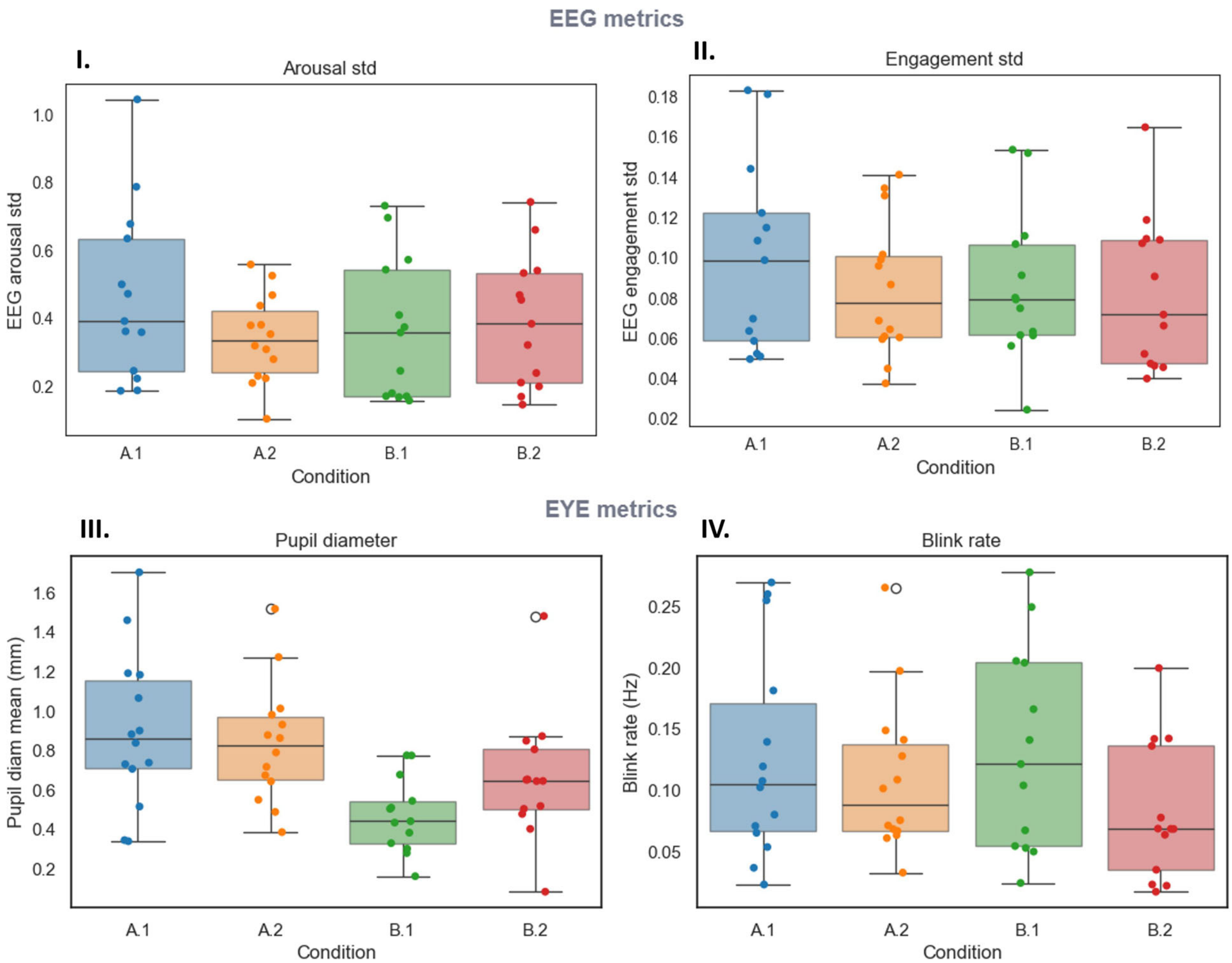

Fig. 8 User state metrics box-plot of subject-wise averages for each condition. **I** & **II** - Subjects presented overall higher EEG engagement and arousal index deviation throughout the trials with interface A.1 the "Fixed control HMI", potentially associated with difficulty in maintaining activation and engagement. **III** & **IV** - Subjects presented overall lower pupil diameter and higher blink rate with interface B.1 the "Single view focus HMI", suggesting lower task engagement and visual attention

lower task completion times observed in Fig. 9IV, while the "Single view focus HMI" (B.1) showed higher use of the side and robot cameras, and the "Explicit view-aligned control HMI" (A.2) exhibited cases of excessive reliance on the side camera, as shown in Fig. 9I. These results suggest that limited use of control frames could be detrimental for the efficiency of tasks with complex manipulation and alignment, and support the conclusion from Hypothesis 1 that multiple control frames are more effective than a single one. In contrast, the visual use of the cameras, measured by fixation duration ratios and gaze entropy, did not directly correlate with performance, as observed in Fig. 10. Although interface B.1 showed the highest fixation ratios in the VE, as shown in Fig. 9II, overall the VE was used less frequently than the cameras, as shown in Fig. 6I, and did not provide effective visual or performance feedback.

Longer task times were significantly correlated with shorter cut phases and extended alignment stages ($r = -0.41$ with the Cut / Task time ratio), suggesting that task delays were primarily due to the alignment stage, highlighting the need for additional manipulation training.

Potential confounders in the statistical comparisons and correlations include subjects' teleoperation experience and initial eye fatigue, assessed via a pre-experiment questionnaire. The level of experience was positively correlated with the mean diameter of the pupil ($r = 0.34$) and the second PC of the characteristics of the reported state of the user ($r = 0.34$), linked to higher perceived performance and lower frustration. Eye fatigue showed a medium-strong correlation with blink rate ($r = 0.48$), possibly influencing its sensitivity as a performance indicator. In longer tasks, blink rate may be less reliable for assessing user state.

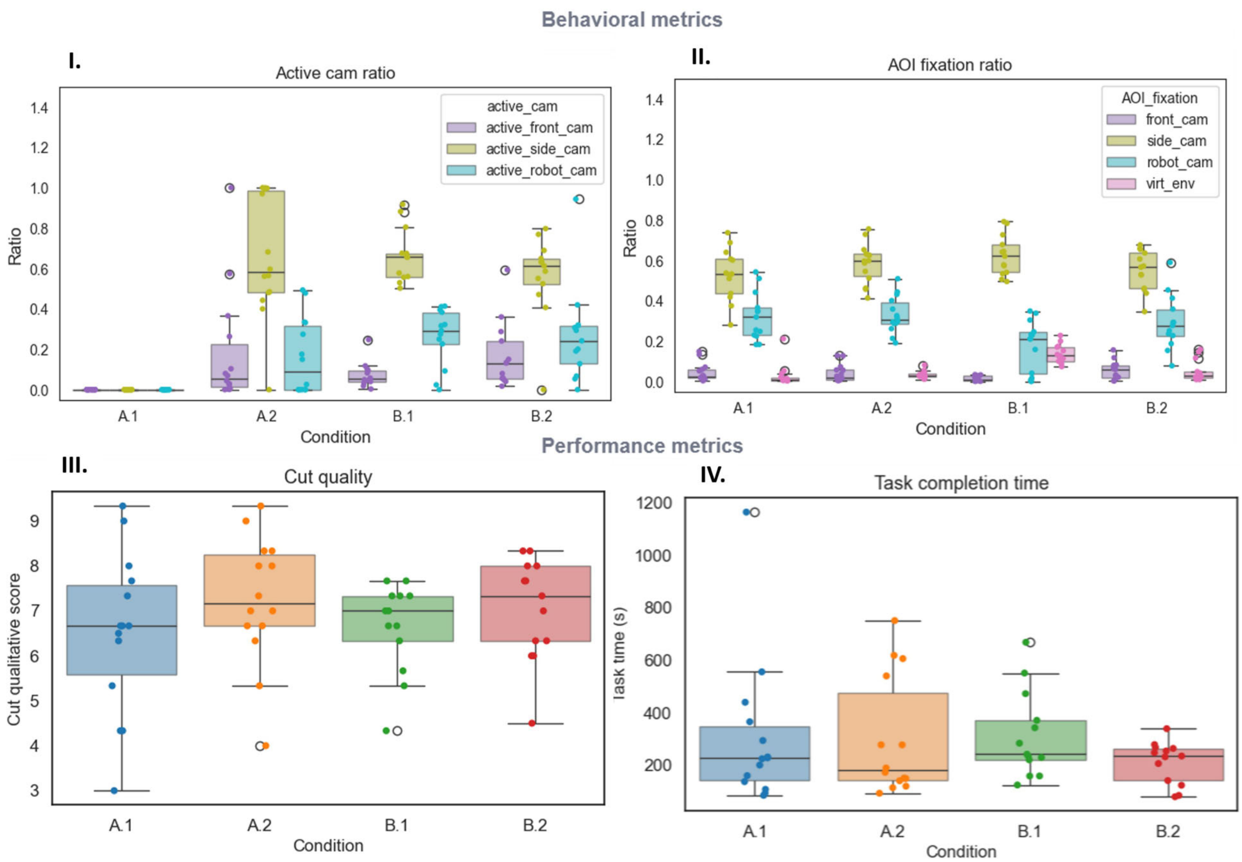

Fig. 9 User behavior and performance metrics box-plot of subject-wise averages for each condition. Subjects presented generally lower task times with interface B.2 the "Multiple view visual hierarchy HMI" as observed in panel **IV**, that promoted a more balanced use of control frames in panel **I**. Interface A.2 presented generally higher cut quality scores as depicted in panel **III**, but some subjects did not made full use of the ability to select different control frames and used only the side camera throughout

7 Discussion

7.1 Virtual Environment

Although the literature suggests that an interactive virtual environment can support situational awareness and depth perception [8, 10–12], our experiments showed that it was not useful for the teleoperation task, as shown in the results of Table 2. Uninterrupted motion has been for the most part not crucial in prior work use cases, typically involving manipulation and pick-and-place of common objects. Cutting tasks with a hot-wire require continuous motion, focus and high-quality performance feedback, which users found better provided by camera views. However, users still need support for situational awareness, collision avoidance, and error prevention, as we observed that collisions occurred due to situational awareness loss, despite the VE's presence.

7.2 User Biometrics

Statistical testing showed significant differences across all metric types except for EEG-based metrics, as observed in Tables 1 and 2. While EEG has been effective for workload classification in tasks with increased demands, our results suggest it is less sensitive to subtle interface design changes with smaller sample sizes. We hypothesize that these results may be due to both inter- and intra-subject variability, difficult to account for in small datasets, as well as the use of a consumer-grade EEG device with limited performance. Additionally, they may reflect the underlying assumptions that informed our selection of the three specific EEG indexes, thereby offering useful insights to inform the design of future EEG-based studies. Given the cost, setup complexity, and data pre-processing required for EEG, Eye-tracking emerged as a more accessible option for interface evaluation.

Mixed-metrics triangulation proved to be effective in testing HMI interaction factors, as different metrics were sensitive to different design elements. The control factor impacted teleoperated task quality and perceived performance, potentially due to the effect of control-view compatibility on mental workload, as show in Table 1. However, the perception interaction factor influenced the use of the interface, visual engagement, and operator frustration, but did not significantly affect the performance, shown in Table 2.

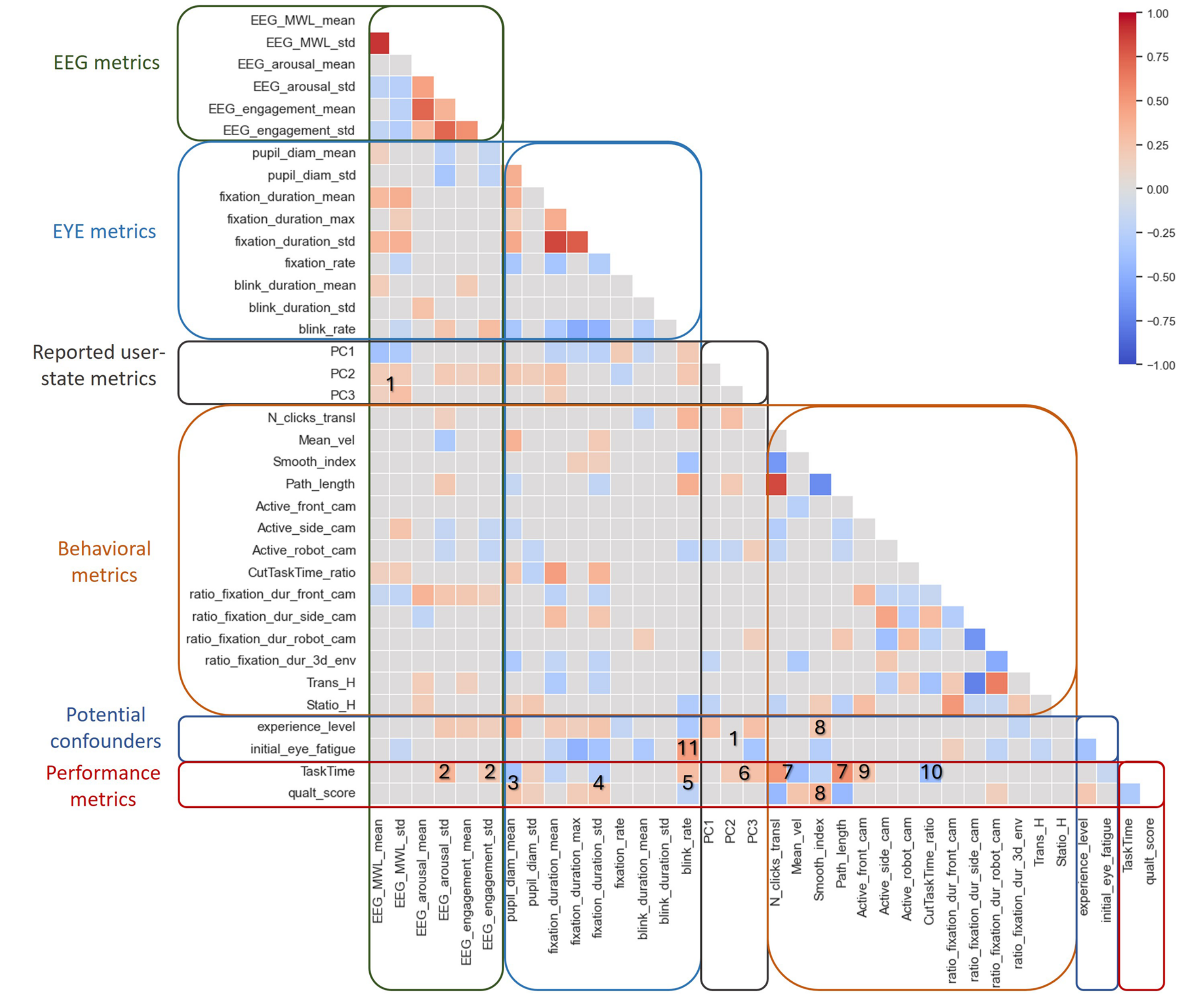

Fig. 10 Trial-based correlation across all subjects and conditions, grouped by main metric groups ($|r| \geq 0.10$, $p < 0.05$). 1) EEG metrics and experience level correlate with perceived user state. 2) Decreased arousal and engagement link to lower task efficiency. 3-4) Adaptive visual attention correlates with shorter task times and higher cut quality. 5) Higher blink rates link to longer, lower-quality tasks. 6) Users did not perceive task demands affecting cut quality. 7-8) Control metrics (e.g., controller clicks, velocity, smoothness) correlate with performance; smoothness reflects skill. 9) Front camera use impacted performance. 10) End-effector alignment affected task time. 11) Eye-tracking reliability may decline with fatigue during longer tasks

Overall, the results indicate that traditional mental workload assessments, employing questionnaires, may be less sensitive than bioindicators in capturing the impact of HMI factors on the user, as observed in the testing of Hypothesis 1 show in Table 1. Furthermore, in the testing of Hypothesis II, the mental workload construct alone does not adequately explain the differences observed in the bioindicators between interfaces, better explained by the concept of engagement (particularly associated with lower blink rates) and arousal (bigger pupil diameters), underlying a higher visual attention state, as shown in Table 2 for condition B.2. These results show once again the benefit of using of mixed-metrics and support the use of the neuroergonomics framework [23] for the evaluation of user state and performance in human-machine interactions.

7.3 Interface Evaluation

Correlation analysis showed that Eye-tracking metrics, supported by self-report assessments and human-machine interaction data, can assess performance states independently of interface configuration, as shown in Fig. 10. Larger pupil diameter and lower blink rates, linked to higher visual attention and arousal, correlated with shorter task times particularly in conditions A.2 and B.2, which also showed better performance efficiency, seen in Figs. 8 and 9. This aligns with Rendon-Velez et al. (2016) [48], highlighting the need for interfaces that balance engagement and workload, especially in repetitive tasks, and for operator assessment methods that evaluate engagement and arousal, in addition to mental workload effects.

Self-reported data favored interface A.2 for user preference, while B.2 had the lowest perceived task demand, visualized in Fig. 5. Control metrics (controller clicks, end-effector velocity, and smoothness index) also indicated better performance in A.2 and B.2. Standard smoothness metrics like trajectory jerk and spectral arc length did not fully capture issues such as lateral deviations from situational awareness loss. Therefore, a task-specific smoothness index, scaled by path length, was proposed to better assess control efficiency, skill, and accuracy for this cutting task. The proposed index can also serve as an indicator of operator proficiency, allowing supervisors to assess whether additional training is required, and to estimate the cut quality that can be expected from an operator's current skill level based on the smoothness and control efficiency exhibited during similar tasks.

7.4 Practical Implications and Applications

The findings discussed in the prior sections suggest that teleoperation interfaces for precision tasks should prioritize perceptually compatible viewpoints, informative real-time performance feedback, and user-state-aware evaluation methods to enhance control efficiency, situational awareness, and error prevention.

Although this study focused on a teleoperated cutting task, the findings generalize broadly to remote and supervisory control domains that require continuous motion, precise manipulation, and high situational awareness. The demonstrated benefits of view-aligned control frames, explicit view selection, and multi-camera feedback translate directly to industrial telemanipulation, remote inspection, hazardous-environment maintenance, and medical or surgical teleoperation, where camera–control compatibility is critical for performance and error avoidance. Similarly, the sensitivity of eye-tracking metrics to operator engagement and arousal, compared with the lower responsiveness of traditional workload questionnaires and consumer-grade EEG, highlights their value for real-time monitoring and interface evaluation in high-precision, screen-based remote tasks, including medical device manufacturing of catheters including skiving, hand gluing, or other fine-grained instruments/catheter guidance scenarios performed under constrained visual conditions. More generally, the mixed-metrics neuroergonomics approach adopted here provides a transferable framework for designing, assessing, and optimizing human–machine interfaces across domains involving continuous control, enabling safer operation, improved usability, and more reliable assessment of operator state during complex remote tasks.

7.5 Limitations

This study presented some limitations, namely the sample size, the number of tasks and conditions tested. The location and duration of the experiments, due to the long setup time of the equipment, constrained the number and experience level (distribution is skewed) of subjects able to participate in the study, and not allowing to follow the recommended sample size from the power analysis (sample of 35 participants needed to detect a large effect size (r = 0.5) with 80% power, as computed with Gpower [47]). Consequently, the statistical power of the study is impacted, particularly for the proposed EEG metrics. Nonetheless, it was still sufficient to detect meaningful effects of HMI factors on the user state using the proposed eye-tracking metrics, even if the experimental conditions were not designed specifically to induce them, and these effects were observed for a short task. The results highlight the practical significance of employing task-independent eye-tracking metrics, as opposed to task-dependent measures, to optimize the design and evaluation of teleoperation interfaces particularly under real-world constraints characterized by limited time and human resources.

Another limitation of this work was the design and use of the VE, limited to a passive role in supporting depth perception and situational awareness, as employed in previous works (see Section 2). The results showed that for such a task, the users need active assistance for collision avoidance and to maintain awareness of the end-effector location and position, without overloading the visual channel (e.g. using haptic or auditory cues).

8 Conclusion and Future Work

We propose a human-in-the-loop teleoperation task for medical device manufacturing, presenting mixed-reality designs for cutting miniature objects. The approach can be adapted to other fine-manipulation tasks. A multimodal human-centered evaluation framework was introduced and tested through a user-study with consumer-grade EEG and Eye-tracking devices, providing valuable insights for telerobotic interface design and evaluation.

Our work supports the argument that typical subjective user metrics and task performance metrics are limited in the analysis of human performance and performance impacting factors. Eye-tracking measures in particular provide a more objective view into the state of operators and how it is affected by HMI factors, having still shown a significant correlation with teleoperation performance across different participants and interfaces.

Despite the limitations of the study, the results show that for fine-manipulation tasks, control-view frame misalignment has a potential for higher impact on teleoperation quality than visual information organization in the interface. Moreover, high-quality, immediate and informative performance feedback is essential for real-time adjustment of uninterrupted irreversible actions, such as cutting tasks.

Future work will explore interaction factors and evaluate user interaction across the task's two stages. The end-effector alignment stage, which required more time and focus on cut-line tracking, may benefit from stage-specific interface optimization. These findings can support the development of teleoperation interfaces for cutting and fine-manipulation in manufacturing and guide human-centered evaluation approaches. Beyond interface design, this work could also facilitate the integration of human-aware AI with human-centered design principles, to achieve the next generation of adaptive telerobotic systems, able to adapt their interface to the user state, enhancing both operator performance and safety in Industry 5.0 environments. Future work will focus on real-time user-state prediction, using fused multimodal sensor data, to leverage the complementarity of multimodal information and capture more relevant features with less noise. This real-time feedback could be used to inform AI-based interface adaptation, an LLM-based agent for user support and guidance, camera/perspective switching, or even autonomy level adaptation (from manual teleoperation, to collaborative, supervised, all the way to full autonomy).

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10846-026-02383-z.

Author Contributions Inês F. Ramos and Keerthi Sagar are the main contributors, contributing equally to the work. All authors contributed to the study conception and design. Philip Long, Maria Chiara Leva and Gabriele Gianini provided the technical guidance, supervision, tools and resources needed for the research. The authors responsible for funding acquisition were Philip Long, Maria Chiara Leva and Ernesto Damiani. All authors read and approved the final manuscript.

Funding Open access funding provided by Università degli Studi di Milano - Bicocca within the CRUI-CARE Agreement. This work was supported by the European Union's Horizon 2020 Research and Innovation Programme under the CISC project (Marie Skłodowska-Curie grant agreement no. 955901), the FineTETHER project (Marie Skłodowska-Curie grant agreement no. 847402), and national capital grant funding program Enterprise Ireland (CE-2020-0221). The work was partially supported by the MUSA-Multilayered Urban Sustainability Action project, funded by the European Union-NextGenerationEU, under the Mission 4 Component 2 Investment Line of the National Recovery and Resilience Plan (NRRP) Mission 4 Component 2 Investment Line 1.5: Strengthening of research structures and creation of R&D "innovation ecosystems", set up of "territorial leaders in R&D" (CUP G43C22001370007, Code ECS00000037); Program "piano sostegno alla ricerca" PSR and the PSR-GSA-Linea 6; Project ReGAInS (code 2023-NAZ-0207/DIP-ECC-DISCO23), funded by the Italian University and Research Ministry, within the Excellence Departments program 2023-2027 (law 232/2016); and-FAIR-Future Artificial Intelligence Research-Spoke 4-PE00000013-D53C22002380006, funded by the European Union-Next Generation EU within the project NRPP M4C2, Investment 1., 3 DD. 341, 15 March 2022.

Data Availability Data will be made available upon request.

Materials Availability Not applicable.

Code Availability Not Applicable.

Declarations

Ethics approval and consent to participate The experiment received ethical approval N° REC-20-52C and all subjects gave their informed consent.

Consent for publication Not Applicable

Conflict of interest/Competing interests Not Applicable

References

1. Rakita, D., Mutlu, B., Gleicher, M.: Remote telemanipulation with adapting viewpoints in visually complex environments. Robotics: Science and Systems XV. (2019)
2. Rea, D.J., Seo, S.H.: Still Not Solved: A Call for Renewed Focus on User-Centered Teleoperation Interfaces. Frontiers in Robotics and AI. **9**, 1–13 (2022). https://doi.org/10.3389/frobt.2022.704225
3. Chen, J.Y.C., Haas, E.C., Barnes, M.J.: Human performance issues and user interface design for teleoperated robots. IEEE Trans. Syst. Man Cybernet. Part C Appl. Rev. **37**(6), 1231–1245 (2007). https://doi.org/10.1109/TSMCC.2007.905819
4. Rakita, D., Mutlu, B., Gleicher, M.: An Autonomous Dynamic Camera Method for Effective Remote Teleoperation. ACM/IEEE Int. Conf. Hum. Robot Interact., pp. 325–333 (2018). https://doi.org/10.1145/3171221.3171279
5. Praveena, P., Molina, L., Wang, Y., Senft, E., Mutlu, B., Gleicher, M.: Understanding Control Frames in Multi-Camera Robot Telemanipulation. In: Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction, pp. 432–440. IEEE Press, (2022)
6. Kent, D., Saldanha, C., Chernova, S.: A Comparison of Remote Robot Teleoperation Interfaces for General Object Manipulation. In: HRI '17: Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction, pp. 371–379 (2017)

7. Lipton, J.I., Fay, A.J., Rus, D.: Baxter's Homunculus: Virtual Reality Spaces for Teleoperation in Manufacturing. IEEE Robot. Automat. Lett. **3**(1), 179–186 (2018). https://doi.org/10.1109/LRA.2017.2737046, arXiv:1703.01270
8. Sun, D., Kiselev, A., Liao, Q., Stoyanov, T., Loutfi, A.: A New Mixed-Reality-Based Teleoperation System for Telepresence and Maneuverability Enhancement. IEEE Trans. Hum. Mach. Syst. **50**(1), 55–67 (2020). https://doi.org/10.1109/THMS.2019.2960676
9. Falcone, S., Englebienne, G., Van Erp, J., Heylen, D.: Toward Standard Guidelines to Design the Sense of Embodiment in Teleoperation Applications: A Review and Toolbox. Human-Computer Interaction. **38**(5–6), 322–351 (2022)
10. Bejczy, B., Bozyil, R., Vaiekauskas, E., Petersen, S.B.K., Bogh, S., Hjorth, S.S., Hansen, E.B.: Mixed reality interface for improving mobile manipulator teleoperation in contamination critical applications. Procedia Manuf. **51**, 620–626 (2020). https://doi.org/10.1016/J.PROMFG.2020.10.087
11. Naceri, A., Mazzanti, D., Bimbo, J., Tefera, Y.T., Prattichizzo, D., Caldwell, D.G., Mattos, L.S., Deshpande, N.: The Vicarios Virtual Reality Interface for Remote Robotic Teleoperation. J. Intell. Robot. Syst. **101**(4) (2021). https://doi.org/10.1007/S10846-021-01311-7
12. Su, Y., Chen, X., Zhou, T., Pretty, C., Chase, G.: Mixed reality-integrated 3D/2D vision mapping for intuitive teleoperation of mobile manipulator. Robot. Comput. Integ. Manuf. **77**, 102332 (2022). https://doi.org/10.1016/J.RCIM.2022.102332
13. Lim, Y., Gardi, A., Sabatini, R., Ramasamy, S., Kistan, T., Ezer, N., Vince, J., Bolia, R.: Avionics Human-Machine Interfaces and Interactions for Manned and Unmanned Aircraft. Prog. Aerosp. Sci. **102**, 1–46 (2018). https://doi.org/10.1016/j.paerosci.2018.05.002
14. Taori, S., Kim, S., Lim, S.: Evaluating and predicting cognitive workload in collaborative manufacturing scenarios: Human-human and teleoperator-robot-human. Available at SSRN 5213995. (2025)
15. Ramos, I.F., Gianini, G., Leva, M.C., Damiani, E.: Collaborative intelligence for safety-critical industries: A literature review. Information **15**(11), 728 (2024)
16. Bjegojević, B., Pušica, M., Gianini, G., Gligorijević, I., Cromie, S., Leva, M.C.: Neuroergonomic attention assessment in safety-critical tasks: EEG indices and subjective metrics validation in a novel task-embedded reaction time paradigm. Brain Sci. **14**(10), 1009 (2024)
17. Pušica, M., Kartali, A., Bojović, L., Gligorijević, I., Jovanović, J., Leva, M.C., Mijović, B.: Mental workload classification and tasks detection in multitasking: Deep learning insights from EEG study. Brain Sci. **14**(2), 149 (2024)
18. Giraudet, L., Imbert, J.P., Bérenger, M., Tremblay, S., Causse, M.: The neuroergonomic evaluation of human machine interface design in air traffic control using behavioral and EGG/ERP measures. Behav. Brain Res. **294**, 246–253 (2015). https://doi.org/10.1016/j.bbr.2015.07.041
19. Li, Y., Elmaghraby, A.S., El-Baz, A., Sokhadze, E.M.: Using physiological signal analysis to design affective VR games. 2015 IEEE International Symposium on Signal Processing and Information Technology, ISSPIT **2015**, 57–62 (2016). https://doi.org/10.1109/ISSPIT.2015.7394401
20. Caiazzo, C., Savkovic, M., Pusica, M., Milojevic, D., Leva, M.C., Djapan, M.: Development of a neuroergonomic assessment for the evaluation of mental workload in an industrial human-robot interaction assembly task: A comparative case study. Machines. **11**(11), 995 (2023)
21. Muñoz-de-Escalona, E., Leva, M.C., Cañas, J.J.: Mental workload as a predictor of ATCO's performance: Lessons learnt from ATM task-related experiments. Aerospace **11**(8), 691 (2024)
22. Amazu, C.W., Demichela, M., Fissore, D., Leva, M.C.: Comparing two control room intervention procedure formats: Preliminary insights from eye tracking measures. Chem. Eng. Trans. **111**, 223–228 (2024)
23. Dehais, F., Lafont, A., Roy, R., Fairclough, S.: A Neuroergonomics Approach to Mental Workload, Engagement and Human Performance. Front. Neurosci. **14**, 268 (2020). https://doi.org/10.3389/fnins.2020.00268
24. McMahan, T., Parberry, I., Parsons, T.D.: Evaluating Player Task Engagement and Arousal Using Electroencephalography. Procedia Manuf. **3**, 2303–2310 (2015). https://doi.org/10.1016/j.promfg.2015.07.376
25. Zhai, W., Liao, J., Chen, Z., Su, B., Zhao, X.: A survey of task planning with large language models. Intell. Comput. **4**, 0124 (2025). https://doi.org/10.34133/icomputing.0124, https://spj.science.org/doi/pdf/10.34133/icomputing.0124
26. Alharasees, O., Kale, U.: Human factors and AI in UAV systems: Enhancing operational efficiency through AHP and real-time physiological monitoring. J. Intell. Robot. Syst. **111**(1) (2024). https://doi.org/10.1007/s10846-024-02188-y
27. Andriella, A., Ros, R., Ellinson, Y., Gannot, S., Lemaignan, S.: Dataset and evaluation of automatic speech recognition for multilingual intent recognition on social robots. In: Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction. HRI '24, pp. 865–869. ACM, (2024). https://doi.org/10.1145/3610977.3637473
28. Fernandez Rojas, R., Debie, E., Fidock, J., Barlow, M., Kasmarik, K., Anavatti, S., Garratt, M., Abbass, H.: Electroencephalographic workload indicators during teleoperation of an unmanned aerial vehicle shepherding a swarm of unmanned ground vehicles in contested environments. Front. Neurosci. **14** (2020). https://doi.org/10.3389/fnins.2020.00040
29. Guo, Y., Freer, D., Deligianni, F., Yang, G.-Z.: Eye-tracking for performance evaluation and workload estimation in space telerobotic training. IEEE Trans. Hum. Mach. Syst. **52**(1), 1–11 (2022). https://doi.org/10.1109/THMS.2021.3107519
30. Bethel, C.L., Henkel, Z., Baugus, K.: Conducting Studies in Human-Robot Interaction, pp. 91–124. Springer, (2020)
31. Leeper, A.E., Hsiao, K., Ciocarlie, M., Takayama, L., Gossow, D.: Strategies for human-in-the-loop robotic grasping. In: HRI'12 - Proceedings of the 7th Annual ACM/IEEE International Conference on Human-Robot Interaction, pp. 1–8 (2012). https://doi.org/10.1145/2157689.2157691
32. Pope, A.T., Bogart, E.H., Bartolome, D.S.: Biocybernetic system evaluates indices of operator engagement in automated task. Biol. Psychol. **40**(1–2), 187–195 (1995). https://doi.org/10.1016/0301-0511(95)05116-3
33. Kim, M.K., Kim, M., Oh, E., Kim, S.P.: A review on the computational methods for emotional state estimation from the human EEG. Comput. Math. Methods Med. **2013** (2013). https://doi.org/10.1155/2013/573734
34. Mijović, P., Ković, V., De Vos, M., Mačužić, I., Todorović, P., Jeremić, B., Gligorijević, I.: Towards continuous and real-time attention monitoring at work: reaction time versus brain response. Ergonomics. **60**(2), 241–254 (2017). https://doi.org/10.1080/00140139.2016.1142121
35. Dehais, F., Duprès, A., Di Flumeri, G., Verdière, K., Borghini, G., Babiloni, F., Roy, R.: Monitoring Pilot's Cognitive Fatigue with Engagement Features in Simulated and Actual Flight Conditions Using an Hybrid fNIRS-EEG Passive BCI. In: Proceedings - 2018 IEEE International Conference on Systems, Man, and Cybernetics, SMC **2018**, 544–549 (2019). https://doi.org/10.1109/SMC.2018.00102
36. Mahanama, B., Jayawardana, Y., Rengarajan, S., Jayawardena, G., Chukoskie, L., Snider, J., Jayarathna, S.: Eye movement and pupil

measures: A review. Front. Comput. Sci. **3** (2022). https://doi.org/10.3389/fcomp.2021.733531
37. Delorme, A., Makeig, S.: EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J. Neurosci. Methods. **134**(1), 9–21 (2004). https://doi.org/10.1016/j.jneumeth.2003.10.009
38. Mullen, T.R., Kothe, C.A.E., Chi, Y.M., Ojeda, A., Kerth, T., Makeig, S., Jung, T.-P., Cauwenberghs, G.: Real-time neuroimaging and cognitive monitoring using wearable dry EEG. IEEE Trans. Biomed. Eng. **62**(11), 2553–2567 (2015). https://doi.org/10.1109/tbme.2015.2481482
39. Pion-Tonachini, L., Kreutz-Delgado, K., Makeig, S.: IClabel: An automated electroencephalographic independent component classifier, dataset, and website. NeuroImage. **198**, 181–197 (2019). https://doi.org/10.1016/j.neuroimage.2019.05.026
40. Mastropietro, A., Pirovano, I., Marciano, A., Porcelli, S., Rizzo, G.: Reliability of mental workload index assessed by EEG with different electrode configurations and signal pre-processing pipelines. Sensors. **23**(3), 1367 (2023). https://doi.org/10.3390/s23031367
41. Marcantoni, I., Assogna, R., Del Borrello, G., Di Stefano, M., Morano, M., Romagnoli, S., Leoni, C., Bruschi, G., Sbrollini, A., Morettini, M., Burattini, L.: Ratio indexes based on spectral electroencephalographic brainwaves for assessment of mental involvement: A systematic review. Sensors. **23**(13), 5968 (2023). https://doi.org/10.3390/s23135968
42. Krejtz, K., Szmidt, T., Duchowski, A.T., Krejtz, I.: Entropy-based statistical analysis of eye movement transitions. In: ETRA '14: Eye Tracking Research and Applications, pp. 159–166. ACM, (2014)
43. Hart, S.G., Staveland, L.E.: Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. Adv. Psychol. **52**(C), 139–183 (1988). https://doi.org/10.1016/S0166-4115(08)62386-9
44. Bradley, M.M., Lang, P.J.: Measuring emotion: The self-assessment manikin and the semantic differential. J. Behav. Ther. Exp. Psychiatry. **25**(1), 49–59 (1994). https://doi.org/10.1016/0005-7916(94)90063-9
45. Hannum, M.E., Forzley, S., Popper, R., Simons, C.T.: Application of the Engagement Questionnaire (EQ) to compare methodological differences in sensory and consumer testing. Food Research International. **140**, 110083 (2021). https://doi.org/10.1016/j.foodres.2020.110083
46. Oh, D.S., Ershad, M., Wee, J.O., Sancheti, M.S., D'Souza, D.M., Herrera, L.J., Schumacher, L.Y., Shields, M., Brown, K., Yousaf, S., Lazar, J.F.: Comparison of global evaluative assessment of robotic surgery with objective performance indicators for the assessment of skill during robotic-assisted thoracic surgery. Surgery. **174**(6), 1349–1355 (2023). https://doi.org/10.1016/j.surg.2023.08.008
47. Faul, F., Erdfelder, E., Buchner, A., Lang, A.-G.: Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. Behav. Res. Methods. **41**(4), 1149–1160 (2009). https://doi.org/10.3758/brm.41.4.1149
48. Rendon-Velez, E., van Leeuwen, P.M., Happee, R., Horváth, I., van der Vegte, W.F., de Winter, J.C.F.: The effects of time pressure on driver performance and physiological activity: A driving simulator study. Transp. Res. Part F Traffic Psychol. Behav. **41**, 150–169 (2016). https://doi.org/10.1016/j.trf.2016.06.013

Inês F. Ramos received her Ph.D. in Computer Science from Università degli Studi di Milano in 2025. Between 2022 and 2025 She has been Early Stage Researcher of the Maria Skłodowska Curie CISC project. She has a background in biomedical engineering, machine learning, and human–machine interaction. Her major research interests include Human Factors and ergonomics in Industry, and human-machine interaction, particularly in teleoperation and robotics contexts.

Keerthi Sagar is a Technologist in the Robotics and Automation Group at Irish Manufacturing Research (IMR). From 2021 to 2024, he was a Marie Skłodowska-Curie Postdoctoral Fellow at IMR, working in collaboration with KUKA Robotics Ireland. He received his PhD in Robotics from the University of Genova, Italy, in 2018. His research focuses on robotic manipulation, robot kinematics, motion planning, and mobile robotics, with applications in advanced manufacturing and automation.

Philip Long received the B.Eng. degree (First Class Honours) in Mechanical Engineering from the University of Galway, Ireland, the European Master's degree in Advanced Robotics from the University of Genova and École Centrale de Nantes, and the Ph.D. degree in Robotics from École Centrale de Nantes, France. He is currently a Lecturer in Robotics & Automation at Atlantic Technological University, Galway (ATU), Ireland. His previous positions include Senior Technologist at Irish Manufacturing Research (IMR), Research Engineer at IRT Jules Verne (Nantes, France), and Researcher at the RIVeR Lab, Northeastern University, USA. His research interests include human–robot collaboration, sensor-based control of robot manipulators, cable-driven parallel robots, humanoid robot motion planning, and agile manufacturing. He has authored more than 30 peer-reviewed publications and holds 4 international patents. He served as Associate Editor for IEEE ICRA (2019–2022) and currently supervises several Ph.D. researchers, postdoctoral fellows, and research engineers.

Maria Chiara Leva is a Senior Lecturer in Health and Safety in Technological University Dublin and the Lead of the Human Factors for Safety and Sustainability Research Centre. Her area of Expertise is Safety Management Systems and Human factors. She is the co-chair of the technical committee on human factors for the European Safety and Reliability Association (ESRA), and former chair of The Irish Human Factors Ergonomics Society and has been working in Ergonomics and Risk Assessment as a consultant since 2008.t

Ernesto Damiani (Senior Member, IEEE) received the Honorary Doctorate degree from the Institut National des Sciences Appliquées de Lyon, France, in 2017, for his contributions to research and teaching on big data analytics. He is currently a Full Professor with the Department of Computer Science, Università degli Studi di Milano, where he leads the Secure Service-Oriented Architectures Research (SESAR) Laboratory. He is also the Acting Dean of the College of Computing and Mathematical Sciences and the Founding Director of the Center for Cyber-Physical Systems, Khalifa University, United Arab Emirates. He is a Distinguished Scientist of ACM. He was a recipient of the 2017 Stephen Yau Award. He has published more than 600 peer-reviewed articles and books. His research interests include cybersecurity, big data, and cloud/edge processing.

Gabriele Gianini received his M.Sc. Degree in Physics (1992) and his Ph.D. Degree in Physics (1996) from the University of Pavia. He is Full Professor at the Department of Informatics Systems and Communication (DISCo) of the Università degli Studi di Milano-Bicocca. He has been Researcher, then Associate Professor at the University of Milan since 2005. He has held visiting positions at a number of international institutions, including INSA de Lyon, University of Passau, CERN, Fermilab. From 2005 to 2012 he has been Adjoint professor at the Free University of Bozen/Bolzano. From 2017 to 2020 he has been Senior Research Fellow at EBTIC/Khalifa University of Science and Technology in Abu Dhabi. Among his research interests are Machine Learning and XAI. From 1992 to present has co-authored over 250 papers published in internationally refereed journals and conferences.