TUDelft

Delft University of Technology

Augmented reality-based telepresence in a robotic manipulation task
An experimental evaluation

de Boer, Thomas A.B.; de Winter, Joost C.F.; Eisma, Yke Bauke

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

The Institution of Engineering and Technology **WILEY**

ORIGINAL RESEARCH

Augmented reality-based telepresence in a robotic manipulation task: An experimental evaluation

Thomas A. B. de Boer | **Joost C. F. de Winter** | **Yke Bauke Eisma**

Department of Cognitive Robotics, Delft University of Technology, Delft, The Netherlands

Correspondence

Joost C. F. de Winter, Department of Cognitive Robotics, Faculty of Mechanical, Maritime and Materials Engineering, Delft University of Technology, Mekelweg 2, 2628 CD, Delft, The Netherlands.
Email: j.c.f.dewinter@tudelft.nl

Abstract

A spectrum of control methods in human–robot interaction was investigated, ranging from direct control to telepresence with a virtual representation of the robot arm. A total of 24 participants used a setup that included a Franka Emika Panda robot arm, Varjo XR-3 head-mounted display, and Leap Motion Controller. Participants performed a box-and-block task using the bare hand (A), and under five gesture-controlled robotic operation methods: direct sight (B), sight via video-feedthrough (C), in a 3D telepresence environment with (D) and without (E) virtual representation of the robot arm, and using a 2D video feed (F). The number of grabbing attempts did not differ significantly between conditions, but local operation (B & C) yielded more transferred blocks than teleoperation (D–F). Teleoperation using a 3D presentation was advantageous compared to teleoperation using a 2D video feed, as demonstrated by lower peak forces and smaller range in gripper heights in conditions D and E compared to condition F, a finding supported by analyses of the head movement activity. Finally, the bare hand yielded the best performance and subjective ratings. In summary, teleoperation using a 3D presentation provided a smoother interaction than teleoperation with a 2D video feed. However, direct human interaction remains a benchmark yet to surpass.

KEYWORDS

data analysis, human-robot interaction, robot dynamics

1 | INTRODUCTION

Augmented reality (AR) is a technology that enhances human perception of the physical world by superimposing digital data streams or virtual objects onto it. Over recent years, AR has seen substantial advancements, such as improved frame rates, resolution, and wearability, making it an increasingly suitable platform for scientific experimentation and practical application [1, 2]. AR has been applied in a diverse array of areas, including medical education and surgery [3, 4], repair and maintenance [5–7], architectural design [8, 9], cultural heritage preservation [10, 11], and remote collaboration [12, 13], amongst others. With regard to information presentation, AR typically overlays context-specific information [14, 15] or enables the integration of digital elements in a real environment [16, 17].

Similar to AR, digital twins, spatial computing, and virtual representations of the work site have been the subject of extensive research over the years, particularly concerning their potential in remote interactions between humans and machines [18–21]. In conventional remote interaction scenarios, such as a human operator managing a crane, the operator must integrate relevant information from 2D sources (e.g. video feeds) and 1D sources (e.g. time-dependent parameters like recorded distance to an object) to create a mental representation required for precise control [22]. A 3D presentation of the work site, on the other hand, potentially allows the human operator to allocate more cognitive resources to meaningful control.

Teleoperation can be useful in situations where robots cannot operate autonomously due to the absence of reliable sensor automation, the need for human oversight, or when

human physical presence is not feasible. One such example involves an astronaut executing a pick-and-place task in space using a teleoperated robotic manipulator. Accomplishing such a task through traditional manual control (e.g. using a joystick) is challenging and necessitates extensive training [23].

Several studies have previously explored the use of AR in gesture-controlled teleoperation, for example, using visual overlays of task instructions [24] or visualisations of the planned motion of the robot arm [25]. Su et al. [26] employed a digital twin approach, immersing the user in a virtual operating space augmented by real-time 3D visual feedback from the robot's work site. This setup yielded favourable results in a pick-and-place task compared to a 2D video feed. Similarly, Peppoloni et al. [27] presented an immersive telepresence system that provided augmented 3D visual feedback to the user, offering the remote environment from the robot's perspective and additional task execution information. Li et al. [28] proposed an AR and digital twin-based multi-robot tele-operation system for manufacturing, incorporating human-in-the-loop control, multi-robot communication, and reinforcement learning for motion planning, validated through various experiments. Similarly, Park et al. [29] proposed a hands-free interaction strategy for human–robot engagement, using multimodal gestures, deep learning for object recognition, and a digital twin for simulating robot actions, and demonstrated the setup's effectiveness in two case studies. Ponomareva et al. [30] demonstrated a system for teleoperated robots in medical scenarios, which included an augmented virtual environment and region-based convolutional neural network for identifying and positioning laboratory instruments. Finally, Shahria et al. [31] explored AR-based teleoperation, Chan et al. [32] used AR in shared manufacturing tasks, and Brizzi et al. [33] focused on AR for embodiment where the robot perspective is reconstructed. AR has also been used in robot programming, the creation of training scenarios, and the development of physical artefacts, enabling the possibility of programming a robot in a workspace without its physical presence and simulating motions before their execution [34–37].

There seems to be an absence of systematic research that examines how human operators perform robotic manipulation tasks using AR interfaces. In the current study, we evaluated the efficacy of a spectrum of control methods, including direct human control of the robot arm while being physically present and teleoperated control. The latter was presented in such a way that it replicates the participant's on-site presence, a concept also referred to as telepresence [38]. Through our multi-method approach, we were able to investigate the effects of individual components, such as the head-mounted display (HMD), remote operation, and a 3D presentation versus a 2D video feed.

It can be expected that teleoperation may lead to performance deficits in comparison to non-teleoperated scenarios. This hypothesis is grounded in the fact that, despite advances in telepresence, teleoperation still induces a time delay and renders a less accurate representation of the task environment, especially concerning depth perception and resolution, compared to a more direct view of the robotic setup. A second hypothesis is that the use of a 3D presentation of the work site in teleoperation generates better task performance relative to teleoperation supplemented merely by a 2D video feed. This hypothesis is based on the notion that stereopsis and motion parallax may be particularly relevant depth cues (e.g. ref. [39]). Stereopsis arises from the horizontal separation between the eyes, enabling the brain to integrate slight differences in the images projected to each eye to derive depth information [40]. Motion parallax, on the other hand, is a monocular cue based on the relative motion of objects at different distances that occurs if we move our head: nearby objects exhibit greater relative motion than more distant ones [41]. Being able to use stereopsis and motion parallax may lead to improved performance on tasks requiring judgement of distances, orientation, and object manipulation as compared to the use of a static 2D video feed.

2 | METHODS

2.1 | Participants

A total of 26 participants took part in the experiment, but two were excluded from the analysis due to measurement errors. The remaining 24 participants (18 males, 6 females) ranged in age from 21 to 34 years (mean: 26.5 years, standard deviation: 3.1 years). Of these participants, 21 were right-handed, two were cross-dominant, and one was left-handed. The two cross-dominant participants chose to use their right hand during the experiment, based on the training task. Fourteen participants did not use any vision aids, eight wore glasses, one wore contact lenses, and one participant who usually wore vision aids opted not to during the experiment. The study received approval from the Delft Human Research Ethics Committee, approval no. 2553, with each participant providing written informed consent before commencement of the experiment.

The participant pool included 15 students from the Faculty of Mechanical Engineering, 5 PhD students and 2 postdocs from the Department of Cognitive Robotics within the Faculty of Mechanical Engineering, and 2 participants from other professions (a gardener and a marketing manager). Participants' experience with virtual reality or mixed reality varied, with 8 having no experience, 9 possessing some experience, 4 having moderate experience, and 3 having professional experience. Regarding teleoperation experience, 16 participants had none, 7 had some experience, and 1 had medium experience, having worked with teleoperation setups in academic environments. In terms of robotics experience, 5 participants had no experience, 6 had some experience, 6 had moderate experience, 5 had extensive experience, and 2 had professional experience.

2.2 | Apparatus

The system was designed to allow participants to control a robot arm remotely through hand motion tracking. Participants were able to manoeuvre the robot arm and manipulate the

gripper (end effector) simultaneously to grab objects. The experiment was designed to enable comparisons between various variants of local and remote operation. Local operation occurred in a shared workspace where the participant sat in front of the robot. Remote teleoperation was implemented using AR, in which the robot's environment was recreated with varying degrees of realism.

Five of the six experimental conditions were executed using a Franka Emika Panda robot arm with Franka Hand [42], in conjunction with a Varjo XR-3 HMD [43] and a Leap Motion Controller [44]. The Panda robot employed is the original version (FER), operating on Libfranka 0.9.2 and software version 4.2.1 [45].

The system, as illustrated in Figure 1, was controlled by a Linux PC running Ubuntu 18.04 with a real-time kernel, equipped with an I7-6700K processor and a GTX2080 graphics card. A ZED2 stereo camera was connected to the Linux PC, and the camera feed was streamed to a local network using Python in 1080p at 30 fps. A Robot Operating System (ROS) environment was established on this PC with Libfranka [46], franka_ros [45], and a custom Cartesian impedance controller [47] to control the robot's end effector. This custom controller was derived from the example_cartesian_impedance_controller provided in the franka_ros package. In the modified controller, the position error between the current and desired pose was capped at a maximum of 0.03 m. This prevented overshoot and overpowering of the robot during large-amplitude movements, while allowing us to employ relatively high stiffness values (translational: 2000 N/m, rotational: 40 Nm/rad, null space: 3 Nm/rad) for precise and responsive control.

A separate computer on the same network was utilised to operate Unity to control the Varjo XR-3, featuring Windows 10, an I9-9900K processor, and an RTX3090TI graphics card. The Leap Motion sensor was also connected to this computer. Within the Unity environment [48], the Leap Sensor recorded the participant's hand location, and this data was subsequently published over the network to the Cartesian impedance controller as direct commands, via the ROS# plugin for Unity [49]. Additionally, the stream from the ZED2 camera was received in Unity using the ZED plugin [50], where a 3D point cloud was computed in real-time and displayed in the scene. Figure 2d displays an overview of the experimental setup.

Four SteamVR base stations 2.0 [51] were situated in the corners of the room to ensure precise spatial tracking of the XR-3 headset. A green screen (chroma key, natively supported by Varjo base, see Figure 2c) was used for accurate projection and positioning of the AR environment. Real-time video imagery of the robot's workspace was provided by a Stereolabs ZED2 stereo camera [52], which was mounted on a tripod and secured to the table of the robot arm (Figure 2a,b) to offer stable recordings and 3D point clouds in the AR environment. Hand-tracking was executed either through a single Ultraleap Leap Motion Controller affixed to the tripod (for the local operation conditions B and C) or by the Varjo XR-3 onboard Leap Motion sensor (for the teleoperation conditions D–F). Video recording for the real-life condition was conducted using a Logitech C920 webcam.

2.3 | System behaviour

The robotic system can be controlled independently in terms of position, orientation, and gripping. The position of the robot's gripper is determined by the position of the user's hand in 3D space. The orientation of the gripper is determined based on the pose of the user's hand. In the zero-pose (Figure 3), the gripper is aligned to point straight down and be perpendicular to the user. To initiate grabbing, the user closes their thumb, index, and middle finger in a pinching movement. Releasing the grip is done by opening these fingers. There are specific thresholds set for this pinching movement to determine whether the gripper should open.

If the robot loses motion tracking, it returns to a centred ready-pose. Upon regaining tracking, it moves back to the tracked input. Regaining motion tracking can be accomplished if the user moves their hand to the centre of the workspace in the zero-pose.

If the robot detects a joint limit or self-collision, it will cease motion and move away from the problematic configuration,
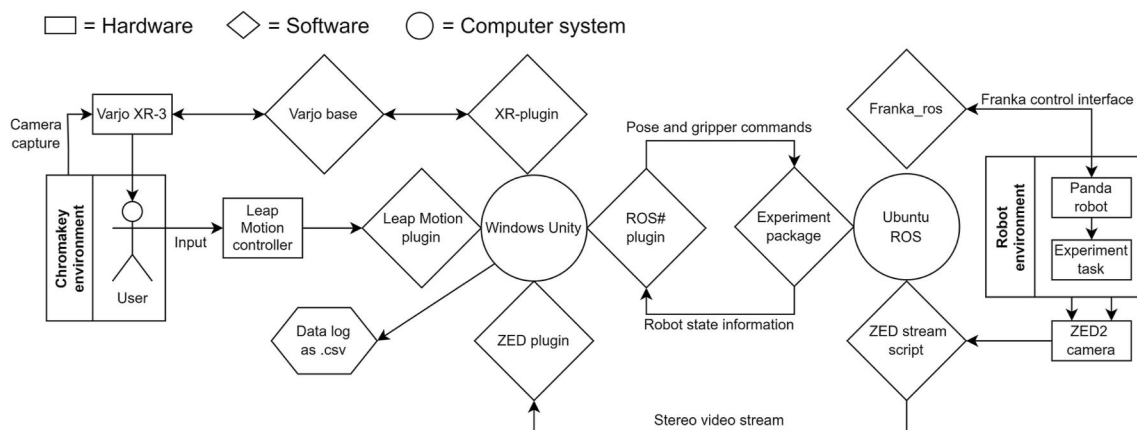

FIGURE 1 System architecture.

FIGURE 2 (a) Franka Emika Panda robot arm featuring the box-and-block task in front. The tripod holding the ZED2 stereo camera and the Leap sensor directed towards the seated participant from below is also visible. (b) Seating location for the participant when controlling the robot in the real environment. (c) Green screen setup with seating position for the participant when controlling the robot in the teleoperation conditions. The Varjo XR-3, connected to a cable boom, is visible on the left side, and two SteamVR base stations are positioned at the top corners of the green screen. (d) Photograph of the entire setup, depicting the experimenter's computer positioned between the robot arm and the green screen.
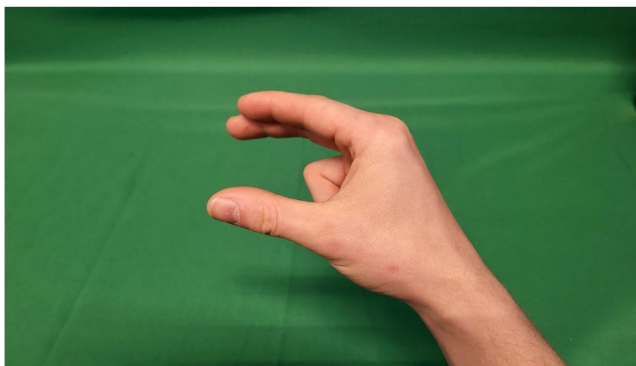

FIGURE 3 Defined zero-pose for providing input to the system.

typically requiring a few seconds. The user should move their hand away from the location where the detection occurred, towards the centre of the workspace, to prevent a joint limit or self-collision from reoccurring.

2.4 | Experimental tasks

2.4.1 | Training task

The training task involved selecting nine shapes from a 3D-printed shape-holder, akin to a child's toy, using the robot arm and placing them on a foam surface adjacent to the holder. The holder and shapes can be found on Thingiverse [53],

with the holder printed at 150% and the shapes at 145% (Figure 4a).

The objective of the training task was to improve the participant's ability to operate the robot arm for picking up and placing objects. Throughout the training task, participants were encouraged to experiment with different approaches and test the system's limits and behaviour. The task was deemed complete when all nine shapes were successfully picked, typically taking around 5 min.

2.4.2 | Experiment task

The experiment task involved performing a modified box-and-block test [54]. This standardised manual dexterity test entailed transferring blocks from one compartment to another within a 60-s timeframe. Participants' scores were based on the number of blocks transferred within this period. The box-and-block test used in this experiment was 3D printed and designed to align with the standardised dimensions of the two boxes (25.4 cm × 25.4 cm) and the 150 cubes (2.5 cm in size) [55]. However, due to the constraints imposed by the size of the robot's gripper, it was decided to exclude the centre divider during the experiment, which deviates from the standardised box-and-blocks test (Figure 4b).

Following the training period and before commencing the experimental task trials, participants were informed about the rules of the box-and-block task. Participants were only permitted to transfer one block at a time, and any accidental picking of two blocks was counted as one block. They were not allowed to throw or flick blocks over the centreline, but dropping blocks in the empty box was permitted once the centreline had been crossed. The task was set up with the full box on the side of the participant's dominant hand.

2.5 | Independent variables

The experiment was conducted using a within-subject design, in which each participant was presented with six conditions. Each condition was performed three times consecutively, with each trial having a time limit of 1 min. The conditions were presented in a counterbalanced order according to a Latin square design. There are six experimental conditions, which can be divided into two categories.

2.5.1 | Local conditions

- *A. Direct control* (Figure 5a): This baseline condition involves the participant executing the task manually, without the assistance of AR or the robot arm. The condition was designed in accordance with the standard box-and-block test [54].
- *B. Gesture-operated* (Figure 5b): In this condition, the participant performed the task by controlling the robot with motion tracking while seated at the same table where the robot was situated, and observing the interaction between the robot and the task directly. This condition served as a baseline for gesture operation with the system to compare the effects of variations in visual interfaces in the other conditions.
- *C. Gesture-operated with HMD* (Figure 5c): In this condition, the participant controlled the robot using motion tracking, as in condition B. However, they wore a HMD in stereo-video passthrough mode while executing the task. This condition evaluates potential influences that the HMD might introduce compared to condition B.

2.5.2 | Teleoperation conditions

D. Teleoperation with three-dimensional mapping and a virtual representation of the robot arm (Figure 5d): In this condition, the participant executed the task in front of a greenscreen with a superimposed virtual environment, while seated in front of a greenscreen. This condition was designed to closely match the user experience of conditions B and C but without being physically present. The robot environment was recreated with a 3D point cloud, as computed from a stereo camera feed, in combination with a virtual representation of the robot arm. The participant experienced this recreated environment virtually within the greenscreen, enabling them to coexist with the recreated environment.

Note that the virtual robot arm replicated the movements of the physical robot. The robot's physical structure is defined in the Unified Robot Description Format file within ROS, and this file is imported into a Unity simulation environment to create the 3D visualisation of the robot arm. Within Unity, the simulation is configured to receive real-time information about
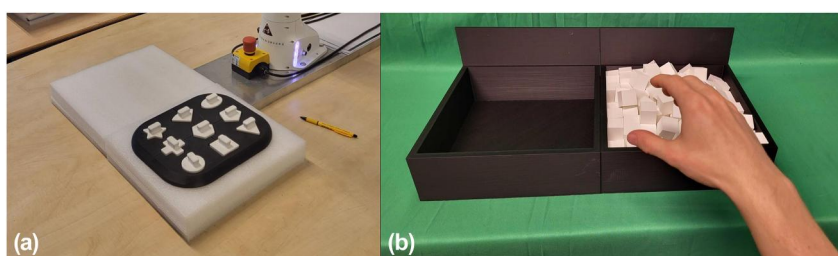

FIGURE 4 (a) Photo of the training task used in the current experiment. (b) Modified box-and-block test as used in the current experiment.
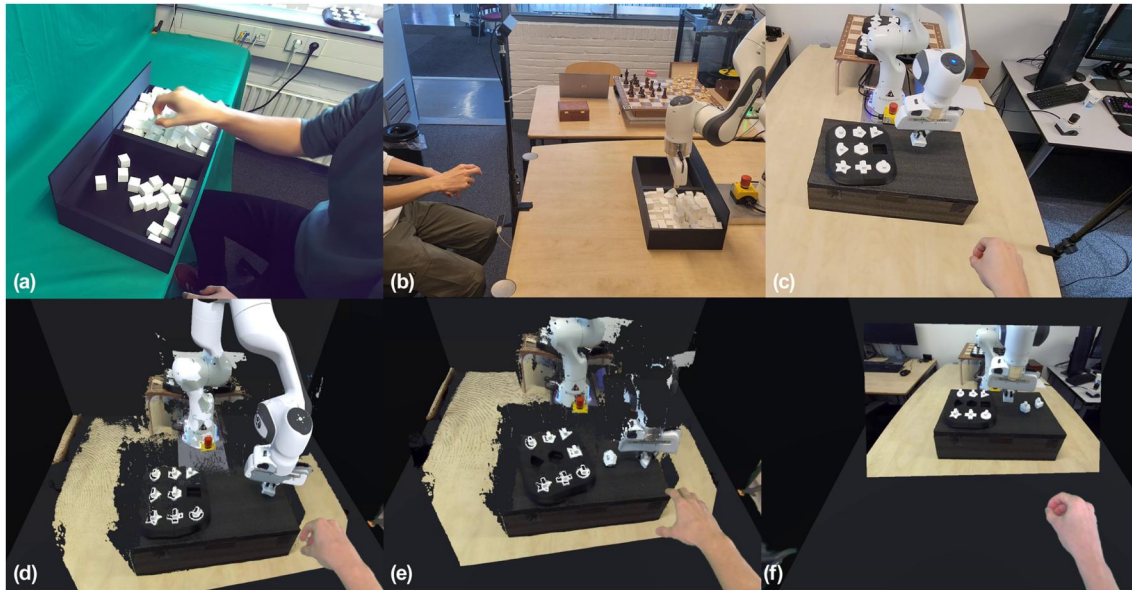
FIGURE 5 Figures (a–f) show the six respective experimental conditions. Figures (a, b) are photographed from the position of the experimenter, while figures (d–f) are screenshots (left eye) of the Varjo XR-3 head-mounted display (HMD). Please note that screenshots (c–f) show the training task, consisting of the 9 shapes that had to be moved. This depiction of the training task is for illustrative purposes only; the actual experiment was conducted using the box-and-block task; the training task was performed without the use of the head-mounted display.

the robot's dynamic state, such as its position and movement, by subscribing to the appropriate ROS topic through a connection using ROS# and rosbridge.

- **E. Teleoperation with three-dimensional mapping** (Figure 5e): This condition was identical to condition D, except that no virtual robot arm was added to the environment, resulting in unrendered parts of the robot when it was outside the field of view or the depth range of the stereo camera. This condition evaluated potential influences of the virtual representation of the robot arm used in condition D.
- **F. Teleoperation with two-dimensional view** (Figure 5f): In the teleoperation with 2D view condition (Figure 5f), participants controlled the robot via motion tracking while seated before a greenscreen, upon which a three-dimensional room was projected, featuring a virtual television screen (81.5 cm × 45.9 cm) approximately 90 cm in front of them. Participants saw the work site through a live feed from the left lens of the ZED2 camera, displayed on the virtual screen. Participants had the ability to adjust their position relative to this screen. This condition served to evaluate the impact of 3D scene mapping (conditions D and E) when compared with a 2D presentation.

2.6 | Dependent variables

The dependent variables in this study are those being investigated and measured to understand their relationship with changes in independent variables. There are 10 dependent variables, divided into three categories.

2.6.1 | Grab variables

- *Grab Attempts*, defined as the total attempts at grabbing a block that resulted in either a failed or successful grab.
- *Failed Grabs*, defined as the number of grabs initiated that did not result in the grabbing of a block.
- *Successful Grabs*, defined as the number of grabs initiated that did result in the grabbing of a block. Note that the total number of attempted grabs is equal to the sum of successful grabs and failed grabs.
- *Transferred Blocks*, representing the actual number of blocks transferred to the empty box within the time limit, manually counted by the experimenters.

The dependent variables *Grab Attempts*, *Failed Grabs*, and *Successful Grabs* were ascertained from the end-effector gripper width using a peak-finding algorithm on the negative width. Figure 6 displays a typical result, where widths smaller than 2 s were used to categorise a full closing of the gripper, indicating a failed grab. A partial closing of the gripper, that is, a width between 2.0 and 3.5 cm, signalled that an object was between the gripper's fingers, indicating a successful grab.

2.6.2 | Gripper movement

- *Movement range.* In addition to information about the number of grab attempts, we also examined how the participants moved the blocks. To this end, we extracted all successful grabs where no hand tracking was lost (see Appendix A). Each block had to be picked up from the tray on the participant's dominant-hand size and deposited into the
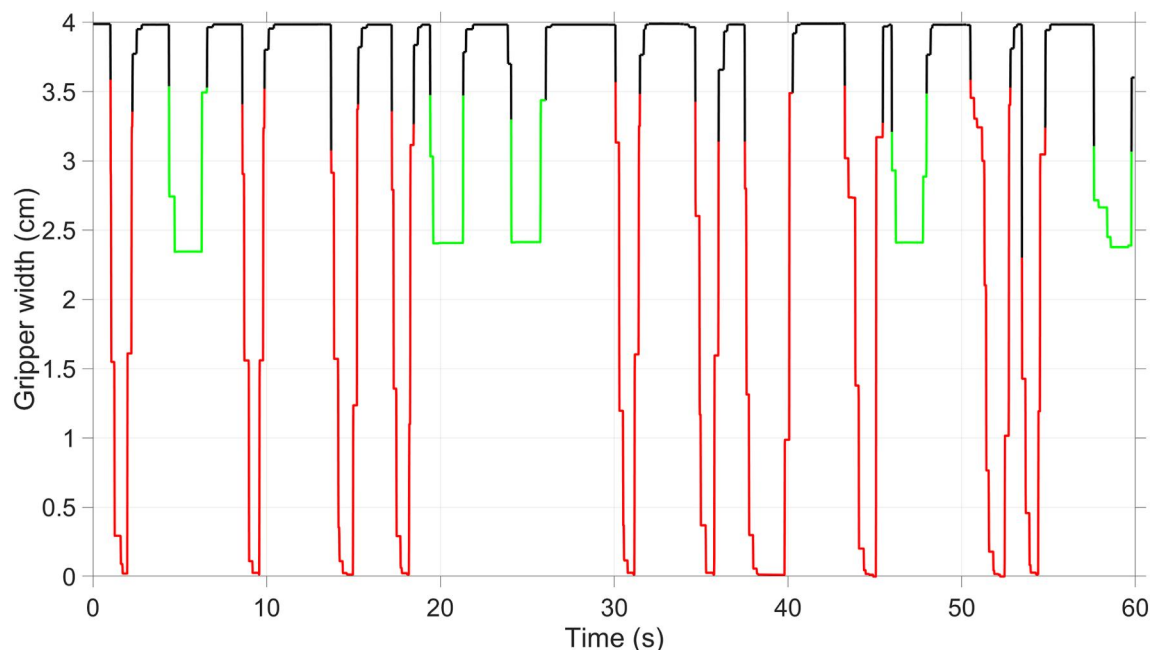
FIGURE 6 Detection example for failed (red) and successful (green) grabs. In this trial, there were five successful grabs.

tray on the opposite side. The range in cm for each separate displacement was determined in the x-, y-, and z-direction, where the x-direction corresponds to left–right movement, the y-direction represents the height above the table, and the z-direction concerns depth, that is, towards versus away from the user. A higher range may be indicative of worse performance, for example, due to deteriorated depth perception. For instance, if the participant lifts the block high with the robot arm (high y-value), or moves it forward and backward (high range in z-direction), this can be seen as a less efficient control method. The ranges were calculated per successful grab and subsequently averaged over grabs of the trial.

- *Peak force,* computed by taking the average of the identified peaks in the resultant force measured at the end effector. The three-dimensional force was provided by franka_ros [45] and calculated from the measured joint torque sensors and the robot kinematics. For a peak to qualify, it must have a height exceeding 10 N. Figure 7 provides a visual representation of detected peaks for a one trial. These peak forces can be attributed to inadvertent collisions with the blocks or other hard objects. We therefore considered the amplitude and frequency of these peaks as an indicator of the dexterity with which the participant controlled the robot arm.

2.6.3 | Head movement

- *Standard deviation (SD) of HMD orientation*, calculated from the pose data of the HMD. The HMD's orientation was captured as a quaternion and converted into a rotation matrix. This matrix was then transformed into a unit vector representing the direction in which the HMD faces, with components u, v, and w. The standard deviation for each component was computed, and the overall standard deviation of the unit vector was determined by taking the square root of the sum of the standard deviations of the components. The metric describes the extent to which participants moved themselves, and could provide an indication for the employment of motion parallax cues or another strategy to gauge depth.

- *Mean HMD angular speed.* In addition to variation in the orientation of the HMD, the extent to which participants moved their heads was also considered. This was calculated by determining the angular difference of the orientation for all consecutive samples and then averaging across all samples. This measure provides an indication of the rotational mobility of the participants' heads during the trial.

2.6.4 | Subjective variables

- Perceived easiness for each trial, recorded using a Single Ease Question (SEQ; ref. [57]), ranging from −3 (very difficult) to 3 (very easy).
- Perceived usability of the condition, recorded using a Usability Metric for User Experience (UMUX; ref. [58]), ranging from −3 (strongly disagree) to 3 (strongly agree).

2.7 | Procedures

The present study was administered by a pair of investigators: the primary researcher introduced the procedures and oversaw the telerobotic system, whilst the assistant reset the tasks and disseminated the surveys; jointly they counted the relocated blocks and noted any salient observations.
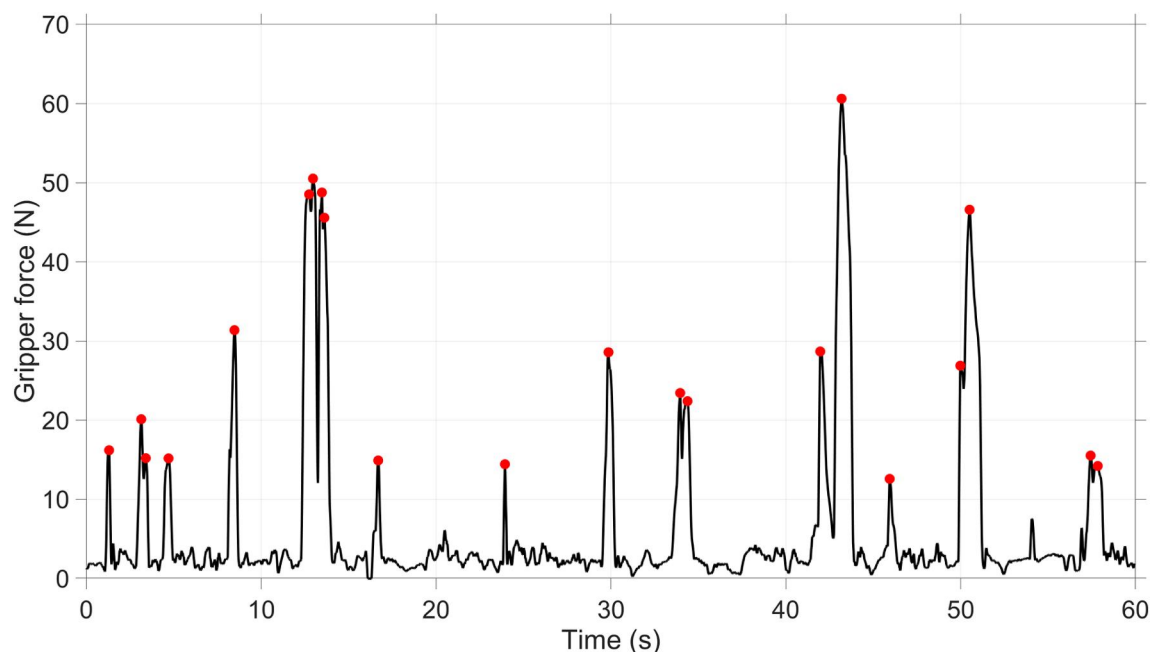
FIGURE 7 Peak force detection in the resultant force signal acting on the gripper.

2.7.1 | Training task

During the training session, the experimenter provided contextual information and instructions. The participant was given the freedom to adjust the chair as needed and was asked to roll up their sleeves to ensure optimal hand tracking. The participant was reminded to maintain a safe distance from the robot and to view it as a tool to increase task performance.

Participants were instructed on the proper hand positioning for motion tracking, adopting a designated zero-pose (Figure 3). The experimenter also clarified that only the index, middle, and thumb fingers should be used for grabbing.

The participant received information about potential scenarios that could temporarily disrupt hand tracking and the system's response in those cases. If tracking was lost, the robot would automatically return to a ready position; once tracking was reestablished, the robot would resume responding to the participant's hand movements as before. Participants learned that tracking typically recovers quickly and that it could be reliably reacquired by moving one's hand to the centre of the workspace into the zero pose. By exploring the limits of the Leap Motion's tracking capabilities, participants discovered how loss of tracking could occur. They then practiced reacquiring tracking using the Leap Motion Control Panel's Diagnostic Visualiser, which displayed the raw sensor feed overlaid with a tracked 3D hand model.

The participant was then given control of the robot to experience its translation and rotation in response to their input. The experimenter also explained the protocol for self-collisions or joint-limit situations, and the participant observed how the robot would react.

Finally, the participant completed the actual training task: manoeuvring the robot into specific poses and using the gripper

to pick and place nine different objects. During training, the participant was allowed to ask questions and explore the system's capabilities through various approaches.

2.7.2 | Experimental task

Subsequent to the training session, the experimental task was prepared and the task objectives, along with its rules, were explained to the participant. The participant received instructions that they would be granted three attempts per condition to transfer as many blocks as possible from one box to the other within a 1-min time limit. The experimental task was reset at the beginning of each trial. Upon completion of the third trial, the participant was presented with the option to take a brief break if they so choose. The entire experiment took about 70 min to complete (see Table 1 for an overview).

2.8 | Statistical analysis

The scores of the dependent variables were computed per trial and subsequently averaged over the three trials of the corresponding experimental condition (A–F). To compare the conditions, direct comparisons through paired-samples t-tests were conducted at the level of participants ($n = 24$).

Comparisons of conditions B and C are of interest to evaluate whether the feedthrough of condition C is disadvantageous compared to the direct view of the robot arm in condition B. Furthermore, it is of interest to assess whether there might be a disadvantage between remote teleoperation (D–F) and the more direct forms of robotic operation (B–C). Moreover, a comparison of condition D versus condition E

T A B L E 1   Procedure overview and duration of the experiment trials.

Phase	Experimenter's role	Duration
Pre-experiment	Welcome participant and direct them to be seated at the table	1 min
	Provide participant with informed consent form to read and sign	1 min
	Request participant to complete pre-experiment questionnaire	2 min
Training	Illustrate experiment and setup to the participant, including control instructions	5 min
	Facilitate participant's control input practice on the training task. Enquire about participant's progress and offer guidance as needed	5 min
Conditions A–F	Arrange the condition in accordance with the Latin square	1 min
	First trial	1 min
	Request participant to verbally respond to SEQ questionnaire and reset setup	30 s
	Second trial	1 min
	Request participant to verbally respond to SEQ questionnaire and reset setup	30 s
	Third trial	1 min
	Request participant to verbally respond to SEQ questionnaire and reset setup	30 s
	Direct participant to complete UMUX questionnaire	3 min
Post-experiment	Engage in a one-on-one discussion regarding the participant's experience	5 min

allows for an assessment of the two forms of 3D presentation, that is, with and without a virtual representation of the robot arm. Lastly, comparing condition F with conditions D and E investigates whether the 3D presentation (conditions D & E) is beneficial compared to the 2D feed (condition F). Note that condition A did not yield robot-related or HMD data, as it was manually operated. However, condition A did yield results for the self-reports (SEQ, UMUX) and the number of blocks transferred, which the experimenters counted manually.

An alpha value of $0.05/2 = 0.025$ was chosen because the interest typically lies in comparing 1 or 2 pairs of conditions for the same hypothesis. A visual presentation using averages and 95% confidence intervals, which is appropriate for a within-subjects design [59] was used. In all figures, non-overlapping confidence intervals are statistically significant ($p < 0.05$).

3 | RESULTS

3.1 | Grab counts

Figure 8 provides boxplots showing the number of grab attempts, which are further categorised into failed and successful grabs. Additionally, Figure 8 shows the number of transferred blocks as counted by the experimenters after each trial. It is important to note that condition A (completely manual without a robotic arm) is not included due to the lack of available measurement data for grab attempts. In condition A, participants transferred an average of 74.0 blocks (standard deviation: 9.41 blocks), which is significantly and substantially higher than the average of 4.4 to 6.9 blocks transferred in conditions B–F.

A few interesting patterns emerge from the data. First, there is no significant difference in the number of grab

attempts between conditions B–F, suggesting that participants did not feel restricted in their attempts to grab, irrespective of the condition. However, differences exist in the number of successful and unsuccessful attempts, with the local conditions B and C resulting in greater success compared to the teleoperation conditions D–F. In terms of the total number of blocks transferred, conditions B–C also outperformed conditions D–F. There was no statistically significant difference between conditions B and C, or between conditions D, E, and F for any of the four grab-related measures.

3.2 | Gripper movement

Figure 9 presents a heatmap of the x versus y coordinate of the gripper, thus offering a frontal view. A characteristic pattern is evident, explained by the fact that first a block is grabbed from the right box, which is then lifted and released above the left receptacle. There are also differences between conditions to be seen, where for example, condition F appears to show a greater spread in the trajectories than conditions B and C.

These differences have been quantified by means of the range for the three different directions, as depicted in Figure 10. The lateral amplitude, that is, in x-direction, is approximately 21 cm, and does not significantly differ between conditions. However, in the vertical direction (y-direction), there are statistically significant differences, with condition B having significantly lower values than conditions D–F, and condition C having lower values than D and F. Furthermore, condition F yielded a higher range than both condition D and condition E. Also in the z-direction there are statistically significant differences, with B having a lower value than D and F. In short, it appears that direct view (condition B) yielded a
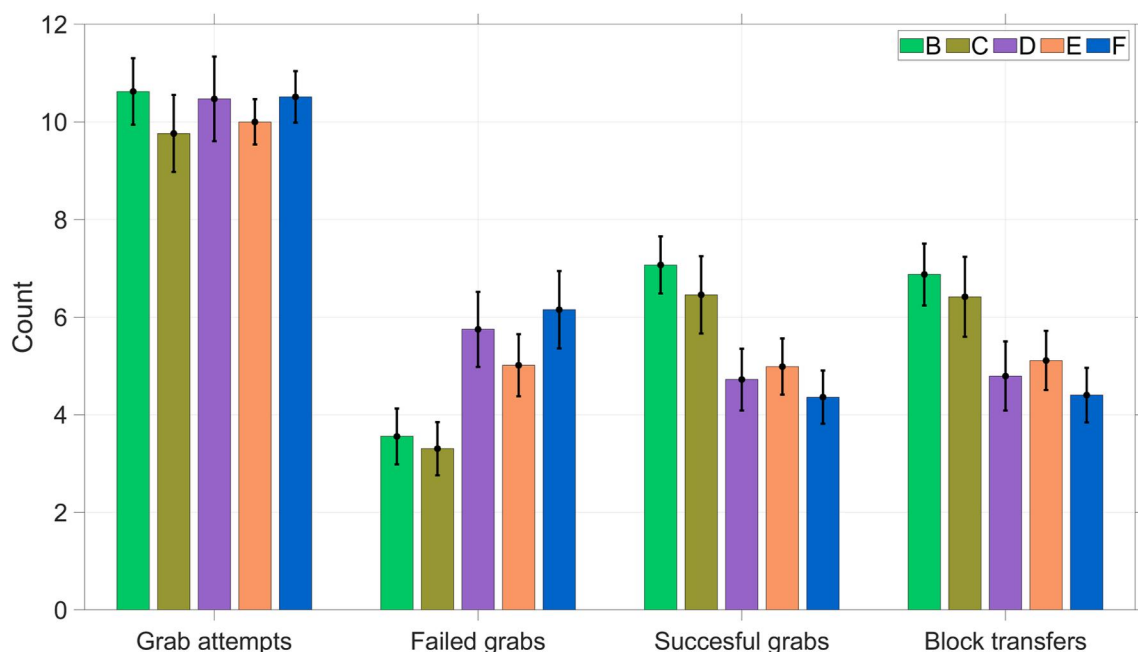
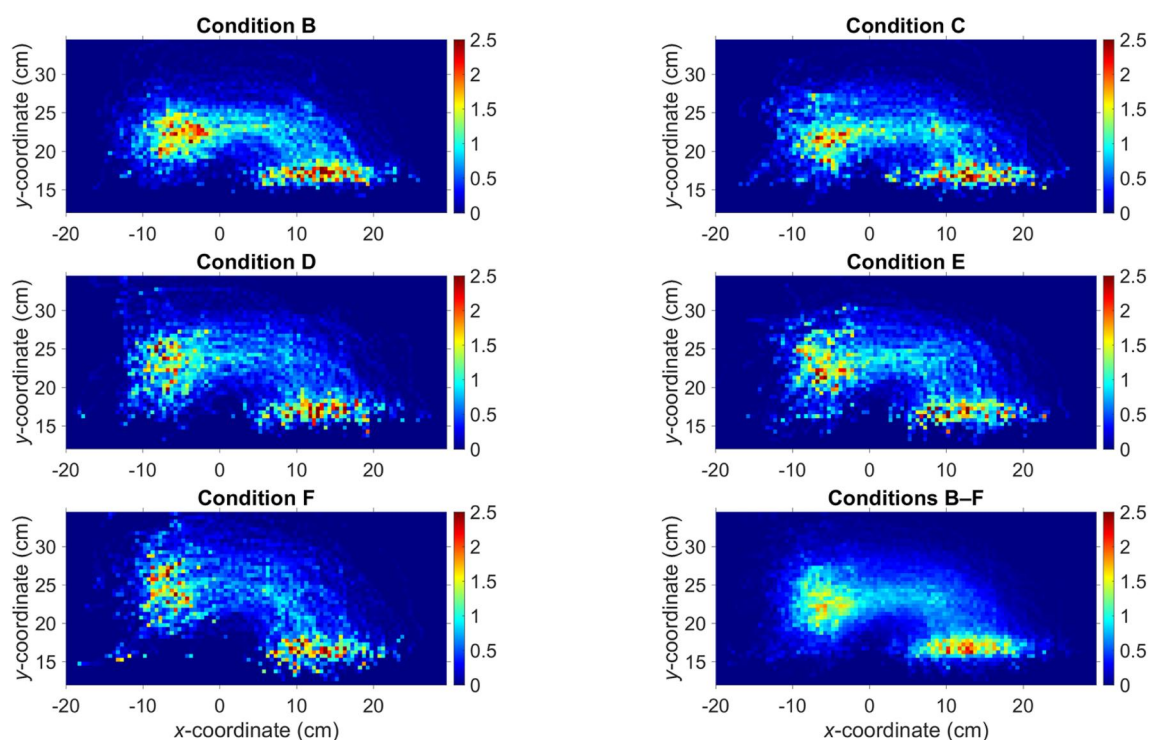**FIGURE 8** Grab-related means and 95% confidence intervals.

FIGURE 9 Heatmap of gripper x- and y-coordinates of all trials per robotic condition and for the five robotics conditions combined. The resolution for creating this heatmap was set at 0.5 cm. Colour coding from blue to red corresponds to the number of measurement samples in that cell. The sum of the values of all cells equals 1000.

more efficient movement of the robot than the teleoperation conditions, and in particular the 2D video (condition F).

Regarding the peak forces (see Figure 11), there was a comparable trend. Here, condition B yielded a significantly lower peak force, as well as fewer such peaks, than conditions D–F, and condition C yielded significantly lower peaks and fewer peaks than conditions D and F. Also, condition F yielded a significantly higher peak force than condition E.
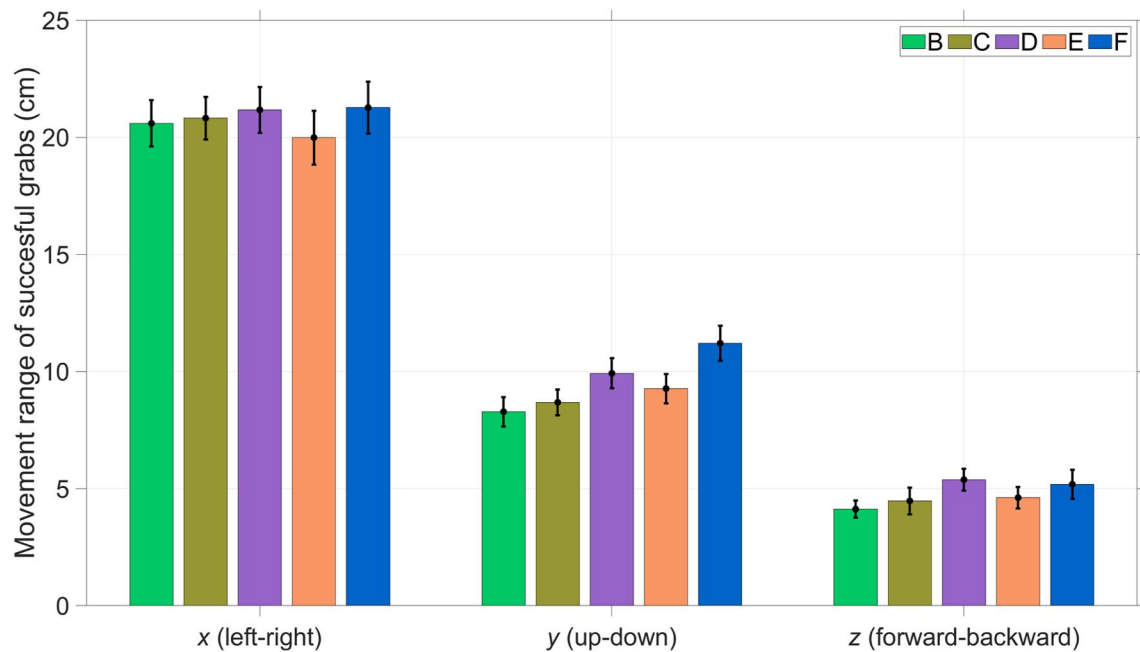
FIGURE 10 Means and 95% confidence intervals of movement range of successful grabs.
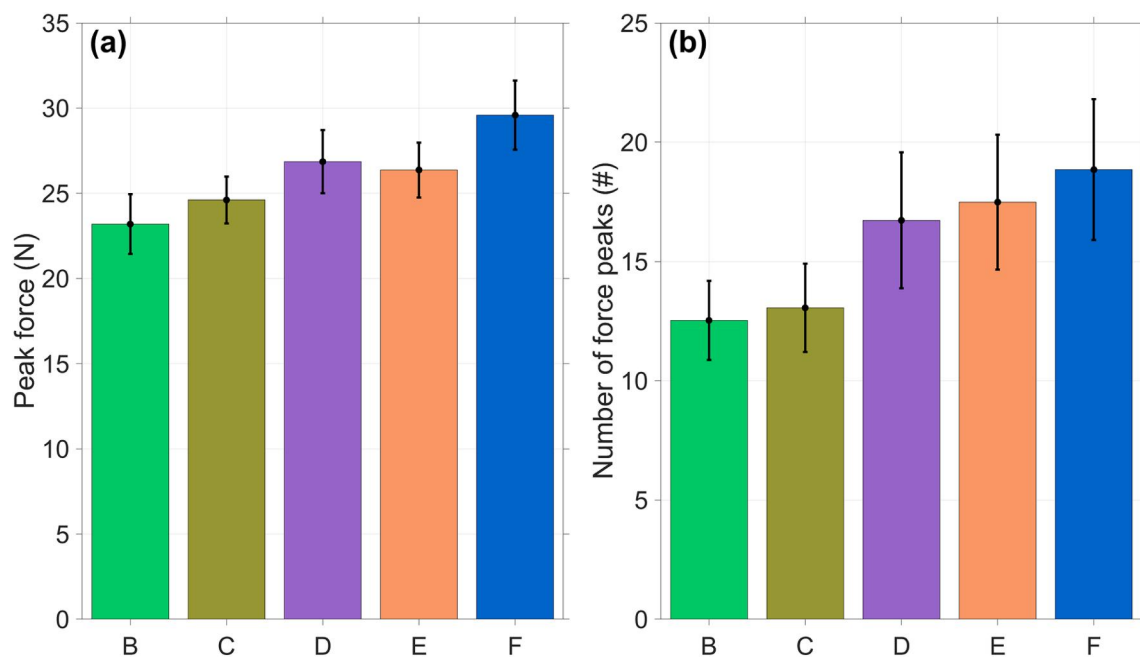

FIGURE 11 (a) Means and 95% confidence intervals of mean peak force. (b) Means and 95% confidence intervals of number of force peaks.

3.3 | Head movement

The analysis of head movements does not provide information about how well participants performed the task, but provides information about their strategies for visually exploring the task domain. In this regard, a noteworthy finding is that condition F, teleoperation using 2D video feed, yielded a statistically significantly reduced variability of head orientation (Figure 12a). Moreover, a local video feedthrough (condition C) of the robot

arm resulted in statistically significantly higher head movement compared to the other conditions (Figure 12b).

3.4 | Subjective variables

The results of the Single Ease Question (SEQ; Figure 13) reveal a clear pattern, with participants finding direct interaction between their hands and the blocks (A) the easiest,
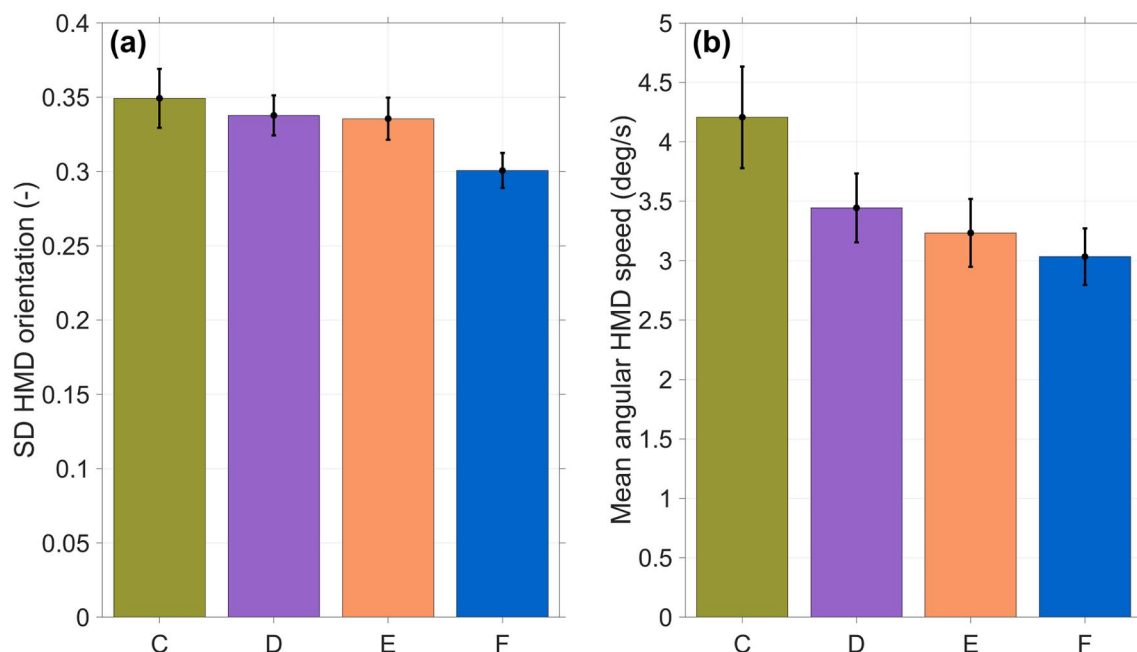
FIGURE 12 (a) Means and 95% confidence intervals of standard deviation (SD) of head orientation angle. (b) Means and 95% confidence intervals of head movement angular activity.
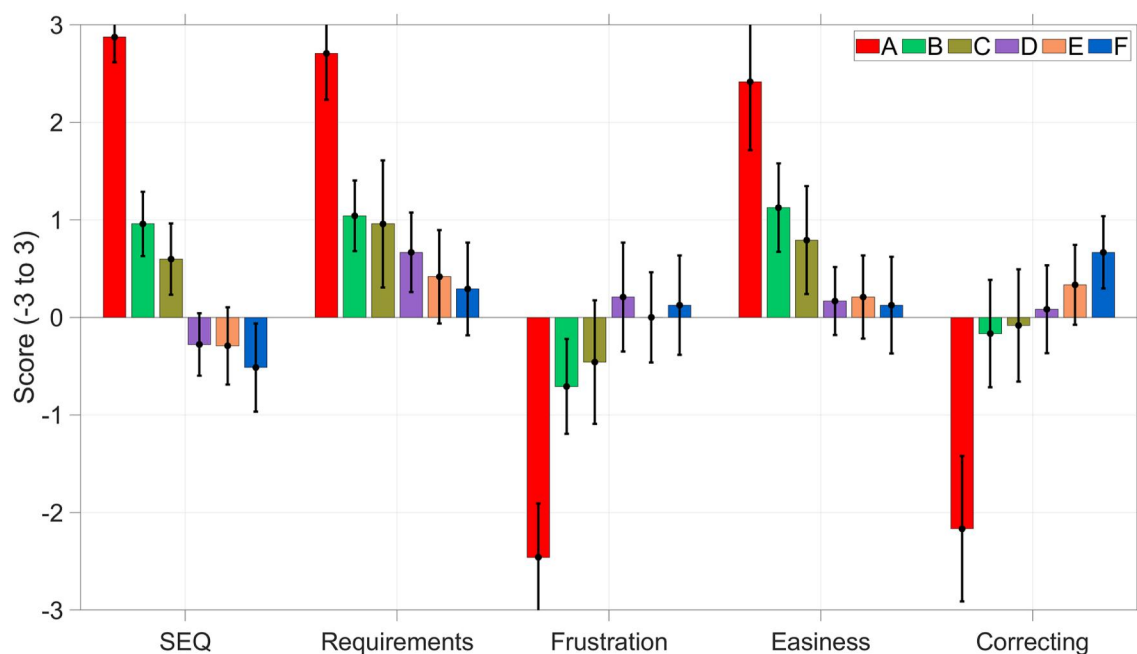

FIGURE 13 Means and 95% confidence intervals of responses to the Single Ease Question (SEQ) and the four items of the Usability Metric for User Experience (UMUX).

followed by the local conditions (B & C), and then the teleoperation conditions (D–F). With regard to local operation, there were no significant differences between conditions B and C, and with regard to the teleoperation conditions, there were no significant differences between conditions D, E, and F.

A similar trend is present for the UMUX, where participants found condition A, the direct interaction, to best meet their requirements, cause the least frustration, be the easiest to use, and require the fewest corrective interactions (see Figure 13). Once again, there was a slight advantage for the local conditions (B & C) over the teleoperation conditions (D–F). Also with regard to these metrics, there were no statistically significant differences between B and C, nor between D, E, and F.

4 | DISCUSSION

This work presented the results of a human–robot interaction experiment using the box-and-block task, a widely used assessment for measuring unilateral gross manual dexterity [60, 61]. Participants were seated in front of the box, and their task was to transfer as many blocks as possible from one compartment to another in 60 s, using only one hand.

A key finding was that teleoperation (conditions D–F) yielded lower performance and a less satisfactory user experience than local robot operation (B & C). A possible explanation for the diminished performance in teleoperation could be imperfect resolution or rendering, or that the participants missed veridical depth cues in the environment. Additionally, using a high-speed camera, we established that the visual latency—from motion inception to visual presentation—in conditions D, E, and F, was approximately 125 ms, predominantly attributable to the ZED2 camera. This is contrasted by the 20 ms latency observed in condition C, where a direct video feed from the Varjo XR-3 was used.

At the same time, we found that 3D telepresence, either with virtual representation of the robot arm (condition D) or without it (condition E), enjoyed advantages over teleoperation using a 2D video feed (condition F). This was evident from the movement amplitude during the transfer of blocks, with participants lifting the blocks higher when relying on the 2D video. There were also trends that condition F caused more collisions, for example, with blocks or the edge of the receptacle, which seems to represent a less coordinated form of interaction. An explanation for these findings can be found in theories such as animate vision [62], purposive/active vision [63], and sensorimotor contingency [64], which highlight the intertwinedness of action and perception. Specifically, participants in condition F were unable to access stereopsis and motion parallax cues, which are essential for perceiving depth (e.g. [39, 65]). In contrast, conditions D and E led to an increase in head movements among the participants compared to condition F, a behaviour that could be interpreted as a strategy to enhance depth inference, thereby potentially improving their task performance.

Another key finding was that condition A, which involved the use of the hand without a robot arm, produced the best performance. Participants achieved an average of 74.0 blocks transferred, a result comparable to the 77 blocks described in the literature [66]. In comparison, the average number of blocks transferred per minute ranged between 4.4 and 6.9 in the other experimental conditions, which confirms that direct task execution remains unparallelled compared to the use of a robot arm. An explanation is that the human arm and hand have evolved over millions of years, adapting to accomplish tool use with exceptional dexterity and sensitivity (e.g. [67–69]). The biomechanics of the human hand, comprising many bones, muscles, and tendon configurations, as well as advanced tactile receptors, endow the hand with capabilities that remain difficult to reproduce in robotic systems [70–72]. This is further illustrated in Appendix A, where we demonstrate that there is a delay in reproducing the hand movement of the robot arm, averaging at about 290 ms.

5 | CONCLUSIONS AND OUTLOOK

In conclusion, this study investigated the application of AR in teleoperation tasks involving robotic manipulation. The results showed that direct human engagement with the task (condition A) yielded the highest performance, as participants transferred significantly more blocks compared to the other conditions. Among the teleoperation conditions, a 3D image, either with virtual representation of the robot arm (condition D) or without it (condition E) enabled smoother interaction than 2D video teleoperation (condition F). However, all teleoperation modalities were outperformed by the local control conditions (B and C). Furthermore, teleoperation interface impacted head movement patterns. Specifically, the 2D video feed (condition F) led to reduced head motions compared to other modalities. Overall, these findings indicate that while AR-based teleoperation shows promise, it has yet to match the efficiency of direct human–environment interaction.

These findings have theoretical relevance, where the observed performance deficits in teleoperation compared to direct human interaction can be attributed to the inherent limitations of current teleoperation systems, such as time delays, which hinder the ability of the human operator to accurately perceive objects in the remote environment. Furthermore, the advantage of using a 3D imagery in teleoperation, as opposed to a 2D video feed, lies in the improved depth perception made possible by the 3D presentation. This allows for more accurate manipulation of objects, alleviating some of the limitations faced in traditional teleoperation. However, the superior performance of direct human interaction with the environment remains unmatched and can be credited to the evolutionary adaptation of the human hand, which over millions of years, has developed exceptional dexterity and sensitivity.

Moving forward, it is recommended that researchers continue studying and developing robotic systems that more closely mimic the capabilities of the human hand, drawing inspiration from human biomechanics and tactile receptors. Further inquiry into active vision in teleoperation, like examining the role of motion parallax, should also be pursued, as this could allow for more effective teleoperation systems. Additionally, research should focus on pragmatic aspects like minimising time delays and improving hand-tracking. For example, teleoperation (conditions D–F) involved visual time delay (see Appendix A), while conditions B and C sometimes lost hand tracking. For a seamless comparison between interaction types, future work should improve and equalise these factors.

AUTHOR CONTRIBUTIONS

Thomas A. B. de Boer: Conceptualization; data curation; formal analysis; investigation; methodology; resources; software; validation; visualization; writing – original draft; writing – review & editing. **Joost C. F. de Winter**: Conceptualization; data curation; formal analysis; methodology; resources; software; validation; visualization; writing – original draft; writing – review & editing. **Yke Bauke Eisma**: Conceptualization; data curation; investigation; methodology; project administration;

resources; software; supervision; validation; writing – original draft; writing – review & editing.

ACKNOWLEDGEMENTS

CONFLICT OF INTEREST STATEMENT

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

DATA AVAILABILITY STATEMENT

- The ROS package for Panda control is available here: https://github.com/Thomahawkuru/cartesian_panda.
- The Unity project is available here: https://github.com/Thomahawkuru/panda_teleoperation_experiment.
- The experimental data and MATLAB scripts used for analysing the data and for creating the figures in the paper are available here: https://doi.org/10.4121/8e458097-bc95-4a62-9b03-6f36fc8faedb.

ORCID

Joost C. F. de Winter 🔟 https://orcid.org/0000-0002-1281-8200

REFERENCES

1. Arena, F., et al.: An overview of augmented reality. Computers 11(2), 28 (2022). https://doi.org/10.3390/computers11020028
2. Makhataeva, Z., Varol, H.A.: Augmented reality for robotics: a review. Robotics 9(2), 21 (2020). https://doi.org/10.3390/robotics9020021
3. Gerup, J., Soerensen, C.B., Dieckmann, P.: Augmented reality and mixed reality for healthcare education beyond surgery: an integrative review. Int. J. Med. Educ. 11, 1–18 (2020). https://doi.org/10.5116/ijme.5e01.eb1a
4. Lungu, A.J., et al.: A review on the applications of virtual reality, augmented reality and mixed reality in surgical simulation: an extension to different kinds of surgery. Expet Rev. Med. Dev. 18(1), 47–62 (2021). https://doi.org/10.1080/17434440.2021.1860750
5. Egger, J., Masood, T.: Augmented reality in support of intelligent manufacturing – a systematic literature review. Comput. Ind. Eng. 140, 106195 (2020). https://doi.org/10.1016/j.cie.2019.106195
6. Fernández Del Amo, I., et al.: A systematic review of augmented reality content-related techniques for knowledge transfer in maintenance applications. Comput. Ind. 103, 47–71 (2018). https://doi.org/10.1016/j.compind.2018.08.007
7. Ong, S.K., Siew, C.Y., Nee, A.Y.C.: Augmented reality in maintenance: a review of the state-of-the-art and future challenges. In: Nee, A.Y.C., Ong, S.K. (eds.) Springer Handbook of Augmented Reality, pp. 575–595. Springer, Cham (2023). https://doi.org/10.1007/978-3-030-67822-7_23
8. Milovanovic, J., et al.: Virtual and augmented reality in architectural design and education. In: Proceedings of the 17th International Conference CAAD Futures 2017, Istanbul, Turkey (2017)
9. Song, Y., Koeck, R., Luo, S.: Review and analysis of augmented reality (AR) literature for digital fabrication in architecture. Autom. ConStruct. 128, 103762 (2021). https://doi.org/10.1016/j.autcon.2021.103762
10. Boboc, R.G., et al.: Augmented reality in cultural heritage: an overview of the last decade of applications. Appl. Sci. 12(19), 9859 (2022). https://doi.org/10.3390/app12199859
11. Trunfio, M., et al.: Innovating the cultural heritage museum service model through virtual reality and augmented reality: the effects on the overall visitor experience and satisfaction. J. Herit. Tourism 17, 1–19 (2022). https://doi.org/10.1080/1743873X.2020.1850742
12. Pidel, C., Ackermann, P.: Collaboration in virtual and augmented reality: a systematic overview. In: De Paolis, L., Bourdot, P. (eds.) Augmented Reality, Virtual Reality, and Computer Graphics. AVR 2020, pp. 141–156. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58465-8_10
13. Wang, P., et al.: AR/MR remote collaboration on physical tasks: a review. Robot. Comput. Integrated Manuf. 72, 102071 (2021). https://doi.org/10.1016/j.rcim.2020.102071
14. Laera, F., et al.: Augmented reality for maritime navigation data visualisation: a systematic review, issues and perspectives. J. Navig. 74(5), 1073–1090 (2021). https://doi.org/10.1017/S0373463321000412
15. Limbu, B.H., et al.: Using sensors and augmented reality to train apprentices using recorded expert performance: a systematic literature review. Educ. Res. Rev. 25, 1–22 (2018). https://doi.org/10.1016/j.edurev.2018.07.001
16. Araujo, B., et al.: Snake charmer: physically enabling virtual objects. In: Proceedings of the TEI'16: Tenth International Conference on Tangible, Embedded, and Embodied Interaction, Eindhoven, The Netherlands, pp. 218–226 (2016). https://doi.org/10.1145/2839462.2839484
17. Hertel, J., et al.: A taxonomy of interaction techniques for immersive augmented reality based on an iterative literature review. In: Proceedings of the IEEE International Symposium on Mixed and Augmented Reality, Bari, Italy, pp. 431–440 (2021). https://doi.org/10.1109/ISMAR52148.2021.00060
18. Delmerico, J., et al.: Spatial computing and intuitive interaction: bringing mixed reality and robotics together. IEEE Robot. Autom. Mag. 29(1), 45–57 (2022). https://doi.org/10.1109/MRA.2021.3138384
19. Elbasheer, M., et al.: Shaping the role of the digital twins for human-robot dyad: connotations, scenarios, and future perspectives. IET Collaborative Intelligent Manufacturing 5(1), e12066 (2023). https://doi.org/10.1049/cim2.12066
20. Malik, A.A., Brem, A.: Digital twins for collaborative robots: a case study in human-robot interaction. Robot. Comput. Integrated Manuf. 68, 102092 (2021). https://doi.org/10.1016/j.rcim.2020.102092
21. Tsokalo, I.A., et al.: Remote robot control with human-in-the-loop over long distances using digital twins. In: Proceedings of the IEEE Global Communications Conference, Waikoloa, HI (2019). https://doi.org/10.1109/GLOBECOM38437.2019.9013428
22. Top, F., Pütz, S., Fottner, J.: Human-centered HMI for crane teleoperation: intuitive concepts based on mental models, compatibility and mental workload. In: Harris, D., Li, W.C. (eds.) Engineering Psychology and Cognitive Ergonomics. HCII 2021, pp. 438–456. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-77932-0_34
23. Liu, A.M., et al.: Predicting space telerobotic operator training performance from human spatial ability assessment. Acta Astronaut. 92(1), 38–47 (2013). https://doi.org/10.1016/j.actaastro.2012.04.004
24. Balatti, P., et al.: An augmented reality interface for improving task performance in close-proximity teleoperation. In: Proceedings of the Institute for Robotics and Intelligent Machines Conference, Rome, Italy (2019)
25. Rosen, E., et al.: Communicating and controlling robot arm motion intent through mixed-reality head-mounted displays. Int. J. Robot Res. 38(12-13), 1513–1526 (2019). https://doi.org/10.1177/0278364919842925
26. Su, Y.-P., et al.: Mixed-reality-enhanced human-robot interaction with an imitation-based mapping approach for intuitive teleoperation of a robotic arm-hand system. Appl. Sci. 12(9), 4740 (2022). https://doi.org/10.3390/app12094740
27. Peppoloni, L., et al.: Augmented reality-aided tele-presence system for robot manipulation in industrial manufacturing. In: Proceedings of the 21st ACM Symposium on Virtual Reality Software and Technology, Beijing China, pp. 237–240 (2015). https://doi.org/10.1145/2821592.2821620
28. Li, C., et al.: AR-assisted digital twin-enabled robot collaborative manufacturing system with human-in-the-loop. Robot. Comput. Integrated Manuf. 76, 102321 (2022). https://doi.org/10.1016/j.rcim.2022.102321

29. Park, K.B., et al.: Hands-free human-robot interaction using multimodal gestures and deep learning in wearable mixed reality. IEEE Access 9, 55448–55464 (2021). https://doi.org/10.1109/ACCESS.2021.3071364

30. Ponomareva, P., et al.: GraspLook: a VR-based telemanipulation system with R-CNN-driven augmentation of virtual environment. In: Proceedings of the 20th International Conference on Advanced Robotics, Ljubljana, Slovenia, pp. 166–171 (2021). https://doi.org/10.1109/ICAR53236.2021.9659460

31. Shahria, M.T., et al.: A novel framework for mixed reality–based control of collaborative robot: development study. JMIR Biomed. Eng. 7(1), e36734 (2022). https://doi.org/10.2196/36734

32. Chan, W.P., et al.: Design and evaluation of an augmented reality head-mounted display interface for human robot teams collaborating in physically shared manufacturing tasks. ACM Trans. Hum. Robot Interaction 11(3), 1–19 (2022). https://doi.org/10.1145/3524082

33. Brizzi, F., et al.: Effects of augmented reality on the performance of teleoperated industrial assembly tasks in a robotic embodiment. IEEE Trans. Hum. Mach. Syst. 48(2), 197–206 (2017). https://doi.org/10.1109/THMS.2017.2782490

34. Bambušek, D., et al.: Combining interactive spatial augmented reality with head-mounted display for end-user collaborative robot programming. In: Proceedings of the 28th IEEE International Conference on Robot and Human Interactive Communication, New Delhi, India (2019). https://doi.org/10.1109/RO-MAN46459.2019.8956315

35. Kaarlela, T., et al.: Digital twins utilizing XR-technology as robotic training tools. Machines 11(1), 13 (2022). https://doi.org/10.3390/machines11010013

36. Peng, H., et al.: RoMA: interactive fabrication with augmented reality and a robotic 3D printer. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Montréal, Canada (2018). https://doi.org/10.1145/3173574.3174153

37. Quintero, C.P., et al.: Robot programming through augmented trajectories in augmented reality. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, Madrid, Spain, pp. 1838–1844 (2018). https://doi.org/10.1109/IROS.2018.8593700

38. Draper, J.V., Kaber, D.B., Usher, J.M.: Telepresence. Hum. Factors 40(3), 354–375 (1998). https://doi.org/10.1518/001872098779591386

39. Gibson, J.J.: The Ecological Approach to Visual Perception. Houghton Mifflin, Boston (1979)

40. Poggio, G.F., Poggio, T.: The analysis of stereopsis. Annu. Rev. Neurosci. 7(1), 379–412 (1984). https://doi.org/10.1146/annurev.ne.07.030184.002115

41. Rogers, B., Graham, M.: Motion parallax as an independent cue for depth perception. Perception 8(2), 125–134 (1979). https://doi.org/10.1068/p080125

42. Emika, F.: Panda. The sensitive, interconnected lightweight robot for everybody. https://web.archive.org/web/20180325091804/; http://franka.de (2018)

43. Varjo: Varjo launches XR-3 and VR-3 headsets to bring the next generation of immersive technology to every workplace. https://varjo.com/press-release/varjo-launches-xr-3-and-vr-3-headsets-to-bring-the-next-generation-of-immersive-technology-to-every-workplace (2020)

44. Ultraleap: Leap motion controller. https://www.ultraleap.com/product/leap-motion-controller (2023)

45. Emika, F.: Libfranka changelog—Franka control interface. https://frankaemika.github.io/docs/libfranka_changelog.html (2017)

46. Emika, F.: Libfranka—Franka control interface. https://frankaemika.github.io/docs/libfranka.html (2017)

47. Thomahawkuru: cartesian_panda. https://github.com/Thomahawkuru/cartesian_panda (2023)

48. Unity Technologies: Unity. https://unity.com (2023)

49. Bischoff, M.: ros-sharp. https://github.com/siemens/ros-sharp (2021)

50. Stereolabs: ZED Unity plugin. https://github.com/stereolabs/zed-unity (2022)

51. HTC: SteamVR base station 2.0. https://www.vive.com/eu/accessory/base-station2 (2023)

52. Stereolabs: Meet ZED 2. https://www.stereolabs.com/zed-2 (2023)

53. Javicom: Baby's first shape puzzle. https://www.thingiverse.com/thing:4694175 (2020)

54. Figueiredo, S.: Box and block test (BBT). https://strokengine.ca/en/assessments/box-and-block-test-bbt (2011)

55. Thomahawkuru: Box-and-blocks test. https://www.thingiverse.com/thing:5915827 (2023)

56. Emika, F.: franka-ros. https://github.com/frankaemika/franka_ros (2017)

57. Sauro, J., Dumas, J.S.: Comparison of three one-question, post-task usability questionnaires. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Boston, MA, pp. 1599–1608 (2009). https://doi.org/10.1145/1518701.1518946

58. Finstad, K.: The usability metric for user experience. Interact. Comput. 22(5), 323–327 (2010). https://doi.org/10.1016/j.intcom.2010.04.004

59. Morey, R.D.: Confidence intervals from normalized data: a correction to Cousineau (2005). Tutorials Quant. Methods Psychol. 4(2), 61–64 (2008). https://doi.org/10.20982/tqmp.01.1.p042

60. Chen, H.-M., et al.: Test-retest reproducibility and smallest real difference of 5 hand function tests in patients with stroke. Neurorehabilitation Neural Repair 23, 435–440 (2009). https://doi.org/10.1177/1545968308331146

61. Trombly, C.A.: Occupational Therapy for Physical Dysfunction. Williams & Wilkins (1983)

62. Ballard, D.H.: Animate vision. Artif. Intell. 48(1), 57–86 (1991). https://doi.org/10.1016/0004-3702(91)90080-4

63. Aloimonos, J.: Purposive and qualitative active vision. In: Proceedings of the 10th International Conference on Pattern Recognition, Atlantic City, NJ, pp. 346–360 (1990). https://doi.org/10.1109/ICPR.1990.118128

64. Maye, A., Engel, A.K.: Using sensorimotor contingencies for prediction and action planning. In: Ziemke, T., Balkenius, C., Hallam, J. (eds.) From Animals to Animats 12: 12th International Conference on Simulation of Adaptive Behavior, pp. 106–116. Springer, Berlin (2012). https://doi.org/10.1007/978-3-642-33093-3_11

65. Yu, K., et al.: EyeRobot: enabling telemedicine using a robot arm and a head-mounted display. Curr. Dir. Biomed. Eng. 6(1) (2020). https://doi.org/10.1515/cdbme-2020-0019

66. Mathiowetz, V., et al.: Adult norms for the box and block test of manual dexterity. Am. J. Occup. Ther. 39(6), 386–391 (1985). https://doi.org/10.5014/ajot.39.6.386

67. Kivell, T.L.: Evidence in hand: recent discoveries and the early evolution of human manual manipulation. Phil. Trans. Biol. Sci. 370(1682), 20150105 (2015). https://doi.org/10.1098/rstb.2015.0105

68. Marzke, M.W.: Tool making, hand morphology and fossil hominins. Phil. Trans. Biol. Sci. 368(1630), 20120414 (2013). https://doi.org/10.1098/rstb.2012.0414

69. Navarro, J., Hancock, P.A.: Did tools create humans? Theor. Issues Ergon. Sci. 24(2), 206–232 (2023). https://doi.org/10.1080/1463922X.2022.2076954

70. Billard, A., Kragic, D.: Trends and challenges in robot manipulation. Science 364(6446) (2019). https://doi.org/10.1126/science.aat8414

71. Hodson, R.: How robots are grasping the art of gripping. Nature 557(7704), S23–S25 (2018). https://doi.org/10.1038/d41586-018-05093-1

72. Kim, U., et al.: Integrated linkage-driven dexterous anthropomorphic robotic hand. Nat. Commun. 12(1), 7177 (2021). https://doi.org/10.1038/s41467-021-27261-0

APPENDIX A

Data quality checks

Figure A1 provides information pertaining to hand-tracking (Figure A1a) and time delay (Figure A1b).

Figure A1a shows that conditions B and C saw an increase in hand-tracking errors. For these conditions, the external Leap Motion tracker was used for motion tracking, whereas for conditions D, E, and F, the internal Leap Motion sensor of the headset was used. Apparently, the difference in sensor view angles resulted in an increased experience of tracking errors.

Figure A1b shows that the commanded motion-input to robot-movement delay is similar for all conditions, as calculated by a cross-correlation analysis.
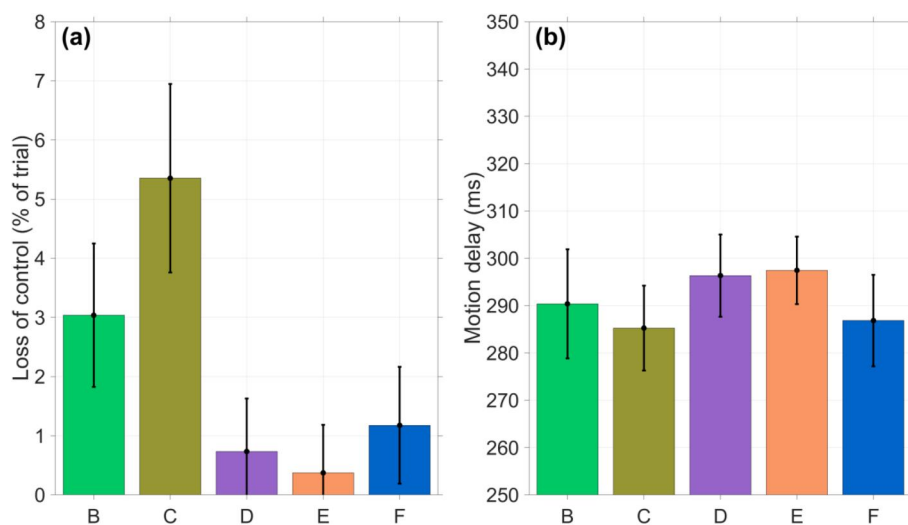

FIGURE A1 Results of data quality checks.