A Comprehensive User Study on Augmented Reality-Based Data Collection Interfaces for Robot Learning

Xinkai Jiang
xinkai.jiang@partner.kit.edu
Karlsruhe Institute of Technology
Karlsruhe, Germany

Paul Mattes
paul.mattes@kit.edu
Karlsruhe Institute of Technology
Karlsruhe, Germany

Xiaogang Jia
xiaogang.jia@partner.kit.edu
Karlsruhe Institute of Technology
Karlsruhe, Germany

Nicolas Schreiber
nicolas.schreiber@kit.edu
Karlsruhe Institute of Technology
Karlsruhe, Germany

Gerhard Neumann
gerhard.neumann@kit.edu
Karlsruhe Institute of Technology
Karlsruhe, Germany

Rudolf Lioutikov
lioutikov@kit.edu
Karlsruhe Institute of Technology
Karlsruhe, Germany

ABSTRACT

Future versatile robots need the ability to learn new tasks and behaviors from demonstrations. Recent advances in virtual and augmented reality position these technologies as great candidates for the efficient and intuitive collection of large sets of demonstrations. While there are different possible approaches to control a virtual robot there has not yet been an evaluation of these control interfaces in regards to their efficiency and intuitiveness. These characteristics become particularly important when working with non-expert users and complex manipulation tasks. To this end, this work investigates five different interfaces to control a virtual robot in a comprehensive user study across various virtualized tasks in an AR setting. These interfaces include Hand Tracking, Virtual Kinesthetic Teaching, Gamepad and Motion Controller. Additionally, this work introduces Kinesthetic Teaching as a novel interface to control virtual robots in AR settings, where the virtual robot mimics the movement of a real robot manipulated by the user. This study reveals valuable insights into their usability and effectiveness. It shows that the proposed Kinesthetic Teaching interface significantly outperforms other interfaces in both objective and subjective metrics based on success rate, task completeness, and completion time and User Experience Questionnaires (UEQ+).

CCS CONCEPTS

• **Human-centered computing → Human computer interaction (HCI)**.

KEYWORDS

Augmented reality (AR), Robot Interface, Learning from Demonstration

HRI '24, March 11–14, 2024, Boulder, CO, USA

ACM ISBN 979-8-4007-0322-5/24/03…$15.00
https://doi.org/10.1145/3610977.3634995

ACM Reference Format:
Xinkai Jiang, Paul Mattes, Xiaogang Jia, Nicolas Schreiber, Gerhard Neumann, and Rudolf Lioutikov. 2024. A Comprehensive User Study on Augmented Reality-Based Data Collection Interfaces for Robot Learning. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction (HRI '24), March 11–14, 2024, Boulder, CO, USA.* ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3610977.3634995

1 INTRODUCTION

Teaching robots new skills and tasks through demonstrations are essential goals of robot learning and human-robot interaction. The challenge of learning tasks from demonstrations has received much attention through Imitation Learning [11], Learning from Demonstrations [39] and Inverse Reinforcement Learning [8] approaches.

An important prerequisite for such approaches is the quality of the data and, hence, the data collection process itself. This requirement becomes even more important due to the high demand for data required by recent learning methods. Prominent approaches focus on collecting demonstrations from various sources such as online videos [33] or dedicated first person videos [13]. However, collecting specific task demonstrations in real world experiments harbours several challenges, e.g., reproducibility issues due to changing objects and object poses, cumbersome and slow resetting of experimental setups and inaccurate measurements due to sensor noise. Virtualisation of the experiment, including the objects and the robot itself, alleviates many of these challenges as it allows for highly reproducible and controllable settings that can be quickly reset and repeated. However, the advantages of collecting demonstrations with a virtual robot highly depend on the efficiency and intuitiveness of the control interface. This work investigates and evaluates several interfaces to control the virtual robot, ranging from hand tracking to a physical robot platform. While some approaches utilize screens to virtualize experiments [23], leveraging augmented or virtual reality (AR) offers significant advantages over screens [31], by providing an immersive experience that allows intuitive control over the perspective of the virtual environment.

While research has highlighted the advantages of AR headsets over screens for visualization purposes [31], there has not yet been a study investigating the comparative performance of different interaction methods within AR environments for collecting task demonstrations for robots. This paper closes this gap by presenting

(a) Hand Tracking (b) Virtual Kinesthetic Teaching (c) Gamepad (d) Motion Controller (e) Kinesthetic Teaching

Figure 1: The participants were asked to solve the same task, for instance, the Cup Inserting, with all five interfaces. The top row shows a participant collecting demonstrations using the different interfaces. The bottom row shows the virtualized environment (here the Cup Inserting task) as it is presented to the participant via the HoloLens 2.

a comprehensive study of different ways to interact with virtual experiments for the sake of collecting demonstrations in an AR setting. The study compares five different options to interact with the virtual robots, i.e., inside-out Hand Tracking, Kinesthetic Teaching of a virtual robot, end-effector control via Gamepad, end effector control via Motion Controller and Kinesthetic Teaching via a physical robot. Kinesthetic Teaching is a common method for demonstrating motions on a real physical system, however, we are not aware of an application of Kinesthetic Teaching to control a virtual robot in such a context. Yet, we believe that this is a very intuitive and efficient interface which is also confirmed by our study. We refer to these different approaches as interaction interfaces and describe them in detail in Section 3.

A total of 35 participants took part in the study, whereas each participant utilized every interface to collect up to 3 demonstrations across 3 tasks with varying difficulty levels. The performance of the interfaces was evaluated across several dimensions including objective measures, such as success rate, task completeness, and completion time of the demonstrated tasks as well as subjective measures surveyed via the well-established modular extension of the User Experience Questionnaire (UEQ+) [28].

The conducted study discloses that combining physical Kinesthetic Teaching with augmented reality provides a powerful yet intuitive system to efficiently collect demonstrations in virtual environments. The study further reveals that this system significantly outperforms any other interface with respect to both the objective and subjective measures resulting in an effective yet intuitive system. Such an interface could also be applied to control a real robot via teleoperation, however, this is part of future work.

In summary, the contributions of this paper are twofold. First, a comprehensive study of different interaction interfaces to collect task demonstrations using a virtual robot in an AR setting. Second, introducing a new interaction interface leveraging a physical robot platform for controlling a virtual robot in virtual experimentation.

2 BACKGROUND

While there has been work regarding both interaction interfaces and AR for Robotics, there has not yet been a study on the efficiency and intuitiveness of interaction interfaces in virtual environments.

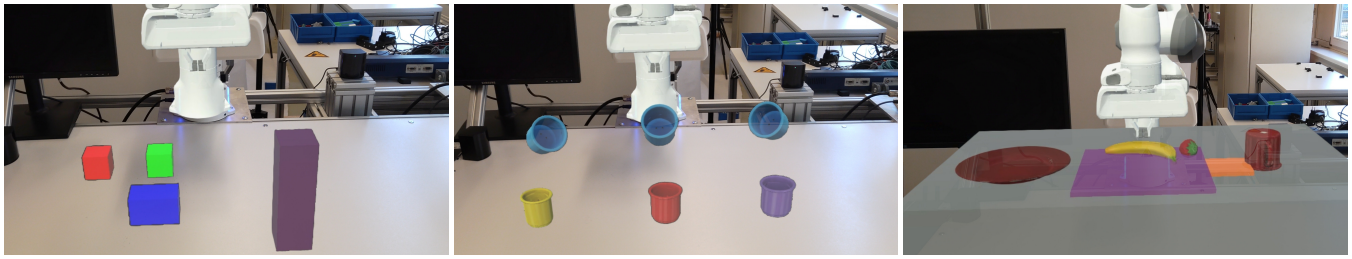
2.1 Robot Interface

Given the goal of efficiently collecting demonstrations in a virtual environment, we identified five promising interaction interfaces found in literature, namely (inside-out) Hand Tracking, Virtual Kinesthetic Teaching, Gamepad Control, Motion Controller and (Physical) Kinesthetic Teaching.

Hand Tracking uses sensors, usually cameras, to track the hand of the user and map the robot state to the hand. This interface has successfully been applied in teleoperation [5, 15] and shared-control telemanipulation [36, 38] scenarios. Virtual Kinesthetic Teaching interfaces allow the manipulation of a virtual robot directly using the participant's hands. This interface has been utilized to teleoperate physical robots in bi-manual [40] and digital twin [18] settings. Gamepads have been widely used to control physical robots while adding very little system complexity. Specifically in the area of teleoperation, Gamepads have been used to control robots directly [3] or to create trajectory templates [27]. In addition, there has been some work investigating the combined effects of Gamepads and AR in industrial robot programming [35]. Recently, Motion Controllers have become increasingly popular in both robot learning and teleoperation [37]. This interface has been used to demonstrate complex tasks on a mobile manipulator platform [34] and efficiently teleoperate a robot by separating position and orientation control into separate controllers [21]. Furthermore, haptic cues of the Motion Controller and AR visual cues can significantly improve teleoperation performance [22]. Kinesthetic Teaching commonly refers to the manipulation of a physical robot for the purpose of collecting demonstrations directly on that platform [49]. However, it also provides a very intuitive and straightforward demonstration interface for teleoperation systems [39]. While there has been some preliminary work investigating physical Kinesthetic Teaching controlling a virtual twin [44], it does not leverage the advantages of combining this interface with an AR system.

2.2 Human-Robot Interaction

AR systems have shown great promise in the area of Human-Robot Interaction. A comparison of virtual and physical robots with respect to deictic gestures has shown several advantages of mixed



(a) Box Stacking (b) Cup Inserting (c) Practical Manipulation

Figure 2: Participants were asked to control a virtual robot via various interfaces to solve three tasks with varying difficulties. (a) Box Stacking task: Stack 3 boxes vertically within the purple area. (b) Cub Inserting task: Insert 3 cups with varying diameters into larger cups in particular orientations. (c) Practical Manipulation task: Move the fruits onto the plate, subsequently move the plate into the purple area and finally flip the mug and position it on the orange area.

reality approaches [14], while other work has leveraged AR systems to investigate the human behavior in robot learning tasks [16, 30].

2.3 AR & VR for Robotics

The ability to directly render and immerse users in a 3D virtual environment makes AR technologies a promising tool for collecting task demonstrations. In combination with haptic feedback, these technologies have been shown to provide higher teaching efficiency than common GUIs [31]. AR-assisted robot learning frameworks for surgery tasks have been shown to reduce the workload demand on the users when compared to traditional Kinesthetic Teaching methods [12]. Mixed Reality visualization of potential robotic arm trajectories can enhance the accuracy and speed of collaborating users [41]. Similarly, AR can be leveraged to efficiently communicate robot motion and improve overall task performance compared to non-AR baselines [47]. AR has further been leveraged in approaches for constrained Learning from Demonstration [26] as well as sim2real RL [6]. AR is also used to improve the automation of manufacturing robots [42] and it helps robot development and research by providing visual debugging tools [17]. Other approaches use virtual reality for teleoperation including robot arm [32, 43], mobile robot [43], bi-manual robot arm [9, 18, 24], humanoid robots [7, 19] and surgery robots [2].

3 TECHNICAL DETAILS

This study investigates various interaction interfaces in the context of their effectiveness and intuitiveness for gathering demonstrations for a virtual robot in AR environments. To achieve this, a novel framework has been developed, that seamlessly integrates a game engine responsible for rendering the virtual environment on AR headsets with a physics simulator, enabling the simulation of both the virtual robot and manipulated objects.

Within this framework, five distinct interaction interfaces were implemented, including a physical Panda robot manipulator by Franka Emika, allowing for direct control of the virtual robot. This framework significantly facilitates the design of virtual tasks, presents them in an intuitive manner via augmented reality, and can efficiently collect demonstrations through various interaction interfaces. The complete source code and a detailed description, of how to set up and use the framework and all interfaces including the Panda robot for controlling its virtual counterpart can be found at https://github.com/intuitive-robots/IRXR-Unity.git. The framework serves as the foundation of the conducted user study, which aims to determine the most efficient and intuitive interaction interface for the collection of task demonstration in virtual environments based on both objective metrics, such as success rate, task completeness, and completion time and subjective metrics evaluated via UEQ+.

3.1 Virtualization

3.1.1 Physics Simulator. The framework deploys the MuJoCo physics simulator [46], which is widely utilized in robotic algorithms and simulations. The virtual environments, including various manipulatable objects and a Panda robot, were implemented within the simulator and combined into several meticulously designed scenarios. Additional data loggers were implemented that record state information of the virtual robot and objects, including position, velocity, acceleration and orientation.

3.1.2 Augmented Reality Platform. The virtual scene is presented to the user via the Microsoft HoloLens 2 [25]. Leveraging the Unity Engine [45], a custom AR application specifically tailored for the HoloLens 2 was developed. This AR application is capable of real time message passing via WebSocket to and from the simulator running on a desktop PC. The HoloLens 2 renders all virtual elements, including the robot and the objects, in real time, providing users with an immersive and interactive experience.

3.1.3 AR Alignment. The alignment between the virtual environment and the real world is achieved by incorporating a trackable QR code, that represents the world coordinate frame of the virtual environment. This approach allows the easy alignment of elements from the virtual to the physical realm, such as the virtual robot and the real robot in the physical Kinesthetic Teaching interface.

3.2 Interaction Interfaces

This study investigates five different interaction interfaces with increasing hardware demands beyond an AR Headset. In order to maximize the effectiveness and intuitiveness of each interface a pre-study, described in Section 4.2.1, with six selected users, was

conducted and feedback and suggestions from each user were leveraged to improve the interfaces. To avoid bias, these six users did not participate in the subsequent user study.

3.2.1 Hand Tracking. The inside-out Hand Tracking (HT) interface, shown in Figure 1(a), uses two scene cameras of the AR Headset for hand tracking and gesture recognition. To increase the intuitiveness, the gripper of the virtual robot is aligned with the index finger and thumb of the tracked hand. Given the end effector pose, the robot configuration is determined using an IK solver. The participants can intuitively control the robot's movements by moving their hands around. A pinch or release motion triggers the closing and opening of the gripper. A significant limitation of this approach is the need to always maintain a clear view of the hand since the inside-out tracking requires the hand to be within the view of the headset.

3.2.2 Virtual Kinesthetic Teaching. Similar to the Hand Tracking interface the Virtual Kinesthetic Teaching (VT), shown in Figure 1(b), also allows for the direct control of the virtual robot without additional hardware. However, rather than directly mapping the hand to the end effector this interface turns the virtual robot into an interactable virtual module. The participants can move the robot by grabbing the virtual end effector. Releasing the end effector stops the tracking. Stretching and squeezing gestures trigger the gripper to close or open. The robot configuration is determined using an IK solver given the current end effector pose. Since this interface also relies on Hand Tracking and gesture recognition it suffers from the same limitation as the Hand Tracking interface. However, it has the big advantage, compared to Hand Tracking, that the control of the virtual robot can be paused at any time by releasing it which, for instance, allows for comfortable re-orientation of the hands.

3.2.3 Gamepad. The Gamepad (GP) interface, shown in Figure 1(c), uses a Microsoft Xbox controller to manipulate the virtual robot. Similar to the control of aerial robots [48], participants control the end effector pose by pushing and pulling the thumb-sticks. Subsequently, an IK solver is used to compute the configuration of the virtual robot. Additional buttons reset the end effector pose and open or close the gripper.

3.2.4 Motion Controller. The Motion Controller (MC) interface, shown in Figure 1(d), uses a Vive Pro MC 2.0 and 4 SteamVR base stations 2.0, which precisely measure the MC's position and orientation in real time. The end effector pose of the virtual robot is mapped to the pose of the MC and an IK solver is used to compute the robot configuration. The virtual gripper is opened and closed by holding and releasing the triggers of the controller. A limitation of this interface is the need for an additional tracking system.

3.2.5 Kinesthetic Teaching. The Kinesthetic Teaching (KT) interface, shown in Figure 1(e), allows participants to directly control a physical version of the virtual robot. The physical robot transmits joint positions and velocities to the virtualization framework in real time, mapping the configuration from the physical robot to the virtual one. The virtual environment is aligned with the real world, such that the virtual and the real robot overlap exactly. While this interface provides the most detailed control of the virtual robot it has the limitation that access to the actual physical robot is required.

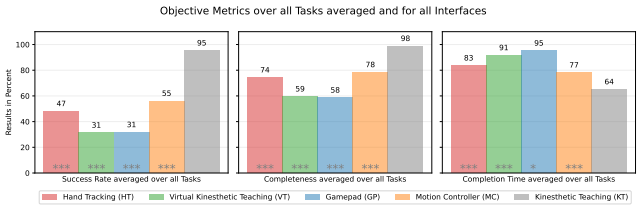


Figure 3: This graph shows three objective metrics averaged across all tasks, with each subgraph corresponding to one specific metric. From this graph, KT took the first place in success rate and completeness and has the lowest average completion time. The stars inside the bars correspond to statistical significance compared to KT (*: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$).

4 USER STUDY DESIGN

In order to assess the efficiency and intuitiveness of various interfaces for collecting demonstrations, we designed a comprehensive user study. The study aims to evaluate each interface thoroughly and establish meaningful comparisons among them. The study was granted ethical approval, participation was voluntary and informed consent was given by all participants and guardians if necessary.

4.1 Questionnaire Design

Each participant was asked to answer two types of questionnaires. One regarding their background with respect to the interfaces and one to asses each individual interface in the context of controlling a virtual robot. The background questionnaire includes 7 questions and aims at theoretical and practical past experiences of participants with respect to physical robots, AR/VR/MR devices, and the GP. The multiple choice questions identify potential positive influences and biases during task execution and helped to avoid subjective scale measuring [20]. Further details are provided in Appendix A.1.

The control questionnaire measures the subjective assessment of participants regarding the usage of each of the 5 different interfaces. The questionnaire itself consists of five UEQ+ scales [28], including attractiveness, efficiency, perspicuity, dependability, and novelty. Attractiveness focuses on the likeability of the interface, efficiency measures how well the participants think they performed, perspicuity indicates how easy it is to learn the interface, dependability reflects if the interface responds predictably and consistently to the input and commands of the participant, and novelty measures if the participant thinks that the interface is original. Each scale presents four pairs of contrastive adjectives along with a scale ranging from one to seven, where four is neutral.

4.2 Study Procedure

4.2.1 Pre-Study. Before the actual user study, a pre-study was conducted to collect initial feedback regarding the interfaces and to investigate the different metrics. The participants performed all three tasks using all interfaces and filled out the same questionnaires as in the actual user study. Further, free-form feedback about the usage of the different interfaces was collected. Based on this feedback, all five interfaces were optimized to function as flawlessly

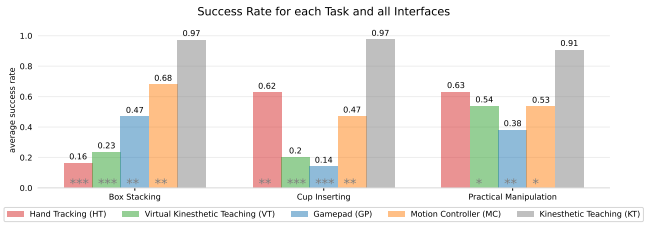


Figure 4: This graph shows the success rates for the different tasks averaged over all demonstrations. KT consistently maintains the highest success rate of over 90% in all three tasks. The poor GP performance in task 2 indicates that it is not suitable for tasks that require precise control of the orientation of the end effector. The stars inside the bars correspond to statistical significance compared to KT (*: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$).

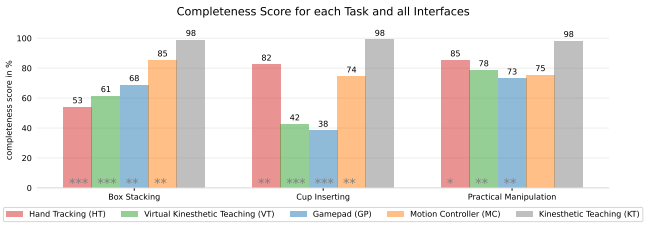


Figure 5: This graph shows the task completeness for each interface across the different tasks averaged over all tasks. KT consistently provides high completeness of 98%, showing that this interface is reliable and easy to use. While MC maintains a relatively high completeness between 74% to 85% it can not compete with KT. The stars inside the bars correspond to statistical significance compared to KT (*: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$).

and intuitively as possible. The participants executed tasks very slowly to ensure successful demonstrations. This behavior had two negative side effects. First, the efficiency measure became diluted since all interfaces were efficient if the tasks were performed slow enough. Second, due to the length of the study cognitive fatigue and decreased engagement resulted in a strong bias against whatever interface appeared later in the study. To remove these biases, a time limit per task and a random task assignment were introduced. Each participant of the proper study will attempt to solve only one randomly selected task with all five interfaces in a given time. Through these measures, the overall study time per participant was reduced to one hour, while introducing a gamification effect that reduced cognitive fatigue and increased engagement across interfaces.

4.2.2 User Study. The user study started with participants filling out the background questionnaire. Afterwards, each participant was randomly assigned one task and provided with a corresponding video explaining the task objectives. A randomized order of the five interfaces prevented potential biases. Before each interface usage, the participants had one minute to get familiar with the corresponding interface. Subsequently, participants performed three demonstrations with each interface, allowing for potential improvements over time. Each demonstration either finished because the task was completed or the time limit had been reached, indicating successful and unsuccessful demonstrations respectively. After the completion of three demonstrations with one interface, participants were asked to fill out the control questionnaire, to indicate their impressions and experiences regarding the corresponding interface. The control questionnaire included free-form feedback, where participants had the possibility to write down additional thoughts about the interface.

4.3 Metrics

4.3.1 Objective Metrics. The study evaluates the interfaces along three objective metrics, *task success*, *task completeness* and *task completion time*. *Task success* indicates if the entire task was successfully finished with the given interface or not. In contrast, task completeness represents how many sub-tasks of the task were completed as a percentage of the task. Task completion time represents either the time required if the task was completed successfully or the task specific time limit if the task was not finished fast enough. Failing a sub-task did not lead to the immediate failure of the whole task. Participants were able to recover from it, leaving exceeding the time limit as the only failure condition. The maximum time value was chosen based on the pre-study results and ensures that a single participant can conclude the entire study within one hour to reduce cognitive fatigue. Without the time limit, every task can be finished with every interface, making completion time the exclusive comparable objective metric.

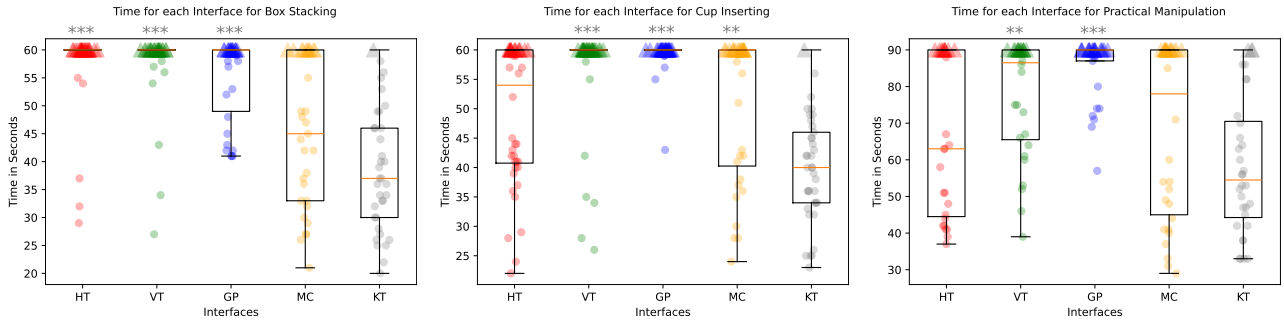
4.3.2 Subjective Metrics. The subjective metrics are based on the UEQ+ catalogue. The selected modules include attractiveness, efficiency, perspicuity, dependability, and novelty. Each participant filled out a questionnaire for each interface, providing a direct indicator of their subjective impressions and overall experiences regarding the various interfaces. In the context of subjective metrics, our analysis is grounded in the responses gathered from the interface assessment questionnaire administered during the user study. This metric serves as a direct indicator of the participant's subjective impressions and overall experience with the interfaces.

4.4 Study Tasks Design

The user study included three different tasks, box stacking, cup stacking and practical manipulation, to evaluate each interface in several dimensions: basic manipulation skills, flexibility, precision and proficiency. All three tasks were evaluated on the objective metrics, described in Section 4.3.1.

4.4.1 Box Stacking Task. This task assesses the basic pick and place capabilities of the interfaces. The participants were asked to place and stack three boxes (two cubes and one cuboid) within the target area, as shown in Figure 2(a). Each successfully stacked cube contributes 0.3 to the completeness score, while the cuboid adds 0.4 to the score. The time limitation to finish this task is 60 seconds.

4.4.2 Cup Inserting Task. This task was designed to evaluate the flexibility and precision of the interfaces with respect to dexterous motion. All cup models used in this task are sourced from the YCB [4] library. Participants were asked to insert three cups with



(a) Task 1 (b) Task 2 (c) Task 3

Figure 6: This graph shows the time spent for each demonstration per interface per task. The dots and triangles represent successful and unsuccessful demonstrations respectively. The yellow line in each box shows the mean time of every demonstration. The time of each failed demonstration is counted as the maximum time limit. The demonstrations with KT take the least time and have relatively small variance. The stars above the box plots correspond to statistical significance compared to KT (*: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$).

sizes of 55mm, 60mm and 65mm into three tilted 75 mm cups, as seen in Figure 2(b). This task assesses the ability to perform 3D manipulation using the interface. Inserting the 55mm, 60mm and 65mm cups successfully adds 0.25, 0.35, and 0.4 to the completeness score respectively. The time limitation for this task is 60 seconds.

4.4.3 Practical Manipulation Task. This task was designed to evaluate the comprehensive manipulation ability of each interface in a longer sub-task sequence. It consists of five steps, as shown in Figure 2(c), including placing a banana on a plate, placing the strawberry on the same plate, pushing the plate into a target area, flipping a mug, and placing it in a specific location on the table. Each successful step adds 0.2 to the completeness score. The time limitation for this task is 90 seconds.

4.5 Participants

The user study included 35 participants aged between 15 and 30, including 6 females and 29 males. Each participant used all 5 interfaces three times for a randomly assigned task. 42 demonstrations had to be discarded due to system failure, error records, or hardware issues. 483 valid human demonstrations were performed with the different interfaces. The statistics of the demonstrations of each task and each interface are shown in Appendix B.

5 RESULTS & ANALYSIS

5.1 Objective Metrics

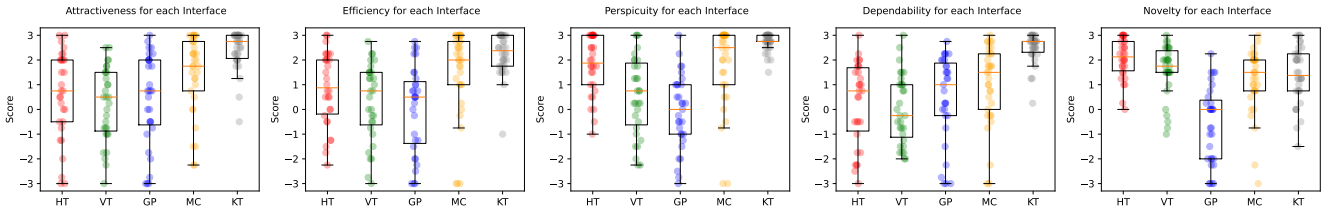
Given that all objective metrics follow a similar distribution with ties between the interfaces for each participant, the Mann-Whitney U test [29] was used to analyze significant differences between interfaces. To avoid dependencies across demonstrations of the same participant, the three demonstrations for each interface were averaged. A general comparison was conducted over the average of all tasks, where all different interfaces were compared to each other, resulting in 10 dependent statistical tests. The more fine-grained comparison, where every task was analyzed independently, was only conducted for KT and every other interface, resulting in 4 dependent statistical tests per task. The Benjamini–Hochberg procedure [1], a false discovery rate method, was applied for statistical correction in regards to the different amount of dependent statistical tests, to control the increase in type I errors. All p-values and effect sizes can be found in Appendix C and Appendix D.

Table 1: This table illustrates the p-values and effect sizes (in brackets) for all tasks averaged comparing Kinesthetic Teaching to all other interfaces (* : $p < 0.05$).

Interfaces	Completeness	Success	Time
GP	<0.001(0.85)	<0.001(0.83)	<0.001 (0.96)
HT	<0.001(0.69)	<0.001(0.67)	<0.001 (0.55)
MC	<0.001(0.55)	<0.001(0.53)	0.010*(0.39)
VT	<0.001(0.91)	<0.001(0.87)	<0.001 (0.81)

5.1.1 Success Rate. The average success rate over all three tasks shows that KT was able to significantly outperform the other four interfaces, as can be seen in Figure 3 and Table 1. All four statistical tests show a p-value of less then 0.001 and effect sizes of above 0.67, except for a value of 0.53 compared to MC. This indicates that the observed effect is not only statistically, but also practically significant. Figure 4 displays the success rate achieved by the different interfaces with respect to all three tasks separately, where KT achieves a success rate above 90% across all tasks. The statistical tests for all tasks separately also reveal that KT significantly outperforms all other interfaces, as seen in Table 2, except when compared to HT in Task 3, where the p-value is slightly above 0.05. Again the effect sizes are almost always above 0.5 which indicates a high probability to observe this effect outside the study.

5.1.2 Task Completeness. The task completeness results are shown in Figure 5. Figure 3 shows the results averaged over all tasks. The results confirm that KT outperforms the other interfaces, with a very high completeness of 98% across all tasks. The performed statistical tests for the average over all tasks reveal the same findings as for



(a) Attractiveness (b) Efficiency (c) Perspicuity (d) Dependability (e) Novelty

Figure 7: This graph shows the subjective metrics along the five questionnaire scales. KT positively stands out, boasting a significant score with relatively low variance compared to the other interfaces, except for novelty. These results indicate that most of the participants are very satisfied with this interface.

Table 2: This table illustrates the p-values and effect sizes (in brackets) for all tasks separately comparing Kinesthetic Teaching to all other interfaces ($^{*}: p < 0.05$, $^{**}: p < 0.01$).

Metric	HT	VT	GP	MC
T1 Comp.	<0.001 (0.82)	<0.001 (0.99)	0.003** (0.59)	0.008** (0.51)
T1 Succ.	<0.001 (0.82)	<0.001 (0.99)	0.003** (0.59)	0.008** (0.51)
T1 Time	<0.001 (0.88)	<0.001 (0.91)	<0.001 (0.81)	0.150 (0.35)
T2 Comp.	0.002** (0.64)	<0.001 (0.89)	<0.001 (1.00)	0.002** (0.69)
T2 Succ.	0.002** (0.64)	<0.001 (0.89)	<0.001 (0.99)	0.001** (0.69)
T2 Time	0.060 (0.44)	<0.001 (0.83)	<0.001 (1.00)	0.010** (0.65)
T3 Comp.	0.040* (0.51)	0.002** (0.84)	0.008** (0.66)	0.050 (0.43)
T3 Succ.	0.050 (0.48)	0.010* (0.71)	0.008** (0.66)	0.050* (0.43)
T3 Time	0.390 (0.27)	0.007** (0.75)	<0.001 (0.93)	0.340 (0.24)

the success rate, shown in Table 1. KT is significantly better than any other interface with p-values below 0.001. The same holds for the separately tested tasks, shown in Table 2. The only exception can be found in task 3 when compared to the MC, where the p-value is again slightly above 0.05. The calculated effect sizes indicate a medium to large (0.43 - 1.0) observable effect for experiments under non-laboratory conditions.

5.1.3 Task Completion Time. The mean completion time for the different interfaces is shown in Figure 6. The KT interface allowed for the fastest task completion times as most trials could be completed within the given time limit while other interfaces failed to do so. The statistical tests reveal again statistical significance in terms of task average, with p-values below 0.001, except for the comparison between KT and MC where the p-value is 0.01, as shown in Table 1. The separated tasks reveal non-significance between some interfaces and KT, including MC in task 1 (p = 0.15), HT in task 2 (p = 0.06), and HT (p = 0.39) and MC (p = 0.34) in task 3, as shown in Table 2. In all other interface and task comparisons KT is still significantly faster, indicating the overall dominance of KT across all tasks.

5.2 Subjective Metrics

The subjective metrics are shown in Figure 7. The study revealed that KT reaches a higher score with a relatively small variance across four out of five scales. Looking at Figure 7, KT outperforms the other interfaces on attractiveness, efficiency, perspicuity, and dependability. On the novelty scale, KT performs on par with the other interfaces, with the exception of GP which performs worse than the rest. The efficiency scale further identified MC as a very good second option after KT. The perspicuity scale revealed that MC and HT have a higher ease of understanding and clarity than VT and GP. HT was further considered the most novel interface, in stark contrast to GP. These subjective metrics offer valuable insights into the user perception and preferences associated with each interface, providing a holistic understanding of the strengths and weaknesses of each interface.

5.3 Background Analysis

The background assessment indicates that there was no significant influence of the participant's background regarding the task execution. Interestingly, prior experience with a GP positively influenced performance across all interfaces, not limited to the GP interface alone, as shown in Appendix A.2. This observation suggests that the skills and familiarity gained by using a GP, for instance by playing computer games, could be beneficial when using various interfaces. Additionally, the background assessment explored if there was a connection between the regular usage of physical robots and the success rate when using the KT interface. However, the analysis showed no significant difference, again shown in Appendix A.2. Finally, the background assessment in combination with the strong performance of the KT interface, indicates that little to no prior experience with real robots is required to efficiently collect demonstrations in a virtual setting.

6 DISCUSSION

The study presented in this paper, revealed several advantages and disadvantages of each interaction interface with respect to both, the objective and subjective metrics.

The GP interface achieved the worst average performance of all interfaces, as shown in Figure 3, resulting from its poor performance in Task 2, compared to Task 1 and 3, as shown in Figure 5. The poor Task 2 performance indicates that dexterous manipulation, e.g., the positioning of one cup inside another with a particular orientation is a significant limitation of this interface. At the same time, GP offers an easy and cheap way to control a virtual robot, without complex software, and can be used most effectively by people familiar with GP interfaces, as shown in Appendix A.2.

Participants see VT as an innovative approach to robot interaction, as shown in Figure 7(e), but it performed only slightly better than the GP interface on task average performance in Figure 3. Feedback from the participants indicated that a substantial contribution to the subpar performance of the VT interface was the lack of haptic or physical feedback. The participants were not able to feel the weight or inertia of the virtual robot, leading to a similar effect as the size-weight illusion [10]. Some participants, additionally, reported task failures due to the instability of the hand tracking and gesture recognition system, reported in Figure 7(d), resulting in an overall worse performance on all tasks using the VT interface.

HT exhibits substantial potential as an interface for robot interaction. It stands out by delivering a performance that beats the GP and VT on task average, as seen in Figure 3, and even outperforms MC in Task 2 and 3, as shown in Figure 5. HT is based on inside-out tracking and has the significant limitation that the participants have to maintain a clear view of their hand during the tasks. The questionnaires revealed that participants perceived HT as the most novel interface for controlling robots, as seen in Figure 7(e).

The MC interface was perceived to be efficient and user-friendly given the subjective metrics, as seen in Figure 7(b) and Figure 7(c). Participants particularly appreciated the simplicity of gripping objects by merely holding the trigger of the MC, which was deemed more convenient in comparison to KT. The MC also exhibited commendable completeness scores across all three tasks, with a success rate exceeding 74%. With the increasing prevalence of AR consumer products, MC deployment and integration have become more accessible. Its notable efficiency, high success rate, and ease of use make it also a great candidate for data collection in virtual settings.

The user study revealed KT to be the overall best interface for controlling virtual robots in the aspects of efficiency and intuitiveness. It outperformed other interfaces with an almost perfect succession rate of 95% across all tasks, compared to the 55% of the second highest interface, the MC, as seen in Figure 3. This high success rate makes KT a great interface to record human trajectories reliably and successfully in virtualized experiments. Furthermore, the participants only needed 64% of the maximum time to finish tasks on average, as shown in Figure 3. This high success rate combined with the little to no prior experience with the physical robot, indicates that KT provides a fast learning curve. KT is on average the most attractive and efficient interface, as seen in Figure 7(a) and Figure 7(b), suggesting that people not only like controlling with a real robot but also feel more effective doing so. Almost all participants agreed on the clear and straightforward use of KT, as shown in Figure 7(c), as well as its dependability, shown in Figure 7(d). These factors are important for non-expert users, as clear usage and dependability make it easier to actually succeed with a new interface during task execution. The only issue is the novelty aspect, shown in Figure 7(e), where HT and VT are seen as more novel, instead of using a real robot to control a virtual one.

The main disadvantage of KT is the need for the physical robot as a control interface for the virtual robot, which is expensive and not necessarily available for other researchers. The robot used in this study was a Panda by Franka Emika, which is a common robot in many research laboratories, alleviating the availability problem. If a physical robot is no option, the study identified the MC interface as a very good alternative to KT. However, it still requires the MC setup, including the lighting house system. The cheapest and most readily available solution appears to be the GP, as it performed well on stacking and manipulation tasks, as seen in Figure 5 Box Stacking and Practical Manipulation.

Some participants experienced difficulties with the end effector of the physical robot, as closing the grippers required more effort and its mobility could be affected by the joint configuration. Some feedback suggested that incorporating a physical button, such as the ones in Motion Controllers or Gamepads, would significantly enhance the user experience for closing the gripper.

A limitation of the study is the young age bracket of participants, 15-30 years. Age could influence the intuitive use of different interfaces, like GP and MC, as they are in general much more familiar with these devices than older people. Similarly, participants with frequent GP usage prior to the study, appear to be more efficient with all five interfaces in general. This observation raises the question if more practice with any of the interfaces increases the general performance. Hence, the results could be affected by an increased amount of demonstrations per participant. However, both of these limitations are only minor concerns, as the age group reflects the main AR target group and more practice will in general always increases results. Indeed, less experienced users were preferable for this study, since it investigated the intuitive use of the interfaces, where performance given less practice is more informative.

7 CONCLUSION & FUTURE WORK

This paper, presented a comprehensive study with 35 participants using different interaction interfaces for controlling a virtual robot in an AR setting. The results highlight the outstanding performance of the KT interface across almost all objective and subjective metrics. If the high hardware requirements of KT can not be fulfilled, the study identified MC as a strong alternative to control virtual robots, with lower cost and setup requirements. While the requirements for GP are even lower, GP is only recommendable in simple tasks such as pick and place, whereas MC also performs well in 3D manipulation tasks. KT can also be transferred to industry applications. Creating demonstrations using the actual robot can be used to collect data on the exact same hardware using the same objects as the production setup. Similarly, the system is also applicable in teleoperation settings to collect human demonstrations over long distances or in hazardous environments. Future work can build on the findings of this study to collect human demonstrations in a fast and efficient way virtually or in real life. A follow-up study could investigate the usage of real objects during demonstrations and transfer the presented work to control real robots.

ACKNOWLEDGMENTS

The presented work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 448648559. Xinkai Jiang and Xiaogang Jia acknowledge the support from the China Scholarship Council (CSC). The authors would like to thank Linus Witucki from Institute of Control Systems (IRS) in Karlsruhe Institute of Technology (KIT) for his support during the study.

REFERENCES

[1] Yoav Benjamini and Yosef Hochberg. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* 57, 1 (1995), 289–300. http://www.jstor.org/stable/2346101

[2] Pierre Berthet-Rayne, Konrad Leibrandt, Gauthier Gras, Philippe Fraisse, André Crosnier, and Guang-Zhong Yang. 2018. Inverse Kinematics Control Methods for Redundant Snakelike Robot Teleoperation During Minimally Invasive Surgery. *IEEE Robotics and Automation Letters* 3, 3 (Jul 2018), 2501–2508. https://doi.org/10.1109/LRA.2018.2812907

[3] A. Bushman, M. Asselmeier, J. Won, and A. LaViers. 2020. Toward Human-like Teleoperated Robot Motion: Performance and Perception of a Choreography-inspired Method in Static and Dynamic Tasks for Rapid Pose Selection of Articulated Robots. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*. 10219–10225. https://doi.org/10.1109/ICRA40945.2020.9196742

[4] Berk Calli, Arjun Singh, Aaron Walsman, Siddhartha Srinivasa, Pieter Abbeel, and Aaron M. Dollar. 2015. The YCB object and Model set: Towards common benchmarks for manipulation research. In *2015 International Conference on Advanced Robotics (ICAR)*. 510–517. https://doi.org/10.1109/ICAR.2015.7251504

[5] Yu-Wei Chao, Wei Yang, Yu Xiang, Pavlo Molchanov, Ankur Handa, Jonathan Tremblay, Yashraj S. Narang, Karl Van Wyk, Umar Iqbal, Stan Birchfield, Jan Kautz, and Dieter Fox. 2021. DexYCB: A Benchmark for Capturing Hand Grasping of Objects. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, Nashville, TN, USA, 9040–9049. https://doi.org/10.1109/CVPR46437.2021.00893

[6] Andre Cleaver and Jivko Sinapov. 2023. Demonstrating TRAinAR: An Augmented Reality Tool that Helps Humans Teach Robots. In *Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, Stockholm Sweden, 878–880. https://doi.org/10.1145/3568294.3580207

[7] Stefano Dafarra, Kourosh Darvish, Riccardo Grieco, Gianluca Milani, Ugo Pattacini, Lorenzo Rapetti, Giulio Romualdi, Mattia Salvi, Alessandro Scalzo, Ines Sorrentino, Davide Tomè, Silvio Traversaro, Enrico Valli, Paolo Maria Viceconte, Giorgio Metta, Marco Maggiali, and Daniele Pucci. 2022. iCub3 Avatar System. arXiv:2203.06972 (Mar 2022). http://arxiv.org/abs/2203.06972 arXiv:2203.06972 [cs].

[8] Neha Das, Sarah Bechtle, Todor Davchev, Dinesh Jayaraman, Akshara Rai, and Franziska Meier. 2021. Model-Based Inverse Reinforcement Learning from Visual Demonstrations. In *Proceedings of the 2020 Conference on Robot Learning*. PMLR, 1930–1942. https://proceedings.mlr.press/v155/das21a.html

[9] Joseph DelPreto, Jeffrey I. Lipton, Lindsay Sanneman, Aidan J. Fay, Christopher Fourie, Changhyun Choi, and Daniela Rus. 2020. Helping Robots Learn: A Human-Robot Master-Apprentice Model Using Demonstrations via Virtual Reality Teleoperation. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*. 10226–10233. https://doi.org/10.1109/ICRA40945.2020.9196754

[10] Murray DJ, Ellis RR, Bandomir CA, and Ross HE. 1999. Charpentier (1891) on the size-weight illusion. *Percept Psychophys* (Nov 1999). https://pubmed.ncbi.nlm.nih.gov/10598479/

[11] Pete Florence, Corey Lynch, Andy Zeng, Oscar A. Ramirez, Ayzaan Wahid, Laura Downs, Adrian Wong, Johnny Lee, Igor Mordatch, and Jonathan Tompson. 2022. Implicit Behavioral Cloning. In *Proceedings of the 5th Conference on Robot Learning*. PMLR, 158–168. https://proceedings.mlr.press/v164/florence22a.html

[12] Junling Fu, Maria Chiara Palumbo, Elisa Iovene, Qingsheng Liu, Ilaria Burzo, Alberto Redaelli, Giancarlo Ferrigno, and Elena De Momi. 2023. Augmented Reality-Assisted Robot Learning Framework for Minimally Invasive Surgery Task. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*. 11647–11653. https://doi.org/10.1109/ICRA48891.2023.10160285

[13] Kristen Grauman, Andrew Westbury, Eugene Byrne, Zachary Chavis, Antonino Furnari, Rohit Girdhar, Jackson Hamburger, Hao Jiang, Miao Liu, Xingyu Liu, Miguel Martin, Tushar Nagarajan, Ilija Radosavovic, Santhosh Kumar Ramakrishnan, Fiona Ryan, Jayant Sharma, Michael Wray, Mengmeng Xu, Eric Zhongcong Xu, Chen Zhao, Siddhant Bansal, Dhruv Batra, Vincent Cartillier, Sean Crane, Tien Do, Morrie Doulaty, Akshay Erapalli, Christoph Feichtenhofer, Adriano Fragomeni, Qichen Fu, Abrham Gebreselasie, Cristina Gonzalez, James Hillis, Xuhua Huang, Yifei Huang, Wenqi Jia, Weslie Khoo, Jachym Kolar, Satwik Kottur, Anurag Kumar, Federico Landini, Chao Li, Yanghao Li, Zhenqiang Li, Karttikeya Mangalam, Raghava Modhugu, Jonathan Munro, Tullie Murrell, Takumi Nishiyasu, Will Price, Paola Ruiz Puentes, Merey Ramazanova, Leda Sari, Kiran Somasundaram, Audrey Southerland, Yusuke Sugano, Ruijie Tao, Minh Vo, Yuchen Wang, Xindi Wu, Takuma Yagi, Ziwei Zhao, Yunyi Zhu, Pablo Arbelaez, David Crandall, Dima Damen, Giovanni Maria Farinella, Christian Fuegen, Bernard Ghanem, Vamsi Krishna Ithapu, C. V. Jawahar, Hanbyul Joo, Kris Kitani, Haizhou Li, Richard Newcombe, Aude Oliva, Hyun Soo Park, James M. Rehg, Yoichi Sato, Jianbo Shi, Mike Zheng Shou, Antonio Torralba, Lorenzo Torresani, Mingfei Yan, and Jitendra Malik. 2022. Ego4D: Around the World in 3,000 Hours of Egocentric Video. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, New Orleans, LA, USA, 18973–18990. https://doi.org/10.1109/CVPR52688.2022.01842

[14] Zhao Han, Yifei Zhu, Albert Phan, Fernando Sandoval Garza, Amia Castro, and Tom Williams. 2023. Crossing Reality: Comparing Physical and Virtual Robot Deixis. In *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, Stockholm Sweden, 152–161. https://doi.org/10.1145/3568162.3576972

[15] Ankur Handa, Karl Van Wyk, Wei Yang, Jacky Liang, Yu-Wei Chao, Qian Wan, Stan Birchfield, Nathan Ratliff, and Dieter Fox. 2020. DexPilot: Vision-Based Teleoperation of Dexterous Robotic Hand-Arm System. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*. 9164–9170. https://doi.org/10.1109/ICRA40945.2020.9197124

[16] Erin Hedlund, Michael Johnson, and Matthew Gombolay. 2021. The Effects of a Robot's Performance on Human Teachers for Learning from Demonstration Tasks. In *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, Boulder CO USA, 207–215. https://doi.org/10.1145/3434073.3444664

[17] Bryce Ikeda and Daniel Szafir. 2022. Advancing the Design of Visual Debugging Tools for Roboticists. In *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 195–204. https://doi.org/10.1109/HRI53351.2022.9889392

[18] Inmo Jang, Hanlin Niu, Emily C. Collins, Andrew Weightman, Joaquin Carrasco, and Barry Lennox. 2021. Virtual Kinesthetic Teaching for Bimanual Telemanipulation. In *2021 IEEE/SICE International Symposium on System Integration (SII)*. 120–125. https://doi.org/10.1109/IEEECONF49454.2021.9382763

[19] Steven Jens Jorgensen, Murphy Wonsick, Mark Paterson, Andrew Watson, Ian Chase, and Joshua S Mehling. [n. d.]. Cockpit Interface for Locomotion and Manipulation Control of the NASA Valkyrie Humanoid in Virtual Reality (VR). ([n. d.]).

[20] David Kent, Carl Saldanha, and Sonia Chernova. 2017. A Comparison of Remote Robot Teleoperation Interfaces for General Object Manipulation. In *2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI*. 371–379.

[21] Achyuthan Unni Krishnan, Tsung-Chi Lin, and Zhi Li. 2022. Design Interface Mapping for Efficient Free-form Tele-manipulation. In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 6221–6226. https://doi.org/10.1109/IROS47612.2022.9982149

[22] Tsung-Chi Lin, Achyuthan Unni Krishnan, and Zhi Li. 2022. Comparison of Haptic and Augmented Reality Visual Cues for Assisting Tele- manipulation. In *2022 International Conference on Robotics and Automation (ICRA)*. 9309–9316. https://doi.org/10.1109/ICRA46639.2022.9811669

[23] Tsung-Chi Lin, Achyuthan Unni Krishnan, and Zhi Li. 2023. The Impacts of Unreliable Autonomy in Human-Robot Collaboration on Shared and Supervisory Control for Remote Manipulation. *IEEE Robotics and Automation Letters* 8, 8 (Aug 2023), 4641–4648. https://doi.org/10.1109/LRA.2023.3287039

[24] Jeffrey I. Lipton, Aidan J. Fay, and Daniela Rus. 2018. Baxter's Homunculus: Virtual Reality Spaces for Teleoperation in Manufacturing. *IEEE Robotics and Automation Letters* 3, 1 (Jan 2018), 179–186. https://doi.org/10.1109/LRA.2017.2737046

[25] lolambean. 2023. HoloLens 2 hardware. https://learn.microsoft.com/en-us/hololens/hololens2-hardware

[26] Matthew B. Luebbers, Connor Brooks, Carl L. Mueller, Daniel Szafir, and Bradley Hayes. 2021. ARC-LfD: Using Augmented Reality for Interactive Long-Term Robot Skill Maintenance via Constrained Learning from Demonstration. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*. 3794–3800. https://doi.org/10.1109/ICRA48506.2021.9561844

[27] Karthik Mahadevan, Yan Chen, Maya Cakmak, Anthony Tang, and Tovi Grossman. 2022. Mimic: In-Situ Recording and Re-Use of Demonstrations to Support Robot Teleoperation. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*. ACM, Bend OR USA, 1–13. https://doi.org/10.1145/3526113.3545639

[28] Schrepp Martin and Jorg Thomaschewski. 2019. Design and validation of a framework for the creation of user experience questionnaires. *IJIMAI* 5, 7 (2019), 88–95.

[29] Patrick E McKnight and Julius Najab. 2010. Mann-Whitney U Test. *The Corsini encyclopedia of psychology* (2010), 1–1.

[30] Nina Moorman, Erin Hedlund-Botti, Mariah Schrum, Manisha Natarajan, and Matthew C. Gombolay. 2023. Impacts of Robot Learning on User Attitude and Behavior. In *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, Stockholm Sweden, 534–543. https://doi.org/10.1145/3568162.3576996

[31] James F. Mullen, Josh Mosier, Sounak Chakrabarti, Anqi Chen, Tyler White, and Dylan P. Losey. 2021. Communicating Inferred Goals With Passive Augmented Reality and Active Haptic Feedback. *IEEE Robotics and Automation Letters* 6, 4 (Oct 2021), 8522–8529. https://doi.org/10.1109/LRA.2021.3111055

[32] Abdeldjallil Naceri, Dario Mazzanti, Joao Bimbo, Domenico Prattichizzo, Darwin G. Caldwell, Leonardo S. Mattos, and Nikhil Deshpande. 2019. Towards a Virtual Reality Interface for Remote Robotic Teleoperation. In *2019 19th International Conference on Advanced Robotics (ICAR)*. 284–289. https://doi.org/10.1109/ICAR46387.2019.8981649

[33] Mandela Patrick, Yuki M. Asano, Polina Kuznetsova, Ruth Fong, Joao F. Henriques, Geoffrey Zweig, and Andrea Vedaldi. 2021. On Compositions of Transformations in Contrastive Self-Supervised Learning. In *2021 IEEE/CVF International*

Conference on Computer Vision (ICCV). IEEE, Montreal, QC, Canada, 9557–9567. https://doi.org/10.1109/ICCV48922.2021.00944

[34] Adam Pettinger, Cassidy Elliott, Pete Fan, and Mitch Pryor. 2020. Reducing the Teleoperator's Cognitive Burden for Complex Contact Tasks Using Affordance Primitives. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 11513–11518. https://doi.org/10.1109/IROS45743.2020.9341576

[35] Camilo Perez Quintero, Sarah Li, Matthew KXJ Pan, Wesley P. Chan, H.F. Machiel Van der Loos, and Elizabeth Croft. 2018. Robot Programming Through Augmented Trajectories in Augmented Reality. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 1838–1844. https://doi.org/10.1109/IROS.2018.8593700

[36] Daniel Rakita, Bilge Mutlu, and Michael Gleicher. 2017. A Motion Retargeting Method for Effective Mimicry-Based Teleoperation of Robot Arms. In *2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI*. 361–370.

[37] Daniel Rakita, Bilge Mutlu, and Michael Gleicher. 2018. An Autonomous Dynamic Camera Method for Effective Remote Teleoperation. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, Chicago IL USA, 325–333. https://doi.org/10.1145/3171221.3171279

[38] Daniel Rakita, Bilge Mutlu, Michael Gleicher, and Laura M. Hiatt. 2018. Shared Dynamic Curves: A Shared-Control Telemanipulation Method for Motor Task Training. In *2018 13th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 23–31.

[39] Harish Ravichandar, Athanasios S. Polydoros, Sonia Chernova, and Aude Billard. 2020. Recent Advances in Robot Learning from Demonstration. *Annual Review of Control, Robotics, and Autonomous Systems* 3, 1 (2020), 297–330. https://doi.org/10.1146/annurev-control-100819-063206

[40] Frank Regal, Young Soo Park, Jerry Nolan, and Mitch Pryor. 2023. Augmented Reality Remote Operation of Dual Arm Manipulators in Hot Boxes. arXiv:2303.16055 (Mar 2023). http://arxiv.org/abs/2303.16055 arXiv:2303.16055 [cs].

[41] Eric Rosen, David Whitney, Elizabeth Phillips, Gary Chien, James Tompkin, George Konidaris, and Stefanie Tellex. 2020. Communicating Robot Arm Motion Intent Through Mixed Reality Head-Mounted Displays. In *Robotics Research (Springer Proceedings in Advanced Robotics)*, Nancy M. Amato, Greg Hager, Shawna Thomas, and Miguel Torres-Torriti (Eds.). Springer International Publishing, Cham, 301–316. https://doi.org/10.1007/978-3-030-28619-4_26

[42] Inês Soares, Marcelo Petry, and António Paulo Moreira. 2021. Programming Robots by Demonstration Using Augmented Reality. *Sensors* 21, 17 (Sep 2021), 5976. https://doi.org/10.3390/s21175976

[43] Patrick Stotko, Stefan Krumpen, Max Schwarz, Christian Lenz, Sven Behnke, Reinhard Klein, and Michael Weinmann. 2019. A VR System for Immersive Teleoperation and Live Exploration with a Mobile Robot. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 3630–3637. https://doi.org/10.1109/IROS40897.2019.8968598

[44] Fouad Sukkar, Victor Hernandez Moreno, Teresa Vidal-Calleja, and Jochen Deuse. 2023. Guided Learning from Demonstration for Robust Transferability. arXiv:2302.03901 (Feb. 2023). http://arxiv.org/abs/2302.03901 arXiv:2302.03901 [cs].

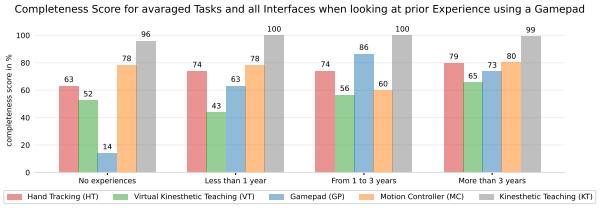
[45] Unity Technology. 2023. *Unity*. https://unity.com/

[46] Emanuel Todorov, Tom Erez, and Yuval Tassa. 2012. Mujoco: A physics engine for model-based control. In *2012 IEEE/RSJ international conference on intelligent robots and systems*. IEEE, 5026–5033.

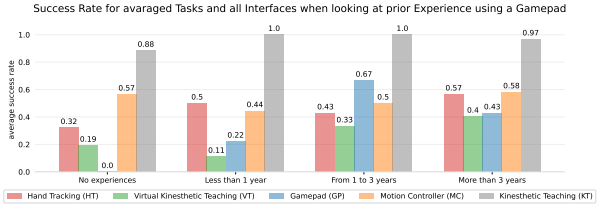
[47] Michael Walker, Hooman Hedayati, Jennifer Lee, and Daniel Szafir. 2018. Communicating Robot Motion Intent with Augmented Reality. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, Chicago IL USA, 316–324. https://doi.org/10.1145/3171221.3171253

[48] Michael E. Walker, Hooman Hedayati, and Daniel Szafir. 2019. Robot Teleoperation with Augmented Reality Virtual Surrogates. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, Daegu, Korea (South), 202–210. https://doi.org/10.1109/HRI.2019.8673306

[49] Sebastian Wrede, Christian Emmerich, Ricarda Grünberg, Arne Nordmann, Agnes Swadzba, and Jochen Steil. 2013. A User Study on Kinesthetic Teaching of Redundant Robots in Task and Configuration Space. *Journal of Human-Robot Interaction* 2, 1 (Mar 2013), 56–81. https://doi.org/10.5898/JHRI.2.1.Wrede

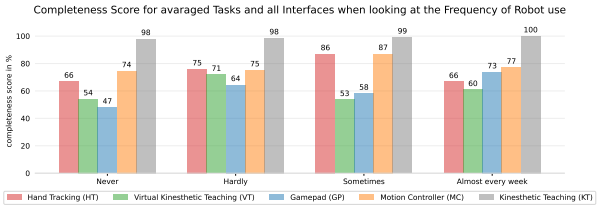


(a) Completeness Experience Gamepad Usage

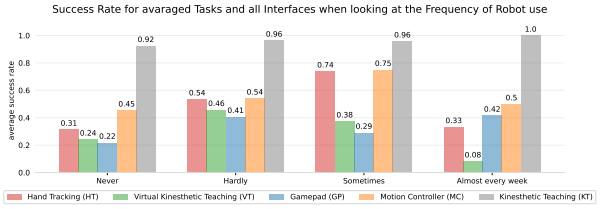


(b) Success Experience Gamepad Usage

Figure 9: The values are averaged over all tasks and only differentiate between interfaces and different answers. Participants with a higher frequent usage of Gamepads actually achieve higher results.

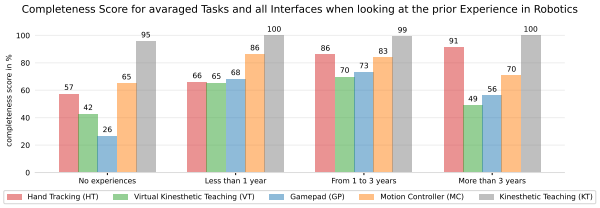


(a) Completeness Frequency Robot Usage

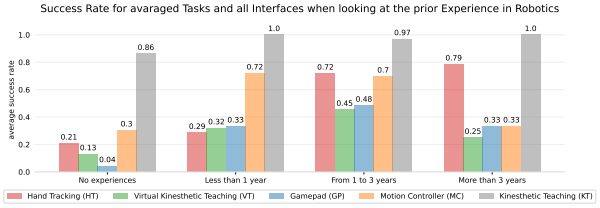


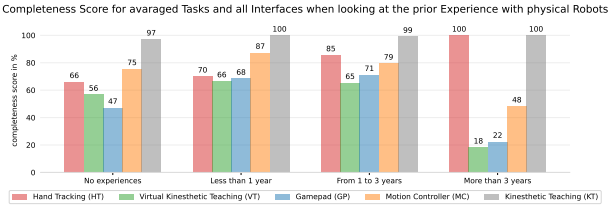
(b) Success Frequency Robot Usage

Figure 10: The values are averaged over all tasks and only differentiate between interfaces and different answers.

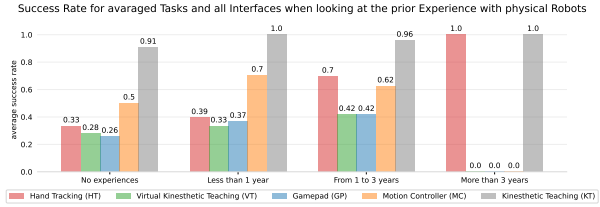


(a) Completeness Experience Robotics



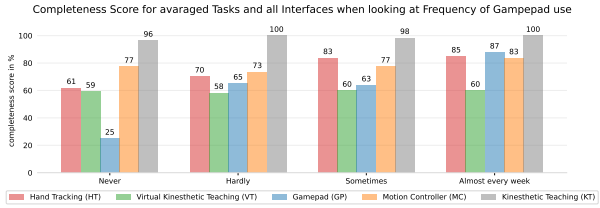


(a) Completeness Experience physical Robot Usage

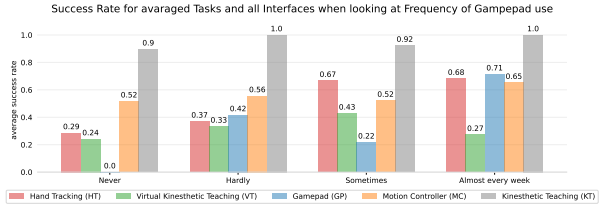


(b) Success Experience physical Robot Usage

Figure 12: The values are averaged over all tasks and only differentiate between interfaces and different answers.

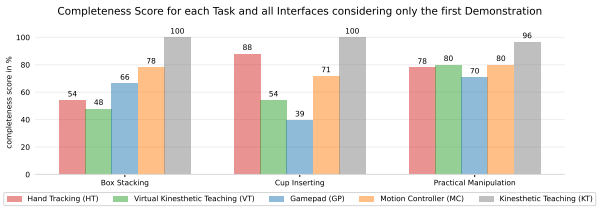


(a) Completeness Frequency Gamepad Usage

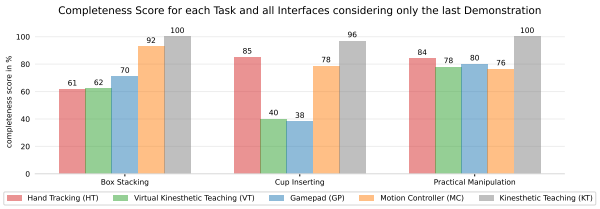


(b) Success Frequency Gamepad Usage

Figure 8: The values are averaged over all tasks and only differentiate between interfaces and different answers. Participants with a higher frequent usage of Gamepads actually achieve higher results.

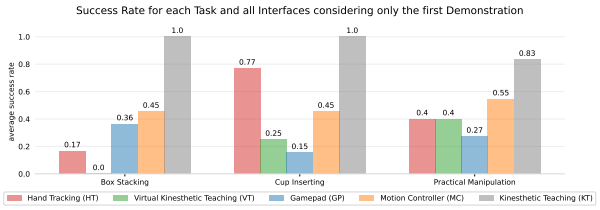


(a) Completeness First Try

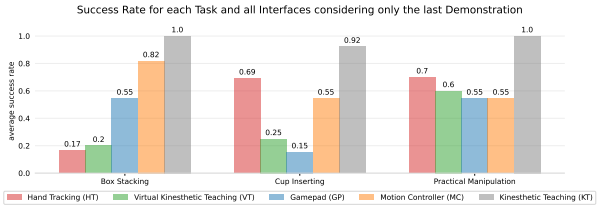


(b) Completeness Last Try

Figure 13: This graph illustrates the distribution of the time spent on the first demonstration created by each participants with each interface from Task 1 to Task 3.



(a) Success Rate First Try



(b) Success Rate Last Try

Figure 14: This graph illustrates the distribution of the time spent on the last demonstration created by each participant with each interface from Task 1 to Task 3.

Table 3: The number of demonstrations per task and interface

	Box	Cup	Man.
Gamepad	30	36	29
Hand Tracking	31	40	27
Kinesthetic Teaching	35	37	32
Motion Controller	31	32	30
Virtual Kinesthetic Teaching	30	35	28

A BACKGROUND

A.1 Questionnaire

The questionnaire included the following questions:

- Q1: *"How much experience do you have in robotics?"*
- Q2: *"How much experience do you have in physical robots?"*
- Q3: *"How often do you work with physical robots?"*
- Q4: *"How much experience do you have in AR/VR/MR devices? (e.g, Oculus Quest, HTC Vive, HoloLens, etc.)"*
- Q5: *How often do you use AR/VR/MR devices? (e.g., Oculus Quest, HTC Vive, HoloLens, etc.)*
- Q6: *How much experience do you have in using a Gamepad? (e.g., Joystick/Xbox?)*
- Q7: *How often do you use a Gamepad? (e.g., Joystick/Xbox?)*

The multiple choices for Q1, Q2, Q4, and Q6 are *"No experience"*, *"Less than 1 year"*, *"From 1 to 3 years"* and *"More than 3 years"*, for Q3, Q5, and Q7 they are *"Never"*, *"Hardly (Once or twice a year)"*, *"Sometimes (Around once a month)"* and *"Almost every week"*. All 7 questions are asked in an explicit way, to avoid misunderstandings.

A.2 Analysis

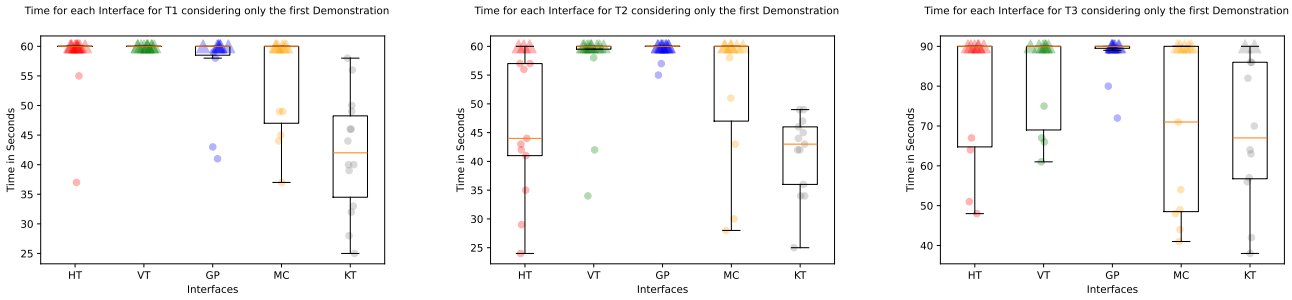
Analyzing the participants background data gives valuable insight for possible conditions that should be met, when high quality demonstrations should be collected. Therefore, two different aspects were investigated, including (prior) Gamepad experience and (prior) physical robot experience. The Gamepad results are displayed in Figure 8 and Figure 9. We can observe an upward trend in success rate and completeness for all interfaces, when looking at participants with a more frequent usage of Gamepads or more prior experience with it. The robot experience results are displayed in Figure 10, Figure 11 and Figure 12 and display no noticeable trends, as they appear more random. Rather low sample sizes for the frequent usage of physical robots additionally complicate interpretation of results.

B COMPARISON OBJECTIVE METRICS FOR FIRST AND LAST DEMONSTRATION

This section captures the results of the objective metrics in regards to the first and last demonstration performed by every participant, all displayed in Figure 13(a), Figure 14(a), and Figure 15 and Figure 13(b), Figure 14(b), and Figure 16. These values show stable performance for KT, regardless of the execution number.

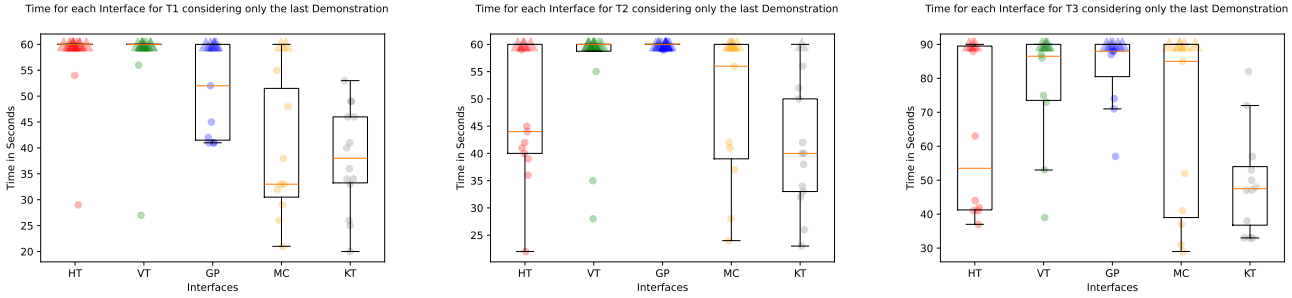
C P-VALUE RESULTS FOR ALL OBJECTIVE METRICS

All calculated p-values are displayed in this section. Table 4 provides the p-values for the completeness metric, where all interfaces are compared with each other. Therefore, the Benjamini-Hochberg procedure was used, to adjust for the increase in type I errors, as 10 different statistical tests were conducted. The same was done for success rate and completion time, both displayed in Table 5 and Table 6. Table 7 shows all p-values in regards to the separate tasks and metrics, now only comparing KT to all other interfaces, resulting in only 4 different statistical tests, still using the Benjamini-Hochberg procedure.



(a) Task 1 first (b) Task 2 first (c) Task 3 first

Figure 15: This graph illustrates the distribution of the time spent on each first demonstration from participants with each interface from Task 1 to Task 3. The dots and triangles represent successful and unsuccessful demonstrations.



(a) Task 1 last (b) Task 2 last (c) Task 3 last

Figure 16: This plot illustrates the distribution of the time spent on each last demonstration from participants with each interface from Task 1 to Task 3. The dots and triangles represent successful and unsuccessful demonstrations.

Table 4: This table illustrates the p-values of completeness in all tasks in regards to the statistical correction procedure Benjamini–Hochberg with 10 different tests (* < 0.05, ** < 0.01, * < 0.001)**

	Gamepad	Hand Tracking	Kinesthetic T.	Motion Controller	Virtual Kinesthetic T.
Gamepad	-	0.09	< 0.001***	0.04*	0.93
Hand Tracking	-	-	< 0.001***	0.55	0.09
Kinesthetic T.	-	-	-	< 0.001***	< 0.001***
Motion Controller	-	-	-	-	0.02*
Virtual Kinesthetic T.	-	-	-	-	-

Table 5: This table illustrates the p-values of success in all tasks in regards to the statistical correction procedure Benjamini–Hochberg with 10 different tests (* < 0.05, ** < 0.01, * < 0.001)**

	Gamepad	Hand Tracking	Kinesthetic T.	Motion Controller	Virtual Kinesthetic T.
Gamepad	-	0.12	< 0.001***	0.06	0.81
Hand Tracking	-	-	< 0.001***	0.59	0.16
Kinesthetic T.	-	-	-	< 0.001***	< 0.001***
Motion Controller	-	-	-	-	0.07
Virtual Kinesthetic T.	-	-	-	-	-

Table 6: This table illustrates the p-values of completion time in all tasks in regards to the statistical correction procedure Benjamini–Hochberg with 10 different test ($^{*} < 0.05$, $^{**} < 0.01$, $^{***} < 0.001$)

	Gamepad	Hand Tracking	Kinesthetic T.	Motion Controller	Virtual Kinesthetic T.
Gamepad	-	0.03^{*}	$< 0.001^{***}$	0.007^{**}	0.36
Hand Tracking	-	-	$< 0.001^{***}$	0.46	0.17
Kinesthetic T.	-	-	-	0.01^{*}	$< 0.001^{***}$
Motion Controller	-	-	-	-	0.04^{*}
Virtual Kinesthetic T.	-	-	-	-	-

Table 7: This table illustrates the p-values for all tasks separately comparing Kinesthetic Teaching to all other interfaces in regards to the statistical correction procedure Benjamini–Hochberg with 4 different test ($^{*} < 0.05$, $^{**} < 0.01$, $^{***} < 0.001$)

Metric + Task	Hand Tracking	Gamepad	Motion Controller	Virtual Kinesthetic T.
Completeness T1	$< 0.001^{***}$	0.003^{**}	0.008^{**}	$< 0.001^{***}$
Success T1	$< 0.001^{***}$	0.003^{**}	0.008^{**}	$< 0.001^{***}$
Time T1	$< 0.001^{***}$	$< 0.001^{***}$	0.15	$< 0.001^{***}$
Completeness T2	0.002^{**}	$< 0.001^{***}$	0.002^{**}	$< 0.001^{***}$
Success T2	0.002^{**}	$< 0.001^{***}$	0.001^{**}	$< 0.001^{***}$
Time T2	0.06	$< 0.001^{***}$	0.01^{**}	$< 0.001^{***}$
Completeness T3	0.04^{*}	0.008^{**}	0.05	0.002^{**}
Success T3	0.05	0.008^{**}	0.05^{*}	0.01^{*}
Time T3	0.39	$< 0.001^{***}$	0.34	0.007^{**}

D EFFECT SIZE RESULTS FOR ALL OBJECTIVE METRICS

This section reports all calculated effect sizes, which indicate if the observed statistical effect is also likely to be observed in a real world scenario. Again, the results are calculated using the Benjamini-Hochberg procedure.

Table 8: This table illustrates the effect sizes of completeness in all tasks in regards to the statistical correction procedure Benjamini–Hochberg with 10 different tests

	Gamepad	Hand Tracking	Kinesthetic T.	Motion Controller	Virtual Kinesthetic T.
Gamepad	-	0.26	0.85	0.34	0.02
Hand Tracking	-	-	0.69	0.09	0.27
Kinesthetic T.	-	-	-	0.55	0.91
Motion Controller	-	-	-	-	0.37
Virtual Kinesthetic T.	-	-	-	-	-

Table 9: This table illustrates the effect sizes of success in all tasks in regards to the statistical correction procedure Benjamini–Hochberg with 10 different tests

	Gamepad	Hand Tracking	Kinesthetic T.	Motion Controller	Virtual Kinesthetic T.
Gamepad	-	0.25	0.83	0.31	0.03
Hand Tracking	-	-	0.67	0.09	0.22
Kinesthetic T.	-	-	-	0.53	0.87
Motion Controller	-	-	-	-	0.29
Virtual Kinesthetic T.	-	-	-	-	-

Table 10: This table illustrates the effect sizes of completion time in all tasks in regards to the statistical correction procedure Benjamini–Hochberg with 10 different test

	Gamepad	Hand Tracking	Kinesthetic T.	Motion Controller	Virtual Kinesthetic T.
Gamepad	-	0.34	0.96	0.42	0.14
Hand Tracking	-	-	0.55	0.11	0.22
Kinesthetic T.	-	-	-	0.39	0.81
Motion Controller	-	-	-	-	0.32
Virtual Kinesthetic T.	-	-	-	-	-

Table 11: This table illustrates the effect sizes for all tasks separately comparing Kinesthetic Teaching to all other interfaces in regards to the statistical correction procedure Benjamini–Hochberg with 4 different test

Metric + Task	Gamepad	Hand Tracking	Motion Controller	Virtual Kinesthetic T.
Completeness T1	0.59	0.82	0.51	0.99
Success T1	0.59	0.82	0.51	0.99
Time T1	0.81	0.88	0.35	0.91
Completeness T2	1.0	0.64	0.69	0.89
Success T2	0.99	0.64	0.69	0.89
Time T2	1.0	0.44	0.65	0.83
Completeness T3	0.66	0.51	0.43	0.84
Success T3	0.66	0.48	0.43	0.71
Time T3	0.93	0.27	0.24	0.75