

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

Anonymous Author(s)

Affiliation

Address

email

1 **Abstract:** Future versatile, generalist robots need the ability to learn new tasks
2 and behaviors from demonstrations. Technologies such as Virtual and Augmented
3 Reality (VR/AR) allow for immersive, visualized environments and settings that
4 accelerate and facilitate the collection of high-quality demonstrations. However,
5 it is so far unclear which interface is the most intuitive and effective for humans
6 to create demonstrations in a virtualized environment. The intuitiveness and effi-
7 ciency of these interface becomes particularly important when working with non-
8 expert users and complex manipulation tasks. To this end, this work investigates
9 five different interfaces in a comprehensive user study across various virtualized
10 tasks. In addition, this work proposes a so far unexplored interaction interface, the
11 combination of a physical robot for kinesthetic teaching with a virtual environment
12 visualized through augmented reality. The environment, including all objects and
13 a robot manipulator, is virtualized using an AR system. The virtual robot is con-
14 trolled via various interfaces, i.e., Hand-Tracking, Virtual Kinesthetic Teaching,
15 Gamepad, Motion Controller, Physical Kinesthetic Teaching. This study reveals
16 valuable insights into the usability and effectiveness of these interaction interfaces.
17 It shows that our newly proposed intuitive interface for AR control, i.e., using a
18 physical robot as controller, significantly outperforms other interfaces in terms of
19 success-rates and task completeness. Moreover, the results show that the motion
20 controller and hand-tracking are also promising interfaces, in particular for cases
21 where a physical robot is not available.

22 **Keywords:** Robot data collection, Human robot interaction

23 1 Introduction

24 Teaching robots new skills and tasks through demonstrations is an essential goal of the fields of robot
25 learning and human robot interaction. The question how to learn tasks from demonstrations has
26 received much attention through paradigms such as Imitation learning [1], Learning from Demon-
27 strations [2] and Inverse Reinforcement Learning [3].

28 An important prerequisite for such approaches is the quality of the data and, hence, the data col-
29 lection process itself. This requirement becomes even more important given the high data demand
30 of recent learning methods. Prominent approaches focus on collecting demonstrations from vari-
31 ous sources such as online videos [4] or dedicated first person videos [5]. However, demonstrating
32 specific tasks to a particular platform harbours several challenges for real world experiments, e.g.,
33 reproducibility issues due to changing objects and object poses, cumbersome and slow resetting
34 of experimental setups and inaccurate measurements due to sensor noise. Virtualisation alleviates
35 many of these challenges as it allows for highly reproducible and controllable experiments that can
36 be quickly reset and repeated. While some approaches utilize screens to virtualize experiments [6],
37 leveraging augmented or virtual reality (AR/VR) offers significant advantages [7], by providing an
38 immersive experience that allows intuitive control over the perspective onto the virtual environment.

39 While research has highlighted the advantages of AR/VR headsets over screens for visualization
40 purposes[7], there has not yet been a study investigating the comparative performance of different
41 interaction methods within AR/VR environments for collecting task demonstrations for robots. This
42 paper closes this gap by presenting a comprehensive study on different ways to interact with virtual
43 experiments for the sake of collecting demonstrations in an AR/VR setting. This study compares five
44 different options to interact with the virtual environment, i.e., inside-out Hand Tracking, Kinesthetic
45 Teaching of a virtual robot, end-effector control via gamepad, end-effector control via VR Motion
46 Controller and Kinesthetic Teaching via a physical robot. A total of 35 participants took part in the
47 study, whereas each participant utilized every interface to collect up to 3 demonstrations across 3
48 tasks with varying difficulty levels. The performance of the interfaces were evaluated across several
49 dimensions including objective measures, such as success rate and completion time of the demon-
50 strated tasks as well as subjective measures surveyed via the well established modular extension of
51 the User Experience Questionnaire (UEQ+) [8].

52 Our study shows that combining physical Kinesthetic Teaching with AR provides a powerful and
53 intuitive system to efficiently collect demonstrations in virtual environments. The study further re-
54 veals that this system significantly outperforms any other interface with respect to both the objective
55 and subjective measures resulting in an effective and intuitive system that allows for efficient and
56 reliable data collection in virtual environments. Such an interface could also be easily implemented
57 for controlling a real physical robot via tele-operation, yet, this is part of future work.

58 In summary, the contributions of this paper are twofold. First, a comprehensive study of different
59 interaction interfaces for the purpose of collecting virtual demonstrations in an AR/VR setting. Sec-
60 ond, introducing a new interaction interfaces that utilizes a physical robot platform for controlling a
61 (virtual) robot in a virtual experimentation setting.

62 2 Background

63 **Robot Interface** Demonstrations can be broadly categorized into three categories, Kinesthetic
64 Teaching, teleoperation and pure observation. Given, the goal of efficiently collecting demonstra-
65 tions in a virtual environment, we identified five promising interaction interface in the literature.
66 **Hand Tracking** uses sensors, usually cameras, to track the hand of the user and maps the robot state
67 to the hand used in teleoperation [9, 10] and shared-control telemanipulation [11, 12] scenarios.
68 **Virtual Kinesthetic Teaching** interfaces allow the manipulation of a virtual robot directly using
69 the participants hands, for example to teleoperate physical robots in bi-manual [13] and digital twin
70 [14] settings. **Gamepads** have been widely used to control physical robots while adding very little
71 system complexity, specifically in the area of teleoperation [15, 16]. Recently **Motion Controllers**
72 have become increasingly popular in both robot learning and teleoperation[17], especially because
73 of haptic and AR visual cues [18]. **Kinesthetic Teaching** commonly refers to the manipulation of
74 a physical robot for the purpose of collecting demonstrations directly on that platform[19]. How-
75 ever, it also provides a very intuitive and straight forward demonstration interface for teleoperation
76 systems[2]. While there has been some preliminary work investigating physical Kinesthetic Teach-
77 ing controlling a virtual twin [20], it does not leverage the advantages of combining this interface
78 with a AR/VR system.

79 **AR & VR for Robotics** The ability to directly render and immerse users in a 3D virtual environ-
80 nment makes AR/VR technologies a promising tool for collecting task demonstrations. In combina-
81 tion with haptic feedback, these technologies have been shown to provide higher teaching efficiency
82 than common GUIs [7]. It can reduce workload [21], enhance the accuracy and speed of collaborat-
83 ing users [22] and improve overall task performance, compared to non-AR baselines [23].

84 Other approaches use virtual reality to realize teleoperation. including robot arm [24, 25], mobile
85 robot [25], bimanual robot arm [26, 27, 14], humanoid robots [28, 29], and surgery robots [30].
86 Apart from VR-based method,

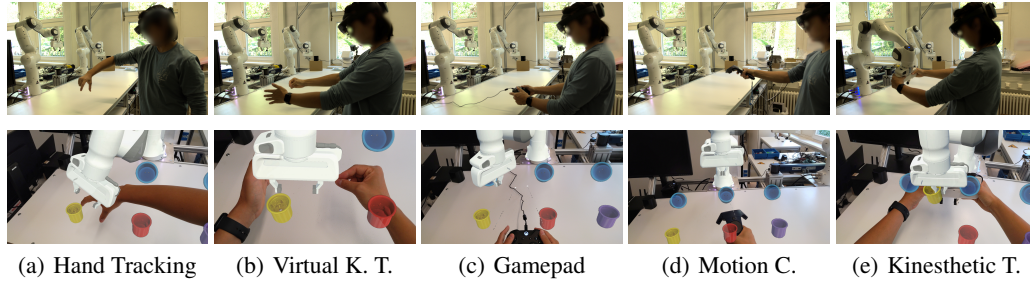


Figure 1: The top row shows a participant collecting demonstrations using the different interfaces from an outside viewpoint. The bottom row shows the virtualized environment as it is presented to the participant via the HoloLens 2.

87 3 Technical Details

88 This study explores diverse interaction interfaces in the context of their effectiveness and intuitive-
 89 ness for gathering demonstrations within AR environments. To achieve this, a novel framework has
 90 been developed, that seamlessly integrates a game engine responsible for rendering the virtual envi-
 91 ronment on AR headsets with a physics simulator, enabling the simulation of both the virtual robot
 92 and manipulated objects.

93 **Physics Simulator** The framework deploys the widely applied and utilized MuJoCo physics
 94 simulator[31]. The virtual environments, including various manipulatable objects and a Franka
 95 Emika Panda Robot were implemented within the simulator and combined into several meticulously
 96 designed scenarios. Additional data loggers were implemented that record state information of the
 97 virtual robot as well as the virtual objects, including position, velocity, acceleration and orientation.

98 **Augmented Reality Platform** The virtual scene is presented to the user via the Microsoft
 99 HoloLens 2 [32], notable for its diagonal field of view (FOV) measuring 96.1 degrees. Leverag-
 100 ing the Unity Engine, a custom AR application specifically tailored for use with the HoloLens 2
 101 was developed. The HoloLens 2 renders all virtual elements, including the robot and the objects, in
 102 real-time, providing users with an immersive and interactive experience.

103 **Interaction Interfaces** This study investigates five different interaction interfaces that have in-
 104 creasing hardware demands beyond an AR/VR Headset. For every interface but Kinesthetic Teach-
 105 ing a IK solver was used to determine the robot configuration, given the end effector pose. The
 106 Inside-out **Hand Tracking** (shown in Figure 1(a)) interface uses two scene cameras of the AR
 107 Headset to track the hand and recognize different gestures. To increase the level of intuitiveness
 108 and immersion, the gripper of the virtual robot is aligned with the index finger and thumb of the
 109 tracked hand. The participants can intuitively control the robot’s movements by moving their hand
 110 around while closing and opening the gripper by executing a pinch or release motion with their
 111 index finger and thumb. Similar to the Hand Tracking interface the **Virtual Kinesthetic Teaching**
 112 (shown in Figure 1(b)) also allows for the direct control of the virtual robot without additional hard-
 113 ware. The participants can move the robot by grabbing the virtual end effector. Releasing the end
 114 effector stops the tracking. Stretching and squeezing gestures trigger the virtual robot to close and
 115 open its gripper respectively. The **Gamepad** interface (shown in Figure 1(c)) uses a Microsoft Xbox
 116 controller to manipulate the virtual robot. The **Motion Controller** interface (shown in Figure 1(d))
 117 uses a Vive Pro Motion Controller 2.0, which precisely measure the Motion Controller’s position
 118 and orientation in real-time. The end effector pose of the virtual robot is mapped to the pose of the
 119 Motion Controller and the virtual gripper is opened and closed by holding and releasing the triggers
 120 of the controller. The **Kinesthetic Teaching** interface (shown in Figure 1(e)) allows participants to
 121 directly control a physical version of the virtual robot. The physical robot transmits joint positions
 122 and velocities to the virtualization framework in real-time, mapping the configuration of the virtual

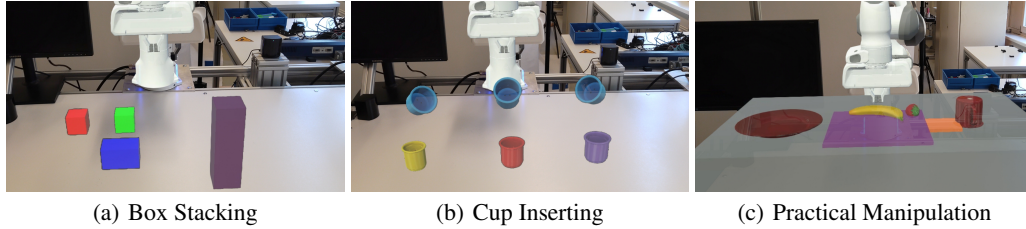


Figure 2: Tasks used in the user study. (a) Box Stacking task: Stack three boxes within the purple area. (b) Cup Inserting task: Insert three cups with varying diameters into larger cups. (c) Practical Manipulation task: Move fruits onto the plate, move the plate into the purple area and flip the mug and position it on the orange area.

123 robot to the physical one. While this interface provides the most detailed control of the virtual robot
 124 it also has the limitation that access to the actual physical robot is required.

125 4 User Study

126 In order to assess the efficiency and intuitiveness of various interfaces for collecting demonstrations,
 127 we designed a comprehensive user study. This study aims to evaluate each interface thoroughly and
 128 establish meaningful comparisons among them.

129 **Questionnaire** The user study consisted of two questionnaires (background and interface assess-
 130 ment) for each participant. The first questionnaire (background) includes seven questions and aims
 131 at the theoretical and practical knowledge of the participant in terms of using physical robots,
 132 AR/VR/MR devices, and the Gamepad. The answers were given as multiple choice in an explicit
 133 way, to avoid subjective scale measuring [33] and misunderstandings.

134 The second questionnaire (interface assessment) measures the subjective assessment from partici-
 135 pants with regard to the usage of the five different interfaces. The questionnaire itself consists of
 136 five scales taken from UEQ+ [8], including **attractiveness**, **efficiency**, **perspicuity**, **dependability**
 137 and **novelty**. Each scale presents four pairs of contrastive adjectives along with a scale ranging from
 138 one to seven, where four is neutral.

139 **Study Procedure** The user study starts with participants filling out the background questionnaire.
 140 Afterwards, each participant was randomly assigned to one task and provided with a corresponding
 141 video to understand the task objectives. The order in which participants use the five interfaces was
 142 randomized, to prevent potential biases. Before each interface usage, the participants had one minute
 143 to get familiar with the corresponding interface. Subsequently, participants used each interface three
 144 times, resulting in a potential learning curve over the three demonstrations. Each demonstration must
 145 be completed within a specific time frame, otherwise the current demonstration was stopped and the
 146 next one started. After the completion of three demonstrations with one interface, participants were
 147 asked to fill out the interface assessment questionnaire, to indicate their impressions and experiences
 148 in regards to the corresponding interface.

149 **Metrics** The study contains objective and subjective metrics. For objective metrics, we look at
 150 task success (did the participant finish the whole task?), task completeness (how many sub-goals did
 151 the participant fulfill?), and required task completion time (how much time did the participant need
 152 to finish the task?). The subjective metrics are based on the interface assessment questionnaire.

153 **Study Tasks Design** The user study included three different tasks, box stacking, cup stacking and
 154 practical manipulation. The **box stacking task** serves to assess the basic pick and place skills of the
 155 interface. Participants were asked to place and stack three boxes in 60 seconds into a target area,
 156 as you can see in Figure 2(a). For each box stacked successfully, this demonstration will give 0.3

Table 1: The number of demonstrations per task and interface

	Task 1	Task 2	Task 3
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

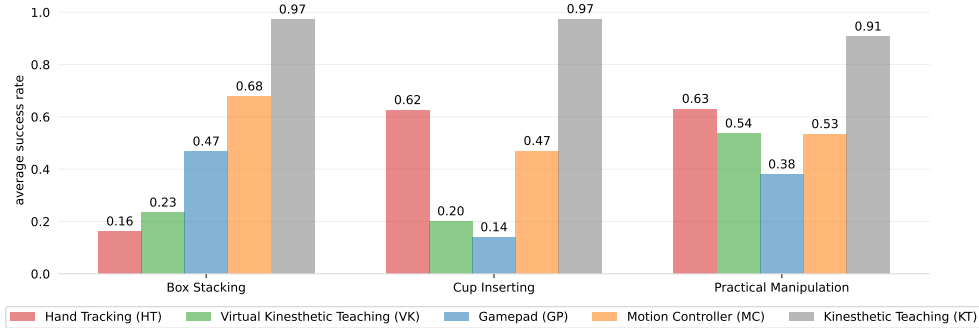


Figure 3: This graph shows the success rates of each interface across the different tasks averaged over all task executions. A task execution is considered successful if all sub-tasks were fulfilled. **Kinesthetic Teaching** consistently maintains the highest success rate of over 0.90 in all three tasks. The poor performance of the **Gamepad** interface in task 2 indicates that this interface is a poor fit for tasks that require precise control of the orientation of the end-effector. Hand tracking reaches the second-highest success rate in tasks 2 and 3, but it exhibits the lowest success rate in task 1. The Motion Controller has a success rate ranging from 0.47 to 0.68 across all three tasks. Virtual Kinesthetic Teaching performs a success rate of approximately 0.2 in tasks 1 and 2, while it achieves a higher success rate of over 0.5 in task 3. In tasks 1 and 3, the Gamepad interface yields success rates of 0.47 and 0.38, respectively. However, it only attains a success rate of 0.14 in task 2.

157 completeness score and a final 0.4 for the last box. The **cup stacking task** is designed to evaluate the
 158 flexibility and precision of the interfaces to do dexterous motion. Participants were asked to insert
 159 three different sized cups in 60 seconds into three corresponding tilted cups, as seen in Figure 2(b).
 160 The completeness gain per cup is 0.25, 0.35, and 0.4. The **practical manipulation task** is designed
 161 to evaluate the comprehensive manipulation ability of each interface in a longer sub-task sequence
 162 and is limited to 90 seconds. It involves five steps (shown in Figure 2(c)), including placing a banana
 163 on a plate, placing a strawberry next to the banana, pushing the plate into a target area, flipping a
 164 mug, and placing it on a specific location. Each successful step will gain a completeness of 0.2.

165 **Participants** The user study included 35 participants from the local and other universities, con-
 166 taining six females and 29 males, aged between 15 and 30. Each participant used all five interfaces
 167 three times for the random assigned task. Some demonstrations had to be discarded due to system
 168 instability, failed records or hardware issues. Finally, we collected 483 valid human demonstrations
 169 which were executed on the different interfaces.

170 5 Result

171 **Success Rate** As depicted in Figure 3, Kinesthetic Teaching has the highest success rate with a
 172 large margin over the other methods, with a success rate of above 90% across all tasks. Given that
 173 the success rate follows a binomial distribution with ties between the interfaces for each participant,
 174 we use a **Mann-Whitney U test** [34] to analyze the significance of the differences in the success
 175 rates. To avoid dependencies across demonstrations of the same participant, we only used the last
 176 of the three demonstrations created by each participant. All following Mann-Whitney U tests are

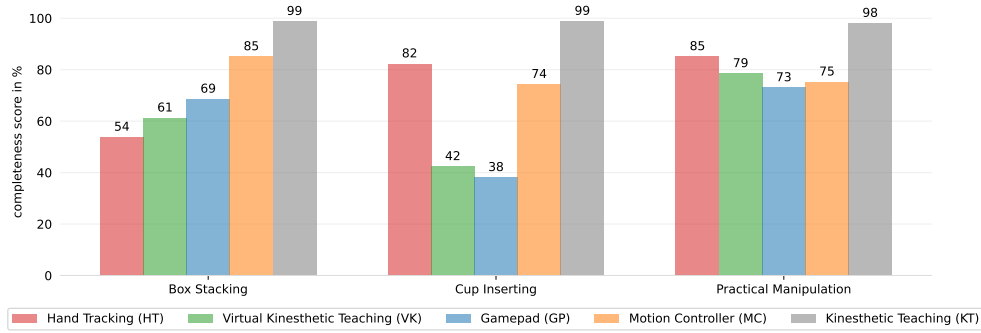


Figure 4: This graph shows the task completeness for each interface across the different tasks averaged over all task executions. The completeness of an execution is based on how many sub-tasks were achieved.

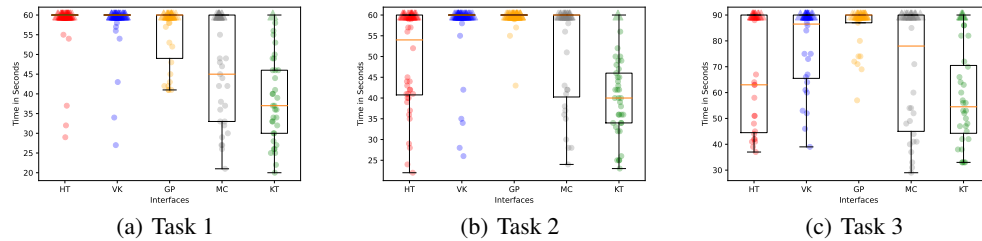


Figure 5: This graph shows the time spent on each demonstration by participants with each interface in each task. The dots and triangles represent successful and unsuccessful demonstrations. The yellow line in each box shows the mean time of every demonstration, and the time of each failed demonstration is counted as the maximum time limitation.

177 computed analogously. In all 3 tasks, the Mann-Whitney U test revealed that **Kinesthetic Teaching**
 178 significantly outperformed the other four interfaces (excluding Motion Controller from task 1 and
 179 Hand Tracking from task 2). Regarding the other interfaces, one key finding is that the **Motion**
 180 **Controller** interface performed reasonable on all tasks, outperforming Virtual Kinesthetic Teaching
 181 and Gamepad on most tasks. **Hand Tracking** performed well on task 2 and 3, even with a margin
 182 over the Motion Controller, but under performed on the first task that required picking the blue box
 183 and reorienting it, which caused problems with the internal Hand Tracking of the Holo-lens 2. The
 184 **Gamepad** and **Virtual Kinesthetic Teaching** showed the worst performance, in particular for tasks
 185 that require a precise control of the orientation of the end-effector (cup inserting).

186 **Task Completeness** For task completeness, the results are shown in Figure 4. They again confirm
 187 our finding that **Kinesthetic Teaching** outperforms all other interfaces. The Kinesthetic Teaching
 188 consistently provides a very high completeness of 98%. A Mann-Whitney U test shows the same
 189 significant difference, as in the success metric, to all other interfaces (again excluding the Motion
 190 Controller from task 1 and the Hand Tracking from task 2).

191 **Task Completion Time** The mean completion time for the different interfaces is shown in Fig-
 192 ure 5. The **Kinesthetic Teaching** interface allowed for the fastest task completion times as most
 193 trials could be completed with this interface while other interfaces had a worse success rate. We
 194 again performed a Mann-Whitney U test to confirm the significance of the results. The test con-
 195 firmed that the **Kinesthetic Teaching** allows for faster task execution than the Gamepad ($p < 0.01$)
 196 and Hand Tracking ($p < 0.001$) in task 1 and task 2, Virtual Kinesthetic Teaching ($p < 0.005$) in all
 197 3 tasks. Moreover, the **Motion Controller** has a significant difference with Gamepad (0.05) in task
 198 1 and task 2, **Hand Tracking** ($p < 0.001$) in task 1.

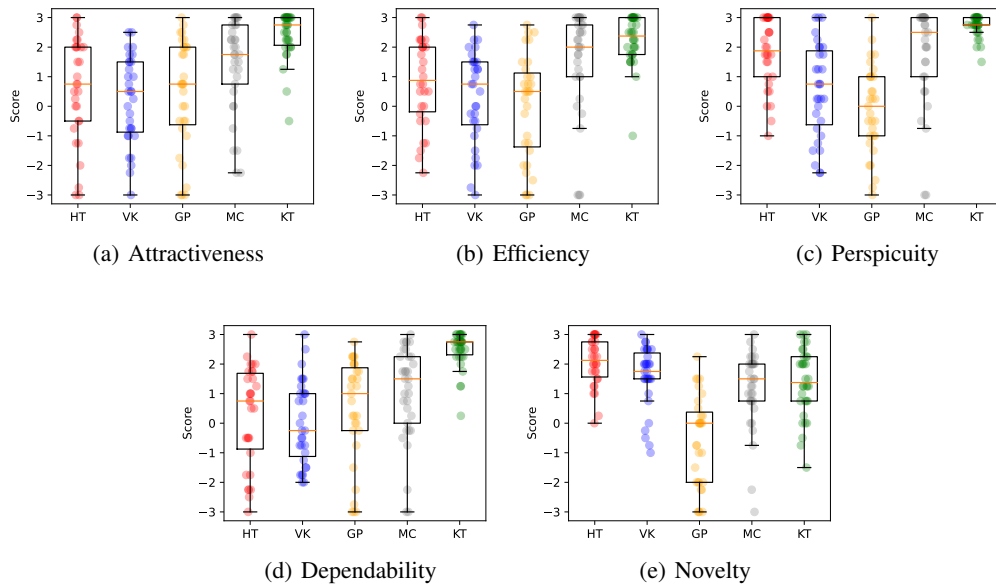


Figure 6: This graph shows the subjective metrics taken from the assessment questionnaire for every interface. including attractiveness, efficiency, perspicuity, dependability, and novelty. **Kinesthetic Teaching** emerged as a standout performer, boasting a significant score with a relatively lower variance over the other four interfaces, except for Novelty. It means that most of the participants are really satisfied with this interface.

199 **Subjective Metrics** The subjective metrics are shown in Figure 6. Our study revealed the follow-
 200 ing insights: (i) **Kinesthetic Teaching** emerged as a standout performer, with a higher score with
 201 relatively small variance over the other four interfaces, except for Novelty, where it was on par with
 202 the other methods. This pattern was consistent in the categories attractiveness and dependability,
 203 with Kinesthetic Teaching exhibiting a noteworthy difference when compared to any of the other
 204 four interfaces, while no significant differences were observed among the latter. The efficiency
 205 scale unveiled a similar trend, however, here also the **Motion Controller** interface showed a sig-
 206 nificant performance difference from the Gamepad, indicating its superior efficiency in comparison
 207 to this particular interface. In terms of Perspicuity, we identified significant differences between
 208 **Motion Controller** and **Hand Tracking** when contrasted with **Virtual Kinesthetic Teaching** and
 209 **Gamepad**, highlighting variations in their ease of understanding and clarity. Lastly, for Novelty,
 210 **Hand Tracking** stood out as the most appealing interface, underlined by its significant difference
 211 from Motion Controller and Gamepad. Conversely, **Gamepad**, emerged as the least novel interface,
 212 significantly differing from all other interfaces in this regard. These subjective metrics offer valuable
 213 insights into the user perception and preferences associated with each interface, providing a holistic
 214 understanding of their strengths and weaknesses.

215 **Background Analysis** We further analyzed the impact of user experience with the Gamepad, and
 216 its high-frequency use, on the performance of the participants. Interestingly, experience with the
 217 Gamepad positively influenced performance across all interfaces, not limited to the Gamepad in-
 218 terface alone. This suggests that the skills and familiarity gained from using a Gamepad (typically
 219 connected with playing computer games) can be transferable and advantageous when navigating
 220 various interfaces. Furthermore, we explored whether there were differences in the success rate
 221 collected through Kinesthetic Teaching based on participants regular physical activity. Here, our
 222 analysis showed no significant difference in the success rates. This result implies that the effective-
 223 ness of Kinesthetic Teaching remains consistent regardless of participants physical activity levels,
 224 highlighting the accessibility and inclusivity of this teaching approach.

225 **Further Discussion Hand Tracking** exhibits substantial potential as an interface for robot inter-
226 action. It stands out by delivering performance that competes favorably with the Gamepad interface.
227 Moreover, it offers the promise of generating high-quality demonstrations and presents a relatively
228 gentle learning curve for users. In our user study, we observed a notable phenomenon: users were
229 more likely to encounter robot singularity when employing Hand Tracking. This observation might
230 be attributed to differences in the robot’s configuration and arm design compared to other interfaces.
231 Further exploration of these factors is warranted to better understand the specific challenges and
232 advantages associated with Hand Tracking in robot interaction contexts.

233 **Virtual Kinesthetic Teaching** represents an innovative approach to robot interaction; however, its
234 performance lags behind that of the Gamepad interface. The participants highlighted a significant
235 limitation: they were unable to perceive the presence of the robot during Virtual Kinesthetic Teach-
236 ing sessions, which contributed to its underwhelming performance. Additionally, the stability of
237 Hand Tracking and gesture recognition emerged as a critical factor influencing the overall user ex-
238 perience with Virtual Kinesthetic Teaching. These technical aspects greatly impacted the interfaces
239 effectiveness and user satisfaction.

240 The **Gamepad** interface presents an economical option for generating human demonstrations. Nev-
241 ertheless, it does come with certain drawbacks, including inefficiency, a steep learning curve, and
242 extensive user experience. Dexterous manipulation, e.g., the positioning of one cup inside another
243 with a particular orientation, as well as complex 3D trajectories required of objects as well as a lim-
244 itation of this interface. Despite these challenges, the Gamepad interface does exhibit a respectable
245 success rate in both Task 1 and Task 3, exceeding 68%. This suggests its capability to handle funda-
246 mental tasks such as object manipulation—particularly picking and placing. Those with more exten-
247 sive experience are more likely to produce high-quality human demonstrations using this interface.
248 Furthermore, it is worth noting that the Gamepad interface is sensitive to users prior experience.

249 **Motion Controller** interface has proven to be both efficient and user-friendly. Participants partic-
250 ularly appreciated the simplicity of gripping objects by merely holding the trigger of the Motion
251 Controller. This feature was deemed more convenient in comparison to Kinesthetic Teaching. The
252 Motion Controller interface also exhibited commendable completeness across all three tasks, with a
253 success rate exceeding 74%. Users also noted that they could provide demonstrations in a shorter
254 amount of time using this interface.

255 **Kinesthetic Teaching** emerged as the most potent interface in our user study. It outperformed other
256 interfaces with its high-precision motion control and a consistently high success rate. However, the
257 main drawback was that users were required to have access to a physical robot to create human
258 demonstrations, which may not be feasible for all potential users. Additionally, some participants
259 encountered difficulties, noting that the end effector of the physical robot was heavy to manipulate
260 and that closing the grippers required more effort.

261 6 Conclusion

262 In this paper, we conducted a comprehensive study with 35 participants of different interaction inter-
263 faces for collecting virtual demonstrations in AR setting. From the result, the Kinesthetic Teaching
264 with AR has a highlighted performance over almost all the metrics. Due to its limitation of complex
265 hardware setting, we also conclude that Motion Controller and Hand Tracking is also a promising
266 way to create high-quality human demonstration. For virtual Kinesthetic Teaching, it was considered
267 as a novelty way but it didn’t have a good performance during this study. Gamepad, as a low-cost
268 way to create human demonstration, has worked well in simple tasks such as pick and place, but it is
269 not feasible for complex 3D manipulation. Our research will help the further studies about learning
270 from demonstration, and may provide them with a novel perspective to create and study how human
271 demonstrations influence the intelligent agent.

272 **References**

- 273 [1] P. Florence, C. Lynch, A. Zeng, O. A. Ramirez, A. Wahid, L. Downs, A. Wong, J. Lee, I. Mor-
274 datch, and J. Tompson. Implicit behavioral cloning. In *Proceedings of the 5th Conference*
275 *on Robot Learning*, page 158–168. PMLR, Jan. 2022. URL [https://proceedings.mlr.](https://proceedings.mlr.press/v164/florence22a.html)
276 [press/v164/florence22a.html](https://proceedings.mlr.press/v164/florence22a.html).
- 277 [2] H. Ravichandar, A. S. Polydoros, S. Chernova, and A. Billard. Recent advances in robot
278 learning from demonstration. *Annual Review of Control, Robotics, and Autonomous Systems*,
279 3(1):297–330, 2020. doi:10.1146/annurev-control-100819-063206.
- 280 [3] N. Das, S. Bechtle, T. Davchev, D. Jayaraman, A. Rai, and F. Meier. Model-based inverse
281 reinforcement learning from visual demonstrations. In *Proceedings of the 2020 Conference*
282 *on Robot Learning*, page 1930–1942. PMLR, Oct. 2021. URL [https://proceedings.mlr.](https://proceedings.mlr.press/v155/das21a.html)
283 [press/v155/das21a.html](https://proceedings.mlr.press/v155/das21a.html).
- 284 [4] M. Patrick, Y. M. Asano, P. Kuznetsova, R. Fong, J. F. Henriques, G. Zweig, and A. Vedaldi.
285 On compositions of transformations in contrastive self-supervised learning. In *2021 IEEE/CVF*
286 *International Conference on Computer Vision (ICCV)*, page 9557–9567, Montreal, QC,
287 Canada, Oct. 2021. IEEE. ISBN 978-1-66542-812-5. doi:10.1109/ICCV48922.2021.00944.
288 URL <https://ieeexplore.ieee.org/document/9710739/>.
- 289 [5] K. Grauman, A. Westbury, E. Byrne, Z. Chavis, A. Furnari, R. Girdhar, J. Hamburger,
290 H. Jiang, M. Liu, X. Liu, M. Martin, T. Nagarajan, I. Radosavovic, S. K. Ramakrishnan,
291 F. Ryan, J. Sharma, M. Wray, M. Xu, E. Z. Xu, C. Zhao, S. Bansal, D. Batra, V. Cartillier,
292 S. Crane, T. Do, M. Doulaty, A. Erapalli, C. Feichtenhofer, A. Fragomeni, Q. Fu, A. Gebre-
293 selasie, C. Gonzalez, J. Hillis, X. Huang, Y. Huang, W. Jia, W. Khoo, J. Kolar, S. Kottur,
294 A. Kumar, F. Landini, C. Li, Y. Li, Z. Li, K. Mangalam, R. Modhugu, J. Munro, T. Mur-
295 rell, T. Nishiyasu, W. Price, P. R. Puentes, M. Ramazanov, L. Sari, K. Somasundaram,
296 A. Southerland, Y. Sugano, R. Tao, M. Vo, Y. Wang, X. Wu, T. Yagi, Z. Zhao, Y. Zhu,
297 P. Arbelaez, D. Crandall, D. Damen, G. M. Farinella, C. Fuegen, B. Ghanem, V. K. Ithapu,
298 C. V. Jawahar, H. Joo, K. Kitani, H. Li, R. Newcombe, A. Oliva, H. S. Park, J. M. Rehg,
299 Y. Sato, J. Shi, M. Z. Shou, A. Torralba, L. Torresani, M. Yan, and J. Malik. Ego4d: Around
300 the world in 3,000 hours of egocentric video. In *2022 IEEE/CVF Conference on Com-*
301 *puter Vision and Pattern Recognition (CVPR)*, page 18973–18990, New Orleans, LA, USA,
302 June 2022. IEEE. ISBN 978-1-66546-946-3. doi:10.1109/CVPR52688.2022.01842. URL
303 <https://ieeexplore.ieee.org/document/9879279/>.
- 304 [6] T.-C. Lin, A. U. Krishnan, and Z. Li. The impacts of unreliable autonomy in human-robot
305 collaboration on shared and supervisory control for remote manipulation. *IEEE Robotics and*
306 *Automation Letters*, 8(8):4641–4648, Aug 2023. ISSN 2377-3766. doi:10.1109/LRA.2023.
307 3287039.
- 308 [7] J. F. Mullen, J. Mosier, S. Chakrabarti, A. Chen, T. White, and D. P. Losey. Communicating
309 inferred goals with passive augmented reality and active haptic feedback. *IEEE Robotics and*
310 *Automation Letters*, 6(4):8522–8529, Oct 2021. ISSN 2377-3766. doi:10.1109/LRA.2021.
311 3111055.
- 312 [8] S. Martin and J. Thomaschewski. Design and validation of a framework for the creation of
313 user experience questionnaires. *IJIMAI*, 5(7):88–95, 2019.
- 314 [9] A. Handa, K. Van Wyk, W. Yang, J. Liang, Y.-W. Chao, Q. Wan, S. Birchfield, N. Ratliff, and
315 D. Fox. Dexpivot: Vision-based teleoperation of dexterous robotic hand-arm system. In *2020*
316 *IEEE International Conference on Robotics and Automation (ICRA)*, page 9164–9170, May
317 2020. doi:10.1109/ICRA40945.2020.9197124.

- 318 [10] Y.-W. Chao, W. Yang, Y. Xiang, P. Molchanov, A. Handa, J. Tremblay, Y. S. Narang,
319 K. Van Wyk, U. Iqbal, S. Birchfield, J. Kautz, and D. Fox. Dexycb: A benchmark for capturing
320 hand grasping of objects. In *2021 IEEE/CVF Conference on Computer Vision and Pattern
321 Recognition (CVPR)*, page 9040–9049, Nashville, TN, USA, Jun 2021. IEEE. ISBN 978-
322 1-66544-509-2. doi:10.1109/CVPR46437.2021.00893. URL [https://ieeexplore.ieee.
323 org/document/9578786/](https://ieeexplore.ieee.org/document/9578786/).
- 324 [11] D. Rakita, B. Mutlu, and M. Gleicher. A motion retargeting method for effective mimicry-
325 based teleoperation of robot arms. In *2017 12th ACM/IEEE International Conference on
326 Human-Robot Interaction (HRI)*, page 361–370, Mar 2017.
- 327 [12] D. Rakita, B. Mutlu, M. Gleicher, and L. M. Hiatt. Shared dynamic curves: A shared-control
328 telemanipulation method for motor task training. In *2018 13th ACM/IEEE International Con-
329 ference on Human-Robot Interaction (HRI)*, page 23–31, Mar 2018.
- 330 [13] F. Regal, Y. S. Park, J. Nolan, and M. Pryor. Augmented reality remote operation of dual arm
331 manipulators in hot boxes. (arXiv:2303.16055), Mar 2023. URL [http://arxiv.org/abs/
332 2303.16055](http://arxiv.org/abs/2303.16055). arXiv:2303.16055 [cs].
- 333 [14] I. Jang, H. Niu, E. C. Collins, A. Weightman, J. Carrasco, and B. Lennox. Virtual kinesthetic
334 teaching for bimanual telemanipulation. In *2021 IEEE/SICE International Symposium on Sys-
335 tem Integration (SII)*, page 120–125, Jan 2021. doi:10.1109/IEEECONF49454.2021.9382763.
- 336 [15] A. Bushman, M. Asselmeier, J. Won, and A. LaViers. Toward human-like teleoperated robot
337 motion: Performance and perception of a choreography-inspired method in static and dynamic
338 tasks for rapid pose selection of articulated robots. In *2020 IEEE International Conference on
339 Robotics and Automation (ICRA)*, page 10219–10225, May 2020. doi:10.1109/ICRA40945.
340 2020.9196742.
- 341 [16] K. Mahadevan, Y. Chen, M. Cakmak, A. Tang, and T. Grossman. Mimic: In-situ recording
342 and re-use of demonstrations to support robot teleoperation. In *Proceedings of the 35th Annual
343 ACM Symposium on User Interface Software and Technology*, page 1–13, Bend OR USA,
344 Oct 2022. ACM. ISBN 978-1-4503-9320-1. doi:10.1145/3526113.3545639. URL [https:
345 //dl.acm.org/doi/10.1145/3526113.3545639](https://dl.acm.org/doi/10.1145/3526113.3545639).
- 346 [17] D. Rakita, B. Mutlu, and M. Gleicher. An autonomous dynamic camera method for effective
347 remote teleoperation. In *Proceedings of the 2018 ACM/IEEE International Conference on
348 Human-Robot Interaction*, page 325–333, Chicago IL USA, Feb 2018. ACM. ISBN 978-1-
349 4503-4953-6. doi:10.1145/3171221.3171279. URL [https://dl.acm.org/doi/10.1145/
350 3171221.3171279](https://dl.acm.org/doi/10.1145/3171221.3171279).
- 351 [18] T.-C. Lin, A. U. Krishnan, and Z. Li. Comparison of haptic and augmented reality visual cues
352 for assisting tele- manipulation. In *2022 International Conference on Robotics and Automation
353 (ICRA)*, page 9309–9316, May 2022. doi:10.1109/ICRA46639.2022.9811669.
- 354 [19] S. Wrede, C. Emmerich, R. Grünberg, A. Nordmann, A. Swadzba, and J. Steil. A user study on
355 kinesthetic teaching of redundant robots in task and configuration space. *Journal of Human-
356 Robot Interaction*, 2(1):56–81, Mar 2013. ISSN 21630364. doi:10.5898/JHRI.2.1.Wrede.
- 357 [20] F. Sukkar, V. H. Moreno, T. Vidal-Calleja, and J. Deuse. Guided learning from demonstration
358 for robust transferability. (arXiv:2302.03901), Feb. 2023. URL [http://arxiv.org/abs/
359 2302.03901](http://arxiv.org/abs/2302.03901). arXiv:2302.03901 [cs].
- 360 [21] J. Fu, M. C. Palumbo, E. Iovene, Q. Liu, I. Burzo, A. Redaelli, G. Ferrigno, and E. De Momi.
361 Augmented reality-assisted robot learning framework for minimally invasive surgery task. In
362 *2023 IEEE International Conference on Robotics and Automation (ICRA)*, page 11647–11653,
363 May 2023. doi:10.1109/ICRA48891.2023.10160285.

- 364 [22] E. Rosen, D. Whitney, E. Phillips, G. Chien, J. Tompkin, G. Konidakis, and S. Tellex. Com-
 365 municating robot arm motion intent through mixed reality head-mounted displays. In N. M.
 366 Amato, G. Hager, S. Thomas, and M. Torres-Torriti, editors, *Robotics Research*, Springer Pro-
 367 ceedings in Advanced Robotics, page 301–316, Cham, 2020. Springer International Publish-
 368 ing. ISBN 978-3-030-28619-4. doi:10.1007/978-3-030-28619-4_26.
- [23] M. Walker, H. Hedayati, J. Lee, and D. Szafrir. Communicating robot motion intent with aug-
 369 mented reality. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-
 370 Robot Interaction*, page 316–324, Chicago IL USA, Feb 2018. ACM. ISBN 978-1-4503-4953-
 371 6. doi:10.1145/3171221.3171253. URL [https://dl.acm.org/doi/10.1145/3171221.](https://dl.acm.org/doi/10.1145/3171221.3171253)
 372 [3171253](https://dl.acm.org/doi/10.1145/3171221.3171253).
- [24] A. Naceri, D. Mazzanti, J. Bimbo, D. Prattichizzo, D. G. Caldwell, L. S. Mattos, and N. Desh-
 374 pande. Towards a virtual reality interface for remote robotic teleoperation. In *2019 19th
 375 International Conference on Advanced Robotics (ICAR)*, page 284–289, Dec 2019. doi:
 376 [10.1109/ICAR46387.2019.8981649](https://doi.org/10.1109/ICAR46387.2019.8981649).
- [25] P. Stotko, S. Krumpfen, M. Schwarz, C. Lenz, S. Behnke, R. Klein, and M. Weinmann. A vr
 378 system for immersive teleoperation and live exploration with a mobile robot. In *2019 IEEE/RSJ
 379 International Conference on Intelligent Robots and Systems (IROS)*, page 3630–3637, Nov
 380 2019. doi:10.1109/IROS40897.2019.8968598.
- [26] J. I. Lipton, A. J. Fay, and D. Rus. Baxter’s homunculus: Virtual reality spaces for teleoperation
 382 in manufacturing. *IEEE Robotics and Automation Letters*, 3(1):179–186, Jan 2018. ISSN
 383 2377-3766. doi:10.1109/LRA.2017.2737046.
- [27] J. DelPreto, J. I. Lipton, L. Sanneman, A. J. Fay, C. Fourie, C. Choi, and D. Rus. Helping
 385 robots learn: A human-robot master-apprentice model using demonstrations via virtual reality
 386 teleoperation. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*,
 387 page 10226–10233, May 2020. doi:10.1109/ICRA40945.2020.9196754.
- [28] S. J. Jorgensen, M. Wonsick, M. Paterson, A. Watson, I. Chase, and J. S. Mehling. Cockpit
 389 interface for locomotion and manipulation control of the nasa valkyrie humanoid in virtual
 390 reality (vr).
 391
- [29] S. Dafarra, K. Darvish, R. Grieco, G. Milani, U. Pattacini, L. Rapetti, G. Romualdi, M. Salvi,
 392 A. Scalzo, I. Sorrentino, D. Tomè, S. Traversaro, E. Valli, P. M. Viceconte, G. Metta,
 393 M. Maggiali, and D. Pucci. icub3 avatar system. (arXiv:2203.06972), Mar 2022. URL
 394 <http://arxiv.org/abs/2203.06972>. arXiv:2203.06972 [cs].
 395
- [30] P. Berthet-Rayne, K. Leibrandt, G. Gras, P. Fraitse, A. Crosnier, and G.-Z. Yang. Inverse kine-
 396 matics control methods for redundant snakelike robot teleoperation during minimally invasive
 397 surgery. *IEEE Robotics and Automation Letters*, 3(3):2501–2508, Jul 2018. ISSN 2377-3766.
 398 doi:10.1109/LRA.2018.2812907.
 399
- [31] E. Todorov, T. Erez, and Y. Tassa. Mujoco: A physics engine for model-based control. In *2012
 400 IEEE/RSJ international conference on intelligent robots and systems*, pages 5026–5033. IEEE,
 401 2012.
 402
- [32] lolambean. Hololens 2 hardware, Mar. 2023. URL [https://learn.microsoft.com/
 403 en-us/hololens/hololens2-hardware](https://learn.microsoft.com/en-us/hololens/hololens2-hardware).
 404
- [33] D. Kent, C. Saldanha, and S. Chernova. A comparison of remote robot teleoperation interfaces
 405 for general object manipulation. In *2017 12th ACM/IEEE International Conference on Human-
 406 Robot Interaction (HRI)*, page 371–379, Mar 2017.
 407
- [34] P. E. McKnight and J. Najab. Mann-whitney u test. *The Corsini encyclopedia of psychology*,
 408 pages 1–1, 2010.
 409