2020 29th IEEE International Conference on
Robot and Human Interactive Communication (RO-MAN)
August 31 - September 4, 2020 - Virtual Conference

Augmented Reality interface to verify Robot Learning

Maximilian Diehl[1], Alexander Plopski[2], Hirokazu Kato[3], and Karinne Ramirez-Amaro[1]

***Abstract*— Teaching robots new skills is considered as an important aspect of Human-Robot Collaboration (HRC). One challenge is that robots cannot communicate feedback in the same ways as humans do. This decreases the trust towards robots since it is difficult to judge, before the actual execution, if the robot has learned the task correctly. In this paper, we introduce an Augmented Reality (AR) based visualization tool that allows humans to verify the taught behavior before its execution. Our verification interface displays a virtual simulation embedded into the real environment, timely coupled with a semantic description of the current action. We developed three designs based on different interface/visualization-technology combinations to explore the potential benefits of enhanced simulations using AR over traditional simulation environments like RViz. We conducted a user study with 18 participants to assess the effectiveness of the proposed visualization tools regarding error detection capabilities. One of the advantages of the AR interfaces is that they provide more realistic feedback than traditional simulations with a lower cost of not having to model the entire environment.**

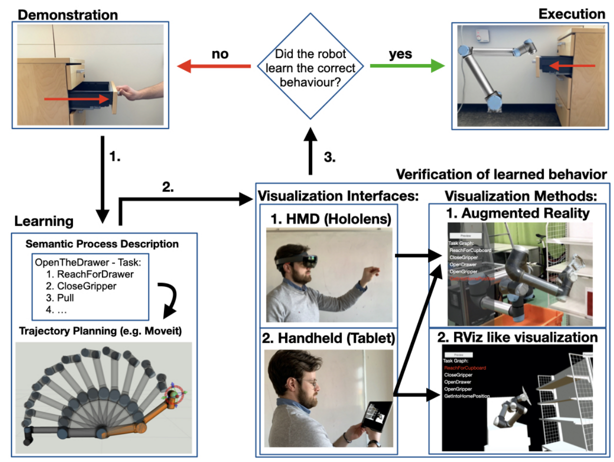

Fig. 1: Structure of our Teaching System, which combines a semantic-based learning system with a verification interface utilizing AR.

I. INTRODUCTION

The prospect of deploying robots in the service domain like hospitals, nursing homes, hotels, or even at home, has boosted research in the area of Human-Robot Collaboration (HRC) [1], [2]. In these situations, bi-directional communication is commonly regarded as the key to an efficient and trust-based interaction between humans and collaborative robots (cobots) [3], [4], [5], [6], [7]. From a human perspective, the robot's ability to communicate its intention is fundamental to fostering trust [8], acceptance [9], and motivation to engage with it. Recently, Edmonds et al. [10] have shown that the visualization method of the robot's intent has a significant impact on the humans' trust level into robots. However, lacking human features, new methods need to be developed to enable humans to understand and evaluate the robots' intentions [11].

State-of-the-art learning systems based on semantic representations [12], [13] allow us to extract a sequential description of the demonstrated process in the form of simple human-understandable labels like *ReachForDrawer*, *CloseGripper*, or *PullDrawer*. However, such learning techniques do not reach a 100% recognition accuracy. Therefore different types of errors can occur, either from the learning process or while executing the learned action. To address both issues, we present a verification tool that enables users to detect errors by coupling the semantic description with a simulation of the task, integrated into the teaching process between the learning and the actual execution on the real robot (Fig. 1).

As a visualization medium, this work explores the usage of Augmented Reality. AR embeds virtual objects into the real environment, which allows for realistic simulations without the need to model the surroundings. Especially in large and changing environments, AR could thus prove to be more practical as traditional simulation tools like RViz or Gazebo, where all objects have to be modeled by hand. In this paper, we evaluate how new technologies like AR, as well as the choice of the visualization device, can facilitate teaching robots new tasks, and thus, in a broader sense, impact the collaboration between humans and robots. Various groups of devices enable the deployment of Augmented Reality. Two of the most important groups are Optical-See-Through Head-Mounted-Displays (OST-HMD) like the Hololens[1] and handheld devices like tablets or phones. We, therefore, developed three representative verification tool designs, which are based on different device/visualization-technology combinations: 1) Hololens + AR simulation, 2) Tablet + AR simulation, and 3) Tablet + RViz like simulation. We performed a user study where the participants were instructed to teach an industrial robot the steps on how to *Open a Drawer*, to compare their effectiveness regarding error detection as well as subjective usability. Using our new verification tools, the

[1]Maximilian Diehl and Karinne Ramirez-Amaro are with the Faculty of Electrical Engineering, Chalmers University of Technology, SE-412 96 Gothenburg, Sweden diehlm@chalmers.se, karinne@chalmers.se

[2]Alexander Plopski is with the Department of Information Science, University of Otago, Dunedin 9016, New Zealand alexander.plopski@otago.ac.nz

[3]Hirokazu Kato is with the Department of Information Science, Nara Institute of Science and Technology, Nara 630-0192, Japan kato@is.naist.jp

[1]https://www.microsoft.com/en-us/hololens

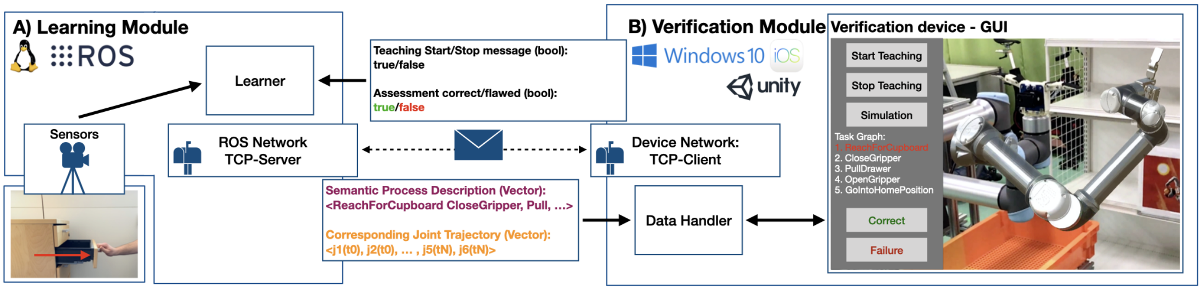

Fig. 2: Communication between the teaching and the verification system.

users then had to assess if the robot would execute the task correctly. The research questions that motivated our study can be summarized as follows:

RQ1: Can AR offer a legitimate alternative or even enhance simulation capabilities compared to RViz or Gazebo?

RQ2: Which device is the most suitable device for AR simulations to verify the results of the learning from demonstration methods?

II. Related Work

Increasing effort is placed into the development of the so-called augmented workspace [5]. To this extent, AR is utilized for a variety of applications in the robotics domain, ranging from teleoperation [14], [15], virtual path planning, to programming [16], [17], [18]. Another typical use case is information overlay, for example, to give instructions [19], [20], or indicate robot parameters and sensor information [21], [22]. The feedback that indicates the robot's movement intention is often provided in the form of arrows or balls. An example of an end-to-end interaction system is demonstrated by Puljiz and Hein [23]. It combines localization and referencing of the robot and its workspace, virtual cell set-up, e.g., through the definition of safety zones, virtual robot programming, and simulation as well as interaction and learning methods. Recently Luebbers et al. [24] started to explore feedback through AR to support humans teaching robots new activities. They suggest visualizing height as well as rotation constraints in connection with a Concept-Constrained Learning from Demonstration (CC-LfD) on a Hololens. If one of the constraints is violated at a particular part of the trajectory, it is highlighted in red color. Another approach virtually displays the coordinate frames of the robot and end-effector force through arrows [21]. Additionally, the robot's knowledge structure is visualized, and the ability to reteach particular actions like grasp or pinch by navigating a virtual robot gripper is provided.

One main difficulty that arises with LfD is that debugging the learned behavior gains complexity. Studies that show that AR can improve, for example, collision detection capabilities [25], clearly specify what the user should look out for, namely collisions. However, with LfD, the task itself could also be erroneous at a higher level when the learning failed at one of the subtasks, for example, reaching for the wrong object. In combination with new technologies like AR, this could over-strain especially inexperienced users. This gap motivated our user study approach, where the participants faced a broader range of flawed behaviors due to errors during the learning as well as execution.

III. System Overview

The two main blocks of our teaching system (Fig. 2), A) the learning module, and B) the verification module, are considered as two physically separated entities, ergo running on different devices. In this paper, we assume part A as already given, for example, in the form of a state-of-the-art semantic learning system [12], [13]. These typically generate a human-readable description of the demonstrated task (e.g., *ReachForDrawer*, *CloseGripper*, or *PullDrawer*), along with trajectories that allow the robot to imitate the behavior. Our system is modular in the sense that any learning system that provides this information could be utilized. For the information exchange between the learning and the verification modules, we, therefore, defined four different messages: The learning module dispatches the semantic process description and the corresponding joint trajectory in the form of a vector. To provide the user with the possibility to control the teaching process, we defined a start/stop teaching message and also transfer the final assessment of the verified robot behavior from the verification tool to the learning module. The communication between the verification device and the learning module is handled via a bi-directional TCP Server-Client structure. Unity[2] was chosen for the application development because it facilitates the distribution on several platforms.

We utilize the Vuforia library[3], to set the virtual replica of the robot at the correct position. A marker image is placed in the scene, and the fixed offset to the real robot base was determined manually. Note that in our setup, the robot base is static, and we visually confirmed the alignment quality. Additionally, Vuforia's extended tracking functionality tracks the environment around the marker, which keeps the positioning of the virtual robot accurate even when the marker image is not directly in the camera image of the device (e.g., when moving around the robot). Various groups of devices enable the deployment of Augmented Reality. In the growing research area of AR supported HRC [26], [25], [5], [23],

[2]https://unity.com

[3]https://developer.vuforia.com

(a) OST-HMD: Hololens (b) AR Tablet: iPad (c) RViz like Tablet: iPad

Fig. 3: View of the simulation through all three device-technology setups.

Optical-See-Through Head-Mounted-Displays (OST-HMDs) and, in particular, the Hololens, is one of the most renowned groups of devices. A more accessible and already vital factor of our daily lives is the group of handheld devices, which include the likes of smartphones and tablets. For the OST-HMD + AR scenario, the Hololens (1st gen) is used (Fig. 3a). The two tablet scenarios, AR Tablet (Fig. 3b), and RViz like Tablet (Fig. 3c), are implemented on an iPad 6th generation with a 9.7-inch retina display.

IV. USER STUDY

The user study is designed as a 3x1 within-subject experiment. It was conducted at the Nara Institute of Science and Technology (NAIST) in Japan and approved by the university's Institutional Review Board. In total, 18 participants[4] took part. All of them were students from NAIST. The 12 male and six female participants' age range is between 21 and 32, with a median of 25.5 years. Most of them have no prior knowledge about AR, Hololens, and Robots, although four participants stated 100+ hours in one or several of these fields. The median familiarity with AR is 2 hours, with the Hololens is 0 hours, and with Robots is 0.5 hours. This group of participants, therefore, fits well with our target group of inexperienced users.

Fig. 4: User Study environment from two different angles.

All participants completed 12 teaching cycles, four per device, consisting of the demonstration of *Opening a Drawer*, and the subsequent assessment if the robot learned and executed the task without errors. The teaching environment consisted of an extendable cupboard, commonly found in convenience stores, and a robot manipulator (UR5e[5]) mounted on a base (see Fig. 4). Note that in our experiment, only the virtual robot moved. Correct execution of *Opening a Drawer* would consist of 1) approaching the second highest drawer (because it is the only one that is extendable), 2) closing the gripper around the drawer (the gripper is open in the start position), 3) performing a pulling action, 4) opening the gripper and 5) going back into home position.

We pre-defined 12 unique *Opening a Drawer* executions. Six of them correct and six with different types of errors. They were distributed in a way that participants always saw 2 correct and 2 erroneous executions on each device. The six error cases are permuted, such that each error is equally often displayed on all three devices. To keep the teaching as natural as possible, we lead the participants to believe that the displayed simulation is actually based on what the robot has learned, by installing a Kinect camera, which seemingly captures their demonstrations and telling them that the system will learn from the demonstration. The six errors were designed to occur at different phases during the task execution:

1) For the *GripperDoesNotOpen* error (Fig. 5 a), the robot gripper remains closed after extending the drawer, and before going back to the home position. Also, the corresponding OpenGripper activity is missing in the process description.
2) To open the drawer correctly, a proper pull action is necessary. During the *SmallPull* error (Fig. 5 b), the pulling stops after a few centimeters and leaves into the home position early.
3) The *MissingHandle* error (Fig. 5 c), leads to the wrong goal. The gripper ends up in between two drawers, and, as a result, the pulling action has no effect.
4) During the *HittingBase* error (Fig. 5 d), the robot arm hits its own base while approaching the cupboard.
5) The *ReachWrongOrientation* error (Fig. 5 e) manifests itself in a rotated gripper position while pulling the shelf open. With such a grip, the robot was not deemed to be able to open the cupboard properly.
6) In the *PullToSide* error (Fig. 5 f), the pull action is not straight. While extending the drawer, the gripper is moving not only to the back but also to the left.

A. User Study Procedure

Each participant undergoes the same procedure. At the beginning of the study, the purpose and the study procedure are explained by the conductor. Then, the participants perform four teaching cycles per device. With each new device, they additionally complete a training teaching cycle, where the

[4]As participation reimbursement, each participant received 2000 JPY (around 15 USD).
[5]https://www.universal-robots.com/products/ur5-robot/

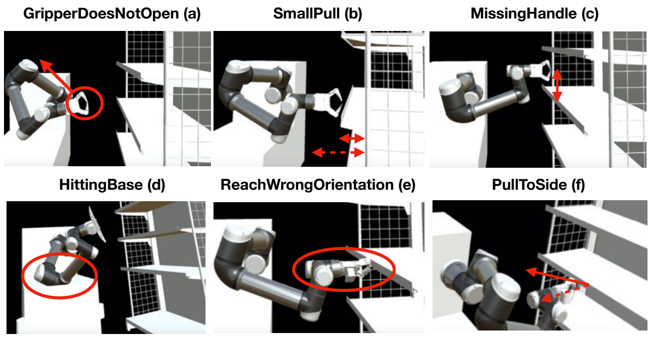

Fig. 5: Visualisation of the 6 different *erroneous* behaviors.

robot, unrelated to the task of opening the drawer, moves into a random goal pose. Each teaching cycle is instructed on the verification devices and consists of the following steps (Please refer to the following video, which displays the performance of such a cycle by one of the participants: https://www.youtube.com/watch?v=NxFbM-L2y6I): A new demonstration can be freely initiated by saying 'Start' and terminated by saying 'Stop'. After the robot execution is seemingly learned from the demonstration, the user watches the simulation through the devices, as often as needed, to derive a decision on whether the task was learned and executed correctly or not. This decision is communicated by pushing a 'Correct' or 'Erroneous' button on the devices, accompanied by a slider that asks for the participants' confidence in the decision (from less to high confidence on a scale from 1 - 10, respectively). Additionally, the verification time and the number of simulations are automatically logged on the device. These four measures will be called objective data in the rest of this paper. A System Usability Scale (SUS) questionnaire [27], is answered after each device to collect data on the subjective valuation. It consists of 10 statements (Table I) that have to be rated from 1-strongly disagree to 7-strongly agree (we used a 7 point Likert scale). After completing all three devices, the user is asked to rank the three methods (into best, middle, worst), and a short interview is conducted where the participants get the chance to give freeform feedback. Overall the experiment took less than one hour per participant.

Q1	I think that I would like to use this verification system frequently.
Q2	I found the verification system unnecessarily complex.
Q3	I thought the verification system was easy to use.
Q4	I think that I would need the support of a technical person to be able to use this verification system.
Q5	I found the various functions in this verification system were well integrated.
Q6	I thought there was too much inconsistency in this verification system.
Q7	I would imagine that most people would learn to use this verification system very quickly.
Q8	I found the verification system very cumbersome to use.
Q9	I felt very confident using the verification system.
Q10	I needed to learn a lot of things before I could get going with this verification system.

TABLE I: SUS questionnaire

B. Hypotheses

Our expectations about the user study results are outlined in the following hypotheses:

H1: The error detection capabilities (H1a), as well as subjective appraisal (H1b), are expected to benefit from the spatial awareness as it is provided through AR.

	assessed as correct			assessed as erroneous		
	HMD	AR	RViz like	HMD	AR	RViz like
actually correct	0.667	0.805	0.861	0.333	0.195	0.139
actually erroneous	0.528	0.583	0.639	0.472	0.417	0.361

TABLE II: Confusion Matrix for assessment success.

H2: The most realistic and natural embeddedness can be reached by using an OST-HMD like the Hololens. Being embedded into the real environment leads to reduced mental strain, especially for inexperienced users. Therefore, we believe that the Hololens would yield the best performance considering error detection capabilities (H2a), as well as subjective appraisal (H2b).

H3: The VR tablet has the advantage that the simulation viewpoint can be adapted through pan and zoom gestures instead of having to move around physically. We expect that the level of embeddedness through the AR tablet is not enough to outweigh the convenience of the VR tablet case.

C. User Study Results

The objective data is statistically evaluated based on the Friedman test [28]. We denote an assessment successful when the participant's decision corresponds to our original classification of a particular robot behavior (e.g., a correct robot trajectory was assessed as *correct* or the robot hitting its base was assessed as *erroneous*). Table II shows the success ratio for each device, marked in green the two fields where the rate should be as close to 1.0 as possible. The true positives ratio of the three designs lies between 0.667 and 0.861; however, no significant difference could be detected. The more critical measure is the true negatives as they represent the potentially harmful robot behaviors. Therefore in the following, only the data of the erroneous scenarios are considered. Fig. 6 displays the results of the objective data averaged over all participants.

None of the measures features a significant difference between the devices. A closer look at the data, however, reveals that despite generally high confidence (over 7.5) among all designs, the assessment success ratio for the flawed cases is below 50%. The SUS score is calculated based on Lewis and Sauro [29]. Two participants accidentally left out one question each (Q4 AR-Tablet, Q8 AR-Tablet), and one participant chose two answers for one question (Q6 Hololens). In these three cases, the median value of 4 was inserted [29]. For the statistical analysis we used the Friedman test and the Wilcoxon signed-rank posthoc test with Bonferroni correction [28] in case of significance ($p < 0.05$).

The score shows a significant usability difference ($p < 0.01$). The posthoc analysis yields a significant usability preference of the Rviz like tablet over the Hololens ($p = 0.04$) but no significant difference for the pairs of HMD - AR ($p = 0.09$) and AR - RViz like ($p = 1.0$). Question wise analysis of the SUS (Table I) reveals statistical significance for Q2 ($p < 0.01$), Q3 ($p < 0.01$) , Q4 ($p < 0.01$) , Q7 ($p < 0.01$) , Q8 ($p = 0.01$) and Q9 ($p = 0.01$). The posthoc analysis

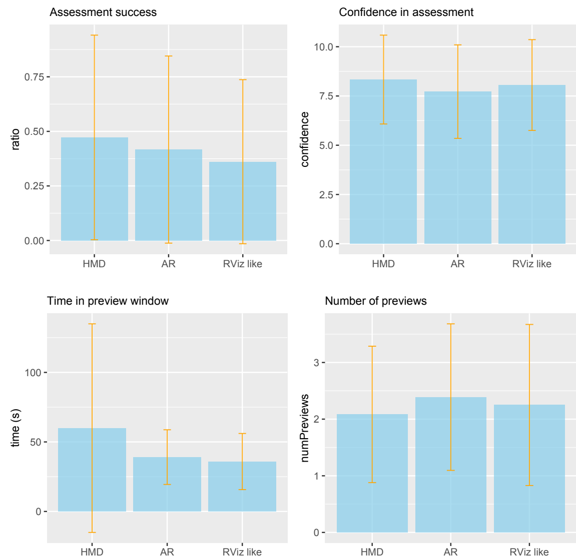

Fig. 6: Mean and standard deviation of the objective measures.

shows significant difference for the pair HMD - RViz like in Q2 ($p < 0.01$) , for HMD - RViz like in Q3 (p = 0.02) and for the pairs HMD - RViz like (p = 0.01) and HMD - AR (p = 0.02) in Q4.

For the evaluation of the ranking data, we counted how often each device was ranked better than the two others, respectively. The results based on the Bradley-Terry Model [30], show no statistically significant difference between the two tablet cases ($\Pr(> |z|) = 0.23$). However, the AR tablet was significantly better ranked than the HMD ($\Pr(> |z|) < 0.01$) and the RViz like tablet was significantly better ranked than the HMD ($\Pr(> |z|) = 0.02$).

During the interview, eight participants commented that the Hololens is uncomfortable to wear because it is too heavy and complained about the small Field of View (FoV). Additionally, five participants thought that it takes more time and training to get used to the interaction with the application on the Hololens, compared to the tablet devices. Three participants, on the other hand, stated explicitly that the Hololens is fun to use and has the potential to be the best device if it was not for its hardware shortcomings. Five participants enjoyed the realism of simulation in the real environment of both AR interfaces. Based on 6 participants, the RViz like tablet is the most convenient to use as it allows to change position through finger gestures without moving physically but misses the real-world experience.

D. Discussion

The subjective analysis shows that the Hololens was not the preferred design as originally hypothesized (H2b). However, the main reasons are hardware related issues like wear-ability and small FoV. As already the next generation of the Hololens promises to be lighter, more comfortable, and equipped with a larger FoV, we believe that this will improve the user experience in the future. Additionally, several participants commented that they preferred the AR-enabled simulation in the real environment as it is more realistic than the RViz like simulation, supporting H1b. The real-world experience, however, comes at the cost of occasional inconvenience to find the right viewing angle manifesting itself in the need to go into difficult positions to get all viewing angles, e.g., under the shelf, as several participants commented in the interview (H3). Objectively, we could not find a significant difference regarding error detection capabilities between the three interfaces (H1a, H2a). However, note that AR can achieve these results at a lower cost as the traditional RViz like simulation, since it is not required to visually model the whole environment. In conclusion, a clear preference for one device over the others could not be detected. Therefore, it is important to also consider the application circumstances for the choice of device. For example, in case the setup allows for a long training phase, users might benefit the most from the Hololens. In other instances, in which the users have to cope instantly with the situation, more intuitive devices like the tablet might be preferable.

Surprisingly, none of the three devices generated an error detection rate of higher than 50%, which is problematic considering that simulation of the robot movements is a common method to detect errors in the execution of a task. We observed two major problems:

1. Many participants put a strong focus on the area and time when the robot is gripping the shelf, disregarding other phases like the approaching or pulling activities. This problem is reinforced by the limited FoV, which makes it challenging to observe the whole robot together with the cupboard if one stands close to the objects. That also applies to the RViz like simulation because it allows the user to zoom into the scene. This problem could be tackled by pointing out possibly dangerous areas and situations through visual indicators or a set of instructions.

2. Some of the participants commented that, occasionally, they observed an unusual behavior but were not sure if it was considered as an error, for example, in the *ReachWrongOrientation* error, where the gripper is rotated 45 degrees during the pulling action. In the current designs, we do not display any information about the consequences of bad actions. Even though the simulations are blended over the real environment, the virtual robot naturally cannot interact with it. If the robot would carry away the virtual replica of the cupboard, also other questions like did the gripper grab the cupboard or the error when forgetting to re-open the gripper before going back to the home position would be easier to detect. Now, while we recognized it as a major advantage of AR, not having to model the whole environment, a compromise could be to include a visual knowledge base of daily objects and then overlay the real cupboard and interact with a stereotypical virtual one. This is considered as future work.

A limitation of the user study is that we did not include a traditional RViz simulation on the computer as a fourth scenario. To keep the structure of the applications as homogeneous as possible, we only incorporated the RViz like case running on the tablet.

V. CONCLUSION

This work investigated the utility of Augmented Reality (AR) to facilitate teaching a robot new tasks by comparing three different interface/visualization-technology combinations, OST-HMD + AR simulation, Tablet + AR simulation, and Tablet + RViz like simulation in a user study. The results suggest that AR produced more realistic visual feedback with lower costs compared to traditional approaches such as RViz. The participants of the study stated explicitly that they enjoyed the realism enabled through the superimposition of virtual simulation over the real environment. Generally, however, inexperienced users seem to be overstrained by the combination of new technology and the difficult task of recognizing previously unknown errors. Amplified by the small FoV, one of the main problems seems to be finding the right focus on the area where the error happens. To improve the error detection rate, we suggest two measures for future work: 1) visual indicators or guidelines that point out possibly dangerous areas and 2) introduction of a visual knowledge base or a similar strategy to indicate an interaction between robot gripper and a prototype of an object like the cupboard without the need to model the actual cupboard. Furthermore, we would like to extend the framework to include the user's feedback during the learning and refinement of the task's future executions.

ACKNOWLEDGMENT

We would like to thank Nicko Reginio Caluya for conducting the experiment, the NAIST robotics laboratory, and in particular Gustavo Alfonso Garcia Ricardez, for kindly providing the hardware for the user study, including the UR5e with base as well as the convenience store shelf. The research reported in this paper has been (partially) supported by Chalmers AI Research Centre (CHAIR).

REFERENCES

[1] S. Schaal, "The new robotics-towards human-centered machines," *HFSP journal*, vol. 1, pp. 115–26, 08 2007.

[2] E. Cha, Y. Kim, T. Fong, and M. Mataric, "A survey of nonverbal signaling methods for non-humanoid robots," *Foundations and Trends in Robotics*, vol. 6, pp. 211–323, 01 2018.

[3] A. Mohseni-Kabir, C. Rich, S. Chernova, C. L. Sidner, and D. Miller, "Interactive hierarchical task learning from a single demonstration," in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, 2015, pp. 205–212.

[4] K. Kawamura, P. Nilas, K. Muguruma, J. A. Adams, and C. Zhou, "An agent-based architecture for an adaptive human-robot interface," in *36th Annual Hawaii International Conference on System Sciences, 2003. Proceedings of the*. IEEE, 2003, pp. 8–pp.

[5] T. Chakraborti, S. Sreedharan, A. Kulkarni, and S. Kambhampati, "Alternative modes of interaction in proximal human-in-the-loop operation of robots," *CoRR*, vol. abs/1703.08930, 2017. [Online]. Available: http://arxiv.org/abs/1703.08930

[6] C. Kardos, Z. Kemény, A. Kovács, B. E. Pataki, and J. Váncza, "Context-dependent multimodal communication in human-robot collaboration," *Procedia CIRP*, vol. 72, pp. 15–20, 2018.

[7] K. Drnec, G. Gremillion, D. Donavanik, J. Canady, C. Atwater, E. Carter, B. Haynes, A. Marathe, and J. Metcalfe, *The Role of Psychophysiological Measures as Implicit Communication Within Mixed-Initiative Teams*, 06 2018, pp. 299–313.

[8] V. Groom and C. Nass, "Can robots be teammates?: Benchmarks in human–robot teams," *Interaction Studies*, vol. 8, no. 3, pp. 483–500, 2007.

[9] M. Coovert, T. Lee, I. Shindev, and Y. Sun, "Spatial augmented reality as a method for a mobile robot to communicate intended movement," *Computers in Human Behavior*, vol. 34, p. 241–248, 05 2014.

[10] M. Edmonds, F. Gao, H. Liu, X. Xie, S. Qi, B. Rothrock, Y. Zhu, Y. N. Wu, H. Lu, and S.-C. Zhu, "A tale of two explanations: Enhancing human trust by explaining robot behavior," *Science Robotics*, vol. 4, no. 37, 2019. [Online]. Available: https://robotics.sciencemag.org/content/4/37/eaay4663

[11] B. Hayes and B. Scassellati, "Challenges in shared-environment human-robot collaboration," 01 2013.

[12] K. Ramirez-Amaro, M. Beetz, and G. Cheng, "Transferring skills to humanoid robots by extracting semantic representations from observations of human activities," *Artif. Intell.*, vol. 247, pp. 95–118, 2017. [Online]. Available: https://doi.org/10.1016/j.artint.2015.08.009

[13] T. Bates, K. Ramírez-Amaro, T. Inamura, and G. Cheng, "On-line simultaneous learning and recognition of everyday activities from virtual reality performances," 09 2017, pp. 3510–3515.

[14] L. Peppoloni, F. Brizzi, C. A. Avizzano, and E. Ruffaldi, "Immersive ros-integrated framework for robot teleoperation," in *2015 IEEE Symposium on 3D User Interfaces (3DUI)*. IEEE, 2015, pp. 177–178.

[15] S. A. Green, J. G. Chase, X. Chen, and M. Billinghurst, "Evaluating the augmented reality human-robot collaboration system," in *2008 15th International Conference on Mechatronics and Machine Vision in Practice*, Dec 2008, pp. 521–526.

[16] C. Quintero, S. Li, M. Pan, W. Chan, H. Loos, and E. Croft, "Robot programming through augmented trajectories in augmented reality," 10 2018, pp. 1838–1844.

[17] J. Chong, S. K. Ong, A. Nee, and K. Youcef-Youmi, "Robot programming using augmented reality: An interactive method for planning collision-free paths," *Robotics and Computer-Integrated Manufacturing*, vol. 25, pp. 689–701, 06 2009.

[18] H. C. Fang, S. K. Ong, and A. Y. C. Nee, "A novel augmented reality-based interface for robot path planning," *International Journal on Interactive Design and Manufacturing (IJIDeM)*, vol. 8, no. 1, pp. 33–42, Feb 2014. [Online]. Available: https://doi.org/10.1007/s12008-013-0191-2

[19] S. Makris, G. Pintzos, L. Rentzos, and G. Chryssolouris, "Assembly support using ar technology based on automatic sequence generation," *CIRP Annals*, vol. 62, no. 1, pp. 9–12, 2013.

[20] R. K. Ganesan, Y. K. Rathore, H. M. Ross, and H. B. Amor, "Better teaming through visual cues: How projecting imagery in a workspace can improve human-robot collaboration," *IEEE Robotics & Automation Magazine*, vol. 25, pp. 59–71, 2018.

[21] H. Liu, Y. Zhang, W. Si, X. Xie, Y. Zhu, and S. Zhu, "Interactive robot knowledge patching using augmented reality," 05 2018.

[22] T. H. J. Collett and B. A. Macdonald, "An augmented reality debugging system for mobile robot software engineers," 2010.

[23] D. Puljiz and B. Hein, "Concepts for end-to-end augmented reality based human-robot interaction systems," 2019.

[24] M. B. Luebbers, C. Brooks, M. J. Kim, D. Szafir, and B. Hayes, "Augmented reality interface for constrained learning from demonstration," in *Proceedings of the 2nd International Workshop on Virtual, Augmented, and Mixed Reality for HRI (VAM-HRI)*, 2019.

[25] E. Rosen, D. Whitney, E. Phillips, G. Chien, J. Tompkin, G. Konidaris, and S. Tellex, "Communicating robot arm motion intent through mixed reality head-mounted displays," *CoRR*, vol. abs/1708.03655, 2017. [Online]. Available: http://arxiv.org/abs/1708.03655

[26] M. Walker, H. Hedayati, J. Lee, and D. Szafir, "Communicating robot motion intent with augmented reality," 02 2018, pp. 316–324.

[27] J. Brooke, *"SUS-A quick and dirty usability scale." Usability evaluation in industry*. CRC Press, June 1996, iSBN: 9780748404605. [Online]. Available: https://www.crcpress.com/product/isbn/9780748404605

[28] D. Pereira, A. Afonso, and F. Medeiros, "Overview of friedman's test and post-hoc analysis," *Communications in Statistics - Simulation and Computation*, vol. 44, pp. 2636–2653, 11 2015.

[29] J. R. Lewis and J. Sauro, "The factor structure of the system usability scale," in *Human Centered Design*, M. Kurosu, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 94–103.

[30] R. A. Bradley and M. E. Terry, "The rank analysis of incomplete block designs — I. The method of paired comparisons," *Biometrika*, vol. 39, pp. 324–345, 1952.