2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
Madrid, Spain, October 1-5, 2018

Comparison of Multimodal Heading and Pointing Gestures for Co-Located Mixed Reality Human-Robot Interaction

Dennis Krupke[1,2,*], Frank Steinicke[1], Paul Lubos[1], Yannick Jonetzko[2], Michael Görner[2] and Jianwei Zhang[2]

***Abstract*— Mixed reality (MR) opens up new vistas for human-robot interaction (HRI) scenarios in which a human operator can control and collaborate with co-located robots. For instance, when using a see-through head-mounted-display (HMD) such as the Microsoft HoloLens, the operator can see the real robots and additional virtual information can be superimposed over the real-world view to improve security, acceptability and predictability in HRI situations. In particular, previewing potential robot actions in-situ before they are executed has enormous potential to reduce the risks of damaging the system or injuring the human operator.**

In this paper, we introduce the concept and implementation of such an MR human-robot collaboration system in which a human can intuitively and naturally control a co-located industrial robot arm for pick-and-place tasks. In addition, we compared two different, multimodal HRI techniques to select the pick location on a target object using (i) head orientation (aka heading) or (ii) pointing, both in combination with speech. The results show that heading-based interaction techniques are more precise, require less time and are perceived as less physically, temporally and mentally demanding for MR-based pick-and-place scenarios. We confirmed these results in an additional usability study in a delivery-service task with a multi-robot system. The developed MR interface shows a preview of the current robot programming to the operator, e. g., pick selection or trajectory. The findings provide important implications for the design of future MR setups.

I. Introduction

Current technologies have paved the way to an era in which robots can co-exist and collaborate with humans in private and public places [1], [2]. For example, autonomous cars are being widely tested and robots can be found in factories, warehouses and in homes nowadays. However, in many occasions the spaces in which humans work are deliberately kept separate from the spaces in which robots operate, since many people do not feel safe or comfortable in close proximity to robots. One essential reason for those concerns is the difficulty of humans to interpret a robot's intent [3]. Humans rely on a rich set of non-verbal cues to interpret the intent and emotions of others human being, whereas most of such cues are not provided by robots. To address these limitations, some researchers have attempted to make robots more expressive [4] or to visualize the robot's planned movements [5]. However, all these solutions impose a number of limitations on the design of the robot, and do not work with traditionally designed robots.

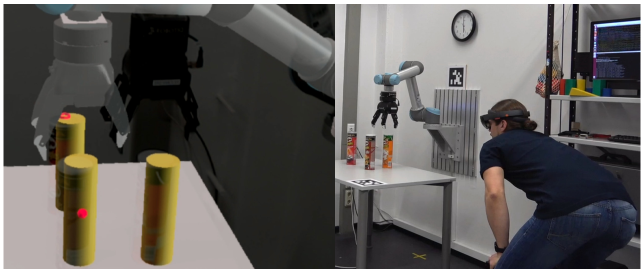

Fig. 1. Illustration of the MR human-robot interaction: A human operator wears an optical-see through HMD, which allows to view the co-located robot and superimposed virtual information such as pick locations or pre-visualization of potential actions of the industrial robot arm.

Mixed reality (MR) [6], [7], [8] opens up new vistas for such human-robot interaction (HRI) scenarios. The visual combination of virtual content with real working spaces creates a MR environment that has enormous potential to enhance co-located HRI [9]. Current technologies such as optical see-through head-mounted-displays (HMDs) such as the Microsoft HoloLens are typical examples of such MR displays.

Using an MR display, a human operator can see the co-located robot in its real physical surroundings, while visual cues can be displayed in the human operator's view and augment the real-world by showing information, which is important for the human-robot collaboration process. Examples include, but are not limited to system states of the robot, potential pick locations in a pick-and-place scenario, previews of potential robot interactions, or proxemic zones for human-robot collaboration [10].

In this paper, we introduce the concept and implementation of an MR human-robot collaboration system allowing users to intuitively and naturally control a co-located industrial robot arm for pick-and-place tasks. First, we present the general concept and then a specific implementation of such an MR human-robot collaboration system. The goal of the system is to provide experts as well as non-experts the capabilities to intuitively and naturally control a robot for selection and manipulation tasks such as pick-and-place. However, the interaction between the human and the robot in such a scenario is still a challenging task. In order to address the first aspect of pick-and-place, the grasping, we compared two different, multimodal techniques to define the grasp point on a target object, i. e., (i) heading or (ii) pointing, both in combination with speech input. Finally, we show how

This research was partially funded by the German Research Foundation (DFG) and the National Science Foundation of China (NSFC) in project Crossmodal Learning, TRR-169.

[*] krupke@informatik.uni-hamburg.de

[1] is with group Human Computer Interaction (HCI), Department of Informatics, University of Hamburg, Germany

[2] is with group Technical Aspects of Multimodal Systems (TAMS), Department of Informatics, University of Hamburg, Germany

978-1-5386-8094-0/18/$31.00 ©2018 IEEE

a pre-visualization of potential robot motions can enhance the human-robot collaboration.

II. Previous Work

A. Mixed Reality for HRI

MR technologies have enormous potential to address some of the challenges mentioned above. The visual combination of digital content with real working spaces creates an mixed environment, which can provide important feedback during HRI. Paul Milgram [11] introduced the term MR based on a *reality-virtuality continuum* as a continuous scale ranging between the real world, i. e., our physical reality, and a completely virtual environment (VE), i. e., virtual reality (VR). The continuum encompasses all possible variations and compositions of real and virtual objects and consist of both *augmented reality (AR)*, where the virtual augments the real, and *augmented virtuality (AV)*, where the real augments the virtual [12]. Obviously, AR and VR serve different user experiences and there is ample space for both of them to coexist. Noticeably, the real world as well as VR are only the end points of this continuum, whereas AR or more general MR provides a certain area of different combinations of real and virtual content [11].

Very recently, MR has been considered in the area of HRI. Researchers at the DFKI [13] showcased three identical robots, which were remotely manipulated by an operator with the aid of a HoloLens HMD. In contrast to our approach, most of the interaction was based on a menu-based system in which the positioning of the robot arm is specified using a 2D heads-up (HUD) display, whereas in our scenarios the selection is performed directly on virtual and real objects respectively.

B. Multimodal MR Selection Techniques

Multimodal interfaces emerged almost 40 years ago when Richard Bolt introduced his seminal *Put-That-There* work in the 1980. His interface combined spoken commands, which are linked with pointing gestures. With current inside-out tracking HMDs like Microsoft's HoloLens, such multimodal interfaces are implemented typically by two different modes: (i) heading-and-commit (HB) interaction in which heading and a select command are combined, and (ii) point-and-commit (H2F) approaches in which pointing gestures are used instead of heading. Both methods are appropriate for high-to-medium level tasks, either for direct selection of surface points at virtual objects, or whole object selection, assuming the prior detection of potential object surfaces by a recognition system. *Heading* of a user's head is defined by the direction in which the head is turned, which only approximates the actual gaze direction, for current MR HMDs. With HB, users target an object with their heading and then commit with a speech command. In the real world, we typically look at an object that we intend to interact with. However, a comparison with eye-tracking-based HMDs pointed out that currently head tracking is more effective and reliable than eye tracking [14]. As the user looks around, the heading defines a ray, which might intersect with both virtual objects and with a 3D mapping mesh to determine what real-world object the user may be looking at.

With point-and-commit, a user can aim with a pointing-capable motion controller at the target object and then commit with a button press or a speech command. In contrast to the heading ray, in this approach the selection ray is attached to the position and orientation of an additionally tracked motion controller. In our implementation, we wanted to omit additional input devices, and therefore track the user's finger with the inside-out tracking capability of the HoloLens.

III. Mixed Reality HRI Prototype

As described above, MR technology has enormous potential to improve user interfaces (UIs) for HRI. The focus of our work is on human-robot collaboration in which an industrial robot arm can be controlled by the MR user to manipulate real-world objects (cf. Fig. 2). In our implementation, the user wears a Microsoft HoloLens and sees a virtual robot, which is displayed superimposed over the real robot. Additional information about objects with which the user can interact is displayed. Actions of the real robot are triggered through either selection gesture combined with speech, which instantly triggers a simulated motion of the virtual robot that serves as a pre-visualization of the robot's actions. If the user is satisfied with the result, the actual actions of the real robot can be initiated.

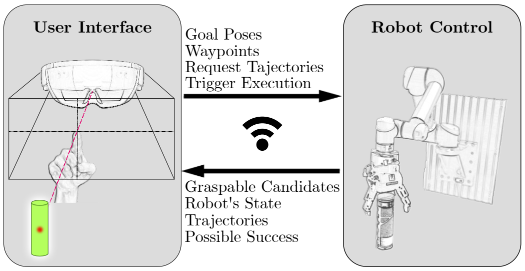

Fig. 2. The concept of the presented human-robot collaboration system is based on a flexible and tetherless setup, which uses a self-contained see-through MR display and an industrial robot in a pick-and-place task.

A. Implementation

The implementation of our MR HRI prototype involves specific hardware, but can be generalized to work with arbitrary robots and different AR and VR HMDs. It follows the principle described in [15]. The usage of the pick-and-place implementation is denoted in Fig. 8

a) Hardware setup: We use an Universal Robot *UR5* industrial manipulator with an attached Robotiq *Adaptive 3-Finger Gripper*, which is controlled by ROS for interaction [16] (Kinetic) on an Ubuntu 16.10 workstation. The UI technology for the MR HRI prototype is implemented for Microsoft's mobile MR headset HoloLens with Unity3D 2017.1.1f. The HoloLens is a self-contained mobile device equipped with inside-out tracking, and it features speech and hand gesture recognition [17]. The HoloLens has a field of view of 30° × 17.5° at a resolution of 1268 × 720. Headset and robot are connected via local network as shown in Fig. 2.

b) Communication: The communication between the headset and the robot control software is implemented using the *ROSbridge* node, provided by the *robot web tools* project [18]. The ROSbridge is a ROS node, which offers a websocket-based communication via the network. The presented system includes an implementation at the Unity3D-side in C# for *Universal Windows Platform* (UWP).

c) Registration and Accuracy: The Microsoft HoloLens places the origin of its coordinate system dynamically and depending on its pose during boot. It automatically applies corrections to the world model and its localization through a SLAM algorithm running on hardware level of the HMD in the background. The stability of HoloLens "holograms" and their world anchors have been evaluated previously and a mean displacement error of $5.83 \pm 0.51\,\text{mm}$ was found [19]. Considering the prototype status of the HoloLens, this is precise enough for pick-and-place tasks with typical objects in tabletop manipulation tasks. Using a single 2D marker approach to align the virtual environment relative to the real one in a short calibration phase, using Vuforia camera-based marker detection [20]. This marker is also used as an infrastructure marker with known transformation from its position to the frame of the planning world, ${}^{W}_{M}T$.

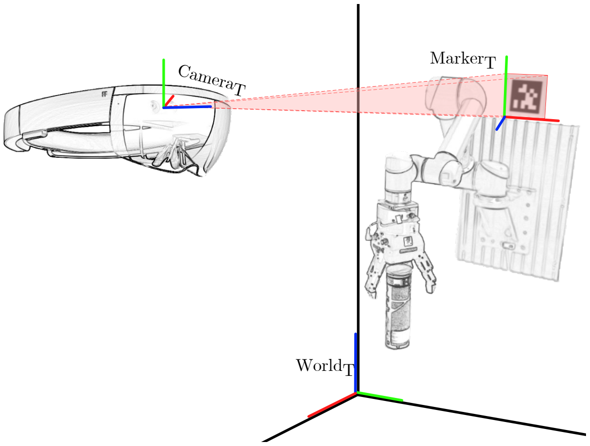

Fig. 3. For transforming poses between different coordinate frames we use a 2D-marker-based approach. The front-facing camera of the mobile AR-headset detects the 6-DoF pose of the marker at a known position in the world model, and thus coordinates within the virtual world of the HMD can be transformed to real world coordinates or other frames.

As shown in Fig. 3 the mobile headset recognizes the marker at the wall using the integrated front-facing RGB camera and computes the marker's 6-DoF pose. By measuring the transformation from the marker position to the origin of the real world coordinate system, we can localize the headset, ${}^{M}_{H}T$. Finally, the conversion between left-handed Unity-coordinate system and right-handed ROS-coordinate system is necessary to communicate pose data between the systems (cf. Equation 1).

To improve the accuracy of the registration process we used a priori knowledge of the lab during marker detection, since the walls are perpendicular to the floor. This way, the rotational tracking only is only necessary to acquire the rotation around the world yaw axis.The resulting reduction of 6-DoF tracking to 4-DoF improved our system's accuracy. We evaluated our system's final accuracy, by positioning virtual objects according to corresponding real world objects and found that our setup can achieve an accuracy of $30 \pm 9.3\,\text{mm}$. The larger inaccuracy, than reported in Vassallo et. al [19], results from the additional inaccuracy of the utilized marker detection with an uncalibrated camera setup. However, with a single marker detection process during startup, we experienced a drift on the vertical axis over time.It can be reduced by repeating the registration marker detection during runtime. On the horizontal plane, the deviation is much smaller ($1.6 \pm 5.4\,\text{mm}$) and we measured no drift over time.

TABLE I

TRANSLATIONAL AND ROTATIONAL MEAN ERROR WITH STANDARD DEVIATION OF THE REGISTRATION IN METERS AND DEGREES.

x-axis	**y-axis**	**z-axis**
$-.0072 \pm .0068$	$.0267 \pm .0114$	$.0040 \pm .0073$
pitch	**yaw**	**roll**
$< .01°$	$2.2050° \pm .3931°$	$< .01°$
euclid 2d	**euclid 3d**	
$-.0016 \pm .0054$	$.0303 \pm .0093$	

$$ {}^{W}_{C}T = {}^{W}_{M}T \cdot {}^{M}_{H}T \cdot {}^{H}_{C}T \qquad (1) $$

Based on this, tracking in the virtual world is applicable to real world tasks, like selecting grasp poses. We implemented heading-based (HB) and head-to-finger-based (H2F) raycast methods to select poses for grasping on virtual objects. Once a pose is selected, HoloLens-built-in voice command recognition is used to trigger actions.

B. Multimodal UI for Pick-and-Place Tasks

In our implementation we extended methods described in [21] for general object manipulation in AR with a mobile headset. The main challenge arises due to the extension of a 2D visualization with classical mouse-keyboard input to a 3D representation with spatial input and output. As described above, the user can freely move around the robot and the objects subject to be picked and placed. The manual specification of robot grasping positions is an open research question. In principle, the inside-out tracking technology allows two different approaches, i. e., (HB) heading-and-commit and (H2F) point-and-commit as described above in section II. In both selection techniques the user perceives visual feedback about the current selection point by means of a red ring-shaped cursor as illustrated in Fig. 1. A commit command can be easily provided via voice, which has the benefit to reduce sensorimotor errors. If the user is satisfied with the current pick location indicated by the red ring-shaped cursor, they can confirm this position with a voice command ("*select*").

a) HB Selection: With our heading-based implementation, a ray is cast from the head position in the forward direction of the HMD. The pose of the user's HMD is used to determine the heading vector, which is the mathematical representation of a laser pointer straight ahead from between the user's eyes towards the forward orientation of the HMD. As the user looks around, this ray can intersect with both virtual objects and with the spatial mapping mesh to determine which real-world object the user may be looking at.

b) H2F Selection: In this technique, the object occluded by the user's finger is the selected object. The finger is tracked with the inside-out tracking capability of the HoloLens. Then, the ray-cast is defined by the position of the user's head (from between the eyes) and the position of the fingertip of index finger.

C. Pre-Visualization of Robot Actions

After the user initiated the confirmation command via speech, the virtual representation of the robot starts the simulation of the picking motion, and shows the corresponding motions to reach the specified pick location. If the user is satisfied with the simulated motion of the robot, they can confirm this action with another voice command ("*confirm*"), and the real industrial robot arm will perform the same actions as visualized in the simulation.

The advantage of such a pre-visualization is manifold. First of all, it allows the user see the motions of the complex robot arm before it actually moves. This is important for increased safety, but also simplifies the robot programming and robot behavior understanding, as described in Section II. The user can see the effective space of the actions and can move themselves or obstacles away from the robot's motion paths, or alternatively specify a different location or different path.

IV. Evaluation

In this section, we describe the evaluation that we performed in order to analyze which of the MR selection techniques described in Section III-B, is more beneficial for our considered HRI setup. While we utilized actual robot descriptions, this experiment was done without an actual robot to test the software beforehand.

A. Hypotheses

In informal pilot evaluations we found out that both, using the finger or the heading to select an objects appears to be natural for users. However, while our HB technique requires only the head for specifying the selection ray, the H2F technique requires the user to position their pointing finger in mid-air within the view of the tracking sensors of the headset. Therefore, we assume that using the finger to select an object requires more effort, is less precise and takes more time than using heading. Hence, we hypothesize that:

H1: HB selection is less demanding (physical, mental and temporal) than H2F.
H2: HB is more accurate than H2F.
H3: HB requires less time than H2F.

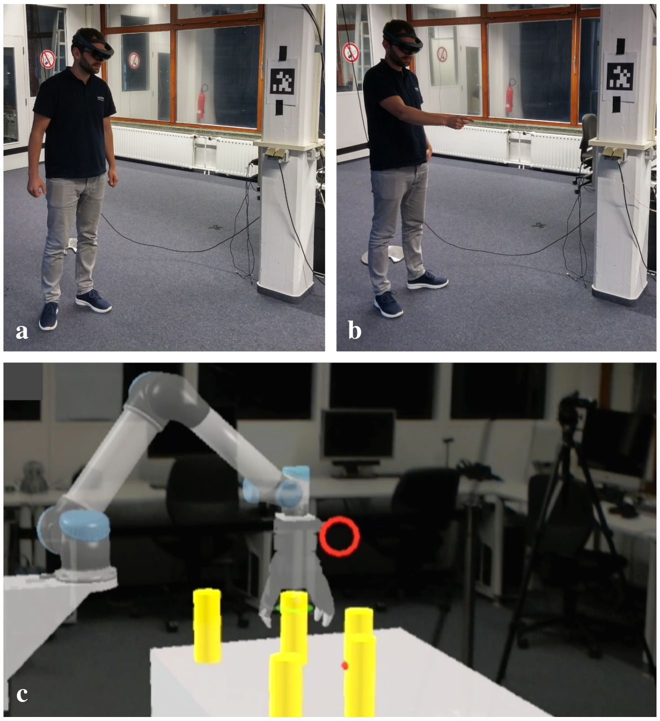

Fig. 4. Participant during the experiment: HB-selection (a), H2F-selection (b) and the view of the participant through the AR headset (c).

B. Participants

We recruited 16 participants (3 female and 13 male, ages 21 to 38, $M = 28.31$). The participants were volunteering employees or students of the local department of computer science. The ratio of female and male participants is represenitive for the members of our department and thus the expected user group. If requested, the students obtained class credit for their participation. 8 of our participants had normal or corrected-to-normal vision and none of them wore glasses during the experiment. 1 participant reported strong eye dominance, but no other vision disorders have been reported. No disorder of equilibrium or a motor disorder such as impaired hand-eye coordination was reported. 3 participants reported prior participation in experiments involving the Microsoft HoloLens. Handedness was not relevant, due to the implementation, which allows either the left or the right hand for gestures. The average total time for the experiment was 25 minutes, the time spent with the HMD worn about 15 minutes.

C. Materials

The main device in this experiment is the self-contained, see-through HMD Microsoft HoloLens [17]. As illustrated in Fig. 4, participants were able to move within a hemispheric area of $3.5\,\mathrm{m} \times 2.5\,\mathrm{m}$. On the edge of this area the marker was placed on a wall, mirroring the actual setup in another lab with an actual UR5 actuator. The virtual part of the visual stimulus was rendered using the Unity engine directly on the HoloLens, following performance optimization to ensure the frame rate was higher than the $60\,\mathrm{Hz}$ display frequency.

D. Methods

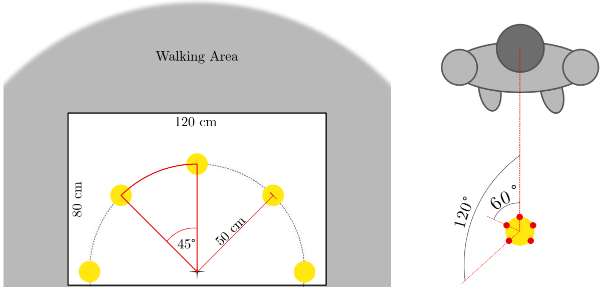

Fig. 5. The arrangement of cylinders on the table and the targets on top of the surface of the cylinders.

As illustrated in Fig. 4, the user saw the real laboratory with a MR overlay of a virtual robot actuator, a virtual table and five virtual, yellow cylinders arranged on a half-circle with a radius of 0.5 m as illustrated in Fig. 5. For each trial, a random yellow cylinder was selected. Using the current participant's position, the target position along the outer edge of the cylinder was determined. The targets were represented by red spheres with a radius of 1.125 cm. The targets appeared sequentially at an angle of ±0, ±60 or ±120 ° in radial coordinates relative to the direct line between the participant and the current cylinder, where ±0 would be the point of the cylinder directly facing the participant. The angles −60 and −120 ° correspond to the points on the left side on the cylinder from the users current position, whereas +60 as well as +120 were displayed on the right side on the cylinder with respect from the users current position (cf. Fig. 5) Positive or negative angles are chosen in a way to direct the user to stay within the hemisphere.To create a more economically valid pick-and-place scenario, we occasionally placed targets on the partly occluded side of the cylinder, which forced participants to walk around the interaction area. Using one of the two selection techniques, the user had to position the gripper at the target sphere. As mentioned above, for both the selection techniques, the users saw a red, ring-shaped cursor which they moved either through rotating their head or by moving their dominant hand's index finger. To avoid learning effects, the experiment was counter-balanced, meaning that half the participants started with the HB selection technique and the other group with the H2F technique and then used the other technique, respectively.

In the experiment, we used a within-subject repeated measures 2 (HB vs. H2F) × 3 (angles) × 5 (cylinder) × 2 (repetitions) design. Before the experiment, all participants filled out an informed consent form and received written instructions on how to perform the task. Furthermore, they filled out a demographic questionnaire before the experiment and the NASA Task-Load Index (TLX) questionnaire [22], Simple Usability Scale questionnaire (SUS) [23] and At-trakDiff usability questionnaire [24] after each condition.

TABLE II
SUMMARY OF TIME, ACCURACY, USABILITY AND TASK-LOAD

	M	SD	T/V Value	p	d-Cohen
TIME HB	5.84	3.72	V(15)=37148	<.001	.33
TIME H2F	7.19	4.45			
ACC HB	.01	.03	V(15)=34748	<.001	.34
ACC H2F	.03	.06			
SUS HB	78.12	17.11	V(15)=40	.15	.52
SUS H2F	68.59	19.54			
TLX HB	39.48	13.66	T(15)=-2.74	<.05	1.03
TLX H2F	53.23	13.03			

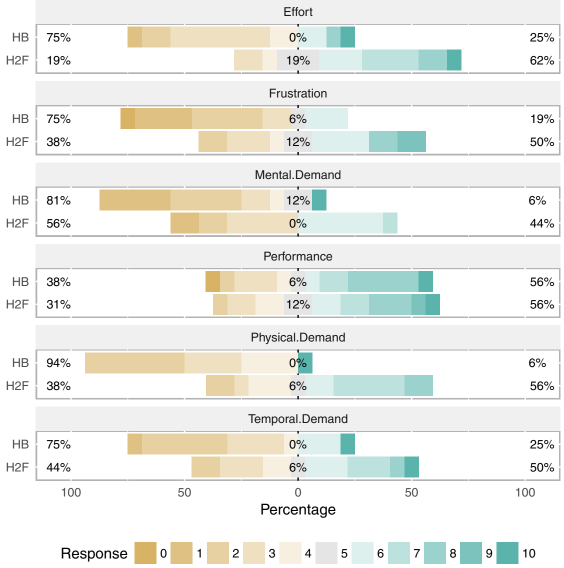

Fig. 6. Comparison of NASA-TLX results show the advantage of the heading-based method.

V. RESULTS

In this section, we summarize the results and analyses of our experiment. All statistical tests were done at the 5% significance level.

a) Performance: Fig. 7 shows the mean euclidean error distance between the selected position by the user and the target position during the trials and the time between the appearance of a new target and the selection by the participant. We tested the results for normality with a Shapiro-Wilk test ($p < .05$) and since they were not normally distributed we used a Wilcoxon Signed-Rank test. The results are shown in Table II. These results confirm H2. Table II and Fig. 7 show that participants selected the targets in shorter time significantly. These results confirm H3. Additionally, we calculated the failure rates of both methods. A failure was defined by selecting a position at the wrong cylinder. For HB the failure rate was 1.29% and for H2F 7.08% in total.

b) Usability: The raw TLX questionnaire results are summarized in Fig. 6. They were normally distributed ($p > .05$) and we used a paired samples T-Test for the analysis.

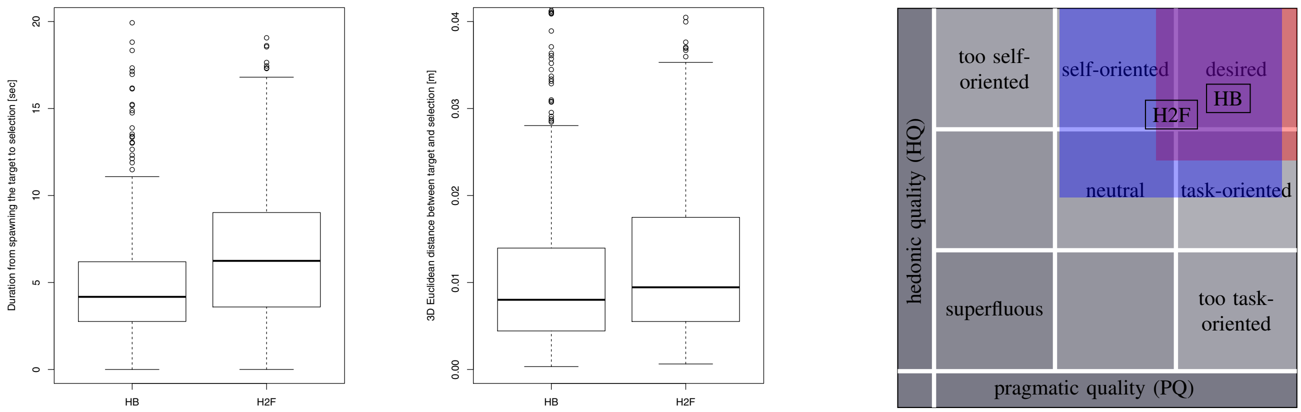

Fig. 7. (left) Selection time HB vs. H2F. (center) Distance error HB vs. H2F. (right) Average values and confidence rectangles for the AttrakDiff questionnaire of the two conditions: (red) HB for the heading-based approach and (blue) H2F for the finger-pointing method.

Taskload for method HB is significantly lower than for H2F and indicates that much less capacity of the operator is used. SUS score for HB is significantly lower than for H2C and denotes an upper B-rank (A starts at 80.3), while H2C is much lower on the accumulated score with a low C-rank. Since the Shapiro-Wilk test for normality showed that the SUS results are not normally distributed, we used a Wilcoxon Signed-Rank test to analyze them. These results confirm H1. The results for the AttrakDiff 2 questionnaire are visualized in Fig. 7 and indicate a tendency that operators prefer the HB method over H2F and that they would use the system in real world tasks.

VI. Validation of the Results

Based on the results described in Section IV, we picked the HB selection method and implemented a fully working prototype, which controls a real robotic arm. After the user triggers a selection by uttering (“*pick*” or “*place*”), our system first requests the planning system to find a robot motion trajectory, and visualizes it in the HoloLens by moving the virtual UR5. The system then waits for another command for re-planning or a (“*confirm pick*” or “*confirm place*”) command, which causes the system to execute the already planned motion.

The virtual model of the robotic arm is used to provide the user with a preview of the motion at the place where it is intended to be executed. By doing so, the user is provided with visual feedback about the possible success and suitability of their input given to the robot. The presented stereoscopic images using a see-through HMD have positive effects spatial on awareness and distance perception. As suggested in [9] we utilized a stereoscopic device. Since, visual perception is our most dominant channel [25], visual cues have the highest impact in decision making during robot operation.

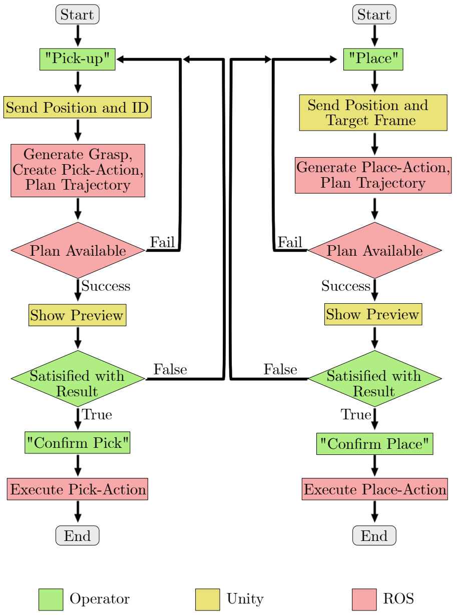

Fig. 8. Procedures of the pick-and-place implementation.

A. Validation Task

The proximity to the robot helps the user to understand the situation presented to the operator. The combination of head orientation and speech control keeps the hands of the operator free, allowing them to support the system with assistive manipulation tasks or handover tasks. In our validation we found the system capable of presenting the user possible problematic situations before they happen. Thus, the user is able to change the strategy before the problematic situation occurs during movement plan execution.

We presented our system to five experts in the fields robotics or 3D user interfaces. Overall, they gave positive feedback, and agreed that it simplifies the interaction in an intuitive way. The main critique was that it would be preferable to see in advance, whether the planner can calculate a trajectory to a point on an object, which can be solved using an approach like [26]. One expert suggested using shorter, one syllable speech commands.

VII. Conclusion and Future Work

In this paper, we presented the general concept and implementation of an MR human-robot collaboration system in which a human can intuitively control a co-located industrial robot arm for pick-and-place tasks[3]. We evaluated two different multimodal HRI techniques to define the pick and place location on a target object using (i) HB and (ii) H2F both in combination with speech. The results show that HB interaction technique is more precise, requires less time and is perceived as less physically, temporally and mentally demanding compared to the H2F techniques for MR-based pick-and-place scenarios. However, there are some limitations, which could be addressed and future work. It remains unknown if the results would also apply for another HMD with larger field of view, allowing the user to lower their hand during the interaction, which would require less physical effort.

Our study was focused on the selection of a single pick point. In future work it might be interesting to investigate if more complex pick poses could be defined with the described setup, for example, by tracking the full hand.

Furthermore, the multitude of information presented to the operator could be extended. The range of the robot as well as the dexterous capabilities in the workspace would be of interest for further improvement of the user's performance. In the implementation the set of objects for manipulation is fixed. Future improvements will integrate object detection and the communication of new candidates between the side of the user interface and the robot.

To evaluate the full working prototype, we conducted a confirmatory expert evaluation, where an actual robot was controlled, gathering overall positive feedback.

We believe that MR setups require more research in the field of HRI and the findings of this paper give important implications for the design of future MR setups and provide some first steps towards novel forms of HRI.

References

[1] M. L. Lupetti, C. Germak, and L. Giuliano, "Robots and Cultural Heritage: New Museum Experiences," in *Proceedings of the Conference on Electronic Visualisation and the Arts (EVA)*, (Swindon, UK), pp. 322–329, BCS Learning & Development Ltd., 2015.

[2] M. Mast, M. Burmester, B. Graf, F. Weisshardt, G. Arbeiter, M. Španěl, Z. Materna, P. Smrž, and G. Kronreif, "Design of the Human-Robot Interaction for a Semi-Autonomous Service Robot to Assist Elderly People," in *Ambient Assisted Living*, pp. 15–29, Springer, 2015.

[3] D. Vernon, S. Thill, and T. Ziemke, "The Role of Intention in Cognitive Robotics," in *Toward Robotic Socially Believable Behaving Systems-Volume I*, pp. 15–27, Springer, 2016.

[4] C. Vandevelde and J. Saldien, "Demonstration of OPSORO - an open platform for social robots," in *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 555–556, March 2016.

[5] F. Leutert, C. Herrmann, and K. Schilling, "A Spatial Augmented Reality System for Intuitive Display of Robotic Data," in *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 179–180, March 2013.

[6] S. Hashimoto, A. Ishida, and M. Inami, "TouchMe: An Augmented Reality Based Remote Robot Manipulation," in *The 21st International Conference on Artificial Reality and Telexistence, Proceedings of ICAT2011*, 2011.

[7] G. Burdea, "Invited Review: The Synergy Between Virtual Reality and Robotics," *IEEE Transactions on Robotics and Automation*, vol. 15, pp. 400–410, Jun 1999.

[8] C. W. Nielsen, M. A. Goodrich, and R. W. Ricks, "Ecological Interfaces for Improving Mobile Robot Teleoperation," *IEEE Transactions on Robotics*, vol. 23, pp. 927–941, Oct 2007.

[9] H. C. Fang, S. K. Ong, and A. Y. C. Nee, "Novel AR-based interface for human-robot interaction and visualization," *Advances in Manufacturing*, vol. 2, pp. 275–288, Dec 2014.

[10] A. Leeper, K. Hsiao, M. Ciocarlie, L. Takayama, and D. Gossow, "Strategies for Human-in-the-loop Robotic Grasping," in *2012 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 1–8, IEEE, 2012.

[11] P. Milgram and F. Kishino, "A Taxonomy of Mixed Reality Visual Displays," in *IEICE Transactions on Information and Systems, Special issue on Networked Reality*, 1994.

[12] F. Steinicke, *Being Really Virtual - Immersive Natives and the Future of Virtual Reality*. Springer, 2016.

[13] DFKI, "Mixed Reality systems for crosssite production in INDUSTRIE 4.0." `https://www.dfki.de/web/news/dfki-cebit-2017/mixed-reality/index_html/`. Accessed: 2018-02-26.

[14] Y. Y. Qian and R. J. Teather, "The Eyes Don't Have It: An Empirical Comparison of Head-Based and Eye-Based Selection in Virtual Reality," in *Proceedings of the 5th Symposium on Spatial User Interaction*, pp. 91–98, ACM, 2017.

[15] D. Krupke, S. Starke, L. Einig, F. Steinicke, and J. Zhang, "Prototyping of Immersive HRI Scenarios," in *Proceedings of CLAWAR 2017: 20th International Conference on Climbing and Walking Robots*, pp. 537–544, CLAWAR Association, 2017.

[16] M. Quigley, K. Conley, B. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, and A. Y. Ng, "ROS: an open-source Robot Operating System," *ICRA workshop on open source software*, vol. 3, no. 3.2, p. 5, 2009.

[17] Microsoft, "Microsoft HoloLens." `https://www.microsoft.com/en-us/hololens/`. Accessed: 2018-02-26.

[18] R. Toris, J. Kammerl, D. V. Lu, J. Lee, O. C. Jenkins, S. Osentoski, M. Wills, and S. Chernova, "Robot Web Tools: Efficient messaging for cloud robotics," in *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 4530–4537, Sept 2015.

[19] R. Vassallo, A. Rankin, E. C. S. Chen, and T. M. Peters, "Hologram stability evaluation for Microsoft HoloLens," in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, vol. 10136, p. 1013614, Mar. 2017.

[20] P. Inc., "Vuforia." `https://www.vuforia.com/`. Accessed: 2018-02-26.

[21] D. Kent, C. Saldanha, and S. Chernova, "A Comparison of Remote Robot Teleoperation Interfaces for General Object Manipulation," in *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 371–379, ACM, 2017.

[22] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research," *Advances in psychology*, vol. 52, pp. 139–183, 1988.

[23] J. Brooke *et al.*, "SUS-A quick and dirty usability scale," *Usability evaluation in industry*, vol. 189, no. 194, pp. 4–7, 1996.

[24] M. Hassenzahl, M. Burmester, and F. Koller, "AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität," in *Mensch & Computer 2003*, pp. 187–196, Springer, 2003.

[25] M. I. Posner, M. J. Nissen, and R. M. Klein, "Visual Dominance: An Information-Processing Account of Its Origins and Significance," *Psychological review*, vol. 83, no. 2, p. 157, 1976.

[26] A. Makhal and A. K. Goins, "Reuleaux: Robot Base Placement by Reachability Analysis," *CoRR*, vol. abs/1710.01328, 2017.

[3] The source code is available on GitHub:
`https://github.com/denniskrupke/holoROS`
`https://github.com/TAMS-Group/hololens_grasp`