Evaluating Visual and Auditory Substitution of Tactile Feedback during Mixed Reality Teleoperation*

Yannick Jonetzko[1], Judith Hartfill[2], Niklas Fiedler[1], Fangwei Zhong[3], Frank Steinicke[2], and Jianwei Zhang[1]

[1] Technical Aspects of Multimodal Systems (TAMS)
[2] Human-Computer Interaction (HCI)
Universität Hamburg, Vogt-Kölln-Straße 30, 22527 Hamburg
{yannick.jonetzko, judith.hartfill, niklas.fiedler, frank.steinicke, jianwei.zhang}@uni-hamburg.de
[3] School of Artificial Intelligence, Peking University and Beijing Institute for General Artificial Intelligence (BIGAI)
Yiheyuan Road 2 and 5, Haidian District, Beijing 100871, P.R. China
zfw@pku.edu.cn

Abstract. Mixed reality (MR) technology has shown enormous potential for real-time human-robot teleoperation. To provide the user with sensor feedback, typically extensive instrumentation is required, such as wearable devices. In this paper, we introduce an MR human-robot teleoperation system based on the Microsoft HoloLens 2 (HL2). The user can directly control the end-effector pose and the gripper of an arbitrary robot arm via hand tracking. Additionally, tactile information can be perceived without wearing data gloves. Therefore, the tactile information gathered by the robot during interaction with objects is substituted and presented to the user visually and auditory. We conducted a pilot user study to analyze the system's usability and the effects of substituted tactile feedback during typical teleoperation tasks. The results show that the system is applicable for teleoperation as the users reached an 87.1 % success rate in the performed manipulation task. The users were satisfied with the easy-to-use teleoperation interface and reported a preference for multimodal tactile substitution modes. Regarding the performance, an improvement in the average applied force could be observed when providing tactile feedback.

Keywords: Teleoperation · Tactile Sensing · Multimodal Feedback · Mixed Reality.

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

1 Introduction

Teleoperation allows the user to control a system or machine over distance. For instance, humans can teleoperate robots in the same laboratory, in another room, in another country, or even in space. However, it requires a lot of expertise and practice to precisely control such robots. In particular, in surgical tasks for medical purposes or handling hazardous materials, precision is an inevitable requirement. To improve the interaction with the robot and provide humans with feedback about the robot's actions, researchers develop easy-to-use systems [26, 29].
One of the most challenging modalities during teleoperation is tactile perception. Realistic haptic or force feedback is often implemented through wearable devices such as data gloves or fingertip devices, which exert vibrations or small deformations to the skin [23]. However, this user instrumentation is often cumbersome and prevents a natural movement of the user's hands. On the other hand, there are approaches without any wearable devices. For instance, Carter et al. [6] present the principle of inducing contactless haptic cues using ultrasound feedback. An acoustically transparent display applies small vibrations to multiple points on the user's hand in an area above the device.
In recent years, the fields of augmented reality (AR) and virtual reality (VR) have been used for robot teleoperation to provide feedback, environment augmentation, and interaction interfaces [7, 12, 17]. Milgram and Kishino [19] summarize the combination of real and virtual environments as Reality-Virtuality continuum, which describes mixed reality as the area where both the real and virtual worlds are mixed. However, MR human-robot teleoperation (e.g. by using the Microsoft HoloLens) often cannot present the above-mentioned tactile information to the user.
This paper presents a novel setup for human-robot interaction with MR-based real-time teleoperation for one- and two-handed manipulation tasks. Additionally, to improve the performance while teleoperating a robot manipulator, we introduce different methods to substitute tactile information gathered by the robot's gripper. The tactile information is presented visually and auditory by the head-mounted display (HMD). To evaluate the new system, we conducted a pilot study with five participants. They performed a task with the robot while receiving different types of uni- and multimodal feedback (see Figure 1).

Hence, the contributions of this work can be summarized as follows:

- Development of an MR-based real-time teleoperation system combining Microsoft HoloLens 2 with arbitrary robot arms
- Introduction of visual and auditory substitution of tactile feedback during MR teleoperation
- Pilot study to evaluate the usability of the MR teleoperation system with the mobile robotic PR2 platform

The remainder of this paper is structured as follows. Section 2 summarizes related work. In Section 3 we introduce the MR system and the different concepts

Fig. 1. The MR teleoperation scenario. A user wears the HL2 and teleoperates the robot arm's end-effector pose (indicated by the sphere). Visually substituted tactile feedback is provided with arrows (condition S).

of visual and auditory substitution of tactile feedback. Section 4 describes the pilot study, which we conducted to evaluate the usability of our approach. Afterward, we discuss the performance and the usability of the setup in Section 5. Section 6 concludes the paper and gives an outlook on future work.

2 Related Work

In recent years, VR and AR have become more and more popular, in particular, in the entertainment and gaming industry, but also for training and simulation as well as for improving human-robot collaboration [9]. Different approaches show the usability of interactive AR robot interfaces and teleoperation [17, 30, 31].
Chan et al. [7] present a multimodal system to teleoperate an industrial robot arm on a predefined trajectory using augmented reality. In their experiment, the users program a trajectory on a table and move the robot afterward along this trajectory while controlling the force applied to the surface by the robot's end-effector. The researchers compare two different force feedback modalities, one of them as a visual arrow indicating the normal force, the other one as haptic vibrations at the user's forearm. As their results show that providing only haptic feedback performs best, they hypothesize that the additional visual feedback leads to cognitive overload.
Contrary to this hypothesis, other researchers determined performance improvements when providing multimodal feedback [5, 13, 20, 25]. Herbst and Stark [13] propose a desktop VR setup. They substitute force magnitudes with visual, auditory, and haptic feedback. In two experiments, participants were asked to sort virtual blocks by their weight and push or pull the block with the least friction out of a stack. They tested each modality individually and in combination. The authors found that a combination of two modalities performs better in terms of execution time and number of transitions compared to the single modalities. The combination of all three modalities did not increase the performance any further.

In the field of medical robotics, different visual presentations of force feedback have been analyzed. Aviles-Rivero et al. [1] compare four different color-based visualizations on a 2D display of a robotic surgical system and found a strong preference for a system with the visual cue as close as possible to the tool, compared to systems with the visual cue in the upper right corner of the display.
Cooper et al. [8] evaluate the effects of substituted feedback in a VR environment with a user study. In their experiment, the participants were asked to perform a wheel change on a car. During the task, they received vibration feedback through tactile gloves, visual cues, and audio information with headphones. They tested each modality individually and in combination and measured the execution time to compare the performance. The subjects performed best when all modalities were used.
Compared to the approaches above, our setup allows teleoperating a 7 Degrees of Freedom (DoF) robotic arm without any wearable devices in an MR environment. To provide tactile information without haptic devices, we use visual and auditory substitutions.

3 System

In this section, we provide an overview of the hard- and software used for this MR-based teleoperation setup. The system enables us to teleoperate the end-effector pose of different robot arms by tracking the user's hand with an HMD. It is not necessary to wear any data gloves or tracking devices on the hands, allowing free hand movements.

3.1 Hardware

Our system consists of two parts: (i) the robotic platform and (ii) the MR head-mounted display HoloLens 2 developed by Microsoft [18]. A robot arm with at least 6-DoF is required for a reliable teleoperation and needs to be controlled using the robot operating system (ROS). To show the usability of the system, we describe it at the example of the mobile platform PR2 and conducted a pilot study as a proof of concept. The platform we used for the pilot study in this work is equipped with one 7-DoF arm with a parallel gripper at the end. This gripper has tactile pressure sensor arrays, one array at each finger [22].

3.2 Communication

Two different environments were integrated in this setup: (i) the robot framework ROS [21] and (ii) the Unity3D game engine [27], which is used on the HoloLens. In recent years, different approaches were developed to combine these environments. The most common one is ROS# [3], which is used in our setup. This framework provides the integration with the Universal Windows Platform (UWP) and handles the communication between ROS and Unity3D.

3.3 Registration

A major challenge in AR setups is the alignment of the virtual environment with the real world. The HoloLens 2 uses its depth sensors to localize itself and place virtual objects at fixed locations in the environment in real-time. Nevertheless, if the virtual objects are to be aligned with real ones, either a known fixed transformation or an object pose recognition algorithm is required. In our setup, we use the fiducial marker tracking library AprilTag [16]. The PR2 detects the marker with its head camera and provides the resulting transform to the Unity MR environment. On the HMD side, the marker is tracked by the front-facing camera of the HL2. Both transforms are transferred into the HMD's coordinate system and align the virtual robot with the real one. Kalaitzakis et al. [14] determine the accuracy of the AprilTag detection algorithm with an average position error of 2 cm, tested with different cameras. As we detect the tag with both the robot and the HMD, the error can add up to 4 cm. Furthermore, when calculating the fiducial marker's pose on the HL2 continuously, the frame rate drops from ~60 fps to ~10 fps and results in an unusable setup. In combination with the marker pose inaccuracy, we decided to register the transform once when starting the system. To minimize the position error afterward, the user can manually adjust position and orientation with a button panel in the virtual environment. This small inaccuracy is no issue in this setup, as the user sees the actual position of the robot's real end-effector during teleoperation.

3.4 Teleoperation

We directly teleoperate the arm via its 6-DoF end-effector pose in Cartesian space using a custom version of the jog_control[4] ROS package. The customization is presented in [28], allowing the teleoperation by an absolute pose, instead of relative position and orientation deltas, with an almost unnoticeable average delay of 0.35 s. The author shows that the framework is independent of the robot platform. As this is the only interface to the robot in the system presented in this paper, our setup can be integrated with different robots as long as it runs ROS. To calculate the end-effector pose, we use the hand tracking algorithm provided by Microsoft in their toolkit[5] for their MR Displays. This algorithm extracts all finger, palm, and wrist joints from the inbuilt depth camera's measurements and provides 25 joint poses overall. In this setup, the first finger and the thumb are used to calculate the goal position and rotation for the teleoperation. We use the knuckle and metacarpal joints for the pose, as their position is more stable than those of the fingertips. To control the gripper's opening state, the distance between the tips of index finger and thumb is directly mapped to the distance between the gripper's fingers. To give some feedback throughout the teleoperation, a small semi-transparent sphere is visualized at the goal pose (see Figure 1).

[4] https://github.com/tork-a/jog_control [Accessed Aug. 2, 2022]
[5] https://github.com/microsoft/MixedRealityToolkit-Unity [Accessed Aug. 2, 2022]

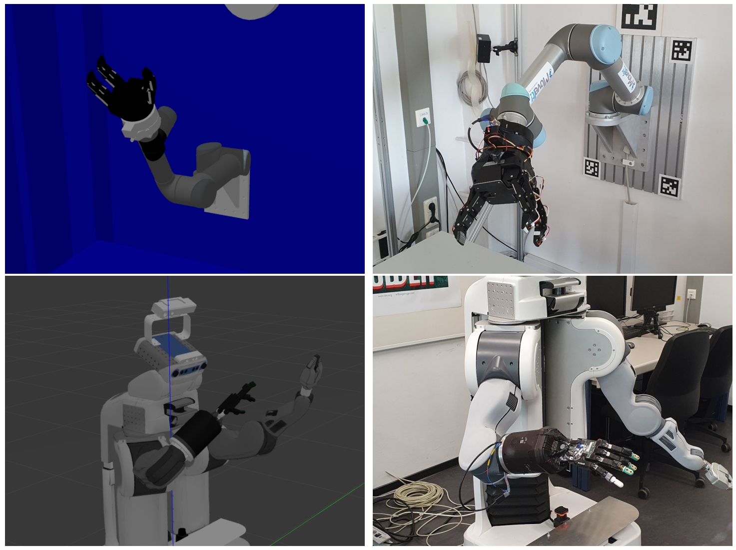


Fig. 2. The teleoperation system integrated on different robotic platforms (top: UR5 + robotic gripper, bottom: PR2 + Shadow Hand) in simulation (left) and on the real hardware (right).

3.5 Tactile Readings

Romano et al. [22] present a method to estimate the applied normal force with the PR2 and the pressure arrays at hand. Each tactile sensor consists of 22 taxels, 15 at the front, two at each side, two at the top, and one at the back. With the 15 front taxels' readings, the applied force is calculated by the sum of the cells' forces. To provide the user with this information, the average force value of both fingers is sent to the MR Display and is substituted by visual and auditory feedback.

3.6 Arbitrary Robots

As mentioned earlier, the teleoperation system we describe and evaluate can easily be integrated on arbitrary robots. We tested it successfully on three different platforms: On a PR2 system using the original left arm, on the right arm of the PR2 with an attached Shadow Dexterous Hand as forearm, and on a UR5 with an attached robotiq 3 finger gripper. We used the original PR2 left arm setup for the pilot study. In addition to Figure 1, Figure 2 shows the other two robots both in simulation and on the real hardware.
We designed the interfaces to be independent of the robot arm and its Degrees of Freedom as well as independent of the gripper so that they can be exchanged easily. In Figure 3, the robot independent parts are marked in blue. Accordingly,

the orange parts need to be updated for the individual robot. We already implemented an interface for full five-fingered dexterous manipulation, which will be part of future work as we do not have a hand controller yet (see the white node in Figure 3). Both the robotiq gripper and the Shadow Hand are currently abstracted as parallel grippers during the teleoperation. The HL2 can track both hands simultaneously, allowing us to teleoperate two arms in parallel. We implemented this on our two-arm PR2 platform. As jog control is taking care of collisions, all movements were collision-free.

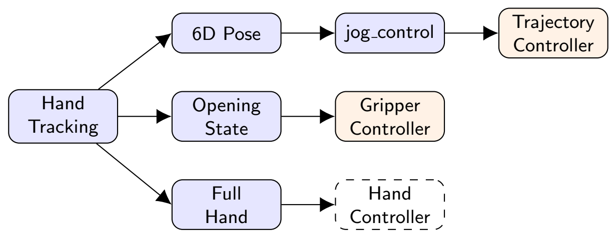


Fig. 3. Abstract teleoperation process for robotic arms and parallel gripper control. Dexterous teleoperation of multifinger hands is not yet supported.

4 User Study

We conducted a pilot study using the mobile robotic PR2 platform to test the user performance and usability of the proposed system to answer the following two research questions:

1. *Does providing tactile information in the form of visual and auditory feedback improve the teleoperation performance regarding execution time, average applied force, and success rate?*
2. *Is the new MR robot setup applicable for teleoperation tasks regarding the usability of non-expert users?*

We designed an object manipulation task to evaluate different levels of tactile substitution methods, including no feedback as well as uni- and multimodal conditions. Furthermore, we used different hardness levels of the manipulated objects to analyze whether this affects the applied force.

4.1 Tactile Substitution

To test the influence of substituted tactile information by visual and audio feedback, the user study was conducted with (i) no feedback, (ii) visual-only, (iii) audio-only, and (iv) multimodal visual + audio feedback.

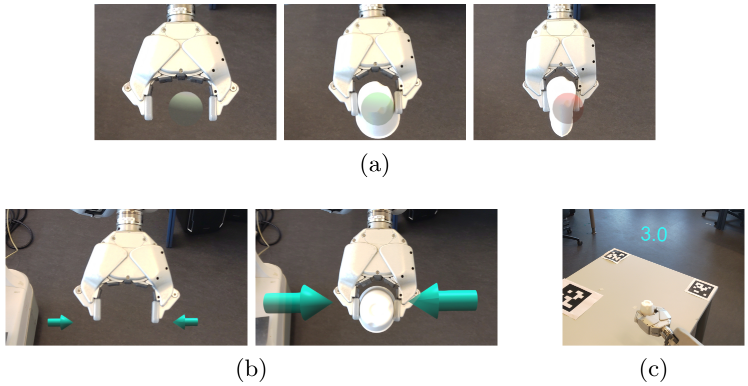


Fig. 4. Visually substituted tactile feedback in the form of (a) a sphere, (b) arrows, and (c) text. Proportional to the measured force, the sphere changes color, the arrow changes size, and the text shows the actual measured value.

Visualization We developed three visualizations that encode the tactile information with different visual properties:

Colored Sphere (*C*) A small sphere is visualized as a position reference for the teleoperation under all conditions. The sphere is placed on top of the gripper to have it as close as possible to the region of interest without occluding the manipulated object. In condition C, the color of the sphere will change proportionally to the applied force from white (0 N) over green (5 N) and yellow (7.5 N) to red (10 N max) (see Figure 4 a).

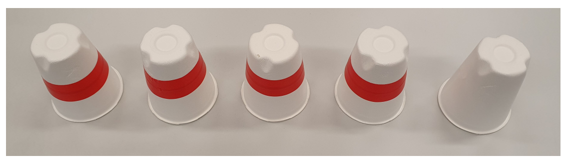
Arrows (*S*) Two arrows are visualized next to each finger of the gripper. The arrows' size will change proportionally to the applied force, with the maximum size at 10 N. To stay comparable with the other substitutions, both arrows are visualized the same size and not individually for each finger (see Figure 4 b).

Text (*T*) The force is visualized as a number in the top center of the user's field of view to avoid occlusion (see Figure 4 c).

Audio (*A*) When the gripper is in contact with both fingers, a constant sine wave is played through the HMD's internal speakers. The frequency varies proportionally to the average measured forces of both fingers between 200 Hz (0 N) and 600 Hz (10 N). Similar to a variometer's feedback, the constant tone starts to beep if the maximum frequency of 600 Hz is exceeded.

4.2 Manipulation Task

In this experiment, the subjects were asked to stack four plastic cups onto a fifth by teleoperating the robot. The participants were instructed to be as fast and precise as possible and apply a minimal amount of pressure to the cup. As it is necessary to maintain the applied pressure to prevent damaging the objects, this task seems reasonable to evaluate the tactile feedback substitution. To simplify the task, all cups were placed in a row, the region where the objects were to be grasped was marked with red to improve comparability, as the bottom of the cups is harder (see Figure 5 a). The task was considered completed when each cup was either stacked or fallen over. To test if the hardness of the cups influences the average applied force or the execution time, each condition was performed with three different cups shown in Figure 5 b. The hardness level is decreasing from cup 1 to cup 3.



(a)

(b)

Fig. 5. The manipulation task. (a) Cup stacking task: The four cups on the left are supposed to be stacked on the cup on the right. The red area indicates the region where the object should be grasped. (b) Different cups were used to test the influence of the object's hardness. The left cup is used in the training phase, the other ones are cups 1 to 3 from left to right with decreasing hardness levels.

4.3 Measures

To evaluate the system and the tactile substitutions, we measured multiple factors. Regarding the performance, execution time, average applied force, and success rate are the factors of interest. The execution time was stopped from the first teleoperated movement until the last cup was stacked or fallen over. When both fingers detected a contact, the force was recorded and averaged afterward. For the success rate, we counted the stacked cups (maximum 4 per trial).
To evaluate the overall user experience of the proposed system, the following standard questionnaires were used: NASA Task Load Index (NASA TLX) [10], System Usability Scale (SUS) [4], Simulator Sickness Questionnaire (SSQ) [15], and AttrakDiff2 [11].

Furthermore, we developed a custom questionnaire to collect feedback on the different substitutions.

4.4 Procedure

At the beginning of each session, the participants were informed about the overall procedure and were introduced to the setup and task. They were also asked to fill out a demographics questionnaire and the first part of the SSQ. Each participant got some warm-up time to get familiar with the robot and grasping a cup and was introduced to the different tactile feedback substitution modes. In this phase, a different kind of cup than the ones in the experiment was used (see Figure 5 b). Each participant completed 24 trials overall, the combination of eight conditions (*No*, *C*, *S*, *T*, *A*, *C+A*, *S+A*, *T+A*) and three cups were shuffled. After 8 and 16 trials, the participants could take a break. Finally, the participants were asked to fill out the remaining questionnaire forms. One session took about 60 minutes and the wearing time of the HMD was approximately 45 minutes.

4.5 Participants

For the pilot study, we recruited 5 participants from the staff of our working groups (1 female, 4 male) aged 23 to 57 ($M = 34.4$, $SD = 13.18$). All of them had normal or corrected-to-normal vision, and none reported a known eye disorder or displacement of equilibrium. Three participants had used an AR headset before, and three reported having experience with hand tracking technology. Four participants had worked with the PR2 robot platform before.

4.6 Results

Performance In Table 1, the average execution times and success rate of all eight conditions are listed. *No* and *T+A* needed the least average execution times with 104 and 102 s, respectively. With *A* and *T*, the participants needed ~111 s. The conditions *S* and *C*, with and without *A*, resulted in the longest average execution times with over 120 s. The success rate is stable over all conditions and ranges from 83.3 % to 91.6 % resulting in an average rate of 87.1 %. Figure 6 shows the average applied force over all cup types for all eight conditions. It is noticeable that the most force was exerted when no feedback was given.
We also measured the difference of the average applied forces between no feedback and any feedback, separated for the three cup types (see Figure 5 b). On cup one, an average force of 5.1 N is exerted, on cup two 4.5 N, and on cup three

Table 1. Average execution time & Success rate

	No	C	S	T	A	C+A	S+A	T+A
time	104.6 s	120.2 s	127.2 s	112.2 s	111.2 s	120.8 s	125.7 s	102.7 s
rate	88.3 %	83.3 %	85 %	88.3 %	86.6 %	85 %	91.6 %	88.3 %

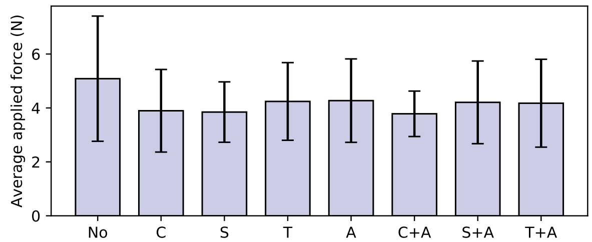


Fig. 6. Average applied force for the individual conditions. *No* is no feedback, *C* is the colored sphere, *S* are the size changing arrows, *T* is the text, and *A* is the audio feedback.

3.0 N. The average saved force is 1.9 N for cup one, 0.44 N for cup two, and 0.82 N for cup three.

Usability Simulator sickness increased in an expected way from 153.38 ($SD=$ 85.8) before the experiment to 228.17 ($SD=171.62$) after the experiment. We found a high SUS mean score of 77 ($SD=13.04$), which can be interpreted to be above average (69.5) [2]. Therefore, the system usability is considered as good.
The results for the NASA TLX are shown in Table 2. Values range from 0-100; higher values refer to a higher task load. The final score consists of six subscales that address different aspects of task complexity. We found especially low values in the overall performance, showing a good user satisfaction and feeling of success.
The results of the AttrakDiff2 questionnaire are summarized in Figure 7. The underlying model of this questionnaire divides attractiveness into a pragmatic quality (PQ), referring to the amount a product enables a person to fulfill a task, and a hedonic quality (HQ) [11]. The HQ describes the degree to which a product stimulates a user or communicates a certain identity. The participants mostly rated the attractiveness of the system positively on both dimensions.
The collected data shows a preference for multimodal tactile substitutions over substitutions using only one modality. Three of the 5 participants liked *S+A* most, while the other two preferred *C+A* over all other conditions. Similarly, *S+A* and *C+A* were stated the most helpful (2 participants each), while only one person found *S* most helpful. The distraction of each tactile substitution mode was relatively low, with mean values of 1.2 (S), 1 (C), 1 (T), and 1.2 (A) (range 0-4). However, each mode was rated distracting to some degree by at least one participant, which suggests personal preference to be an essential factor.

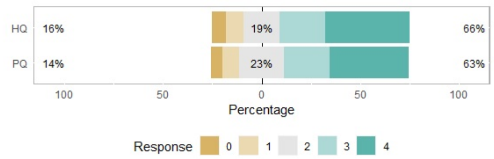


Fig. 7. Percentages for Hedonic Quality (HQ) and Pragmatic Quality (PQ) dimensions of the AttrakDiff2 questionnaire. Responses on 5-point Likert scale.

The participants were asked to rate each visualization against the others on a 5-point Likert scale. We converted the answers into a scoring system, where 0 points were given when no tendency was reported and one and two points were given respectively for each level towards one visualization. C, S, and T scored 14, 10, and 1 point, respectively, showing a clear preference for size and color over text. The participants' agreement on a 5-point Likert scale to the statement "The information from the tactile sensors helped me a lot" was quite mixed, resulting in a mean value of 2.4. Finally, we collected feedback on the difficulty of the different cup types. Three participants perceived cup 1 to be the easiest, while cup 2 was the easiest for the other two.

5 Discussion

5.1 Performance

To answer the first research question (see Section 4), the provided feedback improves the performance, as the average applied force decreases when providing feedback. In contrast to that, we do not observe any difference within the individual feedback conditions. We assume that the participants were already fully occupied by concentrating on the stacking tasks, which means that the differences between the individual modalities and conditions need to be tested with easier tasks. The same applies to the success rate, as it seems that the addition of feedback does not influence it.

Table 2. NASA TLX scores.

	M	SD
Mental Demand	56.7	25.28
Physical Demand	53.4	13.94
Temporal Demand	53.4	18.26
Overall Performance	26.67	9.13
Effort	60.0	19.00
Frustration Level	40.0	22.36
Overall Workload	56.1	11.35

With the three cups, we wanted to test the influence of the object's hardness. As the average applied force is comparable and the force difference for cup 1 (the most rigid cup type) is larger than for the others, we assume that more force is saved when grasping harder objects. Sigrist et al. [24] indicate that too much information can cause perceptual overload in AR. This could explain the comparably long execution time for conditions including S or C. Both augmentations are visualized at the end-effector of the robot. This may attract too much of the user's attention. In contrast to the execution time, none of the two visualizations affect the average applied force compared to the other feedback conditions.

5.2 Usability

The pilot study showed that the proposed system is applicable for teleoperation tasks with regard to the usability of the system and acceptance by non-expert users (see Section 4). The above-average SUS, as well as the high attractiveness ratings, indicate user satisfaction. Although we only recruited participants from our working groups and most of them had a technical background and some experiences with the PR2 platform and AR, none of them was familiar with teleoperating a robotic arm using hand tracking. The different tactile substitution methods were evaluated as helpful, and we found a preference for (redundant) multi- over unimodal methods.

6 Conclusion & Future Work

In this paper, we present an MR human-robot interaction setup, allowing the user to intuitively teleoperate an arbitrary robot arm in real-time. With the pilot user study, we show with the example of a 7-DoF robot arm of a PR2 platform that the system can be used by non-expert users to perform precise manipulation tasks, which is indicated by the success rate of 87.1%. With substituted tactile feedback, the average applied force on grasped objects can be reduced. We assumed more deviation in the performance between the different feedback modalities than we could find with the study. There seems to be a user preference for multimodal substitutions for tactile feedback over unimodal ones. On the other hand, multimodal feedback seems to be less efficient. This contradiction needs to be further investigated. In future studies, we will choose easier tasks to take the users' attention away from the task, allowing for a higher concentration on the provided feedback.

A weak point of the system is the inaccuracy in the registration. This problem does not occur for teleoperation, as the user gets direct position feedback from the real robot and does not rely on any precisely positioned augmentations. Before we can perform collaboration tasks with autonomously moving robots, we need to focus on this problem. Compared to the first version, the new capability of the HoloLens 2 to track full-hand postures opens the potential for more dexterous teleoperated manipulation tasks. This possibility can also be explored in autonomous side-by-side cooperation and interactive, collaborative tasks.

Bibliography

[1] Aviles-Rivero, A.I., Alsaleh, S.M., Philbeck, J., Raventos, S.P., Younes, N., Hahn, J.K., Casals, A.: Sensory substitution for force feedback recovery: A perception experimental study. ACM Transactions on Applied Perception **15**(3), 1–19 (Aug 2018), ISSN 1544-3558, 1544-3965, https://doi.org/10.1145/3176642

[2] Bangor, A., Kortum, P., Miller, J.: Determining what individual sus scores mean: Adding an adjective rating scale. Journal of usability studies **4**(3), 114–123 (2009)

[3] Bischoff, M.: ROS# (Jun 2019), URL https://github.com/siemens/ros-sharp/releases/tag/1.7.0, [Accessed Aug. 2, 2022]

[4] Brooke, J.: SUS-A quick and dirty usability scale. Usability Evaluation In Industry **189**(194), 4–7 (1996)

[5] Burke, J.L., Prewett, M.S., Gray, A.A., Yang, L., Stilson, F.R., Coovert, M.D., Elliot, L.R., Redden, E.: Comparing the Effects of Visual-Auditory and Visual-Tactile Feedback on User Performance: A Meta-analysis. In: ICMI '06: Proceedings of the 8th international conference on Multimodal interfaces, pp. 108–117, Association for Computing Machinery, Banff, Alberta, Canada (2006)

[6] Carter, T., Seah, S.A., Long, B., Drinkwater, B., Subramanian, S.: Ultrahaptics: multi-point mid-air haptic feedback for touch surfaces. In: Proceedings of the 26th annual ACM symposium on User interface software and technology, pp. 505–514 (2013)

[7] Chan, W.P., Quintero, C.P., Pan, M.K., Sakr, M., Van der Loos, H.M., Croft, E.: A multimodal system using augmented reality, gestures, and tactile feedback for robot trajectory programming and execution. In: Proceedings of the ICRA Workshop on Robotics in Virtual Reality, pp. 21–25, IEEE, Brisbane, Australia (2018)

[8] Cooper, N., Milella, F., Pinto, C., Cant, I., White, M., Meyer, G.: The effects of substitute multisensory feedback on task performance and the sense of presence in a virtual reality environment. PloS one **13**(2), e0191846 (2018)

[9] Green, S.A., Billinghurst, M., Chen, X., Chase, J.G.: Human-robot collaboration: A literature review and augmented reality approach in design. International journal of advanced robotic systems **5**(1), 1 (2008)

[10] Hart, S.G., Staveland, L.E.: Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research, vol. 52, p. 139–183. Elsevier (1988), ISBN 978-0-444-70388-0, https://doi.org/10.1016/S0166-4115(08)62386-9

[11] Hassenzahl, M.: The interplay of beauty, goodness, and usability in interactive products. Human–Computer Interaction **19**(4), 319–349 (2004)

[12] Hedayati, H., Walker, M., Szafir, D.: Improving collocated robot teleoperation with augmented reality. In: HRI'18: Proceedings of the 2018

ACM/IEEE International Conference on Human-Robot Interaction, pp. 78–86, Association for Computing Machinery, Chicago, IL, USA (2018)

[13] Herbst, I., Stark, J.: Comparing force magnitudes by means of vibro-tactile, auditory, and visual feedback. In: IEEE International Workshop on Haptic Audio Visual Environments and their Applications, pp. 5–pp, IEEE, Ottawa, ON, Canada (2005)

[14] Kalaitzakis, M., Carroll, S., Ambrosi, A., Whitehead, C., Vitzilaios, N.: Experimental comparison of fiducial markers for pose estimation. In: 2020 International Conference on Unmanned Aircraft Systems (ICUAS), pp. 781–789, IEEE (2020)

[15] Kennedy, R.S., Lane, N.E., Berbaum, K.S., Lilienthal, M.G.: Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. The international journal of aviation psychology **3**(3), 203–220 (1993)

[16] Krogius, M., Haggenmiller, A., Olson, E.: Flexible layouts for fiducial tags. In: 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 1898–1903, IEEE, Macau, China (2019)

[17] Krupke, D., Steinicke, F., Lubos, P., Jonetzko, Y., Görner, M., Zhang, J.: Comparison of multimodal heading and pointing gestures for co-located mixed reality human-robot interaction. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 5003–5009, IEEE, Madrid, Spain (2018)

[18] Microsoft: HoloLens 2—Overview, Features, and Specs — Microsoft HoloLens (Feb 2021), URL https://www.microsoft.com/en-us/hololens/hardware, [Accessed Aug. 2, 2022]

[19] Milgram, P., Kishino, F.: A taxonomy of mixed reality visual displays. IEICE Transactions on Information and Systems **77**(12), 1321–1329 (1994)

[20] Prewett, M.S., Yang, L., Stilson, F.R., Gray, A.A., Coovert, M.D., Burke, J., Redden, E., Elliot, L.R.: The benefits of multimodal information: A meta-analysis comparing visual and visual-tactile feedback. In: ICMI '06: Proceedings of the 8th international conference on Multimodal interfaces, pp. 333–338, Association for Computing Machinery, Banff, Alberta, Canada (2006)

[21] Quigley, M., Conley, K., Gerkey, B., Faust, J., Foote, T., Leibs, J., Wheeler, R., Ng, A.Y.: Ros: an open-source robot operating system. In: ICRA workshop on open source software, vol. 3, p. 5, IEEE, Kobe, Japan (2009)

[22] Romano, J.M., Hsiao, K., Niemeyer, G., Chitta, S., Kuchenbecker, K.J.: Human-inspired robotic grasp control with tactile sensing. IEEE Transactions on Robotics **27**(6), 1067–1079 (2011)

[23] Schorr, S.B., Okamura, A.M.: Fingertip tactile devices for virtual object manipulation and exploration. In: CHI '17: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, pp. 3115–3119, Association for Computing Machinery, Denver, CO, USA (2017)

[24] Sigrist, R., Rauter, G., Riener, R., Wolf, P.: Augmented visual, auditory, haptic, and multimodal feedback in motor learning: A review. Psychonomic Bulletin & Review **20**(1), 21–53 (2013)

[25] Sun, M., Ren, X., Cao, X.: Effects of multimodal error feedback on human performance in steering tasks. Journal of Information Processing **18**, 284–292 (2010), ISSN 1882-6652, https://doi.org/10.2197/ipsjjip.18.284

[26] Tavakoli, M., Patel, R.V., Moallen, M., Aziminejad, A.: Haptics for Teleoperated Surgical Robotic Systems, vol. 1. World Scientific, Singapur (2008)

[27] Unity Technologies: Unity Real-Time Development Platform — 3D, 2D VR & AR Engine (Feb 2021), URL https://unity.com/, [Accessed Aug. 2, 2022]

[28] Wieczorek, F.H.: Universal Teleoperation ROS Interface for Robotic Manipulators. Bachelor's thesis, Universität Hamburg (2020)

[29] Yang, C., Luo, J., Pan, Y., Liu, Z., Su, C.Y.: Personalized variable gain control with tremor attenuation for robot teleoperation. IEEE Transactions on Systems, Man, and Cybernetics: Systems **48**(10), 1759–1770 (2017)

[30] Yew, A., Ong, S., Nee, A.: Immersive augmented reality environment for the teleoperation of maintenance robots. Procedia Cirp **61**, 305–310 (2017)

[31] Zaeh, M.F., Vogl, W.: Interactive laser-projection for programming industrial robots. In: 2006 IEEE/ACM International Symposium on Mixed and Augmented Reality, pp. 125–128, IEEE, Santa Barbara, CA, USA (2006)