2021 18th International Conference on Ubiquitous Robots (UR)
July 12-14, 2021. Gangneung-si, Gangwon-do, Korea

2021 18th International Conference on Ubiquitous Robots (UR) | 978-1-6654-3899-5/21/$31.00 ©2021 IEEE | DOI: 10.1109/UR52253.2021.9494689

# Deploying Multi-Modal Communication Using Augmented Reality in a Shared Workspace

Gabriele Bolano[1], Yuchao Fu[1], Arne Roennau[1], Ruediger Dillmann[1]

***Abstract*— Robots are no longer working isolated in safety fences and Human-Robot Collaboration (HRC) is becoming one of the most promising topic of research to improve the efficiency in many application scenarios. Sharing the same workspace, both human and robot should clearly understand the intentions and motions of each other, in order to enable an efficient and effective interaction. In this work we propose an AR-based system to show the robot planned motion and target to the worker. We focused on representing this information in an intuitive way for inexperienced users. We introduced a multi-modal communication feedback in order to enable the user to agree with or change the robot plan using gestures and speech. The effectiveness of the system has been evaluated with test cases performed by a group of testers with no robotic experience. The results showed that the system helped the user to better understand the robot intentions and planned motion, improving the ergonomics and trust in the interaction. Furthermore, the evaluation included the rating of the different input modalities provided, in order to compare the different ways of communication proposed.**


Fig. 1: The current state of the workspace is visualized in AR. The information about the parts is displayed with markers and text as well as using synthesized speech. The user can interact with the robot and change its current plan using speech or gestures-based commands.

## I. Introduction

In the early years of robotics, robots were forced to work in cells surrounded by fences. However, the actual trend is to make them able to work closely with humans to exploit the skills of both. This has been made possible by the scientific progress in the field of collaborative robots, which are intrinsically safe and allow a close interaction between workers and manipulators. Furthermore, it is really important to make the robot able to perceive its surroundings and take clever decisions based on the state of the workspace. For example, many works in the last years have focused on making the robots able to react to dynamic obstacles, such as humans moving nearby. This is important for safety reasons and also for efficiency, enabling the detection of collision free paths in real time. In order to achieve this, a crucial point is to make them able to predict the human's motion in the same way we do in everyday life.

Anyway, to achieve an ergonomic and efficient collaboration, it is crucial that both participants have a clear understanding of the intentions of the other member. Robots used in manufactory, such as manipulators, don't have a human-like appearance and usually perform their task providing limited information to the user. Natural cues cannot be used by such robots, so it is important to provide an explicit way to communicate their intentions in an intuitive way.

Augmented Reality (AR) is a promising technology to display this information to the user, enabling the visualization of virtual objects and text at the needed position in the real world. The planned robot trajectory can be displayed beforehand to the user, making him able to understand the volume of the workspace which will be occupied by the robot in the near future.

In this work, we propose an AR system to display the swept volume of the planned robot trajectory. In this way the user can understand exactly the volume occupation of the robot planned trajectory, with a consequent reduction of anxiety, which is usually due to a lack of information about the robot intentions. In this way, the ergonomics in the collaboration is improved and the worker can also take better decisions on how to move in the workspace without making the robot change its motion or target. The AR visualization is also used to display information about the state of the parts in the workspace, highlighting the robot target and the workpieces that still need to be worked by the robot. Once this information is represented, the user can also provide a feedback to the robot, in order to confirm or reject the current plan. In this paper we implemented an approach based on gesture and speech commands, with an evaluation of the communication method preferred by a group of test users.

We evaluated the system proposed through a user study performed by 12 people with no robotic experience. The main functionalities and usefulness of the system were rated through a questionnaire.

The structure of this paper is as follows. In Section II, we present the related work on human-robot interaction and communication. In Section III, we describe the system proposed to intuitively represent the robot planned motion

[1]The authors are with FZI Research Center for Information Technology, Haid-und-Neu-Str. 10-14, 76131 Karlsruhe, Germany {bolano, fu, roennau, dillmann}@fzi.de

using AR. In Section IV, we describe the user experiments conducted to evaluate the system. Finally, we provide conclusions and perspectives in Section V.

## II. Related Work

Collaborative robots are intrinsically safe and can work closely to humans in a shared workspace. Safety skins, for example, provide a way to make them stop when getting in contact with an obstacle, with reaction times that allow a safe collaboration and avoid the worker to get injured. Anyway, deploying such methods, the robot is not able to predict a possible collision beforehand and it must stop and wait for clearance. In order to provide a more efficient interaction, it is crucial to make the robot able to predict possible collisions [1], in order to select collision free paths beforehand, in the same way as humans do, without the need to stop when a trajectory cannot be performed because of an obstacle. To make this prediction useful, many approaches are using the current live environment information coming from camera sensors in order to avoid collisions with objects. Furthermore, the collision prediction can also improve the efficiency and ergonomics in the collaboration. These collision checks are computationally expensive and the GPU-Voxels library provide a solution using parallelized algorithm on GPUs in order to make collision prediction in real time and online replanning of the robot motion [2], [3]. However, despite the efforts to make the robot able to dynamically react to changes in the environment and predict the movement of the humans around it, not so many works have focused on the critical issue of making the robot motion easily understandable and transparent to the user.

Research has investigated the problem of representing trajectories for mobile platforms, projecting on the ground the direction of their planned path [4], [5]. In a previous work we studied how to represent the motion of a robotic manipulator using visual information on screen and audio feedback [6]. Anyway, this method has the drawback to draw the user's attention to external devices, making him lose the focus on the task. A further work on the system, explored the use of projector-based AR to display information about the workpieces and the projection of the planned robot motion on the table [7]. This was made to make the collision free areas easily understandable by the human, visualizing the information at the needed position in the workspace. This approach has the disadvantage to lose part of the information about the robot motion, since it allows only the representation of 2D projections. The current developments in Augmented Reality, in particular regarding head-mounted display (HMD), enable the representation of 3D virtual models in the real world, in order to make the information displayed as intuitive as possible [8], [9], [10]. Quintero et al. have proposed an AR interface to ease the robot programming [11]. The authors used the Microsoft Hololens headset to enable the definition of the robot trajectory and execution parameters. In the work from Walker et al., it is highlighted the problem of providing intention information for robots which do not have anthropomorphic features [12]. The authors explored how to represent this information for aerial robots using AR. The proposed designs implemented to display the robot intents were evaluated in a user study, which showed that the AR visualization improved the task efficiency and the user perception of the robot as a teammate.

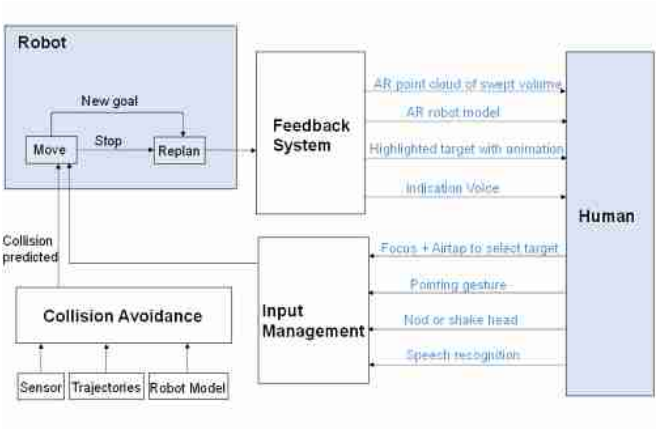



Fig. 2: The overall architecture of the system. In this work we focus on the feedback system which is responsible to provide the user with information about the robot motion and target. Once this information is represented to the user, he is able to accept or change the current robot plan, using different types of inputs which are processed by the input management component.

Regarding effective and intuitive interfaces to convey the user's intentions, gestures and speech are the natural methods used by humans [13], [14], [15]. The research has focused on providing flexible and intuitive interfaces through the use of speech, hand gestures [16] and head gestures [17], [18], [19].

## III. Approach

In this work we developed a system to enable the communication of robot intentions, displaying its motion to the user using AR. For a more complete information about the workspace and the robot target, we added also the representation of the available workpieces and information about their state. We used a system with a robot able to dynamically replan its motion, using the GPU-Voxels library. This library allows fast computation of collisions on high detailed voxels map, in order to provide real time collision predictions and online motion replanning. The live environment occupancy information is captured by 3D cameras positioned around the shared workspace. In our application scenario, an additional camera is positioned on top of the workspace in order to detect the available parts, providing the target position needed by the robot to perform a screwing task.

In Fig. 2 is represented the overall system architecture. The focus of this work is the implementation of the feedback system which is responsible to represent to the user information about the robot motion and target. Furthermore, we developed the input management module, which provides a command interface to collect the input feedback from the user on the current robot plan, in order to change or agree with it.

Fig. 3: The point cloud representing the swept volume of the planned robot motion. This is used to predict collisions with the live environment and it is an useful information to understand the volume that the robot will occupy in the near future.


Fig. 4: The collision points are visualized in AR with a red colored point cloud. These points are the voxels of the trajectory swept volume in collision with the live environment captured by the depth cameras.

### A. Visualization of Robot Motion and Task Information

In order to provide a feedback about the target configuration of the robot, a virtual model can be adopted to represent this information to the user. As depicted in Fig. 1, the virtual model is superimposed on the real robot, showing the final configuration of the robot motion. However, this information does not give a clear and spatial information about the volume of the workspace which will be occupied by the robot in the execution of its next motion. This is particularly relevant because it is the volume used for the detection of collisions with the live environment, to avoid dynamic obstacles such as humans. The 3D swept volume is an important information to make the human understand the volume that the robot will occupy in the next future. Indeed it represents the volume that the worker should avoid to occupy if not necessary. The representation of this information to the user, enables him to avoid useless changes in the robot motion. Furthermore, understanding the intent of the robot, he is then able to feel more comfortable in the interaction and take better decisions on how to move and work in the workspace, with a consequent better ergonomics. For the AR overlay representing the robot information, we used the HoloLens headset. The programming of the virtual information has been implemented in Unity [20]. The communication of the HoloLens and ROS has been developed using the open source software library ROS# [21]. The voxel maps representing the swept volume of the robot trajectory are converted into ROS point cloud messages, which can be imported and represented in Unity. In Fig. 3 is represented in blue the robot swept volume for the current planned trajectory. The information about the workpieces has been implemented using 3D markers, which are displayed at the position detected by the camera system. Every part localized is assigned with a unique id, which is displayed to the user in order to make him able to refer easily to it deploying the speech input interface developed. Different colors and animations are used to represent the current state of the workpieces and to highlight the one that is selected as current target by the robot. Fig. 5 shows the representation of this information for two parts available on the worktable. This includes colored markers to highlight the available parts and the current one selected as target by the robot. Text information enables the user to identify the parts in order to reference them using speech-based commands. With this visual information displayed in AR, the user can understand in an intuitive way the current goal of the robot and the volume that it will occupy in the execution of the trajectory to reach it. The collision free space in the working area can be easily detected, making the worker able to avoid obstructing the robot path. Furthermore, this enables a better comfort in the interaction, due to the clear understanding of the complete robot motion.

Once a collision is detected, the collision area is highlighted with the representation of the colliding voxels in red. These are obtained as result of the intersection between the voxels maps representing the live environment and trajectory swept volume. Fig. 4 shows the situation in which the robot detects a possible collision with the user, showing in red the relevant area.

Synthesized speech is also used to provide further robot information, deploying a different channel of communication. This has been included because, if the user has to focus on a part and cannot draw his attention to the robot motion, he can still understand the current plan of the robot and which target it is aiming for. The robot communicates its intention using simple sentences that are played using the built-in speakers in the HoloLens. “Moving to target one/two” is used to communicate that the robot has found a new executable trajectory. “Moving to target one/two, executing” it means that the execution of the planned trajectory has started once the command from the user is received. In case a possible collision is detected, the robot warns the user with “Warning, possible collision detected”, in order to alert him even if he is not looking in the workspace area.

### B. User Input

In order to enable a flexible interaction, it is important to provide also the user with an input interface in order to change or agree with the robot plan displayed in AR. In this work we propose and evaluate different input modalities based on speech and gestures. In order to evaluate these

Fig. 5: The current state of the workspace is displayed in AR. The information about the parts is represented with markers and text. Synthesized speech is used as well, in order to provide information about the robot target even if the user is not able to visually check the virtual overlay.

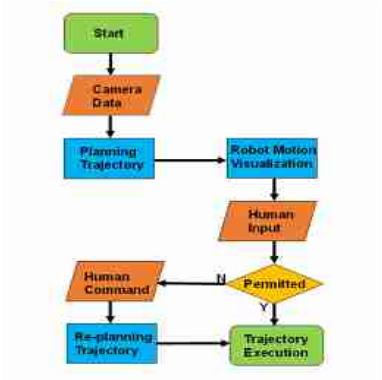



Fig. 6: Flow chart of the communication involved in the interaction system developed. The robot provides information about its planned motion to the human, who can then deny or confirm the execution of the trajectory. If the user denies the robot planned motion, this latter waits for a further human input to select the next target. If the user does not provide any feedback, the robot executes autonomously the planned motion.

methods, when the robot plans a new motion, it shows that to the user as described in the previous section. Before executing the new trajectory, the robot waits for the user input in order to change or agree with the displayed target and planned path. If the user does not provide any input within 2 seconds from the input request, the robot will execute the trajectory displayed in AR. In Fig. 6 is represented the flow chart of the process and the communication involved.

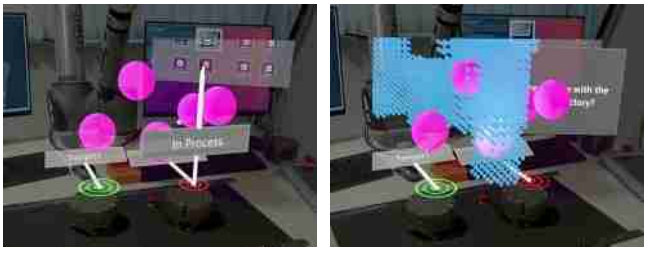
*1) Speech:* In order to change or agree with the robot plan, speech is a natural way of communicate the user's need. The HoloLens headset used to display the robot motion is equipped also with a microphone for voice support. Predefined words and sentences can be recognized and mapped into commands. When the robot creates a new plan, it shows its motion to the user and communicates that using synthesized speech as well. The sentence "Moving to Target One/Two" is used to communicate to the human that the robot found an executable trajectory to a part and waits for the user's confirmation. The user can simply reply with "Yes" or "No" voice commands to confirm or deny the execution of the planned motion. If the planned trajectory is rejected by the human, the robot asks for a target communicating "Select target". The user can then select the desired goal using the input command "Target one/two". Once the command is received and processed, the robot shows the new trajectory and gives a speech-based feedback about the execution of the new trajectory.


Fig. 7: The user performs the pointing gesture input in order to select a target. The direction of the pointing is extracted from the data coming from depth cameras and used to detect the target selected by the user. This is used as target for the robot, which then shows the motion to reach it once a collision free trajectory is found.

*2) Gestures:* We propose the use of different gesture-based inputs to change or agree with the current robot motion or to make the robot replan towards a new goal. The HoloLens headset provides the recognition of predefined gestures such as the Air Tap. A head tracking functionality is also available, which allows the detection of the user's head orientation using IMU and cameras. We developed the following input methods: focus + AirTap, poiting gestures, nod/shake head movement.

The focus + AirTap method uses the basic input mode provided by the HoloLens headset. The human's gaze is tracked in order to get the focus of attention of the user. Once the focus is on a object, the user can select it using the AirTap, which is a predefined gesture which involves the movement of the index finger. However, even if this gesture interface is easy and simple to use, it requires the user to carefully focus his gaze to the desired object and perform the AirTap movement in a way that his detected by the HoloLens. For this reason we proposed also a simpler selection input, which is the pointing gesture. The direction of the pointing is extracted from the point cloud information coming from the 3D cameras around the workspace. The area in front of the user is set to be the area of detection. The points in this area are seen as spatial occupancy of the human arm. The direction of this latter is computed using the least squares method and used to detect which part is selected. Fig. 7 shows a user using the pointing gesture to select a target.

In order to accept or deny the robot plan, we proposed the use of the head tracking to detect nod/shake gestures. The head tracking of the HoloLens is used to get the orientation of the user's head. The nodding and shaking gestures are detected by monitoring the changes in orientation of the head pose, as well as frequency and speed. The parameters to detect these movements have been set after experiments.



(a) Only target. (b) Swept volume.

Fig. 8: Test for the evaluation of the spatial understanding of the robot motion. Virtual obstacles are represented to the users as spheres and need to be judged as colliding or not with the robot motion . In (a) only the target of the robot is visualized. In (b) is represented the same test including the swept volume information.

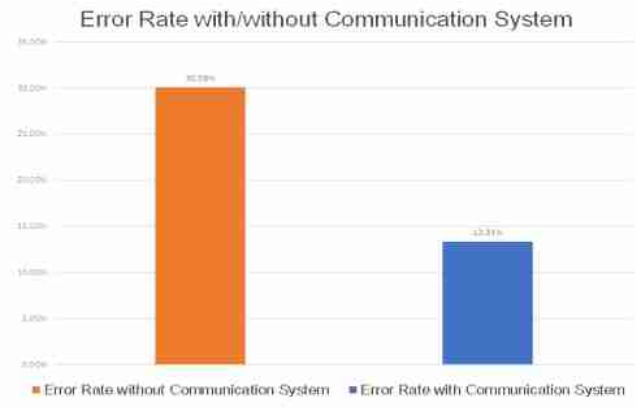



Fig. 9: The results of the test for the evaluation of the spatial understanding of the robot motion. The chart shows that, deploying the swept volume information, the user had a better understanding of which area and obstacles would cause a collision with the planned robot trajectory.

## IV. Evaluation

In order to evaluate the system we deployed a test scenario with a group of 12 users with no experience in robotics and AR. We created two test cases to examine the effectiveness of the system and collect a feedback about the interaction modalities. The first test had the aim to evaluate the spatial understanding of the robot motion displayed to the user in AR. We added virtual obstacles represented as spheres in the AR overlay. The users were asked to state if these objects would cause a collision with the robot planned motion. The test has been performed as first without any information about the robot motion, apart from highlighting its target. After that, the 12 users had to perform the same test with the addition of the trajectory swept volume visualization.

In Fig. 8 is represented the virtual information displayed to the users for this test and Fig. 9 shows the results of the evaluation. As we can observe from the chart, the error rate is lower deploying the visualization of the trajectory swept volume. Without the visual information about the robot motion and its volume occupation, the users had clearly more difficulties in judging the collisions with the virtual obstacles.

In the last test, the overall system and the user perception in the interaction are evaluated. The users had to fill out a questionnaire about the general experience in the interaction

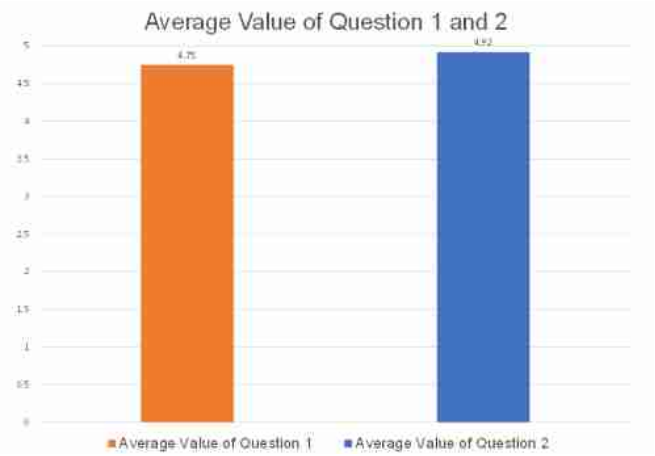



Fig. 10: The results of the questionnaire regarding the understanding of the robot intentions.

with the system. This includes the functionality, accuracy, difficulty to learn and use it. The users had to test the interaction with the robot without any feedback information and then using our communication system. The different inputs method were briefly explained and performed by every tester. At the end of all the tests, the users had to rate 10 sentences using a 5-point Likert scale from “Strongly Disagree” to “Strongly Agree”:

1) The system improved your understanding of the spatial occupancy of the planned motion of the robot.
2) The system improved your understanding of the next robot goal.
3) The head movement input is effective and convenient to command the robot.
4) The speech recognition input is effective and convenient to command the robot.
5) The pointing gesture input is effective and convenient to command the robot.
6) The focus and AirTap gesture input is effective and convenient to command the robot.
7) The system improved your working efficiency by chang-

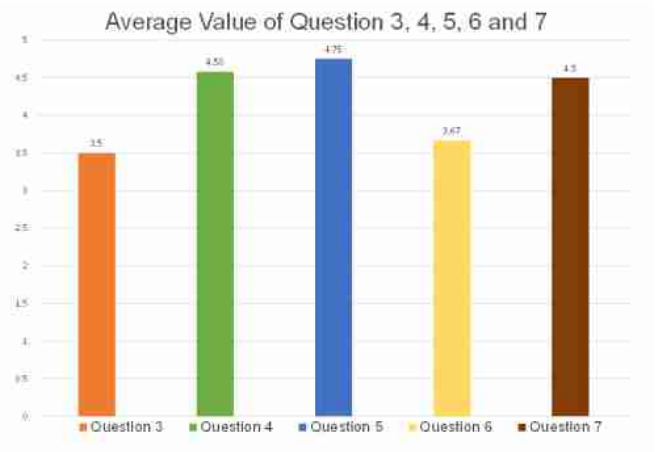



Fig. 11: The results of the questionnaire regarding the input interfaces to command the robot.

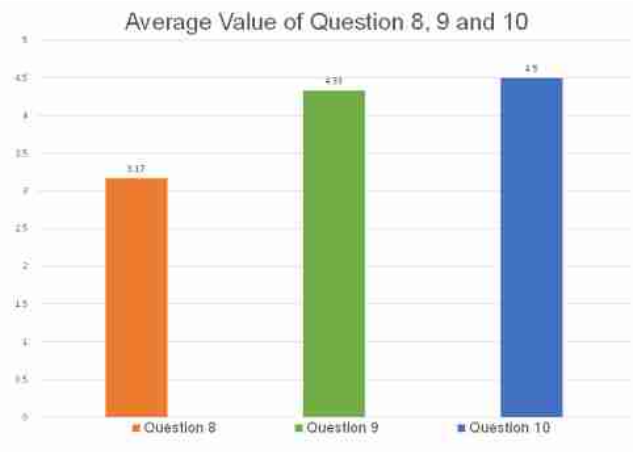



Fig. 12: The results of the questionnaire regarding the trust towards the robot.

ing the robot goal by voice and gestures.

8) It is safe to work in collaboration with the robot without the communication system.
9) It is safe to work in collaboration with the robot with the communication system.
10) The system improved your feeling of safety working along with the robot.

Sentences 1-2 regard the understanding of the robot intentions. The results reported in Fig. 10 show that the users found that our system improved the understanding of spatial occupancy of the planned robot motion and the robot goals.

Senteces 3-7 concern the input interfaces provided to the user in order to deny or agree with the robot motion. As reported in Fig. 11, the preferred methods were the pointing gestures and the speech-based interface, which the users found very intuitive and effective.

Sentences 8-10 deal with the trust of the users towards the robot. The chart in Fig. 12 shows that the users had an improved feeling of safety and comfort in the interaction with the robot using the communication system developed.

## V. Conclusions

The system proposed in this work aims to improve the interaction with a robot in close HRC. The evaluation tests performed with unexperienced users showed that the system improved the understanding of the robot motion in order to make the human more comfortable in the interaction. The use of AR to show the robot planned motion, proved to be an intuitive and effective way to represent the robot plan to the user. The system included also a survey about the feedback input provided by the human using gestures and speech. Regarding the proposed methods to agree or force a change in the robot plan, the evaluation questionnaire showed that the users found the pointing gestures and the speech-based interface as very intuitive and effective methods to give commands to the robot.

The system developed could be improved with the future enhancements in the AR technology, in order to provide a better field of view for the visualization of the virtual overlay. The communication of the robot plan could be further developed including the use of haptics or EMG based interfaces. The use of localization of source and source tracking for the speech commands could also improve the use of the system in manufactoring scenarios with multiple users and noisy environments.

## References

[1] A. Hermann, F. Mauch, K. Fischnaller, S. Klemm, A. Roennau, and R. Dillmann, "Anticipate your surroundings: Predictive collision detection between dynamic obstacles and planned robot trajectories on the GPU," in *2015 European Conference on Mobile Robots, ECMR 2015 - Proceedings*, 2015.

[2] C. Jülg and A. Hermann, "Fast online collision avoidance for mobile service robots through potential fields on 3D environment data processed on GPUs," no. February, 2016.

[3] C. Juelg, A. Hermann, A. Roennau, and R. Dillmann, "Efficient, Collaborative Screw Assembly in a Shared Workspace," in *Intelligent Autonomous Systems 15*. Springer International Publishing, 2018, pp. 837–848.

[4] R. T. Chadalavada, H. Andreasson, R. Krug, and A. J. Lilienthal, "That ' s on my Mind ! Robot to Human Intention Communication through on-board Projection on Shared Floor Space," in *2015 European Conference on Mobile Robots (ECMR)*, no. 1. IEEE, 2015, pp. 1–6.

[5] E. Bunz, R. Chadalavada, H. Andreasson, R. Krug, M. Schindler, and A. J. Lilienthal, "Spatial Augmented Reality and Eye Tracking for Evaluating Human Robot Interaction," 2016.

[6] G. Bolano, A. Roennau, and R. Dillmann, "Transparent Robot Behavior by Adding Intuitive Visual and Acoustic Feedback to Motion Replanning," in *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 2018, pp. 1075–1080.

[7] G. Bolano, C. Juelg, A. Roennau, and R. Dillmann, "Transparent Robot Behavior Using Augmented Reality in Close Human-Robot Interaction," in *2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 2019, pp. 1–7.

[8] J. Guhl and S. T. Nguyen, "Concept and Architecture for Programming Industrial Robots using Augmented Reality with Mobile Devices like Microsoft HoloLens," pp. 2–5, 2017.

[9] M. Ostanin, R. Yagfarov, and A. Klimchik, "Interactive robots control using mixed reality," *IFAC-PapersOnLine*, vol. 52, no. 13, pp. 695 – 700, 2019, 9th IFAC Conference on Manufacturing Modelling, Management and Control MIM 2019.

[10] M. Ostanin and A. Klimchik, "Interactive robot programming using mixed reality," *IFAC-PapersOnLine*, vol. 51, no. 22, pp. 50 – 55, 2018, 12th IFAC Symposium on Robot Control SYROCO 2018.

[11] C. P. Quintero, S. Li, M. K. X. J. Pan, W. P. Chan, H. F. M. V. D. Loos, and E. Croft, "Robot Programming Through Augmented Trajectories in Augmented Reality," pp. 1838–1844, 2018.

[12] M. Walker, J. Lee, and D. Szafir, "Communicating Robot Motion Intent with Augmented Reality," pp. 316–324, 2018.

[13] K. Nickel, "3D-Tracking of Head and Hands for Pointing Gesture Recognition in a Human-Robot Interaction Scenario," 2004.

[14] S. Fujie, Y. Ejiri, K. Nakajima, Y. Matsusaka, and T. Kobayashi, "A Conversation Robot Using Head Gesture Recognition as Para-Linguistic Information," pp. 159–164, 2004.

[15] D. Yongda, L. Fang, and X. Huang, "Research on multimodal human-robot interaction based on speech and gesture," *Computers & Electrical Engineering*, vol. 72, pp. 443–454, 2018.

[16] G. Bolano, A. Tanev, L. Steffen, A. Roennau, and R. Dillmann, "Towards a Vision-Based Concept for Gesture Control of a Robot Providing Visual Feedback," in *2018 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, 2018, pp. 386–392.

[17] N. Rudigkeit, M. Gebhard, and A. Gr, "An Analytical Approach for Head Gesture Recognition with Motion Sensors," pp. 720–725, 2015.

[18] C. Road, "Head Gesture Recognition for Hands-free Control of an Intelligent Wheelchair," pp. 1–10, 2001.

[19] A. Jackowski, M. Gebhard, and R. Thietje, "Head Motion and Head Gesture-Based Robot Control : A Usability Study," vol. 26, no. 1, pp. 161–170, 2018.

[20] Unity, 2019, https://unity.com.

[21] M. Bischoff, "ROS#," 2019, https://github.com/siemens/ros-sharp/releases/tag/v1.5.