

A Multimodal Teach-in Approach to the Pick-and-Place Problem in Human-Robot Collaboration

Niko Kleer

DFKI

Saarland Informatics Campus

Saarbrücken, Germany

niko.kleer@dfki.de

Maurice Rekrut

DFKI

Saarland Informatics Campus

Saarbrücken, Germany

maurice.rekrut@dfki.de

Julian Wolter

DFKI

Saarland Informatics Campus

Saarbrücken, Germany

julian.wolter@dfki.de

Tim Schwartz

DFKI

Saarland Informatics Campus

Saarbrücken, Germany

tim.schwartz@dfki.de

Michael Feld

DFKI

Saarland Informatics Campus

Saarbrücken, Germany

michael.feld@dfki.de

ABSTRACT

Teaching robotic systems how to carry out a task in a collaborative environment still presents a challenge. This is because replicating natural human-to-human interaction requires the availability of interaction modalities that allow conveying complex information. Speech, gestures, gaze-based interactions as well as directly guiding a robotic system count towards such modalities that yield the potential to enable smooth multimodal human-robot interaction. This paper presents a conceptual approach for multimodally teaching a robotic system how to pick-and-place an object, one of the fundamental tasks not only in robotics, but in everyday life. By establishing task and dialogue model separately, we aim to split robot/task logic from interaction logic and to achieve modality independence for the teaching interaction. Finally, we elaborate on an experimental implementation of our models for multimodally teaching a UR-10 robot arm how to pick-and-place an object.

CCS CONCEPTS

• **Human-centered computing** → **Interaction techniques**; *Interaction paradigms*; HCI theory, concepts and models.

KEYWORDS

Human-Robot Collaboration, Human-Robot Interaction, Multimodal Interaction, Interaction Techniques, Pick-and-Place

ACM Reference Format:

Niko Kleer, Maurice Rekrut, Julian Wolter, Tim Schwartz, and Michael Feld. 2023. A Multimodal Teach-in Approach to the Pick-and-Place Problem in Human-Robot Collaboration. In *Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction (HRI '23 Companion)*, March 13–16, 2023, Stockholm, Sweden. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/3568294.3580047>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, and publish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
HRI '23 Companion, March 13–16, 2023, Stockholm, Sweden

© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-9970-8/23/03...\$15.00
<https://doi.org/10.1145/3568294.3580047>

1 INTRODUCTION

Enabling efficient, safe, and natural human-robot collaboration (HRC) has long been an ambitious aim among researchers. Factors such as the robot type, its capabilities, the environment, and the task significantly influence the degree to which a human may collaborate with a robotic system. Therefore, there exist extensive bodies of research that focus on developing methods that allow robotic systems to assist humans during tasks [4, 6, 9, 16, 17, 19].

One task that is of particular importance to humans is to physically relocate an object from an origin to a desired destination. This process is also called pick-and-place. Picking-and-placing objects of arbitrary shape or form is not only a necessity in professional environments such as the industrial field but in a human's everyday life as well. It is due to our dexterous hands and cognitive abilities that we are capable of grasping everyday objects barely facing any challenges. On the other hand, robotic systems limited to their kinematic capabilities and usually rely on a considerable amount of data before gaining the ability to pick-and-place objects. This data may include position and shape of the object or environmental conditions that require collision avoidance procedures while carrying the object. Traditionally, robotic systems are programmed to deal with one clearly specified pick-and-place task. While this approach is suitable in static environments where the system always operates under the same conditions, it lowers the degree to which a human may collaboratively interact with it. To provide robotic systems with the ability to carry out a task in a collaborative manner, they can also be taught interactively [18–20, 23]. This approach enables dealing with constantly changing requirements.

Natural human-robot interaction represents a key aspect in creating a cyber-physical environment that allows humans to smoothly collaborate with robots. This can be achieved by giving humans the opportunity to interact with such systems similar to how they would interact with another human. During such interactions, humans use their voice, hands (e.g. finger pointing) or even gaze to articulate information. Multimodal interaction is not only natural but also enables humans to adapt to a situation at hand by using the modality most suitable for communicating information. However, enabling the use of multiple modalities in the context of robotics commonly requires additional effort in designing and implementing human-machine interaction.

This paper contributes an almost entirely modality independent pick-and-place **task model** that serves as a theoretical framework for human-robot collaborative pick-and-place tasks. In parallel, we design a **dialogue model** for the user to multimodally fill in and control the former to teach the robot task in a natural way.

2 RELATED WORK

In the 1980s, Lozano-Pérez et al. [13] presented HANDEY, a robotic system dedicated towards carrying out pick-and-place operations. The authors describe that the core of a pick-and-place task can be divided into four phases. Given an object that is supposed to be pick-and-placed, a robot needs a suitable grasp, a motion for approaching and grasping the object, a motion towards the destination as well as a motion for opening and retreating the gripper. Based on the pick-and-place robot HANDEY, Jones and Lozano-Pérez [8] also elaborate on the complexity and challenges of grasping. In contrast to HANDEY, our model expands on the definition of a pick-and-place task and is tailored to multimodal human-robot interaction.

To create an environment where humans and robots are able to collaboratively interact with each other, researchers have investigated the challenges that arise during shared workspace-based pick-and-place tasks [1, 6, 9]. Furthermore, the use and potential of modalities such as gestures [2, 9, 12, 19, 20, 22, 23], gaze [17, 24], speech [10, 11, 16, 21, 23], physically guiding a robot [6], and brain computer interfaces (BCIs) [5, 25] has been explored. In some cases, approaches already make use of multiple modalities [7, 11, 23, 24].

To our knowledge, there does not currently exist a generalized definition of the pick-and-place task aimed towards enabling multimodal interaction from arbitrary input modalities. This is because most works explore the potential of one specific modality. The pick-and-place task model we propose is supposed to provide a modality-independent foundation, which enables multimodal interactions with a robot. In contrast to single-modality solutions, multimodality enables users to freely choose or combine modalities for a given task and does not force the use of one concrete modality.

3 MODEL DEFINITIONS

For teaching a robot how to pick-and-place an object, we require a model that describes the task from an abstract perspective, a task model. Additionally, we also need to develop an understanding for how a human may interact with the system to teach the task, a dialogue model. Following, we describe both of these models.

3.1 Pick-and-Place Task Model

Inspired by the literature and after thorough discussions, we determined the following parameters to be significant in order to define a model for a broader scope of pick-and-place tasks.

- The **Object** that is supposed to be grasped.
- The **Pregrasp**, a position close to the object.
- The **Grasp**, a position in which the object can be grasped.
- The **Force** needed for closing the end-effector.
- The **Postgrasp**, a position into which the object was moved.
- The **Destination** for releasing the object.
- The **Release** indicator, determining whether the object may be dropped or must be placed down carefully.

Specifically, our task model depends on the set of input parameters

$$I = \{O, Pre, G, F, Post, D, R\}.$$

We define a pick-and-place task as a sequence of states S_1, \dots, S_n where S_1 represents the initial and S_n the final state after releasing the object. During our attempts to define the state of a robotic system, we noticed that establishing a model that does not distinguish between the state of the robot's arm, end-effector, and gained knowledge about processed parameters may be confusing. Because of this, we decided to define a state as a triple $S = (S_a, S_g, S_m)$ encompassing the state of the robot's arm, gripper and internal memory respectively. Figure 1 shows the visual representation of a state. S_a represents the state of the robot's arm depending on a

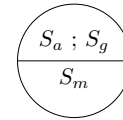


Figure 1: Pick-and-place task state S comprising the state of the robot's arm S_a , gripper S_g , and internal memory S_m .

parameter provided by a human. Initially, this position is defined as P_1 as the arm could be in an arbitrary configuration. Moreover, the position in which the object is released represents the final position P_r . The gripper state S_g indicates whether the robot's end-effector is Open or Closed. Finally, the memory state S_m represents the set of parameters that have already been processed. The requirements of these states are defined by Equation (1), (2) and (3) respectively.

$$S_a = \{a_1, \dots, a_n \mid a_1 = P_1 \wedge a_n = P_r\} \quad (1)$$

$$S_g = \{g_1, \dots, g_n \mid g_1 = O \wedge g_i = O \vee C \forall i > 1\} \quad (2)$$

$$S_m = \{\{m_1, \dots, m_n\} \mid m_1 = \emptyset \wedge m_n = I\} \quad (3)$$

Next, it is important to clarify how the model transitions into another state. Transitions occur when the human interacts with the robot. Naturally, the system needs one of our defined parameters $i \in I$ for continuously progressing with the task. However, a human should not be restricted by this requirement and still have the opportunity to provide an **arbitrary number of parameters at any point during the task**. As a result, we must distinguish between different ways that a human may choose to interact with the robotic system. Specifically, a human may decide to provide an immediately required parameter $p \in I$ or not, leading to different transition functions. In both cases, an arbitrary number of additional parameters $i \in I$ can still be provided. For the sake of simplicity, we formally refer to a set of arbitrary parameters $i \in I$ as $i^k = \{i_1, \dots, i_k\}$. For $k = 0$, $i^0 = \emptyset$. In case a human decides to pass parameters while not including an immediately required parameter $p \in I$, our transition function only causes the robotic system's memory to expand by the parameters provided, as defined by the subsequent transition function.

$$f(i^k \setminus p) \rightarrow S_{m+1} = S_m \cup i_1 \cup \dots \cup i_k.$$

On the other hand, we cannot provide a general transition function in case the human provides an immediately required parameter since the state transition sometimes results in a different outcome as subsequently described in greater detail.

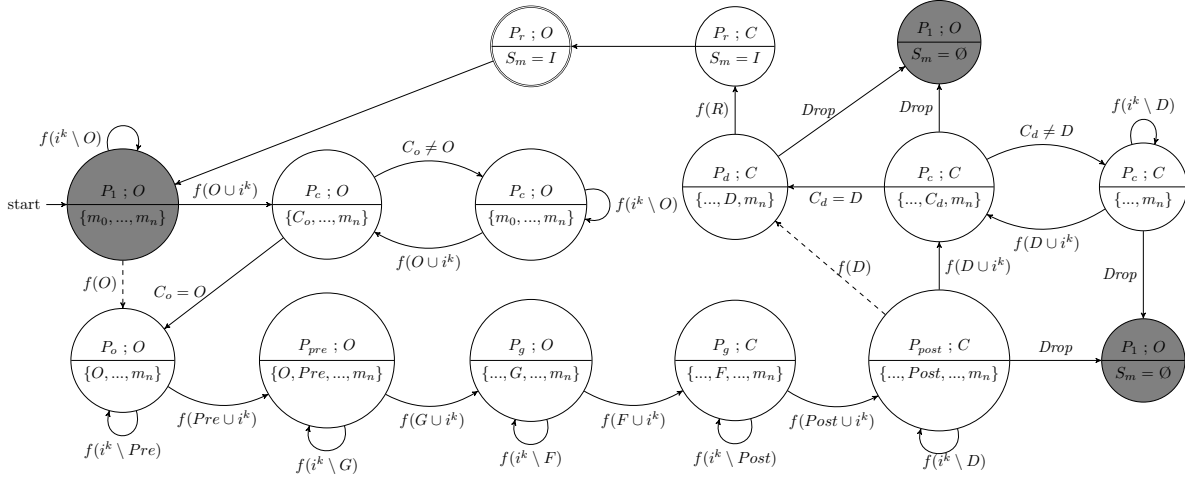


Figure 2: Our pick-and-place task model. The initial and “bad” states have a dark background. All edges, except dashed ones, indicate a modality independent interaction. Dashed edges imply guiding a robotic system into the desired position while circumventing other states. Edges labeled with the term “Drop” indicate that the object was unintentionally dropped.

Object and Destination: Robotic systems sometimes need to autonomously determine which object to pick in a multi-object or cluttered scene. However, when multimodally interacting with the system, the robot might mistakenly approach the wrong object. A pick-and-place model that allows such scenarios should incorporate a confirmation procedure where the approached candidate position C_o is either accepted or rejected. If the candidate position matches the human’s intention, the robot has successfully reached the position P_o . The same concept applies when passing the Destination parameter to the system. Note that a human may still include an arbitrary number of other parameters. As a result, the transition functions for these parameters are defined as following.

$$f(O \cup i^k) \rightarrow \begin{cases} S_{a+1} = P_o \wedge S_{m+1} = S_m \cup O \cup i^k & \text{if } C_o = O \\ S_{a+1} = P_c \wedge S_{m+1} = S_m \setminus C_o \cup i^k & \text{else} \end{cases}$$

$$f(D \cup i^k) \rightarrow \begin{cases} S_{a+1} = P_d \wedge S_{m+1} = S_m \cup D \cup i^k & \text{if } C_d = D \\ S_{a+1} = P_c \wedge S_{m+1} = S_m \setminus C_d \cup i^k & \text{else} \end{cases}$$

Force: As the Force parameter is related to a robot’s end-effector, we require a different transition function. When passing this parameter, the system’s end-effector closes. This is why our transition function for the force parameter is defined as

$$f(F \cup i^k) \rightarrow S_{g+1} = C \wedge S_{m+1} = S_m \cup F \cup i^k.$$

Note that we assume the system to re-open its end-effector after the Release parameter is specified. Finally, in all the other cases, it suffices to define passing an immediately required parameter $p \in I$ and an arbitrary number of additional parameters $i \in I$ as

$$f(p \in \{Pre, G, Post, R\} \cup i^k) \rightarrow S_{a+1} = P_p \wedge S_{m+1} = S_m \cup i^k.$$

It is important to emphasize that these definitions are completely **modality independent**. However, there is one modality that sometimes allows a human to circumvent states in our model, which is by physically guiding the arm. Whenever a human decides to guide the arm into the desired position, there is no specific need for a

confirmation procedure. Therefore, we consider this interaction a special state transition based on the transition function

$$f(p \in \{O, D\}) \rightarrow S_{a+1} = P_p \wedge S_{m+1} = S_m \cup p$$

where no additional parameters are passed. Figure 2 shows a full cycle of our final pick-and-place task model.

3.2 Deriving a Dialogue Model

Based on our task model, we aim to define a dialogue model that allows a human to multimodally teach a robotic system how to pick-and-place an object. For this purpose, one could use an exact reproduction of the task model. However, by focusing on types of interactions that are relevant to the human rather than the exact task-specific requirements in each state, we may significantly lower the total number of dialogue states. On the other hand, the number can also increase due to providing additional ways for interacting with the system. As a result, these dialogue states serve as a natural interface to the task model where state transitions can be triggered through multimodal inputs. This also means that task model specific requirements stay hidden, and a human does not need any knowledge about them. Examining our pick-and-place task model, there exist two types of interactions that can be translated into a general dialogue state, teaching parameters to the system and providing a confirmation. The only other dialogue state that represents an exact match according to the task model is the task initialization. Consequently, our dialogue model can be broken down to the following three general dialogue states.

- (1) **Execute:** A dialogue state in which all task-related requirements are currently satisfied.
- (2) **Request:** This dialogue state adapts to the current situation by querying a required parameter. Furthermore, due to the dynamic nature of our model, this state enables a human to teach an arbitrary number of parameters.
- (3) **Confirm:** Another adaptive dialogue state that generates an utterance matching a human’s previous interaction.

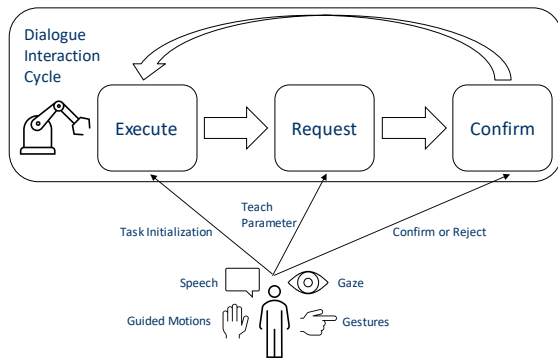


Figure 3: Derived dialogue model representation.

Figure 3 provides an overview of the dialogue model. The dialogue states we have derived construct an interaction cycle where a human may choose an arbitrary modality for progressing in the task. The only limitation to this modality independence lies in what information can be communicated using certain modalities.

4 EXPERIMENTAL SYSTEM

We have implemented an experimental pick-and-place application on a UR-10 robot based on our introduced task and dialogue model (see Figure 4). The task model is implemented into a task-planning software that requires the task description in the JSON format to generate a JSON object-based state chart. By utilizing the MQTT protocol, it further acts as an interface between the UR-10 robot arm and our dialogue model implementation. To this end, we utilize the dialogue platform stepDP. StepDP, the next iteration of SiAM-dp [15], is a domain-independent dialogue platform that allows modelling complex interactions between humans and machines. The system incorporates a classification scheme that maps interactions to a set of rules. A human making the statement

“robot, could you pick this and place it inside the box”

while pointing towards an object corresponding to “this”, could be mapped to a rule that teaches the robot about the **Object** and the **Destination**. In this case, the dialogue platform combines the pointing gesture and speech using a mechanism called multimodal fusion. Our system captures pointing gestures via the Microsoft HoloLens 2 [14]. An index finger pointing gesture with a virtual beam is implemented. For recognizing valid parameters, they were integrated as so-called digital twins. Speech interaction is enabled through the Cerence Studio [3] Natural Language Understanding module. The language model follows an intent-based classification scheme where each intent can be used for firing rules in our dialogue model. Furthermore, the module enables us to define key terms that need to be extracted (e.g. the **Object** or the **Destination**). For simplification purposes, our final dialogue model implementation does not query the **Pregrasp**, **Force**, and **Postgrasp** as individual parameters. As we found these parameters particularly difficult to teach, they are directly embedded into the **Grasp** parameter instead. Some observations that we made during our preliminary experiments with the system are discussed in the next section.



Figure 4: Experimental pick-and-place application.

5 DISCUSSION

There exist several limitations and noteworthy aspects that we need to discuss. Our implementation relies on processing pointing gestures using the Microsoft HoloLens 2, which introduces a dependence on the device that might need to be compensated. We also noticed that, due to the size of the virtual beam, a human has to be precise when pointing towards the digital twin. Additionally, our approach suffers from a scalability issue as we would need a digital twin for each newly introduced valid parameter that a human may point towards. For future implementations in an intelligent environment, one could foresee an image recognition-based approach for detecting such parameters instead. Furthermore, when instructing the robot using speech, the language model we use needs to match the instruction to an intent, meaning that it might need to cover a large vocabulary. Another limitation of our approach is that one cannot currently teach the robot different grasping motions as the grasp applied is based on a parameter’s respective descriptor. This limitation can only be overcome by physically guiding the arm. On the other hand, stepDP’s modality fusion has not caused any issues during our experiments, always enabling the choice of an arbitrary modality for teaching the robot new parameters.

6 CONCLUSION

This paper describes an approach for multimodally teaching a robotic system a pick-and-place task. We elaborated on our nearly modality independent pick-and-place task model and the dialogue model we derive for interacting with the system. Furthermore, we provided an overview of our experimental implementation that incorporates the use of multimodal inputs for interacting with a UR-10 robot. Finally, we discussed several aspects in relation to our preliminary experiments. In order to establish an even more dynamic teach-in procedure, we plan on investigating how multimodal inputs can be used for teaching a robotic system how to pick-and-place different objects. We also aim to put a stronger focus on the aspect of grasping, which turns out to be challenging due to object-based (e.g. shape, size and texture) and environmental constraints (e.g. the surface an object is placed on).

ACKNOWLEDGMENTS

This work is supported by the German Federal Ministry of Education and Research (grant no. 01IW20008) as a part of CAMELOT.

REFERENCES

- [1] Jimmy Baraglia, Maya Cakmak, Yukie Nagai, Rajesh Rao, and Minoru Asada. 2016. Initiative in robot assistance during collaborative task execution. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, Christchurch, New Zealand, 67–74. <https://doi.org/10.1109/HRI.2016.7451735>
- [2] Nico Blodow, Zoltan-Csaba Marton, Dejan Pangercic, Thomas Ruhr, Moritz Tenorth, and Michael Beetz. 2011. Inferring generalized pick-and-place tasks from pointing gestures. *IEEE International Conference on Robotics and Automation (ICRA), Workshop on Semantic Perception, Mapping and Exploration (2011)*, 6.
- [3] Cerence. 2022. *Cerence Studio*. <https://developer.cerence.com/landing/>
- [4] Anthony Cowley, Benjamin Cohen, William Marshall, Camillo J. Taylor, and Maxim Likhachev. 2013. Perception and motion planning for pick-and-place of dynamic objects. In *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, Tokyo, 816–823. <https://doi.org/10.1109/IROS.2013.6696445>
- [5] Angel Perez Garcia, Ingrid Schjølberg, and Serge Gale. 2013. EEG control of an industrial robot manipulator. In *2013 IEEE 4th International Conference on Cognitive Infocommunications (CogInfoCom)*. IEEE, Budapest, Hungary, 39–44. <https://doi.org/10.1109/CogInfoCom.2013.6719280>
- [6] Thorsten Gecks and Dominik Henrich. 2005. Human-robot cooperation: safe pick-and-place operations. In *ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, 2005*. IEEE, Nashville, TN, USA, 549–554. <https://doi.org/10.1109/ROMAN.2005.1513837>
- [7] Enrique Hortal, Eduardo Iáñez, Andrés Úbeda, Carlos Perez-Vidal, and José M. Azorin. 2015. Combining a brain-machine interface and an electrooculography interface to perform pick and place tasks with a robotic arm. *Robotics and Autonomous Systems* 72 (Oct. 2015), 181–188. <https://doi.org/10.1016/j.robot.2015.05.010>
- [8] Joseph L. Jones and Tomás Lozano-Pérez. 1990. Planning two-fingered grasps for pick-and-place operations on polyhedra. In *Proceedings... IEEE International Conference on Robotics and Automation*. IEEE, 683–688. <https://doi.org/10.1109/ROBOT.1990.126063>
- [9] Krishnanand N. Kaipa, Carlos W. Morato, Jiashun Liu, and Satyandra K. Gupta. 2014. Human-robot collaboration for bin-picking tasks to support low-volume assemblies. *Human-Robot Collaboration for Industrial Manufacturing Workshop, held at Robotics: Science and Systems Conference (RSS 2014) (2014)*, 8.
- [10] Wong Sai Kit and Chitturi Venkatratnam. 2016. Pick and place mobile robot for the disabled through voice commands. In *2016 2nd IEEE International Symposium on Robotics and Manufacturing Automation (ROMA)*. IEEE, Ipoh, Malaysia, 1–4. <https://doi.org/10.1109/ROMA.2016.7847799>
- [11] Dennis Krupke, Frank Steinicke, Paul Lubos, Yannick Jonetzko, Michael Görner, and Jianwei Zhang. 2018. Comparison of multimodal heading and pointing gestures for co-located mixed reality human-robot interaction. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, Madrid, Spain, 1–9. <https://doi.org/10.1109/IROS.2018.8594043>
- [12] Hsien-I. Lin and Y. P. Chiang. 2015. Understanding human hand gestures for learning robot pick-and-place tasks. *International Journal of Advanced Robotic Systems* 12, 5 (2015), 49.
- [13] Tomás Lozano-Pérez, Joseph L. Jones, Emmanuel Mazer, and Patrick A. O'Donnell. 1989. Task-level planning of pick-and-place robot motions. *Computer* 22, 3 (1989), 21–29. <https://doi.org/10.1109/2.16222>
- [14] Microsoft. 2022. *Microsoft HoloLens 2*. <https://www.microsoft.com/en-us/hololens/>
- [15] Robert Neßelrath. 2015. SiAM-dp: An open development platform for massively multimodal dialogue systems in cyber-physical environments. (2015).
- [16] J. Norberto Pires. 2005. Robot-by-voice: experiments on commanding an industrial robot using the human voice. *Industrial Robot: An International Journal* 32, 6 (Dec. 2005), 505–511. <https://doi.org/10.1108/01439910510629244>
- [17] Lucas Paletta, Martin Pszeida, Harald Ganster, Ferdinand Fuhrmann, Wolfgang Weiss, Stefan Ladstätter, Amir Dini, Sandra Murg, Harald Mayer, Inka Brijack, et al. 2019. Gaze-based human factors measurements for the evaluation of intuitive human-robot collaboration in real-time. In *2019 24th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*. IEEE, Zaragoza, Spain, 1528–1531. <https://doi.org/10.1109/ETFA.2019.8869270>
- [18] Amit Kumar Pandey, Jean-Philippe Saut, Daniel Sidobre, and Rachid Alami. 2012. Towards planning human-robot interactive manipulation tasks: Task dependent and human oriented autonomous selection of grasp and placement. In *2012 4th IEEE RAS & EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob)*. IEEE, Rome, Italy, 1371–1376. <https://doi.org/10.1109/BioRob.2012.6290776>
- [19] Camilo Perez Quintero, Romeo Tatsambon, Mona Gridseth, and Martin Jagersand. 2015. Visual pointing gestures for bi-directional human robot interaction in a pick-and-place task. In *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, Kobe, Japan, 349–354. <https://doi.org/10.1109/ROMAN.2015.7333604>
- [20] S. Sato and S. Sakane. 2000. A human-robot interface using an interactive hand pointer that projects a mark in the real work space. In *Proceedings 2000 ICRA. Millennium Conference. IEEE International Conference on Robotics and Automation. Symposia Proceedings (Cat. No.00CH37065)*, Vol. 1. IEEE, San Francisco, CA, USA, 589–595. <https://doi.org/10.1109/ROBOT.2000.844117>
- [21] Mohammad Abu Sayed, Nusrat Jahan Prithhee, and Hasan U. Zaman. 2020. Robotic helping hand: A new mechanism for helping disabled people. In *2020 International Conference on Inventive Computation Technologies (ICICT)*. IEEE, Coimbatore, India, 683–688. <https://doi.org/10.1109/ICICT48043.2020.9112576>
- [22] K. N. V. Sriram and Suja Palaniswamy. 2019. Mobile robot assistance for disabled and senior citizens using hand gestures. In *2019 International Conference on Power Electronics Applications and Technology in Present Energy Scenario (PETPES)*. IEEE, Mangalore, India, 1–6. <https://doi.org/10.1109/PETPES47060.2019.9003821>
- [23] Sebastian van Delden, Michael Umrysh, Carlos Rosario, and Gregory Hess. 2012. Pick-and-place application development using voice and visual commands. *Industrial Robot: An International Journal* 39, 6 (2012), 592–600. <https://doi.org/10.1108/01439911211268796>
- [24] Haofei Wang, Xujiang Dong, Zhaokang Chen, and Bertram E. Shi. 2015. Hybrid gaze/EEG brain computer interface for robot arm control on a pick and place task. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, Milan, Italy, 1476–1479. <https://doi.org/10.1109/EMBC.2015.7318649>
- [25] Andrés Úbeda, José M. Azorin, Nicolás García, José M. Sabater, and Carlos Pérez. 2012. Brain-machine interface based on EEG mapping to control an assistive robotic arm. In *2012 4th IEEE RAS EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob)*. IEEE, Rome, Italy, 1311–1315. <https://doi.org/10.1109/BioRob.2012.6290689>