

# How Should We Teach Robots? A Comparison of Kinesthetic, Joystick, and Gesture-Based Teaching

Petr Vanc, Jan Kristof Behrens, Václav Hlaváč, and Karla Stepanova

Czech Institute of Informatics, Robotics and Cybernetics (CIIRC CTU)

Jugoslávských partyzánů 1580/3, 160 00 Dejvice

Email: petr.vanc@cvut.cz

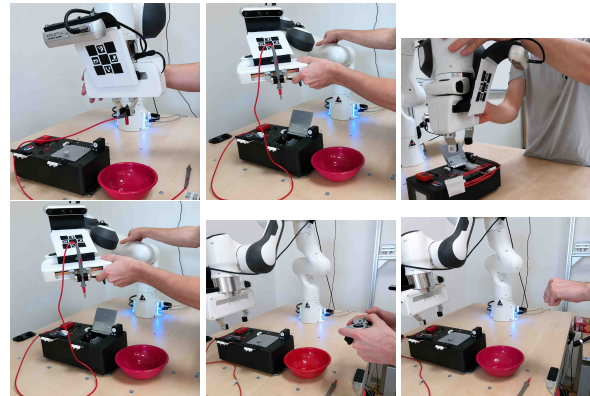
## Abstract

Instructing robots from demonstrations can be done through different teaching modalities, each with different usability and performance trade-offs. This paper compares kinesthetic guidance, joystick teleoperation, and hand gestures in a user study with eight participants. We evaluate replay success, modified NASA-TLX workload, and common teaching errors across three manipulation tasks. Kinesthetic guidance produced the shortest demonstrations, lowest workload, and highest success on the more orientation-sensitive and contact-rich tasks. Joystick teleoperation performed best on simple peg picking. Hand-gesture teaching, although less reliable overall, performed better than expected and in some cases achieved results comparable to kinesthetic guidance.

## 1 Introduction

Robots are still too often programmed by specialists, even when the desired task could be shown more easily than described in code. Learning from Demonstration (LfD) and Programming by Demonstration (PbD) address this problem by allowing users to teach robots through examples rather than manually specifying trajectories or control policies (Argall et al., 2009; Billard et al., 2016; Ravichandar et al., 2020). For non-expert users, however, the question is not only *what* the robot should learn, but also *how* the user should demonstrate it.

The teaching modality strongly shapes the resulting demonstration. Kinesthetic teaching, where the user physically guides the robot arm, is often treated as the most natural option because it gives direct control over the end-effector. Yet it requires safe access to the robot, a compliant control mode, and a user who can comfortably move the robot. These assumptions may fail in remote operation, safety-restricted workspaces, limited-access setups, or accessibility-sensitive scenarios. In such cases, joystick teleoperation or contact-free hand teleoperation may be more practical, although they introduce their own limitations: stepwise control through discrete inputs, hand tracking noise, and difficulty with fine orientation.



**Fig. 1:** Teaching robot manipulation tasks on the Robothon taskboard. The three evaluated tasks are shown in the top row: (left) peg picking, (center) probe measurement, and (right) cable wrapping. The bottom row visualizes the three teaching modalities used in the study: (left) kinesthetic teaching, (center) joystick teleoperation, and (right) direct teleoperation using hand gestures.

This paper compares three teaching modalities for teaching robots from demonstrations: kinesthetic guidance, joystick teleoperation, and hand gestures. We analyze a dataset in which eight participants taught three manipulation tasks under multiple scene variants. Each recorded variant is treated as a demonstration instance. We evaluate demonstration acquisition and replay quality, not learned policy generalization: replay success measures whether the robot can reproduce the recorded behavior under the corresponding demonstrated task condition. This allows us to compare replay success, demonstration duration, user workload measured by a modified NASA-TLX questionnaire, user preference, and common teaching errors.

We address the following research questions:

- How do kinesthetic guidance, joystick teleoperation, and hand gestures compare in replay success and demonstration duration across manipulation tasks?
- What workload trade-offs do these teaching modalities impose on users, especially in terms of

mental demand, physical demand, effort, and frustration?

- Can hand-gesture teaching serve as a viable contact-free alternative when kinesthetic teaching is unavailable, unsafe, or physically demanding?

This comparison is motivated by prior work showing that the form of a demonstration affects both usability and the quality of the learned robot behavior. Demonstrations may be represented as keyframes, trajectories, or movement models, and these choices influence replay, refinement, and generalization (Akgun et al., 2012; Pastor et al., 2009; Calinon and Billard, 2007). Recent work has also compared kinesthetic teaching with gamepad teleoperation for collaborative robot programming, showing that teleoperation can improve some demonstration-quality measures while requiring longer programming time (Dall’Alba and Boriero, 2025). Gesture interfaces have been studied both for direct robot teleoperation and for communicating higher-level human intent, including multi-type gesture sentences that combine action gestures with task parameters (Zhang et al., 2019; de Boer et al., 2023; Du et al., 2012; Vanc et al., 2023). In contrast to these works, our focus is a direct user-centered comparison of kinesthetic guidance, joystick teleoperation, and hand gestures on the same physical manipulation tasks, combining replay success, demonstration duration, user workload, user preference, and observed teaching errors.

Our results confirm that kinesthetic teaching is generally the easiest and most efficient modality, especially for tasks requiring precise alignment. However, the comparison also shows that it is not universally superior: joystick teleoperation performs very well on simple top-down grasping, and hand-gesture teaching performs better than expected. In several cases, gestures provide a viable contact-free alternative when direct kinesthetic teaching is unavailable, unsafe, or physically demanding.

## 2 User Study Design

### 2.1 Robot Platform and Tasks

We analyze a dataset of robot demonstrations collected in a user study on the Robothon Electronic Task Board So et al. (2024), see Fig. 1. The experimental setup used a Franka Robotics Panda robot equipped with a RealSense D455 camera mounted in an eye-in-hand configuration tracking the manipulation process. The Robothon taskboard provides a compact benchmark environment for manipulation tasks involving grasping, tool use, contact, and cable handling.

We considered three manipulation tasks: *Peg pick*, *Probe measure*, and *Cable wrap*. For task details, see Tab. 1. Each task was presented in multiple scene

variants, requiring the user to adapt the demonstrated behavior to the current task-board configuration. For example, the peg could be placed at different locations, the probe measurement point could be accessible or blocked by a door, and the cable could start from a different configuration. These variants make the dataset suitable for comparing teaching modalities across both simple demonstrations and demonstrations requiring adaptation to changed scene conditions. Each demonstration recorded the robot pose, gripper state, and visual observations (camera) during task execution.

### 2.2 Teaching Modalities

As mentioned in Sec. 1, we compare three teaching modalities for providing robot demonstrations: kinesthetic guidance, joystick teleoperation, and hand gestures measured by Leap motion sensor Weichert et al. (2013). All teaching modalities were mapped to a common command representation consisting of robot motion control and gripper control, allowing the recorded demonstrations to be compared independently of the input device, see Tab. 2.

Kinesthetic teaching allows the user to physically guide the robot arm while gravity compensation is enabled. This gives direct control over the robot motion and can support precise end-effector placement, but it requires physical access to the robot. Joystick teleoperation avoids direct contact with the robot, but the user must decompose the desired motion into sequential translation and orientation commands. Hand gestures provide contact-free 6-DoF teleoperation and are therefore useful when direct physical guidance is impractical, although they may be affected by hand-tracking reliability and fine orientation control.

Our hand teleoperation method is *Clutch-based*: The user closes their hand to engage control of the robot end-effector. While engaged, the robot follows the user’s 6-DoF hand motion. Opening the hand disengages control, so the user can reposition their hand without moving the robot. This interaction works similarly to a clutch-based 3D mouse or a drag-and-drop control in space. Additional gestures are used for end-effector orientation. A circular finger motion controls rotation around the end-effector roll axis, while sideways rotation angle of the hand controls the pitch angle. Gesture recognition is based on the methods described in Vanc et al. (2023) and implemented using the Teleoperation Gesture Toolbox<sup>1</sup>. See the paper website<sup>2</sup> for more details and video examples of individual teaching trials using different modalities.

<sup>1</sup>[github.com/imitrob/teleop\\_gesture\\_toolbox](https://github.com/imitrob/teleop_gesture_toolbox)

<sup>2</sup>[imitrob.ciirc.cvut.cz/publications/howtoteachrobot](https://imitrob.ciirc.cvut.cz/publications/howtoteachrobot)

Scenario	Demonstrated behavior	Scene variants	Success criterion	# Var.
<i>Peg pick</i>	Pick a peg and place it into a bowl.	Peg placed at different task-board locations.	Peg is successfully dropped into the bowl.	2
<i>Probe measure</i>	Pick the probe, touch the measurement point, and place the probe into a bowl.	Probe location and accessibility of the measurement point varied; in one variant, the measurement point was blocked by a door.	Probe touches the measurement point and is then dropped into the bowl.	3
<i>Cable wrap</i>	Wrap the cable around the task-board fixture.	Initial cable configuration and starting side varied.	At least two complete loops of cable are wrapped.	2

**Tab. 1:** Manipulation scenarios used in the user study. Each task was presented in multiple scene variants to elicit adapted demonstrations from users.

Modality	Motion control	Gripper control
Kinesthetic	Manual guidance with gravity compensation	Button
Joystick	Gamepad thumbsticks/buttons for translation and orientation	Buttons
Hand gestures	6-DoF hand teleoperation	Hand open/close

**Tab. 2:** Teaching modalities used for robot demonstrations.

### 2.3 Dataset and User Study

Our user study involved eight participants aged 24–40. Seven participants had no prior experience with gesture teleoperation, and five had no prior experience with kinesthetic teaching. Before recording demonstrations, participants were introduced to the task setup and teaching modalities<sup>3</sup>.

We reuse the demonstration dataset from See & Switch (Vanc et al., 2026), but analyze it here from the perspective of teaching modality rather than task branching. During data collection, participants demonstrated the manipulation tasks using each modality. For each task, the environment was presented in multiple scene variants. When the scene configuration changed so that the previous demonstration was no longer appropriate, participants provided an adapted demonstration for the new condition, such as a peg located initially in a different location.

**User Study Procedure** Each participant was first introduced to the system terminology and task setup. To build intuition for the demonstration workflow for our single-arm manipulator, we asked participants to perform the selected tasks single-handedly before actually teaching the robot. We also provided a small set of prac-

tical guidelines to improve safety and consistency during demonstrations, such as avoiding contact with the camera mount and keeping the cable in continuous contact with the task-board fixtures during cable wrapping.

**Collected dataset** Across 8 users and 3 modalities, the two 2-variant tasks and one 3-variant task yielded  $8 \times 3 \times (2 + 2 + 3) = 168$  recorded demonstrations. Participants typically recorded one demonstration per task, modality, and variant. In approximately 3% of cases, participants were allowed to repeat a demonstration and the previous attempt was discarded. Each demonstration variant was replayed and evaluated at least three times, yielding at least 3 roll-outs. Approximately 4% of demonstrations were considered unsafe (i.e., execution could damage taskboard or robot) and therefore contain only a single recorded roll-out; these cases are reported as unsuccessful in the replay-success analysis. In total, the dataset contains more than 600 execution roll-outs.

## 3 Evaluation Measures

We evaluate the teaching modalities from both robot-performance and user-experience perspectives. For each modality, we analyze whether the demonstrated behavior could be replayed successfully, how much time the user needed to produce the demonstration, how demanding the modality was perceived to be, and what types of errors occurred during teaching or replay.

### 3.1 Task Replay Success

Replay success measures whether a recorded demonstration produced the intended task outcome when executed by the robot. Each recorded task variant was treated as a separate demonstration instance, and success rates were computed over the corresponding replay roll-outs. A replay was considered successful when the robot achieved the task-specific goal. The goals are described in Tab. 1.

<sup>3</sup>See the complete user study narration material on our website [imitrob.ciirc.cvut.cz/publications/howtoteachrobot](http://imitrob.ciirc.cvut.cz/publications/howtoteachrobot)

### 3.2 Teaching Effort

We use the demonstration duration as an objective proxy for teaching effort. The duration captures how long the user needed to produce a complete demonstration for a given task variant and modality. Longer demonstrations typically indicate slower motion, additional corrections, hesitation, or extra time needed for end-effector alignment. This measure is not intended to replace subjective workload, but it provides an observable indicator of how efficiently users could express the desired robot behavior.

### 3.3 Subjective Workload and Preferences

Subjective workload was measured using a modified NASA-TLX questionnaire after the demonstrations. Participants rated mental demand, physical demand, effort, frustration, and satisfaction with the achieved teaching quality, each on a 1–7 scale. For mental demand, physical demand, effort, and frustration, lower values are better (i.e., they indicate lower workload). For Performance satisfaction with performance, higher values indicate better perceived teaching quality. We also compute a composite score from mental demand, physical demand, effort, and frustration, and inverted performance satisfaction, calculated as 8 – performance satisfaction.

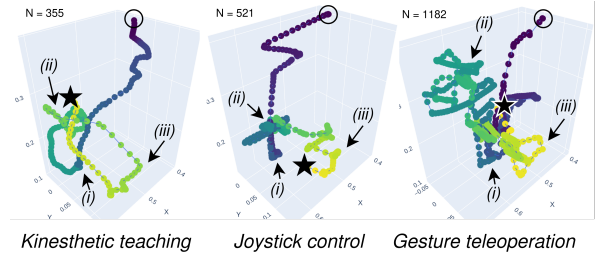
In addition to workload, participants rated the overall difficulty of each task and answered preference questions about the teaching modalities. These included whether they would use each method again, how much they trusted the learned robot behavior, which method felt easiest, which method felt most precise, and which method they would prefer for similar tasks in the future.

### 3.4 Observed Teaching Errors

We further analyzed common errors observed during demonstrations and replay. These included failed grasps, imprecise probe alignment, missing final placement actions, incomplete cable wrapping, unnecessary corrections, unstable motion, excessive orientation adjustments, and loss of hand tracking during gesture control. We also inspected trajectory characteristics across teaching modalities, focusing on whether demonstrations were smooth and direct, segmented into incremental motions, or noisy during fine end-effector orientation, see example differences in Fig. 2.

## 4 Results

Tab. 3 summarizes replay success and average demonstration length for each task and modality. The easiest task was *Peg pick*, where all teaching modalities achieved high replay success. This result is consistent



**Fig. 2:** Example demonstration trajectories for *Probe measure* task.  $\circ$  is demo start.  $\star$  is demo end. (i) is probe pick, (ii) is measurement with probe, and (iii) is probe drop into a bowl. You can see that gesture probe measurement (ii) is messy when the user tries to reorient the gripper and focus on target precision.  $N$  is the number of time steps. Keypoint colors indicate the time encoding.

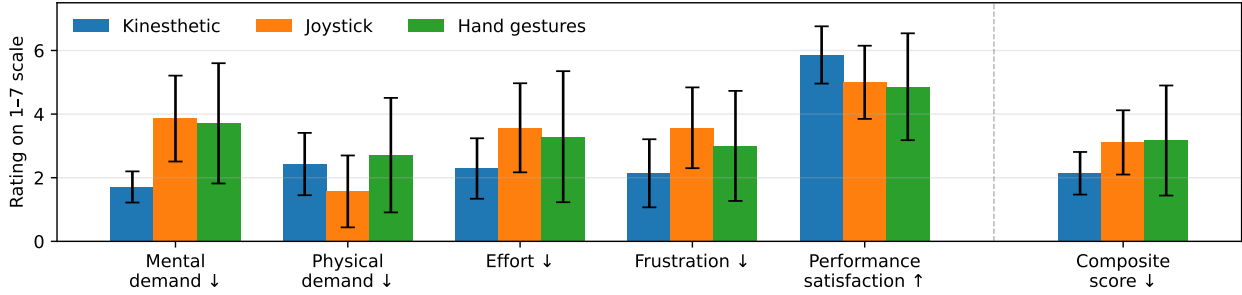
Modality	<i>Peg Pick</i>	<i>Probe measure</i>	<i>Cable wrap</i>
<i>Replay success (%)</i>			
kin	88.0	<b>83.0</b>	<b>58.1</b>
joy	<b>99.2</b>	80.3	53.1
gst	92.7	72.2	48.8
<i>Demonstration length (s)</i>			
kin	<b>17.4 ± 4.2</b>	<b>19.5 ± 3.7</b>	<b>24.0 ± 5.4</b>
joy	26.1 ± 9.2	39.2 ± 18.6	65.3 ± 26.2
gst	38.2 ± 28.7	41.6 ± 13.0	74.8 ± 38.1

**Tab. 3: User study results.** Replay success rate over at least three trials and average demonstration length for each modality and task.

with the task structure: picking a peg mainly required top-down grasping and only limited wrist reorientation. Joystick teleoperation achieved the highest success on this task (99.2%), followed by hand gestures (92.7%) and kinesthetic teaching (88.0%).

The more orientation-sensitive tasks were harder. In *Probe measure*, the user had to align the probe precisely with the measurement point, and kinesthetic teaching achieved the highest replay success (83.0%). In *Cable wrap*, the robot had to maintain a longer contact-rich motion around the task-board fixture. This was the most difficult task across all teaching modalities, with success rates between 48.8% and 58.1%. Kinesthetic teaching again achieved the highest success on this task, but the overall lower values indicate that cable wrapping was difficult to demonstrate reliably for novice users.

Demonstration duration shows a clear difference in teaching effort. Kinesthetic teaching was consistently the fastest modality, with average durations between 19.5 and 24.2 seconds. Joystick and gesture demonstrations became substantially longer for tasks requiring fine orientation control or continuous manipulation. For



**Fig. 3: Modified NASA-TLX workload ratings by modality.** Bars show mean ratings and error bars show standard deviations on a 1–7 scale. Lower values indicate better results for mental demand, physical demand, effort, frustration, and composite score, while higher values indicate better results for performance satisfaction. Composite score is computed from mental demand, physical demand, effort, frustration, and inverted performance satisfaction, where inverted performance satisfaction is calculated as  $8 - \text{performance satisfaction}$ .

*Cable wrap*, joystick demonstrations lasted 61.4 seconds on average, while gesture demonstrations lasted 73.5 seconds. This suggests that direct physical guidance allowed users to produce task trajectories more efficiently, whereas joystick and gesture control required more corrections and slower alignment.

The trajectory examples in Fig. 2 illustrate these modality differences for *Probe measure*, where users had to reorient the gripper precisely while approaching the measurement point. Joystick trajectories were more segmented, reflecting the incremental nature of the control input. Gesture trajectories were more variable, especially during fine end-effector orientation. This was most visible in *Probe measure*, where users had to reorient the gripper precisely while approaching the measurement point.

The most common replay failures were manipulation errors rather than misunderstandings of the task goal. Typical failures included slip grasps, imprecise probe alignment, missing final placement actions, and incomplete cable wrapping. Gesture demonstrations were additionally affected by unstable hand tracking or noisy motion during fine orientation control. Joystick demonstrations were often limited by the need to decompose motion into separate translation and rotation commands, such as when users had to think about which button to press.

Fig. 3 reports the modified NASA-TLX questionnaire results. Kinesthetic teaching received the lowest mental demand ( $1.71 \pm 0.49$ ), effort ( $2.29 \pm 0.95$ ), and frustration ( $2.14 \pm 1.07$ ), and the highest satisfaction with the achieved teaching quality ( $5.86 \pm 0.90$ ). Joystick teleoperation had the lowest physical demand ( $1.57 \pm 1.13$ ), but was rated as more mentally demanding ( $3.86 \pm 1.35$ ) and more frustrating ( $3.57 \pm 1.27$ ) than kinesthetic teaching. Hand-gesture teaching showed a similar overall workload to joystick teleoperation, with mental demand of ( $3.71 \pm 1.89$ ), effort of ( $3.29 \pm 2.06$ ), and frustration of ( $3.00 \pm 1.73$ ), but also showed higher variability between participants.

Task-difficulty ratings followed the replay results. *Peg pick* was perceived as the easiest task ( $1.71 \pm 0.76$ ), *Probe measure* was rated as moderately difficult ( $2.71 \pm 1.38$ ), and *Cable wrap* was rated as the most difficult task ( $4.00 \pm 2.00$ ). User preferences also favored kinesthetic teaching: all participants selected it as the easiest method, five participants selected it as their preferred method for future use, and four respondents rated it as the most precise method. However, two participants preferred hand-gesture teleoperation overall, indicating that contact-free teaching was attractive for some users despite its higher variability.

Open questionnaire responses were consistent with these findings. Participants described kinesthetic teaching as intuitive and capable of expressing all robot motions. Joystick control was described as useful for large motions, but less convenient for precise orientation. Hand gestures were described as more natural than the joystick for some final adjustments, but users also reported hand-tracking failures, missing rotation control around the end-effector, and unintended vertical motion. These increased both the demonstration length and the user’s frustration.

Finally, successful demonstrators produced smoother, more consistent trajectories, avoided unnecessary corrections, and appeared to plan the complete task before moving the robot. Demonstrations with excessive hesitation, repeated reorientation, or missing final actions were less likely to replay successfully. When users repeated demonstrations, later attempts were typically more direct and more reliable Aliasghari et al. (2024).

## 5 Discussion

The results show that kinesthetic teaching remains the strongest baseline for efficient demonstration acquisition when direct physical interaction is possible. It produced the shortest demonstrations, the lowest sub-

jective workload, and the highest perceived precision. This supports the expectation that physically guiding the robot is effective for novice users because it allows direct positioning of the end-effector without translating the intended motion into interface commands.

However, kinesthetic teaching is not universally preferable. It requires safe physical access to the robot, gravity compensation, and sufficient physical ability from the user. It also limits the robot's proprioceptive experience. Furthermore, it requires the user to be aware of the whole robot configuration, not only the end-effector pose. In practice, a user can guide the robot into a configuration that completes the immediate motion but is less suitable for replay or for the following part of the task. This issue was especially relevant in tasks involving contact or constrained motion, such as cable wrapping.

Joystick teleoperation had a different trade-off. It had the lowest physical demand and performed very well on *Peg pick*, where the required motion was mostly translational and top-down. However, it became slower and more demanding for tasks requiring precise orientation, such as probe measurement and cable wrapping. This suggests that joystick control is suitable for simple or remote demonstrations, but less efficient when users must coordinate several translational and rotational degrees of freedom.

Hand-gesture teaching was the most interesting alternative. It was not the best modality overall, but it performed better than expected. Gesture teaching achieved high replay success on *Peg pick* and remained usable for the more difficult tasks. Two participants even selected hand gestures as their preferred future method. This suggests that contact-free hand control can be intuitive for some users, especially when direct kinesthetic guidance is unavailable, unsafe, or physically demanding. Its main limitation is reliability: fine orientation control and hand-tracking errors can introduce noisy trajectories and user frustration.

The comparison also shows that the task strongly influences which modality is appropriate. For simple top-down grasping, joystick and gestures can be competitive with kinesthetic teaching. For tasks requiring precise end-effector alignment, kinesthetic teaching is more efficient. For long contact-rich manipulation, such as cable wrapping, all teaching modalities become difficult, and successful demonstrations depend not only on the interface but also on how well the user plans the complete robot motion.

Overall, the study supports a modality-dependent view of robot teaching from demonstrations. Kinesthetic teaching should be preferred when the robot is safely accessible and precision is required. Joystick teleoperation is useful when physical contact should be avoided and the task does not require complex continuous motion. Hand gestures provide a viable contact-free alternative, particularly for situations where kines-

thetic teaching cannot be used, although improvements in tracking reliability and orientation control are needed.

The main limitation of this study is its small sample size of eight participants and its focus on a single robot platform and task board. The questionnaire results should therefore be interpreted as exploratory. In addition, only a small number of open-ended responses were collected, so qualitative comments are used mainly to support, rather than replace, the quantitative results. Future work should evaluate the same teaching modalities with more users, additional task types, and a more detailed analysis of how users improve over repeated demonstrations.

## References

- Akgun, B., Cakmak, M., Jiang, K. and Thomaz, A. L. (2012). Keyframe-based learning from demonstration: Method and evaluation. *International Journal of Social Robotics*, 4:343–355.
- Aliasghari, P., Ghafurian, M., Nehaniv, C. L. and Dautenhahn, K. (2024). How non-experts kinesthetically teach a robot over multiple sessions: diversity in teaching styles and effects on performance. *Int. J. Soc. Robot.*, 16(11):2079–2105.
- Argall, B. D., Chernova, S., Veloso, M. and Browning, B. (2009). A survey of robot learning from demonstration. *Robotics and Autonomous Systems*, 57(5):469–483.
- Billard, A. G., Calinon, S. and Dillmann, R. (2016). Learning from humans. In *Springer Handbook of Robotics*, pp. 1995–2014. Springer.
- Calinon, S. and Billard, A. (2007). Incremental learning of gestures by imitation in a humanoid robot. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction*, pp. 255–262.
- Dall'Alba, D. and Boriero, F. (2025). Towards an intuitive industrial teaching interface for collaborative robots: gamepad teleoperation vs. kinesthetic teaching. *The International Journal of Advanced Manufacturing Technology*.
- de Boer, T. A., de Winter, J. C. F. and Eisma, Y. B. (2023). Augmented reality-based telepresence in a robotic manipulation task: An experimental evaluation. *IET Collaborative Intelligent Manufacturing*, 5(4):e12085.
- Du, G., Zhang, P., Mai, J. and Li, Z. (2012). Markerless kinect-based hand tracking for robot teleoperation. *International Journal of Advanced Robotic Systems*, 9(2):36.

- Havoutis, I. and Calinon, S. (2019). Learning from demonstration for semi-autonomous teleoperation. *Autonomous Robots*, 43:713–726.
- Pastor, P., Hoffmann, H., Asfour, T. and Schaal, S. (2009). Learning and generalization of motor skills by learning from demonstration. In *IEEE International Conference on Robotics and Automation*, pp. 763–768.
- Ravichandar, H., Polydoros, A. S., Chernova, S. and Billard, A. (2020). Recent advances in robot learning from demonstration. *Annual review of control, robotics, and autonomous systems*, 3(1):297–330.
- So, P., Sarabakha, A., Wu, F., Culha, U., Abu-Dakka, F. J. and Haddadin, S. (2024). Digital robot judge: Building a task-centric performance database of real-world manipulation with electronic task boards. *IEEE Robot. Autom. Mag.*, 31(4):32–44.
- Vanc, P., Behrens, J. K., Hlaváč, V. and Stepanova, K. (2026). See and switch: Vision-based branching for interactive robot-skill programming.
- Vanc, P., Behrens, J. K., Stepanova, K. and Hlavac, V. (2023). Communicating human intent to a robotic companion by multi-type gesture sentences. In *IEEE/RSJ IROS*, pp. 9839–9845.
- Weichert, F., Bachmann, D., Rudak, B. and Fisseler, D. (2013). Analysis of the accuracy and robustness of the leap motion controller. *Sensors*, 13(55):6380–6393.
- Zhang, W., Cheng, H., Zhao, L., Hao, L., Tao, M. and Xiang, C. (2019). A gesture-based teleoperation system for compliant robot motion. *Applied Sciences*, 9(24):5290.