Adaptive Degrees of Freedom for Shared Control of Assistive Robots

Doctoral Thesis
Computer Science

by

Felix Ferdinand Goldau

Supervisor & First Examiner:	Prof. Dr.-Ing. Udo Frese University of Bremen, DFKI
Second Examiner:	Prof. Dr.-Ing. Andreas Hein Carl von Ossietzky University Oldenburg, OFFIS

Submitted: 21st August 2024
Defended: 2nd October 2024

Abstract

We have reached the point where robots prominently enter our daily lives, be it as a versatile tool at work or as an obedient household helper. This development is especially exciting for people with physical limitations, since designated assistive robots harbour a large potential to enhance their users' autonomy and quality of life. Following this line of thought, the field of assistive robotics introduces mechanical assistants to people who would otherwise struggle with activities of daily living. However, this necessitates adequate and potentially personalised control methods. Focussing on wheelchair-mounted robotic arms, this thesis discusses the methods currently applied in the field, evaluates directly applicable manual alternatives, and proposes a novel shared control based on adaptive degrees of freedom. Following a participatory design, each element is developed and evaluated in close collaboration with the target group, thus allowing for appropriate integrations and realistic assessments.

For the contemporary manual analysis, users evaluated the default manufacturer-provided input device in comparison to a gamepad, 3D mouse, and a command-based voice control. Overall, these studies with the target group ($N_1 = 26, N_2 = 15$) show a large potential for improvement of the standard in terms of usability and versatility. During this, especially the necessity of mode switches in the robot's default control was remarked negatively. Instead, the participants expressed an eagerness for personalised adaptability, as well as an explicit willingness to train in the use of more complex but capable systems, such as a 3D mouse.

Heeding this, this thesis introduces the novel shared control approach of *Adaptive Degrees of Freedom*: A camera-based sensor system probabilistically analyses the current situation to generate the most likely directions of robot motion. Subsequently, these directions are mapped onto the user's input device, effectively replacing the classically available cardinal Degrees of Freedom (DoFs) (e.g. *up*, *left*, *roll*, . . .). In the end, this enables users to control a robot along arbitrarily complex DoF with any input device, explicitly including very low-DoF interfaces (e.g. chin joysticks), thereby making robots more accessible for people with very limited mobility. For the user, this feels like the system anticipating their next move without taking over control. Instead, it simply provides them with a selection of movement directions designed specifically for the current situation.

This novel control is mathematically and conceptually introduced with its usability verified in preliminary studies. Preparing for a contemporary data-based realisation, a mixed-reality development framework was developed and used to record an extensive dataset of user controlled robots in assistive settings in simulation and reality. Both the framework and dataset were published open-source and free-of-charge.

The dataset was planned to be used with a state-of-the-art deep-learning neural network to predict DoFs end-to-end based on image data. While this was applicable in an initial 2D baseline scenario, the training of machine-learned models in 3D was unsuccessful. This startling result runs seemingly contradictory to the research community's current achievements using similar methods, which is why this thesis includes an extensive analysis of why this is the case.

As an alternative, the author presents a probabilistic behaviour-based integration that is able to generate the adaptive DoFs. This implementation was evaluated in multiple studies, focussing on its general applicability, human-computer-interaction, and usability. Finally, a study conduced solely with the target group (people with limited upper body mobility, $N_3 = 24$) evaluated the completely integrated system, showcasing high user acceptance with a steep learning curve and high success rates of example trials.

Keywords: Assistive Robotics, Shared Control, Human-Robot-Interface, Participatory Research

Contents

List of Figures

List of Tables

List of Abbreviations

2D	Two-Dimensional
3D	Three-Dimensional
AdaMeKoR	Ein adaptives Mehrkomponenten-Robotersystem für die Pflege
ADL	Activity of Daily Living
AI	Artificial Intelligence
ALS	Amyotrophic Lateral Sclerosis
BCI	Brain-Computer-Interface
CNN	Convolutional Neural Network
DoF	Degree of Freedom
DoF-Adaptiv	Adaptive Freiheitsgradeinbettung als kooperatives Userinterface für einen Assistenzroboter
DORMADL	Dataset of human-Operated Robot arm Motion in Activities of Daily Living
HMI	Human-Machine Interface
MobILe	Physische Mensch-Roboter-Interaktion für ein selbstbestimmtes Leben
PCA	Principal Component Analysis
PoM	Percentage-of-Motion metric
RV	Random Variable
TCT	Task Completion Time
UI	User Interface
VR	Virtual Reality
WMRA	Wheelchair-mounted Robotic Arm

Prelude

Roughly one year after I started working with assistive robotics, my father got seriously sick. Our lives changed drastically, and I somehow gained hands-on experience of a topic which, up to this point, I had barely touched from more than a scholarly perspective. There is one story I'd like to share for this thesis:

My father was an intelligent and extremely well organised person, who was often set in his ways: At home and at work, every object, be it a key, printer cartridge, or simple shoehorn, had defined positions to which they had to be returned whenever not in use. Similarly, there were activities which had to be carried out in certain ways. Some of these he instructed everyone to perform; For example, there was a door which always needed to be closed again immediately in order to avoid heat or insects to come in. Others were more internally automated, resulting in tiny habits, the kind of which we all have. For him, one of these was eating his favourite yoghurt after every day of work.

Then, on the first day my mother and I were allowed to visit him after the surgery, we brought along some of this yoghurt. The operation left him paralysed on his left side, partially affected his mental capacities, and, temporarily, impeded his ability to speak. He was visibly happy to see both us and the yoghurt. At some point my mother removed the lid and I held the cup in place for him to eat one-handedly. However, as my mother walked to throw away the lid, I noticed him getting more and more distressed and following her with his gaze.

Something was not right, but, paralysed as he was, he could neither correct it himself nor communicate with us sufficiently. While already being overburdened from the beginning, we now grew anxious of how to possibly help him. Surprised by his reaction, we remembered him always scratching off the tiny bit of yoghurt sticking to the lid. This was now missing to his routine, stressing him and having an astonishingly large impact on his mental state in this situation. We saved the lid, just before throwing it away, and could marvel at his content of being able to happily complete his routine.

I believe we all have these tiny habits. Some of which we might not even be aware of. But somehow, it is vital to us that certain steps and activities are carried out according to our individual beliefs.

This grows increasingly difficult when receiving care, as it often comes with a loss of independence. Suddenly, every bit of self-determined control is important, as it retains a bit of one's self. We cannot assume approaches of one-size-fits-all or generalised automation to function in this highly individual and personal field. At least, it wouldn't have worked for my father, for whom the yoghurt with lid to scratch bits off remained a highlight up to his last day.

1. Introduction

Even though robots were historically developed, as implied by the linguistic background of *workers*, to perform menial and repetitive labour in industrial settings, their applications have long since extended to regular households and daily living: We use robotic machines to clean our floors, prepare food in the kitchen, and even manipulate patients in hospital beds. Assistive robotics goes one step further by aiming to assist people during activities of daily living, generally focussing on people with motor impairments. However, the general broadening of applications already necessitates adjustments of control strategies, as the usage of robots is no longer limited to trained engineers. For assistive robots in particular, these strategies must also consider variations of physical input devices and software-based support in order to cope with a user's individual physical limitations.

In Germany alone, 7.8 million people live with severe disabilities (9.4 % of the overall population, as of 2021), with approximately 600 thousand having limited functionality, or loss, of at least one arm [53]. For many of these people, assistive robots hold a large potential to increase their independence and overall autonomy, especially if they are otherwise consistently receiving individual care [41]. Figure 1 shows such a device, often integrated as a Wheelchair-mounted Robotic Arm (WMRA), which can usually be directly operated using the wheelchair's default joystick designed for driving [34].

Figure 1: The Kinova Jaco Assistive Robot Arm as WMRA. Courtesy of Kevin Rupp | Frankfurt UAS

However, a robot arm is a conceptually vastly more complex device than a wheelchair. A standard wheelchair can only drive on the horizontal plane by accelerating forwards or backwards, and rotating around its axis; i.e. it has two Degrees of Freedom (DoFs): *drive* and *rotate*. The deployed robotic arms on the other hand can move and rotate arbitrarily in three dimensions, as well as closing their fingers to grasp objects. Aggregating the individual fingers for their intent as a single grasping-DoF, this amounts to seven DoFs for the robot arm. This discrepancy results in an ill-fit when controlling the complex robot with the simple joystick designed for wheelchair use. Usually, this is circumvented by introducing control modes.

With these, the user can, at a single point in time, only move the robot along two DoFs and has to switch modes to reach other options (e.g. they can control translation in one mode, but have to switch to reach rotations). This can be a very slow and tedious process and can cause distress in operators: pouring a glass of water, for example, takes ≈ 500 seconds with ≈ 50 mode switches, thereby spending more than one-sixth of the total execution time with changing modes [20]. Figure 2 shows an overview of the button mapping for the manufacture's 3-DoFs joystick using three modes. Classically, assistive wheelchair joysticks only have two DoFs.

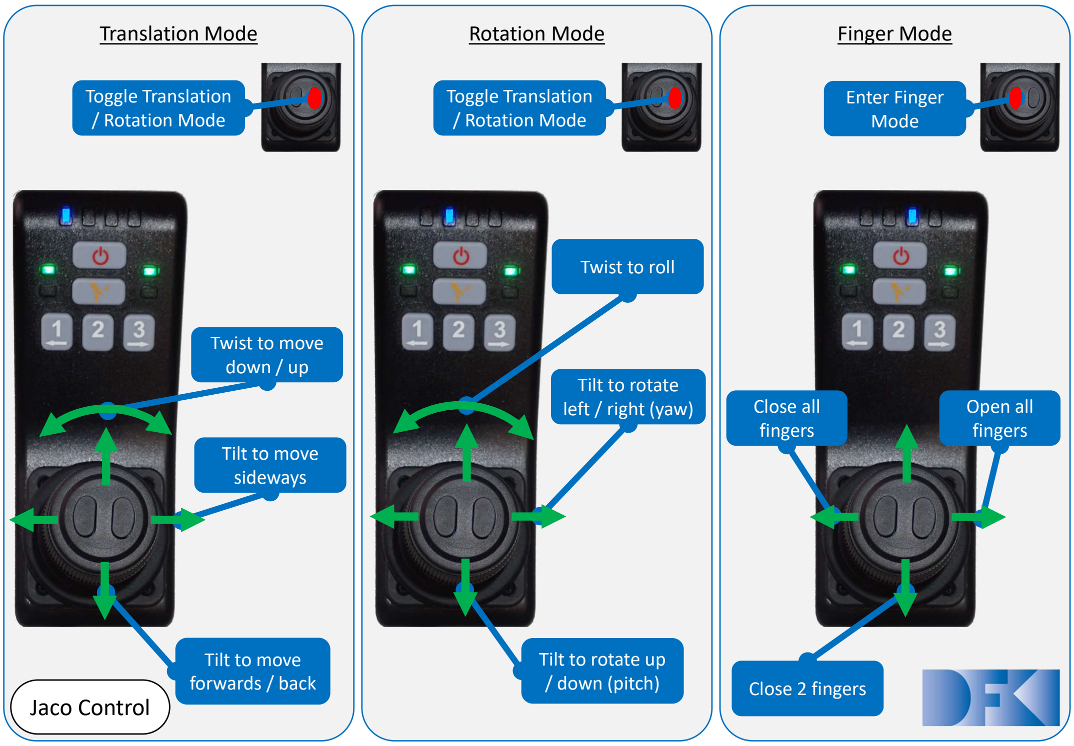


Figure 2: The Kinova Jaco's Default Control Scheme Visualised

Such issues generally do not arise in classic industrial settings where robots are fully automated. However, automation is not a practical solution for assistive robots, as computer-controlled execution of Activities of Daily Living (ADLs) in unknown environments is not only very complex, but would also deprive users of an opportunity for individuality and self-determined autonomy (cf. [28]). As these are basically the essential reasons for people to utilise assistive robots, other solutions need to be examined [36].

The community has developed various different approaches to this problem: On one side, custom input devices were designed that could be adapted to their user's needs and capabilities. For example, one concept uses a Brain-Computer-Interface (BCI) to control a robot arm along predetermined trajectories [52], whereas others attached various body-sensors on a single user to create high-dimensional input data from their remaining mobility and converting this to robot control inputs [30]. On the other side are implementations of partial automation, including time- or space-related hand-overs of controls [11], automated mode-switching [20], as well as simultaneous control shared between the user and a software [56]. In case of the former, the user manually controls the robot up until a software takes over for a certain sub-task [11, 58], whereas the latter involves blending the manual user input with that of a software, e.g. by having the software solely handle the robot's orientation [58].

1.1. Contributions

This thesis experimentally analyses contemporary options for assistive robot control and subsequently presents a novel method of shared control. The initial analysis rests on two extensive studies conducted with care-receiving and care-associated participants. These evaluate four different general purpose input devices for manual control, thereby establishing their requirements and preferences towards such devices. Further, this thesis presents a partially automated system of shared control designed to allow self-determined drinking from a cup for people with tetraplegia. Tetraplegia is the paralysis of all four limbs below the neck. This all leads to the main body of this thesis, which is a novel paradigm for shared assistive control:

Previously, operators would use a low-DoFs input device (e.g. Figure 2) to send a control signal c to the robot. Depending on the current mode D, which is selected by the operator from a pre-determined set of modes $\hat{D}$ (see *translation*, *rotation*, *gripper* in Figure 2), the control signal was used to move the robot along cardinal DoFs (e.g. *up*, *forward*, *yaw*, *gripper*, ...). This can be tedious and time-consuming, as even simple tasks require repeated mode switching.

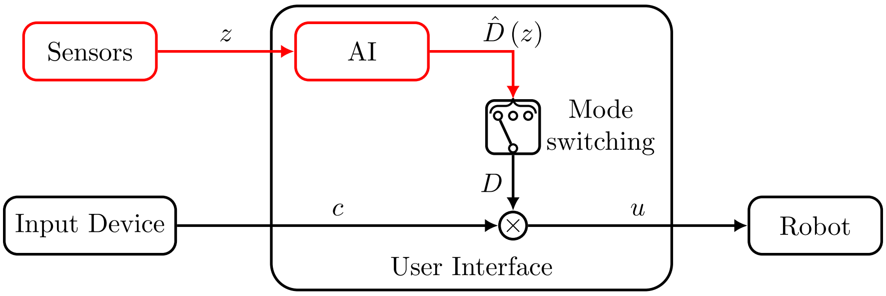


Figure 3: Control Concept for Adaptive DoFs. Adapted from [IV]

The novel concept is shown in Figure 3 and expands (in red) on the classic interaction (black), by making the modes situation-dependant: An Artificial Intelligence (AI)-supported software adaptively analyses the current situation z and generates DoFs $\hat{D}(z)$ of robotic motion that are most likely to be desired by the user. These replace the previously used cardinal DoFs-based modes, such that the user can select to directly map their control input to these most-likely DoFs of robotic motion. The result sees the user controlling the robot along arbitrarily complex high-dimensional motions, while using only a simple input device, for example with only one DoF. To them, this feels like the system anticipating their next move without taking over control. Instead, it simply provides the user with a selection of movement directions designed specifically for the current situation.

A mixed-reality framework for the development and evaluation of such concepts of shared control was developed and published open-source. This framework was used in user-centric studies to evaluate the concept's general applicability and inherent required communication of controllable DoFs for the operator. Further, the framework was used to record an extensive dataset of human-controlled robot motion in assistive settings as preparation for data-based approaches, which was also published free-of-charge.

The dataset was planned to be used with a state-of-the-art deep-learning neural network to predict DoFs end-to-end based on image data. While this was successful in an initial 2D baseline scenario, the results accomplished with machine-learned models in 3D were unsatisfying. This startling result runs seemingly contradictory to the research community's current successes using similar methods. This thesis therefore includes an extensive analysis of why this is the case.

Instead, a probabilistic combination of elemental behaviours was developed and tested in a controlled environment. This successful implementation was finally presented to, and evaluated by, the target audience at a large international trade fair.

In short, the contributions of this thesis are:

- an in-depth experimental evaluation of contemporary control options available for assistive robots [I, VII] with
  - resulting requirements to assistive controls as set by the target group [XI], and

- a novel approach to shared control using adaptive DoFs, including
  - the accessible theoretical and conceptual system framework [XII, IV, VIII, X],
  - a novel free-to-use mixed-reality environment for evaluations of shared control [IX] (Best Paper award winner at EICS 2024[1]),
  - an extensive curated dataset of human-controlled assistive robot motion in realistic and relevant scenarios [III], and
- a CNN-based training of an end-to-end approach [IV], with
  - an in-depth analysis of why this approach failed in 3D, as well as
- an alternative functional and study-proven implementation based on a probabilistic combination of behaviours [XIII, V, VI] (Best Technical Paper award for [V] at PETRA 2024[2]).

Over the course of this thesis, a total of 8 studies were conducted. Of these, 3 focussed extensively on the target group of individuals with limited upper limb mobility, whereas the others gathered mostly preliminary technical results. An overview of all studies, including a short description, number of participants, and participant criteria can be found in Table 1.

Table 1: List of Studies

Name	Description	Number of Participants	Participant Characteristics	Reference
Study 1	Initial Contemporary Study at REHAB (Preliminary)	$N_1 = 26$	Individuals with limited upper limb mobility, care professionals, secondary users	[XI]
Study 2	Contemporary Lab Study	$N_2 = 15$	Individuals with limited upper limb mobility	[XI]
Study 3	Final Behaviour Evaluation at REHACARE	$N_3 = 24$	Individuals with limited upper limb mobility	[VI]
Study 4	Robotic Drinking Assistant: Drinking from the Cup Study	$N_4 = 16$	Able-bodied users and one tetraplegic user	[VII]
Study 5	First Usability Study	$N_5 = 39$	Able-bodied users	[VIII]
Study 6	Second Usability Study	$N_6 = 24$	Able-bodied users	[X]
Study 7	Adaptive DoFs in 2D	$N_7 = 23$	Able-bodied users	[IV]
Study 8	Preliminary Behaviour Study	$N_8 = 24$	Able-bodied users	[V]

In the following, Section 2 discusses alternative approaches to the control of assistive robots as presented by the scientific community, while Section 3 presents contributed studies of contemporary control options, including a proposed partial automated drinking realisation, and the study-based analyses of manual control options.

The novel approach for shared control is introduced with conceptual and systematic derivations in Section 4. Further, implementations regarding the generation of adaptive DoFs are presented in Section 5, which include the end-to-end developed machine-learning variant in Section 5.1 and the alternative behaviour-based realisation in Section 5.2. Finally, Section 6 concludes by discussing general implications of the proposed control, summarises results and outlines subsequent necessary steps to further improve the presented control and make it available to the public.

[1]The 16th ACM SIGCHI Symposium on Engineering Interactive Computing Systems. `https://eics.acm.org/2024/`, last visited 10th December 2024

[2]The PErvasive Technologies Related to Assistive Environments Conference. `https://www.petrae.org/`, last visited 10th December 2024

2. Related Work

Due to an ageing population and a lack of healthcare personnel, the public interest in robotics for care is skyrocketing, thereby leading to a wide range of research in the field. While important work is done to assist caregivers with their tasks (e.g. [3, 26, 37]), this thesis focusses on (re-) enabling people with limited upper limb mobility (often care-receiving individuals) to independently physically interact with their environment for Activities of Daily Living. These ADLs generally refer to the 'basic tasks of everyday life, such as eating, bathing, toileting, and transferring' [14] (cf. [9, 46]). With the assistive robot of this thesis, the applicable ADLs are focussed more on tasks like retrieving objects, drinking with a cup, and opening doors. Further, the target group of this thesis persists mostly of people with limited upper limb mobility, including, for example, people living with tetraplegia, spinal muscular atrophy, and Amyotrophic Lateral Sclerosis (ALS).

Instead of custom devices that serve a singular purpose (e.g. an eating utensil for users with spastics [13, 55]), systems like assistive WMRAs follow a more holistic approach by employing high-dimensional complex machines that aim to be general purpose. The market has a couple of such devices, some of which are even supported by health insurances, including the *Kinova Jaco* [34] (Figure 1) and Assistive Innovations' *iARM* [10, 59], with others in development, such as a 4-DoF robot for users with cerebral palsy [48]. While these machines are often mathematically and mechanically able to perform a variety of tasks, their actual applicability and accessibility depends highly on the control interface. This itself often needs to be adapted to fit with a user's individual requirements [22], which includes custom adjustments of physical input devices as well as introducing supportive software.

As mentioned in Section 1, the default interfaces for classic cardinal control are very manual, often inducing frustration and exhaustion among users [11, 19, 20, 27, 40]. For this reason, the community has developed various Human-Machine Interfaces (HMIs) that span the spectrum from manual control to complete automation and involve multiple different physical input devices. These devices include eye-trackers [51], tongue/chin-joysticks [17], head-controls [27, 36, 51], BCIs [18, 22, 23, 52], laser pointers [63], and touch screens [54]. As most of these devices have far less input-DoFs than the robot has movement-DoFs, numerous proposed systems still use the often frustrating manual mode switch sequences [32]. Those which avoid mode switches on the other side are designed with pre-defined tasks where the user mostly has an administrative purpose, such as supplying *continue*-commands [52] or selecting objects to be further handled autonomously by the robot [63].

Shared control lies in the middle of the spectrum between manual control and automation. It does so by keeping the human in control whilst having an automation software assist during difficult operations. Automated mode switching [20, 29], for example, allows quicker executions and a less tedious interaction by sticking to the established interaction design of cardinal modes and having a software simply perform the mode switch when necessary. Other systems arbitrate the user's control input with a computer generated control, thereby allowing the computer to assist by avoiding obstacles or letting the user control only the translational DoFs of a robot arm, whilst automatically handling rotation [11, 58]. On the far side of the spectrum, users simply indicate the target and let the robotic system automatically handle the grasping and interaction [33, 54].

The latter fusion of simultaneous movement commands is referred to as *control blending* and is also applied by Gopinath et al.: Following an automated mode switching designed to assist the AI to probabilistically isolate the user's goal, the actual motion commands of human and AI are blended once a certainty threshold is surpassed. This aims to assist the user with small adjustments while they otherwise remain limited to the classic cardinal DoFs [16]. Similarly, prior to execution Jain et al. divide their objective into intermediate steps and sub-sequences. These sequences of robotic operations solve parts of the task, are autonomously calculated during runtime, and subsequently suggested to the user. Instead of steering with classic cardinal robot control modes, this allows operators to select the pre-computed automation trajectories and simply control the robot's velocity along such a sequence [30].

Another highly promising example of shared control is made by Losey et al. using a latent action space: A neural network is trained on task demonstrations to embed the necessary complex and high-dimensional motion in a very low-dimensional latent action space. In deployment, the system works directly in this action space by mapping the signal generated from a low-dimensional input device to a latent action dimension. This allows, for example, to have a single joystick axis directly control the complex task of *approaching an object*. However, this requires demonstration and knowledge of a predetermined task and is therefore highly specific [39].

To be successful, any one of these systems involving human-robot cooperation requires clear communication between the partners, as only with that given can a user balance control input and system maintenance. It is especially necessary to effectively communicate the intended assistance provided by (semi-) autonomous system to the user [4]. In case of immediate shared control, this includes well-designed information regarding the robot's (motion) intent, as it is essential for the user to retain awareness and understanding of the level of support provided by the system [44].

2.1. Automation and Personalisation

As the stated purpose of new control interfaces extends from generally higher independence [36] to a partial reintegration of affected people to the workplace [17], a wide variety of applications and user diversity needs to be contemplated during development. This includes user-centric analyses of requirements and preferences for the given controls.

With respect to the spectrum of control (manual control versus autonomous operation), a study by Kim et al. conducted with spinal cord injured subjects showed higher satisfaction for manual controls, even though an autonomous variation required less effort [31]. This coincides with findings from Martinsen et al., who established the users' requirement to personalise their interaction such that personal standards and social norms are met [41], which was later revisited by Canal et al. [8]. One can generally state that no one-size-fits-all solution exists in assistive robotics [24]. As personalised automation is currently not feasible on an individual level, this calls for more manual controls.

If, however, manual control by itself is not possible, the level of assistance should be configurable by the user in order to allow for adjustments to their needs and abilities, as well as increase user satisfaction [28]. A simple way to implement this is to increase speed in safer regions which reduces frustration [32]. Overall, shared control can be the solution to this, as it increases subjectively felt independence [12], while automated solutions (with the users in an administrative or oversighting role) can cause distrust [64], stress and a feeling of losing control [47].

3. Studies of Contemporary Control Interfaces

There is a natural gap between innovations presented by the research community and products available for use by the interested public, as the progression of research-results to market-readiness often requires resources not available to scientists or is not intended in the first place. So, in order to get realistic insights on the current situation and future requirements by affected people that live with limited mobility of their upper limb, one needs to analyse contemporary concepts of control. This does not necessarily limit the options to products already on the market, but instead spans to prototypes close to realistic products that could reasonably be employed in a large scale within a short time frame (e.g. 1–2 years).

Figure 4: Demonstrator of the AdaMeKoR project with the patient-operated smaller robot on the left (main topic of this thesis) and the caregiver-assisting larger robot on the right. Courtesy of DFKI

The project *AdaMeKoR (An Adaptive Multi-Component Robot System for Nursing Care)* [XI][3] aimed to develop a multi-component robotic bed for care (see Figure 4), consisting of a large manipulator to aid caregivers in repositioning patients and a small robot to be controlled by the patients themselves. This thesis focusses on the latter. Motivated by the unsatisfying reception of the classic manufacturer-provided control interface (see Section 1 and Figure 2), this project evaluated 3 additional contemporary concepts of control for the smaller robot, as shown in Figure 5.

As the literature lists mode switches as one of the major concerns of the default joystick (Figure 5a), the assessed interfaces have been selected specifically to mitigate and analyse this concern:

- Deployed by the gaming industry, gamepads (e.g. Figure 5b) are simple and publicly well known input devices that have been designed to be relatively cheap and ergonomic for able-bodied users. With numerous digital buttons and continuous axes (in our case 10 buttons and 8 axes[4]), they easily allow to control all DoFs of the robot without the need for mode switches. However, they require the use of two hands and a fair bit of dexterity.

- De facto originally developed for the control of robots [21], 3D Mice (Figure 5c) are currently best

[3] *Ein adaptives Mehrkomponenten-Robotersystem für die Pflege (AdaMeKoR)* was funded by the German Ministry for Education and Research (BMBF), March 2020 until September 2023. `http://adamekor.de/`, last visited 10th December 2024

[4] *Xbox One Wireless Controller* by Microsoft Corporation. `https://www.xbox.com/en-gb/accessories/controllers/xbox-wireless-controller-adapter-windows`, last visited 10th December 2024

(a) (b) (c) (d)

Figure 5: Contemporary User Interfaces used in Study 1 and 2: Kinova Joystick (a), Gamepad (b), 3D Mouse (c), Voice Control Interface (d). Adapted from [XI]

known for their use in Three-Dimensional (3D) design and construction[5]. Their main input element can be pushed and rotated in arbitrary directions, thus offering 6 DoFs, accompanied by two buttons on the sides. This device is designed to be used single-handedly but requires substantially more dexterity than gamepads.

- Finally, a voice control interface was developed, as it was specifically requested by early participants. To be in-line with the others, this interface was deployed on a readily used smart home device[6] and uses mostly manual pose-relative commands (*up*, *down*, *rotate left*, *close gripper*, . . .).

In this setup, none of the interfaces relies on sensory information of the environment, making them all completely manual as well as capable of general purpose application. Still, each device focusses on specific design principles, such as simple but numerous buttons for the gamepad, or a single complex *button* for the 3D mouse. See the appendix Section B for button mappings and control overviews of all input devices. Exemplarily videos showcasing the Kinova Joystick and 3D mouse are available online[7].

In contrast to many works in literature, the project relied heavily on participatory development and integration of the final user group in order to achieve reliable results. For this purpose, two studies were conducted: An initial study of preliminary experiments at the REHAB trade fair[8] with $N_1 = 26$ individuals with limited upper limb mobility, care professionals, and secondary users (Study 1), and an extensive study with $N_2 = 15$ mobility-impaired participants, carried out in their own homes or a living lab [2] (Study 2).

During both studies, the users controlled the table-mounted assistive robot arm (left in Figure 4) using a subset of the presented interfaces for defined scenarios. These were a simple pick-and-place task for training and a realistic task of pouring water into a cup. After completion, the participants were asked to freely compare the interfaces, thereby indirectly defining requirements towards such systems. As quantitative data analysis for this highly diverse user group was unreasonable and the only important metric for these interactions has to be user acceptance in any way, the within-subject evaluation focussed mostly on qualitative data. Nevertheless, Study 2 includes a novel evaluation strategy for similar surveys, by manually labelling each trial post-study into action sequences (cf. [III]), mindset-states (*Thinking*, *Controlling*, *Being Assisted*), and rating a task's success on a point-scale.

The two studies mainly differed by their scope in terms of time spent per user. In order to attract more trade fair guests during Study 1, the setup was largely unrestricted, with both tasks being relatively undefined and the task of pouring water only feigned with an empty bottle. This led to shorter turnaround times and a higher number of engaged participants. However, this loose design also inhibited most

[5] *SpaceMouse Wireless* by 3Dconnexion GmbH. `https://3dconnexion.com/uk/product/spacemouse-wireless/`, last visited 10th December 2024

[6] *Echo Dot* by Amazon.com, Inc. `https://www.amazon.de/dp/B084J4MZK6`, last visited 10th December 2024

[7] Videos of the Kinova Joystick and 3D mouse available at `https://www.informatik.uni-bremen.de/agebv/DoF-Adaptiv#Comparison_of_Control_Concepts`, last visited 10th December 2024

[8] *REHAB* trade fair. `https://www.rehab-karlsruhe.com`, last visited 10th December 2024

quantitative data recording. For the succeeding Study 2, these issues were addressed by allowing for more time with a single user, a highly structured study design, and realistic pouring of water in the second task. Participants of Study 1 tested all 3 physical control interfaces (a, b, and c in Figure 5), whereas the gamepad was replaced with the voice control in Study 2. Also, participants of Study 2 used only 2 randomly selected devices. This was due to the more time-intensive structured study design and direct user-feedback from Study 1 requesting a voice interface.

The detailed results of both studies can be found in the designated publication [XI]. In summary, they show the 3D mouse to be the clear user-favourite, with this type of step-wise voice control rated worst. While the gamepad was rated second-best in Study 1, it required fine-controlled use of all fingers of two hands — a prerequisite not existing for the other devices and mostly not met by participants of Study 2. In comparison to the classic manufacturer-provided joystick, the main benefit of the other devices was the lack of control modes. The direct mapping of controls (especially for gamepad and 3D mouse) was reported to be extremely helpful, as users could learn motions by heart, akin to muscle memory; Where pushing the joystick of the classic control in one direction can, depending on the mode, result in either a translation, rotation, or change in the gripper, a similar input on one of the other devices always has a consistent result due to the lack of modes. For the 3D mouse in particular, an active user of the arm replied that it 'is easier and requires less mental changes as there are no mode switches' [XI]. With this, they would use the arm daily, instead of asking for assistance by another person, as they currently often do due to the currently existing cumbersome interface.

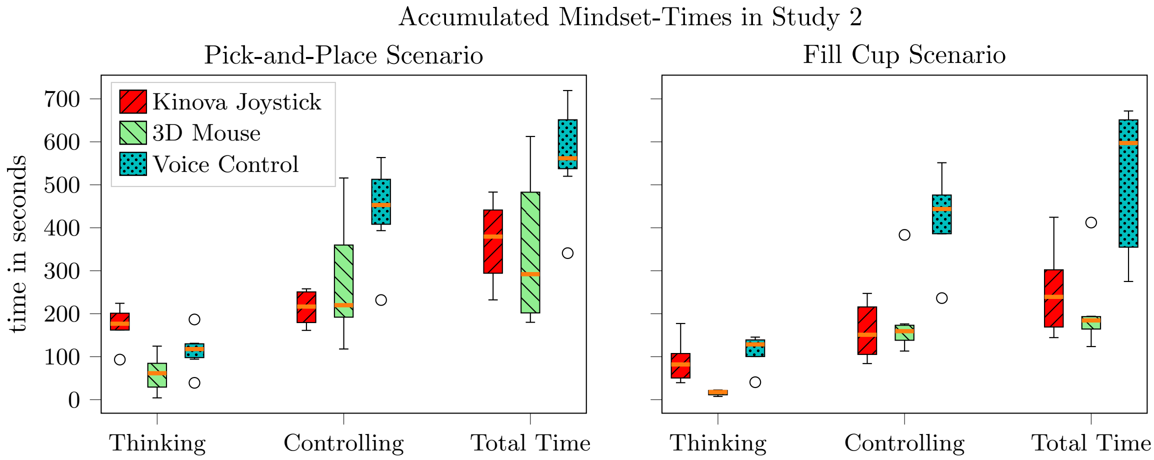


Figure 6: Comparison of Accumulated Mindset-Times in Study 2. Adapted from [XI]

.

The results of Study 2's labelled sequences also showed an interesting dynamic of time spend *Controlling* versus *Thinking* (see Figure 6). In this study, all users started with the pick-and-place scenario, followed by filling the cup. Generally, participants spend a larger percentage of time *Controlling* with the 3D mouse, thus being more explorative and less obstructed by the interface. Interestingly, the time each user spend during each task is distributed very differently, depending on the input device. During the second task, not much change can be seen for the classic control, while the variance of times with the 3D mouse drastically reduces, with its mean being significantly lower than the others. This can be an indicator for a very quick training effect with this more complicated but capable device, as especially the time spend exploratively controlling reduced for this interface.

Regarding the voice control requested by subjects of Study 1, participants were quickly frustrated by the mentally taxing and rushed interaction with this general-purpose design of step-by-step commands. Also, observations during the associated trials showed difficulties of the users verbally breaking down the tasks into commandable steps. This was intriguing, as the same users had no issues with the physical interfaces. So for such a system to be applicable, some kind of autonomy-assistance is necessary.

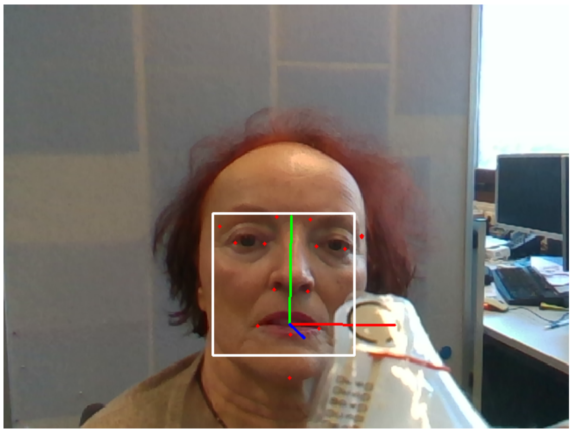
Overall, the studies show a market for more complex user interfaces (e.g. the 3D mouse) that allow the users to improve themselves and gain proficiency with more capable devices. Also, these studies support the results of Holloway that users need to be able to select from various interfaces and no one-size-fits-all exists [24]. Therefore, wherever possible, interface design should focus on user individually instead of further sub-categorising people by symptoms [XI].

3.1. Evaluations in Activities of Daily Living

In order to properly further evaluate strategies of assistive control, they need to be tested in scenarios that are both realistic and relevant to the target audience. As part of the projects *MobILe (Physical Human-Robot Interaction for a Self-Determined Life)*[9], AdaMeKoR, and *DoF-Adaptiv (Adaptive DoF Embedding as Cooperative User Interface for an Assistive Robot)*[10], a collaboration of researchers, care-receiving and care-giving individuals, as well as affected people with limited limb mobility fleshed out relevant situation and scenarios in multiple workshops and interviews [I, XII, II].

In short, the developed scenarios were *Eating and Drinking* (including *Filling a Cup*), *Opening and Closing Doors*, *Preparing Food*, *Shopping for Groceries*, and generally *Picking up Objects* [XII, III]. As the literature also listed eating and drinking as being predominant [9], the latter was isolated to be the main focus of the MobILe project: Developed as an extendable robotic software (akin to a smartphone app), this produced a partially autonomous robotic assistant for drinking [I, VII], the use of which can be seen in Figure 7[11].

(a)



(b)

Figure 7: Assistive Drinking using the Partially-Autonomous Robotic Drinking Assistant. Adapted from [VII] © 2019 IEEE

Since drinking is a highly personalised and contact-intensive task, if not performed with a straw, this robotic assistant assured to keep the user in constant control and attempted to run as a supporting function that could be always-on. Targeted specifically for users with tetraplegia who can only move their head, this device detected the pose of its user's face during approach, and the force of lips during contact.

[9] *Physische Mensch-Roboter-Interaktion für ein selbstbestimmtes Leben (MobILe)* was funded by the German Ministry for Education and Research (BMBF), July 2017 until June 2021. `https://www.interaktive-technologien.de/service/ergebnissteckbriefe/ergebnissteckbriefe-ara/ergebnissteckbrief-mobile`, last visited 10th December 2024

[10] *Adaptive Freiheitsgradeinbettung als kooperatives Userinterface für einen Assistenzroboter (DoF-Adaptiv)* was funded by the German Ministry for Education and Research (BMBF), February 2021 until January 2024. `https://www.interaktive-technologien.de/service/ergebnissteckbriefe/meki/dof-adaptiv`, last visited 10th December 2024

[11] Video of the Robotic Drinking Assistant available as Supplemental Items at `https://ieeexplore.ieee.org/abstract/document/8779521/media#media`, last visited 10th December 2024

Both interactions were used to give the user maximum control during operation and automatic retrieval and abort functionality, if requested by the user.

The robotic assistant was tested in two small studies thereby separately evaluating serving the drink (i.e. engaging in user contact) and actually drinking from the cup. Only the second study regarding the latter (Study 4 with $N_4 = 16$) is part of this thesis. The combined results showed the robots high reliability in safety-critical situations as well as an easy and intuitive concept of control for both sub-tasks [VII].

3.2. Robot Control: Manual Control versus Autonomous Operation

Over the course of this thesis, 3 major studies were conducted with different primary and secondary users of assistive robots. Of these, Study 1 and 2 were previously introduced, whereas Study 3 will mainly be presented in Section 5.2. Each of the studies involved individuals with limited upper limb mobility evaluating different concepts of control with an assistive robot arm on given example tasks. Solely the preliminary Study 1 extended the criteria to include care professionals (e.g. therapists) and secondary users (e.g. caregiving family members).

While the focus of the studies varied and concentrated on the singular interfaces used, they all concluded with open questionnaires regarding the general use of the technology and the participants' preference towards automation. This was done to serve as guidelines for future developments and to assess the often claimed user-aversion towards automation (cf. [31, 58]). For this, the participants could freely elaborate their opinion, before finally positioning their preferred level of automation on a scale from 0 (complete manual control) to 10 (complete automation). The response distributions are shown as box plots in Figure 8. Here, the responses of the different user groups in Study 1 are shown as stacked bars, with care professionals in striped green and secondary users in dotted red.

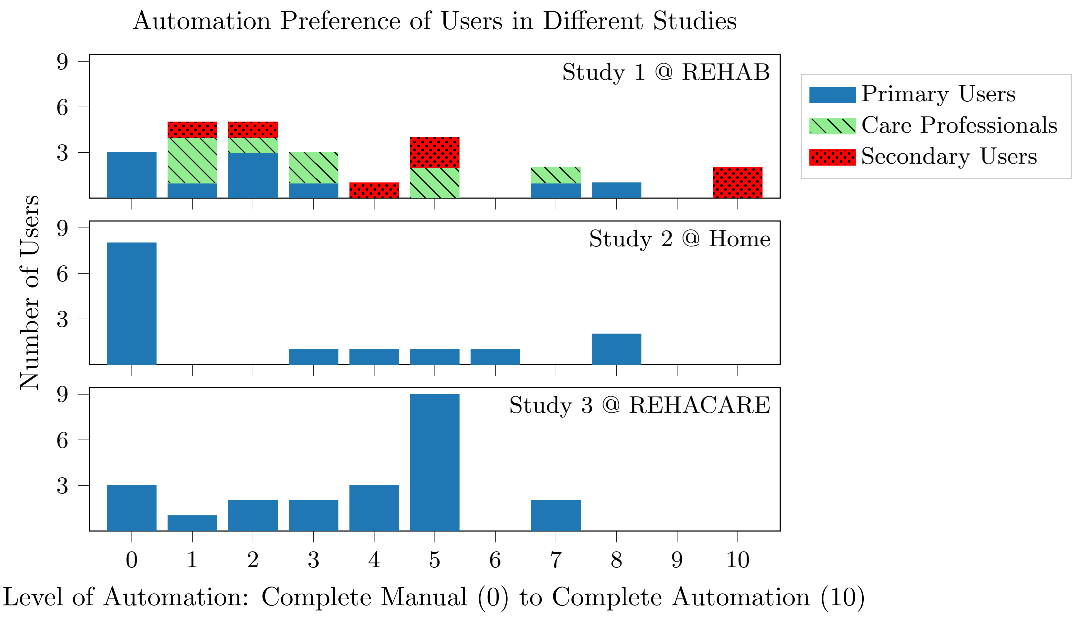


Figure 8: Automation preferences of users in different studies, including the preliminary Study 1 at REHAB ($N_1 = 26$, top), the AdaMeKoR final evaluation Study 2 at users' homes ($N_2 = 15$, centre) and the DoF-Adaptiv final evaluation Study 3 at REHACARE ($N_3 = 24$, bottom, adapted from [VI] © 2024 IEEE)

Though differently distributed over the studies, the data shows a clear preference towards the manual control, with only 2 secondary users opting for complete automation and various users selecting a complete manual control, especially during Study 2. Interestingly, the data of Study 3 has a peak at the midpoint between manual control and complete automation, however with most other responses being evenly spread out over the lower levels of automation [VI]. These differing distributions can most likely be ascribed to the varying input devices and control methods seen by the study participants: Study 1 gave only brief glimpses into 3 manual controls, thereby rather capturing users' initial perspectives, whereas Study 2 provided the subjects with extensive training time, which allowed them to gain proficiency and assertiveness with the devices. Finally, the subjects of Study 3 were introduced to an assistive shared control (see Section 4), which may have given them a better impression for the possible spectrum of automation. As requested, the participants extensively elaborated on their preferences and associated reasons. The following direct quotes were taken in German and freely translated.

Multiple affected primary users (8) stated their desire for independence, which explicitly includes the ability to 'decide *how* something is done'[12]. For some, this coincided with a reluctance of being served, independent of whether this is done by a robot or human. One participant stated that 'if the robot were to do that automatically, [they] could just as well position a nurse there'; which would violate the purpose of the robot being an extension of their paralysed arm[13].

In addition to independence, multiple participants (5) referred to their own fitness, both mentally and physically, being measured and held up-to-date by such applications. One said that 'everything I delegate does not strain me. What does not strain me, lets me expire'[14]. The general idea of this was independently confirmed by the participating therapists and care professionals (7), who explicitly recommended manual control as it supports mental coordination.

Others (7) were cautious of automated controls, deeming them 'cool but prone to errors'[15]. This included cautions regarding the arm behaving unexpectedly, the interface not understanding the user's intent, and the system generally missing necessary customisations, thereby once again missing its target of increased independence. Here, the consensus was often an opt-in option for automation, possibly including the ability to record reoccurring sequences or macros. This was generally represented by mid-range levels of automation (4–6 in Figure 8). During Study 1, this line of thought led to the request for a voice control interface, though also step-wise voice controls were requested twice.

The participants who chose a higher rate of automation often struggled with the manual interfaces. One reported requiring 'a lot of effort for small movements', thereby losing motivation[16], while another saw more opportunities for errors. Finally, the two secondary users opting for complete automation compared this to BCIs and deemed it scary, but beneficial if necessary[17].

In short, the participants require a save manual fallback option in case of any form of automation. Also, every increase of automation needs to be precisely evaluated with reference to a user's remaining capabilities in order to avoid forms of degeneration and retain independence. However, a non-automated support for manual control could greatly benefit usability and error-avoidance of these important devices.

[12] Original German quote: 'Selber entscheiden *wie* etwas klappt.'

[13] Original German quote: 'Wenn der Arm das automatisieren würde, könnte ich mir auch eine Pflegekraft dazustellen. Für mich ist der Arm die Verlängerung meines (gelähmten) Arms.'

[14] Original German quote: 'Alles was ich delegiere, fordert mich nicht. Was mich nicht fordert lässt mich verfallen.'

[15] Original German quote: 'Abläufe programmieren wäre auch cool, aber zu fehlerintensiv.'

[16] Original German quote: 'Viel Aufwand für kleine Bewegung. Wenn ich keinen Becher selber einschenken könnte, hätte ich auch keine Lust das mit dem Roboter zu machen.'

[17] Original German quote: 'Chips sind gruselig, aber es wäre schön Leuten damit zu helfen und Interaktionen zu ermöglichen.'

4. Adaptive DoF Control

As fleshed out in Section 3, a control for an assistive robot needs to be applicable in various different situations (Section 3.1) and keep a focus on the users' own contribution to the interaction. In the best case, users are provided with a sense of self-determined autonomy coupled with a manual fallback option to alleviate their concerns regarding safety and system failures (Section 3.2). Furthermore, such a control needs to be sufficiently modular in order to account for the diversity of physical user capabilities, preferably by being deployable with varying input devices. In contrast, it does not necessarily have to be perfectly intuitive or easy to use off-the-shelf, but can instead evolve with experience; As the users of the previous studies stated, they have sufficient incentive to train and reach proficiency with an interface, especially if the process ends up giving them more options and capabilities. However, simply introducing more complex and capable input devices is not expedient, as many users face physical difficulties with high-dimensional input options (e.g. 3D Mice). For these people, shared control can be extremely beneficial as its perks naturally scale with a user's physical limitations.

Basically, the largest control discrepancy arises when using an input device with a low-dimensional output signal $c \in \mathbb{R}^m$ to operate a robot with a high-dimensional steering interface $u \in \mathbb{R}^n$, i.e. $m \ll n$. (E.g. using a chin-joystick with $m_{\text{chin-joystick}} \leq 2$ to operate the *Jaco* with $n_{\text{Jaco}} = 7$). The shared *Adaptive DoF Control* focuses on bridging this gap by sticking to the established concept of modes, which is well-known by users, and adaptively extending their capabilities based on the current situation [IV]. In the context of this thesis, a mode is a mapping of input-DoFs to robot-commanding output-DoFs, i.e. a matrix $D \in \mathbb{R}^{n \times m}$, such that

$$u = D \times c. \tag{1}$$

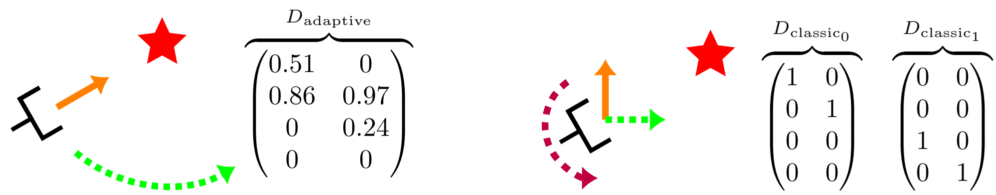
Usually, a mode will be an element of a defined set of modes $\hat{D}$, the sum of which allows users to command the n-dimensional robot to reach any pose in their specifications, only by operating their m-dimensional input device ($m < n$). As an example, Figure 9 shows the manufacturer's set of modes for the *Jaco* in alignment to Figure 2, with $n_{\text{Jaco}} = 7$ and $m_{\text{Kinova Joystick}} = 3$.

$$\begin{array}{c} \\ \text{X-Axis} \\ \text{Y-Axis} \\ \text{Z-Axis} \\ \text{Roll} \\ \text{Pitch} \\ \text{Yaw} \\ \text{Gripper} \end{array}
\overbrace{\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}}^{\text{Translation Mode}}
\overbrace{\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}}^{\text{Rotation Mode}}
\overbrace{\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}}^{\text{Finger Mode}}$$

Figure 9: Static Set of Modes $\hat{D}_{\text{Jaco}}$ of the Default Control on the Kinova Joystick in Figure 2. Adapted from [IV]

As with most available input devices, the embedded DoFs are predetermined, invariable, and aligned to a fixed coordinate system. In the case of the Kinova Joystick, the translational DoFs are relative to the user (i.e. in a world-reference frame), whereas the rotational DoFs are gripper-centric. Within this thesis, such axis-aligned DoFs are referred to as *cardinal* DoFs. The given set of modes allows the robot to be positioned in any arbitrary pose within its reach. However, on the way there, it cannot be moved along arbitrary paths, as users are limited to the mode-imposed subset of cardinal DoFs. While this sounds pedantic, it drastically limits the possible interactions and slows down or effectively rules out certain operations. In practice, non-cardinal motions are often required, and able-bodied users rarely move their arm along cardinal directions. For example, with the 3 modes presented in Figure 9, users have to decide to move the arm either translational or rotational. If the users were now to approach an object, say a cup, with the robot, they would not be able to simply orbit the cup to select an orientation to approach, but need to instead iteratively position the arm to one side and change the mode to reorient the gripper. Similarly, pulling a door is simply not possible, as it requires the arm to move along an arc, i.e. following a motion persisting of synchronised translational and rotational components.

Adaptive DoF Control follows the same principles of modes, but is able to provide DoFs that are dependent on the situation and *non-cardinal*. To allow a wide range of motions, including screw motions, the adaptive DoFs are also relative to the gripper; albeit possibly visualised differently for an easier user understanding. With this, it aims to provide a set of DoFs that is more helpful in the current situation than the classic standard options. A comparative example of a Two-Dimensional (2D) robot is shown in Figure 10, which describes approaching an object (red star) using a 2-DoFs input device. In alignment with the configuration of the Kinova Joystick shown in Figure 9, the exemplary presented translational DoFs are in numerical reference to the world, whereas the rotational DoFs are gripper-centric. In Figure 10a, one DoF of the current adaptive mode D_{adaptive} (bold orange arrow) could represent the straight diagonal path towards the object, with the second DoF (dotted green arrow) allowing the user to orbit the target to approach it from a preferred direction.



(a) Adaptive DoF Control

(b) Classic Control

Figure 10: Comparison of Adaptive DoF and Classic Control for a Simple Approach Task with Numerical DoF-Mappings to their Relative Sides

With the classic static control, two modes (D_{classic_0} and D_{classic_1}) are necessary to reach a similar functionality. This is shown in Figure 10b, where one complete mode of two cardinal DoFs (*up*, bold yellow and *right*, dotted green) is required only to allow the user to command the straight diagonal path towards the object. In addition to that, commanding the orbiting operation requires not only a second mode with the *rotation*-DoF (dashed purple), but also repeated mode switching by the user, as this path requires interactions of all 3 cardinal DoFs.

4.1. Demonstrator

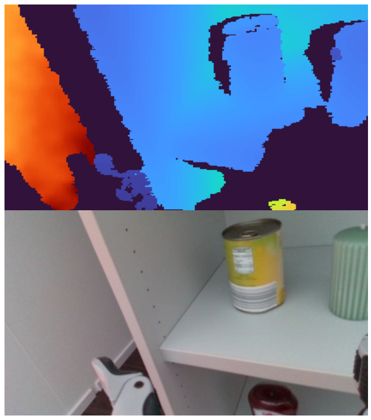
The adaptive DoF control can be thought of as an extension to the classic mode control: In Figure 3, the classic control pipeline is displayed with bold black arrows, with the adaptive DoF control being added on top in red. Basically, it redefines the previously fixed set of available modes $\hat{D}$ to be dependent on the environment, as represented by the measurements z (including e.g. robot states and camera images). The contemporary implementation of this concept sees an AI-system generating the set of modes $\hat{D}(z)$ based on sensor data z.

Figure 11 shows a demonstrator system for this control: A 7-DoFs WMRA[18] with a wrist-mounted colour-and-depth camera[19] to sense its immediate surroundings. This lightweight add-on has basically no impact on the robots capabilities while capturing the most relevant scenes and avoiding clunky external sensors limiting mobility and user acceptance. An embedded computer on the wheelchair holds the AI and communication stack.

[18] *Jaco 2* by Kinova (7-DoF research version) [34]. `https://assistive.kinovarobotics.com/product/jaco-robotic-arm`, last visited 10th December 2024

[19] *RealSense D435i* by Intel. `https://www.intelrealsense.com/depth-camera-d435i/`, last visited 10th December 2024

(a) Demonstrator



(b) Camera Data

Figure 11: Physical Demonstrator used for Adaptive DoF Control (a) with exemplary Camera Data (Depth (top) and Colour (bottom)) (a)

In addition to being a visualisation interface, the main input device of the demonstrator is a set of head-worn smart glasses[20] with an interaction design based on an existing driving system for assistive wheelchairs [45][21]. This original driving system uses distinct head nods for menu-navigation (such as mode selections) and smooth head tilting as continuous joystick input to accelerate (tilt forwards) and rotate the wheelchair (tilt sideways). However, the developers of the drive system cautioned, based on their experience, the use of the forward-tilting interaction for the robot control. Even in their original wheelchair driving interface, driving backwards is only supported separately. With this, the smart glasses effectively only have a single DoF ($m_{\text{glasses}} = 1$) for the adaptive control. Alternatively, the modular system integration also allows to replace the glasses as input device, e.g. with a classic joystick, and use the glasses only for visualisation.

4.2. Control Synergy

In application, adaptive DoFs only expand on the cardinal DoFs of the classic control. This assures a fallback option for the user in the inevitable case of miss-detections or completely unknown situations. Also, once a user has selected a mode, its associated DoFs stay fixed. While this may initially seem counterintuitive, as the system thereby looses opportunities for on-the-fly error correction, it is essential for usability: A system constantly adapting its active DoFs would be unpredictable and therefore unsafe for the user, as sudden changes in the controlled DoFs can cause the robot to jerk and behave irrationally.

Instead, an accompanying visualisation informs the user about their options for new adaptive DoFs. This was mostly arrow-based, with details depending on the specific User Interface (UI). An example can be seen in Figure 13. The user can, at any point, decide to remain in their currently selected mode, switch to newly updated adaptive DoFs, or fall back to classic cardinal DoFs. An example interaction is shown in Figure 12, again with a 2D robot, but this time with two red target stars and only a single adaptive DoF. The interaction is shown at three points in time: At t_0, the robot is far to the left of the targets leading the user to command driving closer towards them, either using the suggested adaptive DoF, or the cardinal *right*. At t_1, the adaptive system suggests a rotation to select a target, which the user initially ignores in favour of continuing moving right. Just shy of t_2, the user has selected the still suggested rotation and turned left, which in itself leads to a new adaptive DoF at t_2 that allows a target-oriented approach.

[20]*Google Glass EE2* by Google. `https://developers.google.com/glass-enterprise`, last visited 10th December 2024
[21]*munevo DRIVE* by munevo. `https://munevo.com/en/munevo-drive`, last visited 10th December 2024

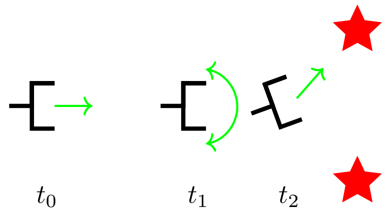
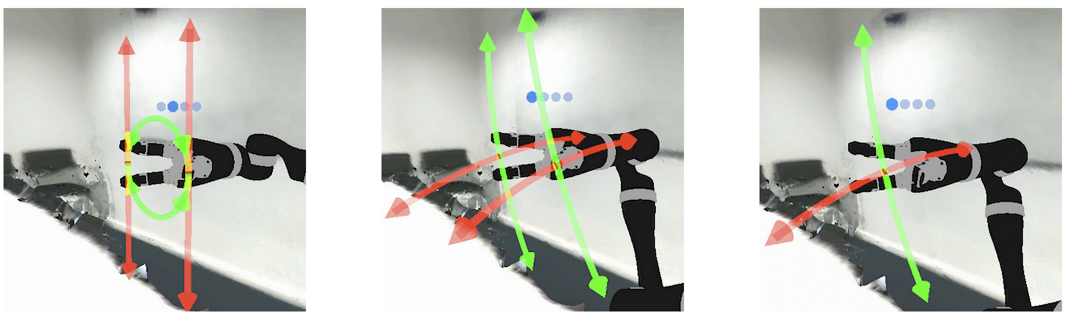


Figure 12: Example of Adaptive DoF Control: Two red star-shaped targets and a 2D-robot at 3 points in time. For each point in time, a green arrow represents the currently suggested adaptive DoF. Adapted from [V]

This level of AI is obviously not sufficient for human-independent robotic autonomy. However, as discussed in Section 3.2, that would not be preferable in any case. Instead, the AI collaborates with the user and basically empowers them to accomplish the task at hand easier and faster.

4.3. Usability

This novel concept of control was preliminarily tested in two studies (Study 5 [VIII], and Study 6 [X]) within a custom 3D Virtual Reality (VR) simulation environment [IX], which included a setup similar to the demonstrator shown in Figure 11. Focussing on the control's general usability, both studies examined human-robot interfaces and, more explicitly, communication of robot intent (i.e. suggested DoFs). Especially for such a novel control, a clear and, if possible, unambiguous UI is essential to avoid the mode- and button-induced confusions reported in Study 2 (see Section 3). For these trials, the adaptive DoFs were not supplied by a general-purpose system (as presented in Section 5), but instead simply scripted for the given situations.



(a) (b) (c)

Figure 13: DoF Visualisations in VR: Classic (a), Double Arrow (b), and Single Arrow (c). Reused from [VIII]

The first usability study (Study 5 [VIII]) assessed the general merit of adaptive DoF control, compared it to the classic cardinal variation, and explored two visualisation options shown in Figure 13. During this, the VR-controller's joystick was used as a 2D input device ($m_{\text{Study 5}} = 2$) for the system with an additional button for mode switching. The UI therefore displays 2 DoFs (red and green arrows in Figure 13) and a selection of 4 modes (blue circles in Figure 13). The study involved $N_5 = 39$ able-bodied participants in a remote study designed within-subject, during which each participant performed a simple pick-and-place task 24 times (8 classic + 8 per adaptive visualisation type) with the order of controls being randomised. For each trial, the Task Completion Times (TCTs) and number of mode switches were measured. In addition, the users provided a subjective workload measurement for each control and visualisation after completion.

The results show no significant differences in subjective workload or TCTs, although there seems to be a training effect, as the 2-arrowed adaptive variation was the overall fastest, if participants initiated with the single-arrowed version. The number of mode switches on the other hand showed significant differences, as there was a clear reduction with the adaptive controls independent of visualisation type.

Subjectively, users reported difficulties understanding the 3D arrows and ascribed this in part to poor angles or them being concealed by the robot. In addition, participants reported the adaptive arrows to include more rotations and thereby being slightly harder to understand than the classic ones, eventually leading to an increase of mental load [VIII].

The second usability study (Study 6) [X] followed a similar setting and scenario, but examined options to visualise changes in the suggested DoFs: Where the UI previously displayed two currently controllable active DoFs, this study explored the option to visualise DoFs suggested by the system but not yet selected or controllable by the user. In order to avoid visual clutter during this, the study reduced the input device to only control a single DoF ($m_{\text{Study 6}} = 1$). Further, by following an arrow-based visualisation similar to Figure 13c, the *new* DoF could simply be represented with the second colour.

For a similar pick-and-place task, the study compared the classic control with two variations of adaptive control: One with a continuous visual update of the new DoF, and one where the new DoF was only shown once it deviated from the current DoF to a significant degree. In case of the latter, the user was also informed of the change by an audible signal.

Study 6, involving $N_6 = 24$ able-bodied participants, saw significant reductions of TCT, mode switches, and workload when using either of the adaptive controls, with no significant differences among the two variants. A subjective user-ranking of the 3 controls was not significant. Again, users responded to require a bit of training for the adaptive control; however they also stated that eventually 'routine set in fast'.

Regarding the comparison of the threshold-based and continuous visualisation, user responses varied. For some, the threshold-based variant was simplest, as they trusted the system, reportedly did not think as much and just acted when a new DoF was available. For others, this variant 'felt too directed', with the continuous alternative giving them more freedom. Lastly, the qualitative user feedback indicated that some users occasionally felt a lack of control when the adaptive suggestions did not match their expectations. In these cases, a fallback option of a classic control could be a valuable addition. This was intentionally omitted during these two studies to focus on the adaptive DoFs. [X]

4.4. Comparison to Literature

In comparison with the literature, adaptive DoF control interaction is fairly similar to controlling with latent action spaces [39], as both approaches provide the user with the option to control situation-dependent high-dimensional movements with low-dimensional input devices. For the latent action space-control, a conditional autoencoder is trained on demonstrations of a task, with the autoencoder's latent space dimensionality being predetermined to be equal to that of the eventual low-DoFs input device (m). During training, the encoder maps the demonstrated robot motion to the low-dimensional subspace based on the current state, with the decoder reconstructing the motion from the subspace and the state. In application, the decoder is used standalone, with the user's m-dimensional control input being directly inserted to the latent space. The decoder then aims to reconstruct the robot motion based on the current state and control-input-defined latent space.

However, due to the black box nature of the latent-action-space-decoder, the robot's motion does not necessarily scale with user input, but can instead cause arbitrary movements, even given the authors' defined control measures of *Latent Consistency* and *Latent Scaling* [39]. This can cause lower usability and generally less predictable interaction for the user. In addition, the creation of the latent action space involves a manual alignment of input-DoF to action-space-DoF and is scaled to a single scenario and input device-dimensionality m. This makes it less generalisable than a situation aware adaptive DoF control.

In that sense, the adaptive control is closer to the mode switch assistance by [16], where possible user intentions are isolated and modes are suggested, based on which best disambiguates targets. However, this assistance is designed with a focus on automation of tasks, as the system blends control once it is sufficiently certain of the user's intent. This stands in vast contrast to the adaptive DoF control, which sees the user in sole control at every point in time. Further, the mode switch assistance is limited to cardinal DoFs, thereby limiting manual interaction options.

Overall, the adaptive DoF control has the potential to embed general concepts of robotic interaction and utilise them to provide the users with powerful techniques that are applicable in a multitude of situations. These concepts can be amalgamations of various different motions based on the current situation, such as approaching objects, avoiding obstacles, or operations specific to a task or certain held object (e.g. pouring liquid). Further, the integration of complete sets of modes avoids the control from being limited to a specific input device-dimensionality (m), but can instead be used with any continuous or discrete input device, without requiring retraining or adjustments. During all of this, the user remains in control, as no motion is carried out at any time without an explicit user input.

5. Generating Adaptive DoFs

The usefulness of this control is inherently dictated by the quality and suitability of the provided adaptive DoFs for the current situation. A successful evaluation therefore depends on an appropriate DoFs prediction. In a perfect world, these generated DoFs directly correlate to the user's intent, with the primary (i.e. most likely) DoF allowing them to immediately achieve their desired result. In reality however, any system can only estimate the user's intent based on sensory information of the environment.

Modelling this, let U be the Random Variable (RV) representing the n-dimensional direction of robot motion desired by the user. Also, let Z be the random variable describing the current situation as measured by the systems sensors (e.g. camera data). This allows to define

$$U|Z = z \tag{2}$$

to follow a conditional probability distribution. As such, this random variable models the direction of motion U desired by the user given the current situation $Z = z$. It is assumed that U and Z exist. This section explores two distinct approaches of initially modelling $U|Z = z$ and subsequently using it to extract adaptive DoFs.

It was initially planned to generate the adaptive DoFs end-to-end, using a state-of-the-art neural network processing image data. Surprisingly though, considering the recent successes and great performances of the community with AI, the results for this specific application were unsatisfying. Therefore, an analysis evaluated why this is the case, examining the task, model capabilities, as well as the available data. All of this is presented in Section 5.1.

As an alternative, a functional generation of adaptive DoFs was successfully developed based on the probabilistic combination of (simple) elemental behaviours. The complete system is showcased in Section 5.2, where its technical viability and general user acceptance were initially tested in a controlled lab study, followed by an extensive study with the target group in the semi-unstructured environment of a trade fair.

Both variants interpret the environment based on image-data and produce complete and holistic sets of DoFs; meaning that the sets' DoFs are respectively orthogonal and thus span the complete n-dimensional space of the robot's motion.

5.1. End-To-End Approach using Deep Learning

As evident by the current state of the art in image processing, machine-learning models such as Convolutional Neural Networks (CNNs) have immense capabilities of feature analysis, context understanding, and information extraction. As such, it seems obvious to directly generate DoFs from images using CNNs. However, this is not an established or trivial task, as the requirements implied by the control concept do not fit to existing methodologies. This includes model architectures, output-definitions, metrics, as well as available data.

5.1.1. Proof of Concept in 2D

As proof of concept and in order to gather necessary insights for this approach, a 2D simulation environment was developed [IV]. This simulation consists of a simple robot with $n_{\text{2D}} = 4$ DoF (*up*, *right*, *rotate*, *grasp*), 2 boxes, and a target mark (Figure 14), with the intentionally loosely set task of bringing the boxes to the target. During this task, the boxes could be triggered to have a physical spike to give the user an incentive to grasp from a defined side and thus diversify the interaction. The associated input device comprises 4 binary keyboard buttons (i.e. $m_{\text{keyboard}} = 2$), with a mode switch being triggered by idle time.

In alignment with the *Jaco*'s control, two modes are available respectively for both classic and adaptive control.

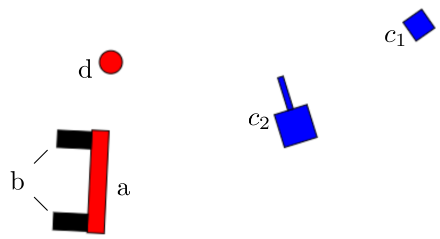


Figure 14: 2D Simulation Environment, including the Robot (a), its Gripper (b), Blue Boxes (c_1 and c_2, the Latter with a Spike), and a Red Target Circle (d). Adapted from [IV]

The data necessary for supervised learning was generated using a gamepad with 4 continuous input axes (similar to 5b) to allow operators to synchronously command arbitrary robot motion. With this, researchers recorded separate datasets for the scenario with and without spikes, where each sequence started with randomised poses for each simulation element. As defined by the simulation task, the operators further commanded the robot to grasp the boxes one after another and retrieve them to the target. A statistical dataset overview can be seen in Table 2. The data recorded consists of sequences of image-action pairs, where the image is a robot-centric top-down view of the situation (similar to Figure 14) and the action is the user-commanded relative n_{2D}-dimensional motion of the robot. The supplementary video[22] provides an overview of the interactions of all components of the 2D simulation with both classic and adaptive control interfaces.

Table 2: Statistics of the 2D Dataset

Dataset Name	Spikes	Operators	Sequences	Data Points
2D Default	No	2	392	29 927
2D Spiked	Yes	3	488	28 075

The datasets were used to train a simple CNN [IV] capable of generating adaptive DoFs based on image data available at runtime. Hereby, the model's designed purpose is not specifically simply predicting the apparently *correct* direction of robotic movement (i.e. that controlled by the user during training), but instead suggesting DoFs most likely to be controlled by the user, given the current situation (see Equation 2). This distinction is designed to allow a probabilistic treatment of the model and data.

As such, the model aims to describe the probability distribution $U|Z = z$ of DoFs most likely to be commanded, given the current situation (see Equation 2). For this purpose, $U|Z = z$ is modelled to follow a multi-variate normal distribution $\mathcal{N}(\mu_{\text{end-to-end}}, \Sigma_{\text{end-to-end}})$ with mean vector $\mu_{\text{end-to-end}} \in \mathbb{R}^n$ and the symmetric, positive definite covariance matrix $\Sigma_{\text{end-to-end}} \in \mathbb{R}^{n \times n}$. In this approach, the dataset's labels (n-dimensional motion vectors) are seen as samples of this distribution [IV].

For safety reasons, the robot arm should never move without input by the user. This coincides with the adaptive DoF control placing the user as sole origin of motion. As such, no robot movement is performed without explicit user input. This also means, that an adaptive DoF only needs to encode a directional line (e.g. left and right), without any specific preference therein. The ultimate *sign* of movement along this line, as well as the velocity of motion itself, are then directly defined by the user's control input (e.g. direction and inclination of a joystick). This is identical to classic cardinal control. With this, the

[22]Video of 2D simulation available at `https://www.informatik.uni-bremen.de/agebv2/downloads/videos/GoldauPetra21.m4v`, last visited 10th December 2024

model's output only needs to describe this directional line to be a valid DoF and is therefore independent of amplitude and sign. In practice, this means that $U|Z = z$ is zero-centric, i.e. $\mu_{\text{end-to-end}} = (0, \ldots, 0)^T$, and the model only needs to estimate the covariance matrix $\Sigma_{\text{end-to-end}}$.

To do this, the model includes a novel custom head as shown in Figure 15, which was specifically designed for adaptive DoFs: The CNN backbone (e.g. based on MobileNet [50]) and head of the model [IV] convert the sensor-generated image data to a set of motion vectors $Y \in \mathbb{R}^{n \times k}, Y = (y_1, y_2, \ldots, y_k)$, with $y_i \in \mathbb{R}^n$. A custom layer interprets these vectors as sigma-points (deterministic samples) of the zero-centric distribution described by $\Sigma_{\text{end-to-end}}$ and estimates $\Sigma_{\text{end-to-end}}$ as

$$\Sigma_{\text{end-to-end}} = \frac{1}{k} \sum_{i=1}^{k} \left(\frac{y_i}{\|y_i\|_2} \right) \left(\frac{y_i}{\|y_i\|_2} \right)^T. \tag{3}$$

Further normalisation and numerical stabilisation (see [IV]) assures $\Sigma_{\text{end-to-end}}$ to be positive semi-definite with a trace of $\operatorname{tr}(\Sigma_{\text{end-to-end}}) = 1$. The latter is used to stabilise the loss function during network convergence.

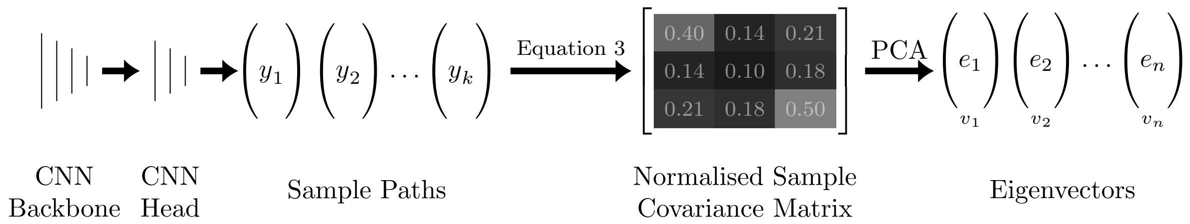


Figure 15: End-to-End Model Architecture Overview with Exemplary Data of $n = 3$. The covariance's values are colour-coded, with lighter areas having higher absolute values.

At this point, the covariance should describe the distribution of likeliness of motion over all possible directions in n-dimensional space. The model head now uses a Principal Component Analysis (PCA) to isolate main axes of the distribution, effectively generating a list of directions (DoFs) sorted by their eigenvalues (i.e. importance), which can directly be used as a complete mode set $\hat{D}$.

During training, a loss based on the Mahalanobis distance (later termed Mahalanobis-loss) was deployed to measure the representation of the label motion within the predicted covariance. Effectively, this punishes the network for generating covariances with small extents in the desired direction. Coupled with the aforementioned ($\operatorname{tr}(\Sigma_{\text{end-to-end}}) = 1$)-normalisation, this leads to the desired effect.

In addition, this architecture experimented with a *Moment Layer* as new type of pooling layer: CNNs can often be split into two parts, one operating on structured image-data (e.g. with convolutional operations) and one working on flattened dimensionless data (e.g. with fully connected layers). Usually, the transition between these two parts is performed by either flattening the image-data or pooling the image dimensions to a single element per channel. However, the former option results in a large flattened vector that makes further operations expensive, whereas the latter loses all positional information in the images, which is necessary for adaptive DoFs.

Similar to convolutional layers, the novel *Moment Layers* expand on a concept from classic computer vision: Based on Hu-Moments [25], this layer embeds information from image-like data in a flattened vector using a position-dependent weighted average. In this, all operations are strictly linear, leading to simple gradients and a quick convergence during training. As these moments are often applied in classic computer vision and are still being researched [65], they hold a large potential for CNNs. As such, the layer processes image-based data and retains positional information while still drastically reducing the size of the output vector.

Evaluation in 2D A user-centric evaluation (Study 7) of 2 successfully trained networks (one with spikes on the boxes and one without), aimed to serve as a proof of concept of the adaptive DoF control and to compare it to a classic control alternative. Here, 4 binary buttons were used to steer the robot along a total of 2 DoFs (2 buttons per DoF, one positive and one negative). In addition, an idle time of 5 seconds without user input caused a mode switch, thus directly exchanging the controllable DoFs.

For the classic control, this meant a cycling of 2 cardinal DoFs in 2 modes. For the adaptive control, each such idle time triggered an update of the adaptive DoFs based on the current situation and provided the user with those newly generated DoFs. At this stage, the adaptive DoF control did not have an option to return to previously suggested DoFs or fall back to classic controls.

The evaluation itself was conducted as an unsupervised within-subject online study with $N_7 = 23$ able-bodied participants. Here, the participants were instructed with introductory videos explaining the interfaces and control methods, followed by them completing each box scenario (with and without spikes, see Figure 14) 6 times, 3 of which with every control. In the end, the participants evaluated their experiences in a questionnaire. In order to gain an impression of training effects, a small subset of users (4) were given a more extensive training session of 10 minutes afterwards, consequently repeating the experiment.

The results show a significant reduction in measured task execution times. This was confirmed by the users who subjectively felt the adaptive control to be faster, though more complex. After the additional training, the involved users rated the adaptive control to be faster and easier, though still not quite as easy as the classic control [IV].

5.1.2. Deployment in 3D

Generating adaptive DoFs in a realistic 3D environment is exceedingly more difficult than in case of the previously shown proof-of-concept 2D environment. With respect to the demonstrator in Section 4, this transition increases the robot's DoFs to $n_{\text{Jaco}} = 7$ and involves more complex sensory systems creating both colour and depth images of diverse environments. However, general concepts, such as the requirement for a high-dimensional input device for data generation, remain. With 7 available DoFs and more opportunities for different approaches of tasks, handling the robot during recording also needs to be timely and smooth. This is especially relevant, as the group of data-generating users needs be expanded in comparison to the 2D example in order to capture a glimpse of the variety of individual robot interactions.

As an assistive robot arm would, in a perfect world, behave similar to that of an able-bodied person, the idea used for data generation was to capture the natural interaction of an able-bodied user's arm

Figure 16: Screenshot from the *AdaptiX* 3D Simulation Environment. Adapted from [III] © 2023 IEEE

and transfer this to a robot wherever possible. This way, operators can concentrate on the task at hand and are not distracted or limited by their own input device. For this, a 3D tracking and high-resolution simulation environment [IX] was developed, in which a digital twin of the demonstrator's robot arm directly follows the movements of a hand-held VR-controller[23]. As such, it allows for arbitrary movements by the user, quick recordings, and associated higher numbers of recordings and participants.

Figure 16 shows a screenshot of the simulation environment build on a realistic high-quality gaming engine[24]. With this, it is reasonable to assume a good transferability of the data to reality. This tool includes realistically behaving models of the assistive robot arm, the colour-and-depth camera, and various 3D household objects. The arm can be controlled programmatically, follow a VR controller, or a real robot in a setting of mixed-reality. This is complemented by a custom grasping system and realistically behaving doors with handles. The resulting general purpose mixed-reality framework was published free and open-source [IX][25]. The mixed-reality interactions were also partially used for the aforementioned preliminarily interaction tests of Study 5 and 6 [VIII, X].

Nevertheless, at least a small subset of data recorded with the real system is required for additional training in order to bridge the simulation-reality-gap. For this, the real demonstrator was used and controlled with a 3D mouse (e.g. Figure 5c). However, where the simulation environment allows for untrained individuals to record data, the 3D mouse requires the assignment of trained engineers for smooth and purposeful operations. This, coupled with the lower speed of a real robot, limits the recordable amount of real data.

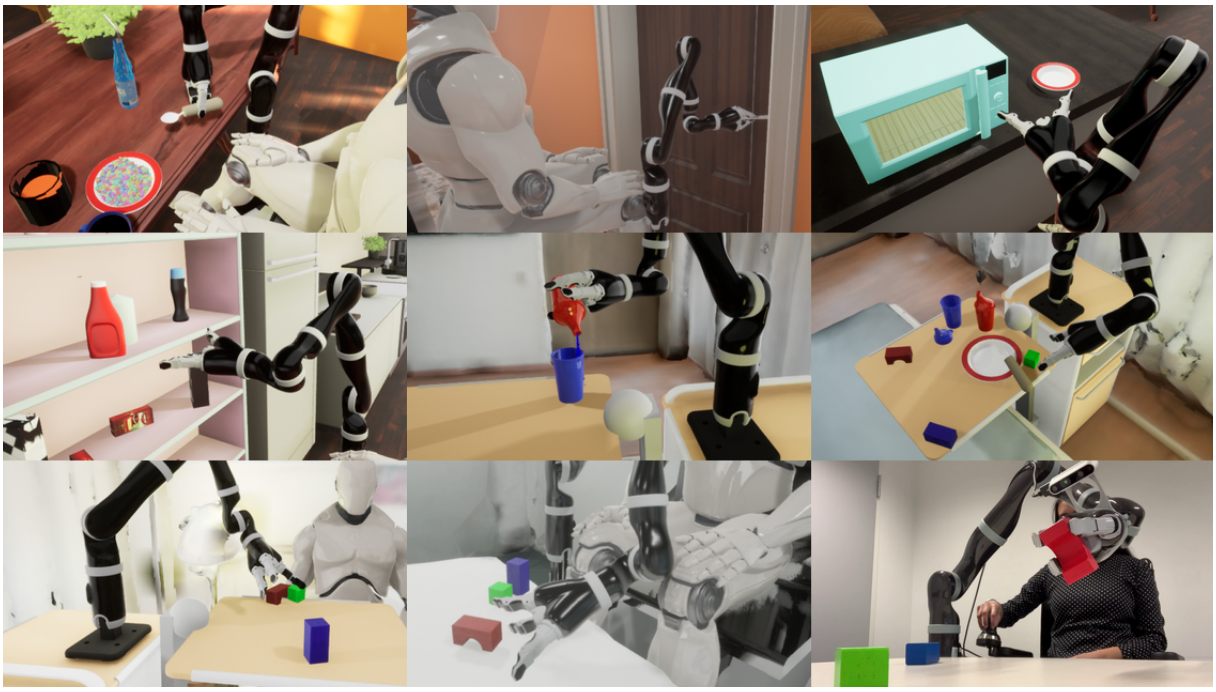


Figure 17: Overview of Scenarios in 3D. Reused from [III] © 2023 IEEE

To achieve a good representation of relevant scenarios, variations of the participatorily developed scenarios, as listed in Section 3.1, were implemented in simulation and reality. Screenshots are shown in Figure 17, with videos of all scenarios being shown in the associated supplementary video[26].

[23]*VIVE Pro 2* by the HTC Corporation. `https://www.vive.com/us/product/vive-pro2-full-kit/overview/`, last visited 10th December 2024

[24]*Unreal Engine* by Epic Games. `https://www.unrealengine.com`, last visited 10th December 2024

[25]*AdaptiX* project. `https://adaptix.robot-research.de`, last visited 10th December 2024

[26]Video of simulated scenarios available as Supplemental Items at `https://ieeexplore.ieee.org/document/10341459/media`, last visited on 10th December 2024

Table 3: Statistics of the *DORMADL* Dataset [III] © 2023 IEEE

Data Origin	Operators	Sequences	Data Points	Hours of Data
3D Simulation	38	2 973	659 k	15.4
Reality	4	986	502 k	13.9

The complete 3D Dataset of human-Operated Robot arm Motion in Activities of Daily Living (DORMADL), combines data of both simulation and reality and was published free-of-charge [III][27]. Its statistics are shown in Table 3. This large dataset embeds over a million data points in approximately 3 000 individual sequences, i.e. recordings. Each time-stamped data point includes the robot's current pose, motion, and action label, as well as sensory information of colour and depth data generated by the wrist-mounted camera (see Figure 11). In this context, action labels are manually tagged sub-sequences within a recording. They describe necessary sub-tasks performed by the robot and are generally composed of a verb and an object (e.g. *Grasping a cup*). Table 4 provides an overview of available action labels.

Table 4: Overview of Action Components. Reused from [III] © 2023 IEEE

verbs	Approach, Grasp, Let go, Push, Retrieve
objects	block, book, bottle, candle, cap, cup, door/handle, food, fork, microwave door, milk carton, plate, spoon, tea
singular verbs	Align *[cup]*, Close *[door]*, Fill *[cup]*, Press *[handle]*, Pull *[door]*
stand-alone actions	Discard, Drinking, Eating, Idle

In contrast to other datasets, the DORMADL dataset partitions the scenarios into small short-term sub-activities (termed actions in this context, see Table 4) instead of more general activities, such as e.g. eating with knife and fork (cf. [1, 6, 49]). Furthermore, each recorded scenario represents a continuous task execution, including intermediate steps to objects, whereas other datasets only store pairs of images to grasp or object poses (cf. [7, 15, 38, 61]). Lastly, the dataset's high-dimensional robot motion was generated by human operators with purposeful interactions of the task in mind and without external limitations of motions otherwise induced by obstructive interfaces.

5.1.3. Training Adaptive DoFs in 3D

The multi-facetted structure of the DORMADL dataset allows the training of very different models on the same data. Among others, these can incorporate distinct model heads for motion and action labels, and either be single-shot models or include sequential data with a time-dimension. Following the same principles as in 2D, a number of models with different sizes and architectures were trained on varying subsets of the dataset. In general, they build upon pre-existing backbones (e.g. *MobileNet V2* [50]) and mostly incorporated the previously described head for DoF-based motion prediction (Figure 15) and/or a simple classification prediction head for the action labels. The actual selection of heads depended on the attempted scope of the model.

Quality Criteria In order for the adaptive control to be a reasonable (at least functional) alternative to classic cardinal control, certain criteria must be met. Once deployed to affected, these will have to be user-centric and be mostly qualitative, such as the felt reliability (i.e. how able the user feels to complete the task with this control), consistency (i.e. seemingly continuous and predictable adaptive DoFs), and relatability (i.e. the user's ability to understand reasons behind DoFs suggestions).

Before that, more quantitative values can be acquired by analysing prediction metrics during model convergence. For this, the Mahalanobis-loss is used as well as a custom *Percentage-of-Motion metric (PoM)*.

[27] *DORMADL* dataset. `https://www.kaggle.com/f371xx/dormadl`, last visited 10th December 2024

The latter represents the percentage of the true motion (i.e. that operated during data recording) that is controllable using only the first $f \in \mathbb{N}$ adaptive DoFs, where $f < n$. It does so by calculating the squared relative length of the true n-dimensional motion projected into the f-dimensional subspace. This subspace is spanned by the respective adaptive DoFs as defined by the set $E_f = \{e_1, \ldots, e_f, v_1, \ldots, v_f\}$ of predicted eigenvectors e_i and associated eigenvalues v_i. As such, the PoM of true motion u for the first f adaptive DoFs, as defined by E_f, is given by

$$\mathrm{PoM}_f(u, E_f) = \left(\frac{||\mathrm{projection}\,(u, E_f)||_2}{\|u\|_2}\right)^2. \tag{4}$$

As the f eigenvectors form an orthogonal basis and $\|u\|_2 = \|e_i\|_2 = 1$, this can be reduced to

$$\mathrm{PoM}_f(u, E_f) = \left(\left\|\sum_{i=1}^{f} < u, e_i > \cdot e_i\right\|_2\right)^2 \tag{5}$$

$$= \sum_{i=1}^{f} < u, e_i >^2, \tag{6}$$

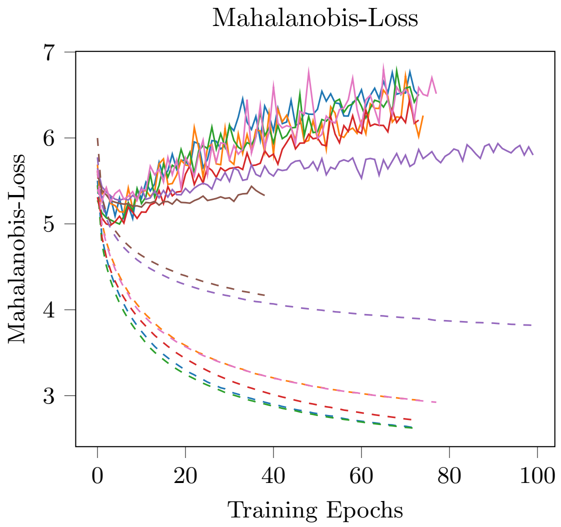
with $< \cdot, \cdot >$ as the scalar product of the two vectors.

In combination, these two metrics (Mahalanobis-loss and PoM) can be thought of as the equivalence to a Cross-Entropy-loss and Accuracy-metric in classification problems: The loss function describes the well-behaved numerical measurement used during model training, whereas the metric represents the data in a more human-readable form and comes closer to describing the actual engineer-targeted result. For the PoM, a value of 100 % would represent the user's ability to perfectly retrace the labelled movement, solely using f DoFs. If the dimensionality m of the input device is known during training, this metric can be designed such that $f = m$, thereby indirectly providing an inverse indicator on the necessary number of mode switches.

The baseline value for the Mahalanobis-loss is equal to the vector dimensionality n, as this would be the case for a perfectly spherical covariance. Theoretically, a perfect covariance representing only the true motion would have a Mahalanobis-loss value of 1. Likewise, the baseline value expected for the PoM lies at $1/n$, i.e. a fair distribution of motion along all DoFs, with an optimum at 100 %. For the 3D case, the baseline would therefore be a loss value of $n_{\mathrm{Jaco}} = 7$ and a PoM value of $1/n_{\mathrm{Jaco}} \approx 14\,\%$ per dimensionality f.

Achieved Results Numerous CNNs with a variety of architectures, backbones, heads, and hyperparameters were trained on the 3D motion data. Surprisingly however, considering the most recent general successes of machine learned AI-models, the results turned out to be poor and underwhelming. Often, the performance on test data was below the expected baselines. In practice, once a reasonable improvement could be seen on the training data during convergence, metrics on the test data diverged. Figure 18 gives an impression of this behaviour by exemplarily displaying the progression of Mahalanobis-Loss and PoM for six different models. Here, the plots of a single model's training have the same colour, with a solid line representing the evaluation on the test data and a dashed line showing the process' results on training data. Clearly, the trainings shown suffered from overfitting. Not shown in the graphs are smaller networks which did not converge.

Using the DORMADL dataset's multiple scenarios and labelled actions, it is possible to compare the achieved results within the data to evaluate if certain conditions explicitly lead to the issues. This should highlight whether specific scenarios or actions extremely over- or underperform and thereby distort the average numbers. Figure 19 shows the evaluated Mahalanobis-loss on predictions on the training and



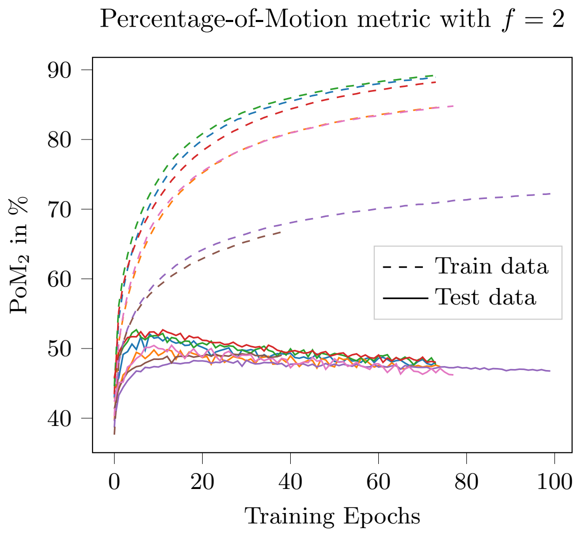


Figure 18: Validation Divergence during Model Training

test set, separated by scenario. For this, a late-stage model was used (i.e. after overfitting) in order to let extremes get more pronounced. However, the data does not show any major outliers, thereby disproving an imbalance regarding scenarios. Figure 33 in the appendix shows the equivalent data of Mahalanobis-loss per labelled action in the dataset. While the results shown there are more diverse than those of the scenarios, it is still inconclusive and has no major outliers. Some actions perform better than others, but this generally aligns for both test and training sets.

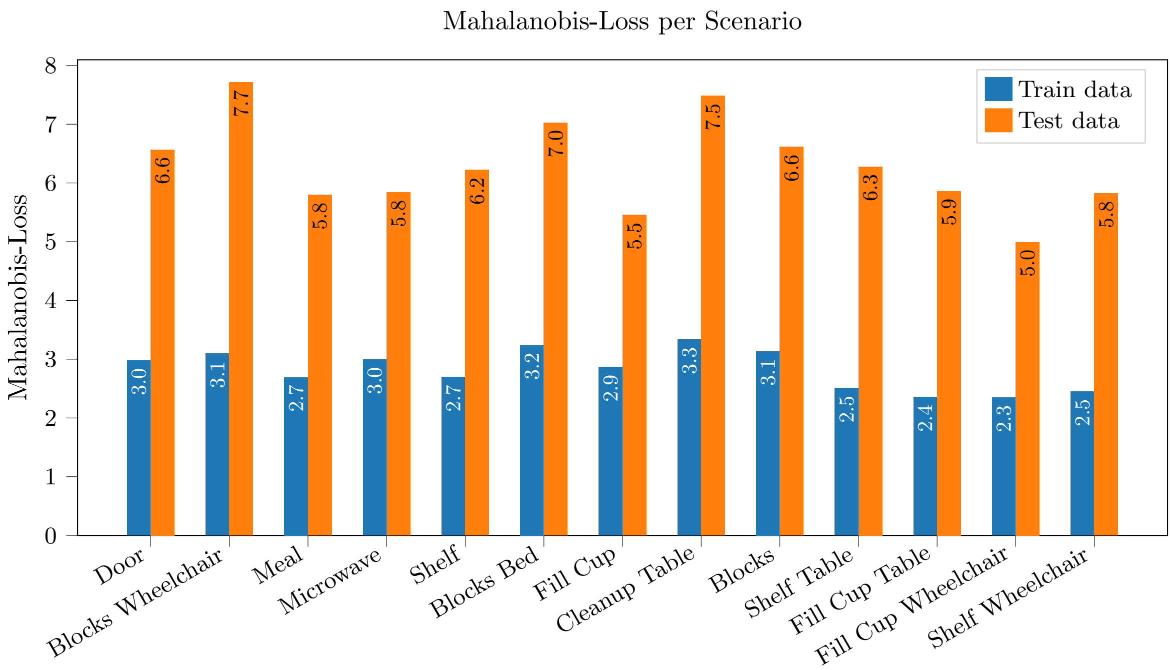


Figure 19: Mean Mahalanobis-Loss per Recording and Scenario for Test and Training Data

Running trained models on a live system to attempt an initial measurement for the qualitative metrics also turned out less than ideal: The proposed DoFs where partially shaky or erratic which led to the operator not being able to grasp the intent behind the DoFs. This naturally accumulated to being an unreliable control in total.

5.1.4. Analysis of Trainability in 3D

Based on the unexpectedly poor results achieved during training, the question arose whether the task of generating usable DoF in these settings is solvable in itself, given the provided data. However, empirically proving non-trainability is difficult. Instead, the following is an attempt to verify the complete operation chain, including kinematic transformations and robot configurations, data logging and preprocessing, as well as network architectures and associated functions. For this, the process was decomposed into 5 essential elements to be thoroughly tested in isolation. These are 1. pre-processing and preparation of data, 2. the network's input, 3. the network's ability to gather semantic information from the data, 4. the network's output definition and associated loss, and 5. the full chain.

Data Pre-Processing and Preparation (1.) The foundation for any supervised learning is the quality and preparation of available data. Inconsistencies in the provided data can directly lead to failed trainings. In this context, the network attempts to learn a distribution of motion-directions given image data of the environment. For this, the necessary motion-image-pairs are directly supplied by the used DORMADL dataset.

However, the relative camera motion could not be measured directly during dataset generation, but had to instead be calculated post-recording based on timestamps and poses in world coordinates. This, in addition to data smoothing and handling of missing data, required some non-trivial data processing. Even small issues in this pipeline can cause the output motion to significantly deviate from the truth.

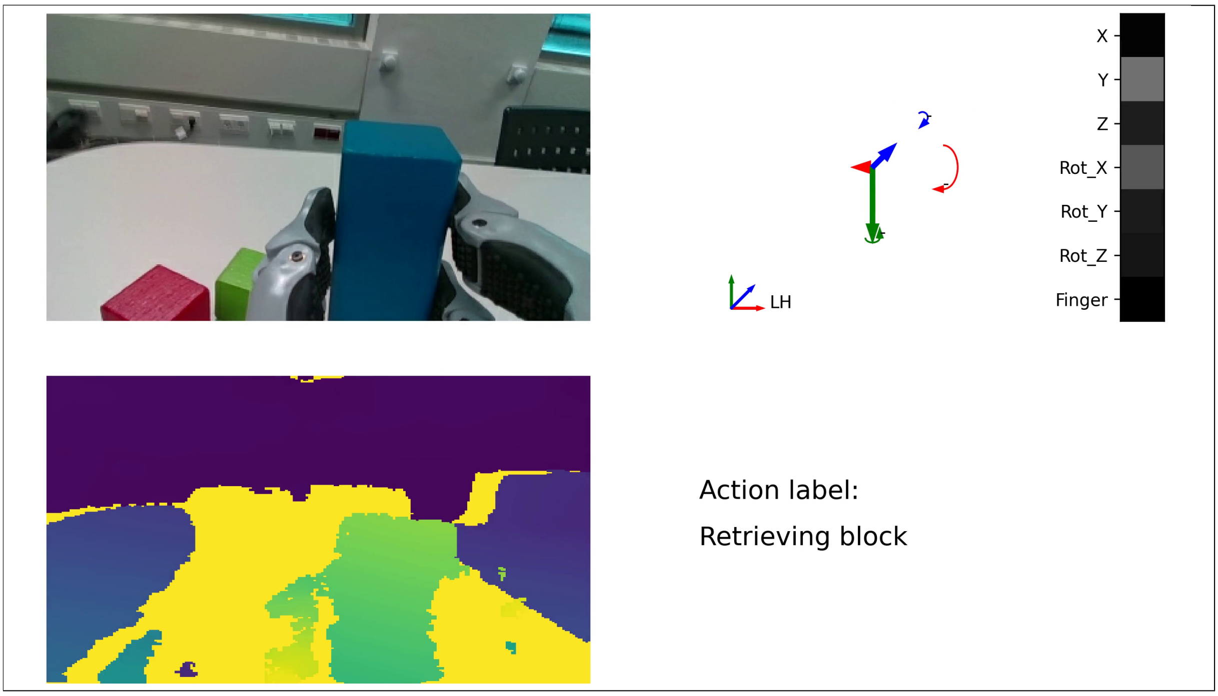


Figure 20: Motion Verification Tool: Dimensionally Isolated Direction Arrows. Adapted from [III] © 2023 IEEE

Luckily, the validity of the final motion-image pairs can be verified fairly simply as the data is sequenced and can be run as a video. This makes it possible for researchers to visually inspect the images and assess the general direction of motion. Two distinct verification tools were implemented by two independent researchers: One of these (Figure 20) visualises a single data point completely and in isolation, individually representing the dimensions of motion (except grasping) as arrows in a coordinate system (top right). This allows to subjectively inspect the directions of motion and composition of the vector. In the situation shown in Figure 20, the user is placing down the blue box by lowering the arm (green arrow) in addition to a slight pitch-rotation (red curved arrow).

The other tool (Figure 21) shows the motion vector from the dataset as arrows overlaid with the image. If everything is correct, the actual motion in the image corresponds to these arrows. Several arrows start at the same position, because the motion depends on the unknown depth there. This allows to visually verify the correctness of camera calibration, coordinate transforms and motion vector computation. In the case of Figure 21, the user is panning to the right, coupled with a slight left-wise yaw rotation.

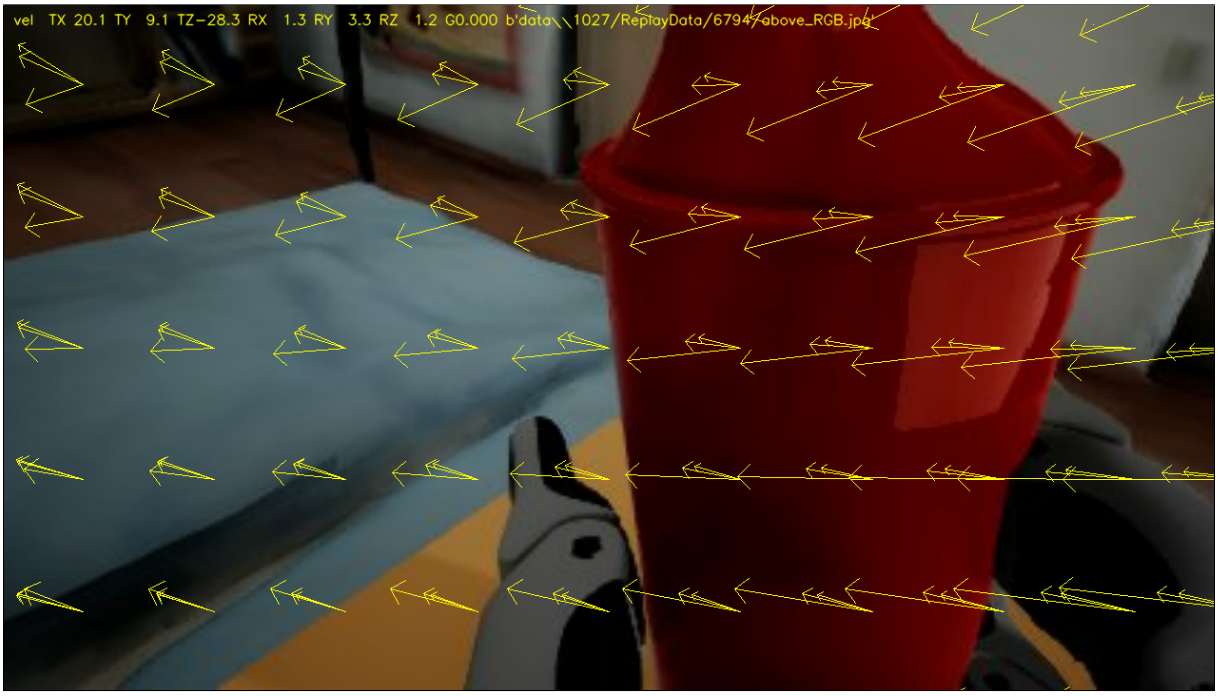


Figure 21: Motion Verification Tool: Overlaid Pixel Vectors

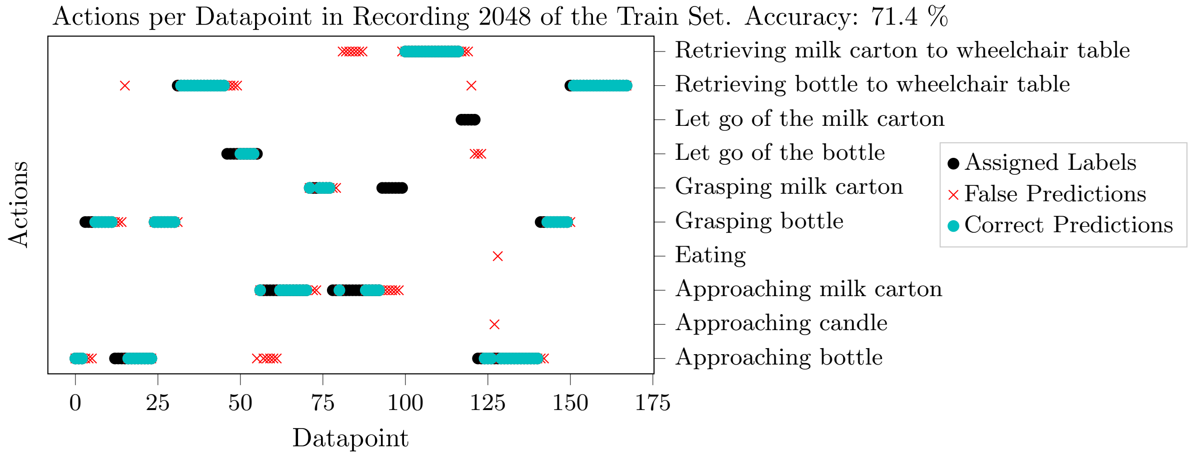
Using these tools, both researchers independently verified the plausibility of the image-motion-pairs on multiple recordings in the dataset. Example videos of both tools can be found online[28]. Following this, the data preparation is rejected as cause for the poor results.

Network Input and Subsequent Semantic Understanding (2., and 3.) Once data is prepared, the next element in the chain is having a model that can extract semantic information and feeding the data into it. Due to the black-box nature of neural networks, a loss of information in these areas can occur without major indications to the engineer. In this case, the network would not be able to identify the situation $Z = z$. This can be caused, for example, by vanishing gradients, a poorly defined network architecture, or a faulty input encoding.

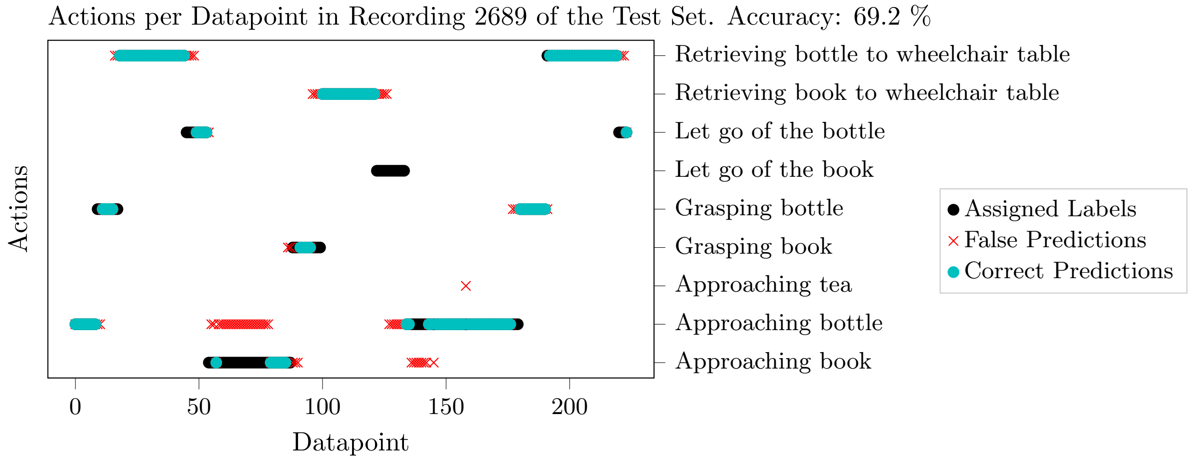
In order to specifically examine these elements on the network and data without being interfered by the output-definitions, a simple task was trained using a well-known approach: Predicting the action labels (see Table 4) based on image data. For this, a softmax head was added to the existing architectures and trained alongside the motion-prediction-head of Figure 15 with a crossentropy-loss.

Figure 22 shows the model's predictions for two recordings on a timeline. Correctly predicted data is drawn as lighter turquoise circles, whereas red crosses show prediction errors next to the true labels as black circles. The graphs show very good predicted values, with errors mostly occurring at the beginning or end of actions. This is reasonable, as the data was labelled manually and action transitions could often

[28]Videos for the motion verification tools available at `https://www.informatik.uni-bremen.de/agebv/DoF-Adaptiv#Motion_Verification_Tools`, last visited 10th December 2024



(a) Training Set Recording



(b) Test Set Recording

Figure 22: Plot of Predicted Actions for Recordings of the Training Set (top) and Test Set (bottom)

not be perfectly defined to fixed points in time. The longer miss-detection in Figure 22b (*Approaching bottle* instead of *Approaching book*) was analysed in video and ascribed to both objects being adjacent.

In short, the action label prediction was fairly successful. There was some hope that the introduction of this additional information would assist the overall model convergence by contextualising the motion into a bigger picture. This turned out not to be the case, as the metrics on the predicted motions did not significantly change.

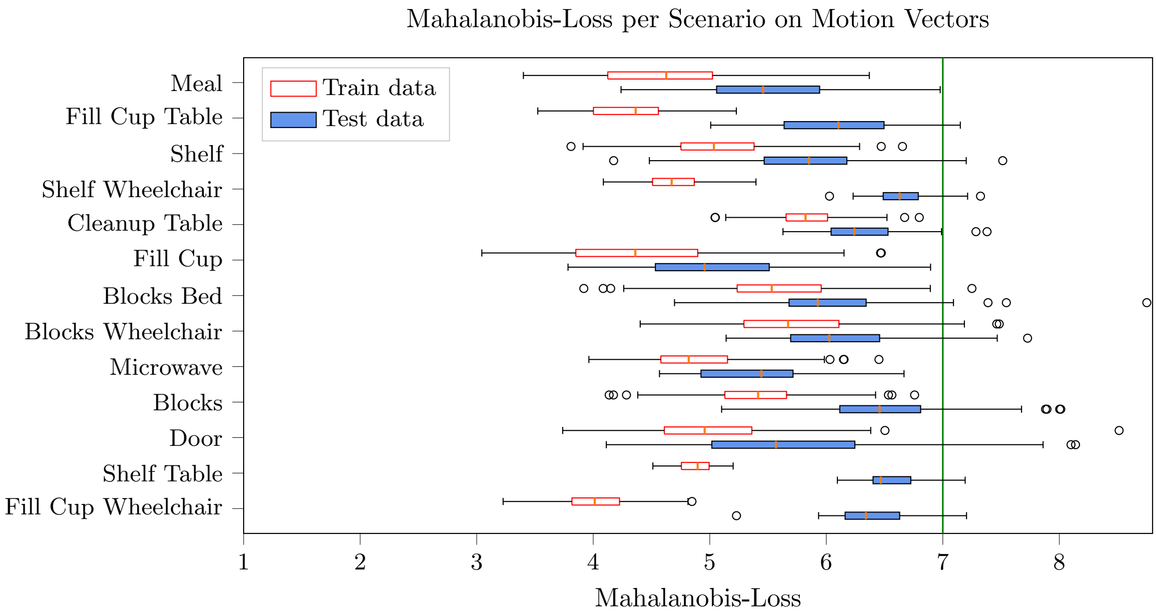
Nevertheless, this process showed the model being capable of gathering information from input (2.) and successfully interpreting it semantically (3.) to produce conclusive results. This basically verifies the complete network architecture, except for the motion-prediction head that ran parallel to the one tested.

Network Output Definition and Loss (4.) The model's motion-head (Figure 15) is a novel design that only indirectly predicts the provided labels by preparing a covariance matrix and having the Mahalanobis-Loss bridge the remaining mathematical gap. As all elements in this prediction head are non-standard in deep learning, it is possible that this general idea of covariance-based handling of vector-based labels is not functional in deep learning or that the given implementation is faulty. In this case, vanishing gradients or a poorly designed loss-derivative could hinder convergence for example.

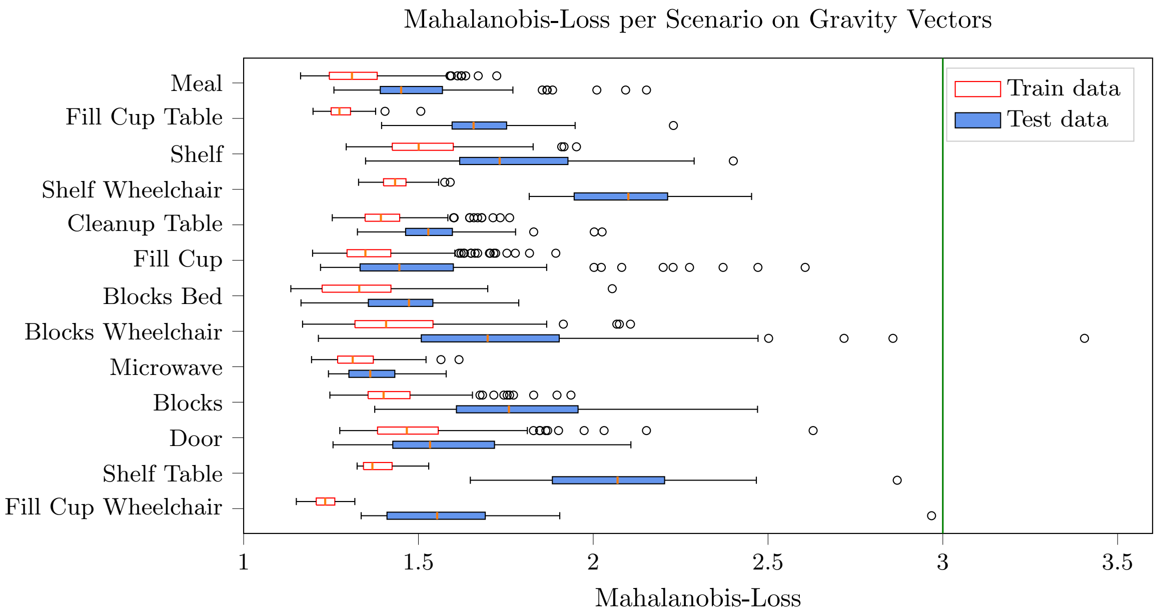
Once again, a simpler task was trained on the same models in order to examine the interaction. In this case, the model architecture stayed the same, with only the datasets preprocessing being slightly adjusted:

Instead of providing the camera's motion vector, it outputs the inclination vector, i.e. the direction of gravity relative to the camera. With this, the data is conceptually very similar, as both the motion and the gravity-direction are vector-based and calculated based on available robot poses, only that the direction of gravity only has $n_{\text{gravity}} = 3$. It is essential however, that the direction of gravity does not depend on the user, and also requires substantial image understanding. As such, the pipeline up to this point can be tested in its entirety without needing to take the data-generating users into account.

Figure 23 compares the Mahalanobis-loss results of two models per scenario. One of these (Figure 23a) was classically trained on motion data, whereas the other (Figure 23b) was trained with the gravitation vectors. A green vertical line represents the expected baseline value n for each plot, as the dimensionality of the prediction vectors need to be taken into account. The best possible value of 1 remains unchanged. The graphs are aligned to one another with respect to their dimensionality.



(a) Predicting Motion Vectors



(b) Predicting Gravitation Vectors

Figure 23: Predicted Motion Vectors (top) versus Predicted Gravitation Vectors (bottom)

The data clearly shows good predictions for the model trained on gravitation-vectors. Even taking the reduction of complexity into account, the results are vastly better than those on motion-data.

This thereby shows the promising functionality of the applied output definition and loss combination, which was further endorsed by the functional 2D proof of concept in Section 5.1.1.

Overall Reasonability of Approach (5.) With the points listed above, the complete training pipeline was examined: 1. The data preparation was visually verified with two independent tools by two independent developers. 2. The models can load the data correctly and 3. extract semantic information from it, as shown by the successfully predicted actions and gravitation vectors. Finally, 4. the handling of vector-based labels with a covariance-output and Mahalanobis-loss was shown to be effective on data similar in shape and origin to the desired motion data, only missing the influence of the data-generating users.

As such, the data itself has to be taken into account once more. Here, the theory is, that if the technical model architecture is functional, then the issue must lie in the data itself. For this, the recordings were visually reevaluated by multiple researchers, who attempted, step-by-step, to predict the motion the camera would be taking next.

This task turned out to be exceedingly difficult: While the researchers would generally be able to predict the overall goal, the precise motion at a certain point would often elude them. The variability in the motion controlled during data generation was simply too high. For example, given an obstacle in the path to an object, the goal of reaching the object is absolutely clear, whereas the path around the object is completely free: Some users would pull their arm back and make a large motion, while others move above the obstacle and jet another group circumvented it horizontally. Interestingly, this aligns with the results of the model being able to predict the actions, but not motion directions.

This analysis was done on multiple recordings in different scenarios, always with similar results. Individual situations were often very similar, however approached by the users with vastly differing motions. Coupled with the technical verification of the training, it has to therefore be concluded that the direction of movement, as performed during this training, is too diverse and cannot be predicted end-to-end with such a machine-learning model generalised over all users.

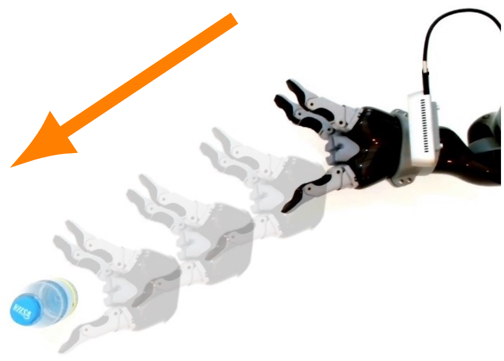
Based on this, no further attempts were made to train other types of AI models, such as transformers for example, on the problem. With these kinds of issues, a completely new approach is necessary to both incorporate recorded motion data and eventually model a prediction output. This is, however, outside the scope of this thesis.

5.2. Behaviour-Based Approach

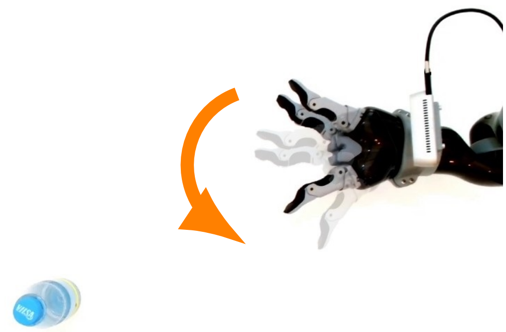
A more analytical approach to generating adaptive DoFs was developed by probabilistically combining custom designed *behaviours*. These are fairly simple and elemental robotic operations, such as *Heading towards light* or *Pulling back after collisions*, and directly generate motion based on sensory input. Prominently used by W. Grey Walter's tortoise robots [60], they are studied in the field of behaviour-based robotics, where alternating between simple sequences can create complex robotic interactions [5]. Generally, a single behaviour is an isolated elemental operation that only serves a distinct purpose and does not aim for conceptual completeness by itself. This makes designing and implementing individual behaviours fairly simple.

In comparison to directly generating the adaptive DoFs end-to-end with a deep-learning-based setup as in Section 5.1, this approach allows for custom probability regulation, memory management, and a more reliable evaluation of capabilities. In terms of specific implementations, this variant also allows the representation of directed DoFs, which can ease computation and error detection during development.

Figure 24 displays the isolated effects of two exemplary behaviours without interaction. Based on the known (e.g. sensor-measured) relative pose of the water bottle, one behaviour (24a) may represent the purely translational approach, whereas another (24b) solely reorients the gripper towards the goal.



(a)



(b)

Figure 24: Two Classic Behaviours of (a) *Approaching* an Object and (b) *Rotating towards* it. Represented top down using a Jaco robot with ghosts and arrows to show motion. Adapted from [V]

For assistive robotics, given sufficiently vast behaviours and knowledge of user intent, the behaviours' underlying operations could theoretically be directly used as adaptive DoFs (though not necessarily strictly fulfilling the definitions of Section 4). In that case, the interaction would be similar to switching between automated trajectories, as in [30]. For the general case however, it is neither feasible to introduce a sufficiently large set of behaviours, nor to assume the user's intent to an adequate degree of certainty. Often enough, users will be indecisive or even explicitly desire to command motions only possible by superimposing multiple behaviours. Therefore, a simple distinct activation-triggering of behaviours is not feasible for a shared robotic control.

Instead of such a binary selection, a set of behaviours B can be handled probabilistically, allowing for an encompassing statistical analysis to generate the set's combined distribution, and eventually adaptive DoFs. Conceptually, these DoFs represent the directions of movement most prominent over all behaviours of B. To showcase this, a second approachable object (a book) is added in Figure 25, thus triggering an additional *Approach*-behaviour. Based purely on these two equally likely behaviours, resulting adaptive DoFs could point towards the objects' combined centre point (bold green) and orthogonal direction of major distinction (dashed green). These directions are most prominent over B, as each behaviour includes a large component going forwards and a lesser component moving sideways.

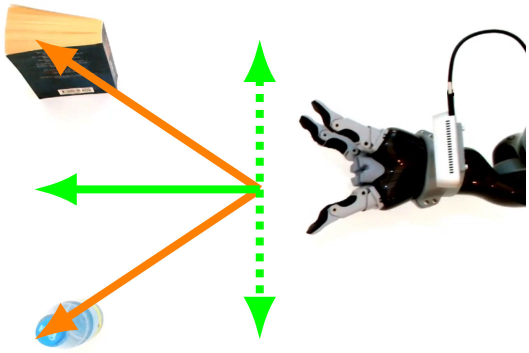


Figure 25: Combining Two *Approach*-Behaviours (orange) into Most Prominent Directions of Movements (green). Adapted from [V]

5.2.1. Probabilistic Combination of Behaviours

Probabilistically speaking, this approach assumes that the user aims to control a motion composed of one or more behaviours from the finite set of available behaviours B. This question of which behaviour the user desires is described by the random variable G, which is dependent on the current situation $Z = z$ and is assumed to exist. Further, for each behaviour $b \in B$, the random variable V_b describes the direction of motion U required by the user to perform the behaviour's underlying operation in the current situation $Z = z$ given the desired behaviour $G = g$. V_b is assumed to exist and follow a multivariate normal distribution h_b with mean $\mu_b \in \mathbb{R}^n$ and covariance $\Sigma_b \in \mathbb{R}^{n \times n}$ for a fixed $Z = z, G = g$:

$$V_b \sim h_b = P\left(U|Z = z, G = g\right) \tag{7}$$

$$V_b \sim \mathcal{N}(\mu_b, \Sigma_b). \tag{8}$$

Conceptually, this now depicts each individual behaviour's application with an associated random variable and multivariate probability distribution, thus allowing to treat them as mixture modes of $U|Z = z$ as a mixture distribution. In this behaviour-based variant of generating adaptive DoFs, $U|Z = z$ is the random variable describing the DoFs most likely to be commanded, given the current situation (see Equation 2). It can be constructed using V_b over all behaviours B as:

$$U|Z = z \sim \sum_{b \in B} P\left(V_b|Z = z\right) \cdot \alpha_b, \tag{9}$$

with α_b as the scaling factor of the mixture mode, in this case modelling the likeliness of the user to follow the associated behaviour

$$\alpha_b = P\left(G = b|Z = z\right). \tag{10}$$

In the implementation, a behaviour's distribution h_b is represented by a set of sigma points (deterministic samples) Q_b, similar to their usage in unscented Kalman filters [57]. These define the behaviour's normal distribution as $\mathcal{N}(\mu_b, \Sigma_b)$ (see Equation 8) with

$$\mu_b = \frac{1}{|Q_b|} \cdot \sum_{q \in Q_b} q, \quad \text{and} \tag{11}$$

$$\Sigma_b = S\left(Q_b, \mu_b\right) \tag{12}$$

where $S(Q, \mu)$ is the general covariance of a set of sigma points Q with defined reference point μ:

$$S(Q, \mu) = \frac{1}{|Q|} \cdot \sum_{q \in Q} (q - \mu)(q - \mu)^T. \tag{13}$$

Figure 26 shows these sigma points in different colours for 3 exemplary fictitious behaviours.

As shown in [V], the expected value $\mu_{\text{behaviour-based}}$ and covariance $\Sigma_{\text{behaviour-based}}$ of the mixture distribution can be directly calculated from the sigma points Q_b of all behaviours $b \in B$. The resulting distribution is shown as a dashed ellipse in Figure 26 and can be calculated as

$$\mu_{\text{behaviour-based}} = \frac{1}{\sum\limits_{b \in B} \alpha_b} \cdot \sum_{b \in B} \alpha_b \cdot \mu_b, \text{and} \tag{14}$$

$$\Sigma_{\text{behaviour-based}} = \frac{1}{\sum\limits_{b \in B} \alpha_b} \cdot \sum_{b \in B} \alpha_b \cdot S(Q_b, \mu_{\text{behaviour-based}}). \tag{15}$$

The resulting mixture distribution now represents the likeliness of all directions, summed over all behaviours. Based on this, adaptive DoFs can be selected such that the m-dimensional space spanned by the m primary DoFs encompasses the largest share of the distribution. In other words, the DoFs are chosen to minimize the expected square distance from the distribution to the m-dimensional space [V]. This is usually achieved by applying a PCA on the covariance at a defined centre of $u_0 = \mu_{\text{behaviour-based}}$.

However, as discussed in Section 5.1, a non-zero default value is not applicable in the context of such a

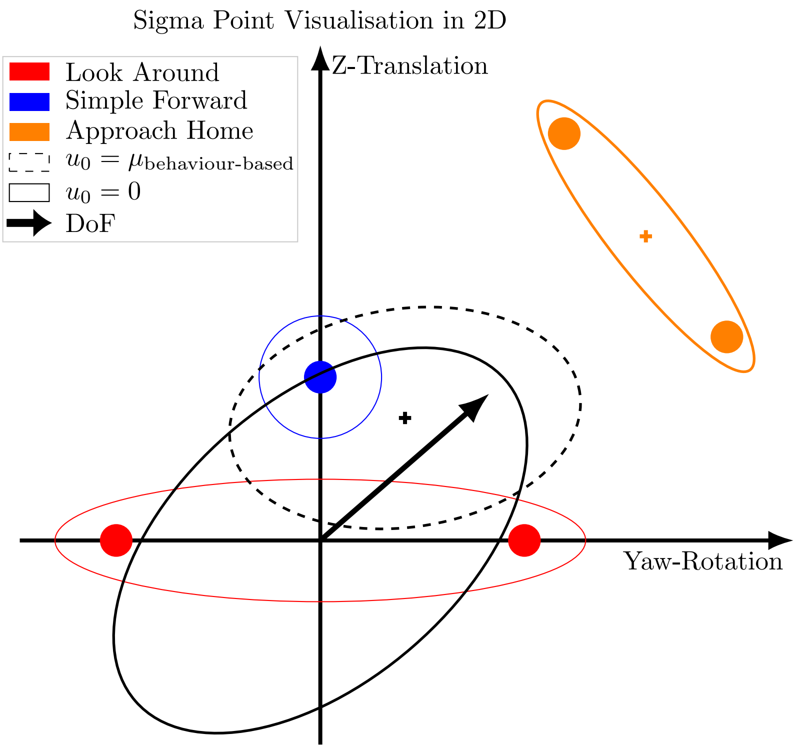


Figure 26: Illustration Representing $U|Z = z$ as a Mixture Distribution. The dots are the sigma-points of three fictitious behaviours (red, blue, orange), with the ellipses showing the corresponding mean and covariance. The black dotted ellipse shows the overall mean and covariance of the mixture, the bold ellipse corresponds to S from Equation 16, and the black arrow shows the single optimal DoF. Mixture weights are represented by the distance of the sigma points to the origin. Adapted from [V]

control interface, as it would result in robot motion without user input (i.e. the robot would move on its own in the direction u_0 if the user gave an input of $c = 0$). Instead, $u_0 = 0$ can be enforced by adding $\mu_{\text{behaviour-based}}\mu_{\text{behaviour-based}}{}^T$ to the covariance $\Sigma_{\text{behaviour-based}}$ before calculating the eigendecomposition (see [V]):

$$D = \text{eigen}_{1:m}(S), \quad u_0 = 0, \quad S = \Sigma_{\text{behaviour-based}} + \mu_{\text{behaviour-based}}\mu_{\text{behaviour-based}}{}^T. \tag{16}$$

The result is shown in Figure 26 as a bold ellipse, with the primary DoF (i.e. the largest eigenvalue of S in Equation 16) shown as an arrow. Overall, the figure shows a simplified example of the mathematical combination of three behaviours using sigma points.

This procedure allows for simple implementation and integration of an arbitrary number of behaviours, with an associated statistical generation of usable adaptive DoFs. The full software control loop can be found in the appendix Figure 34.

5.2.2. Behaviour Design

For this initial realisation, behaviours were developed focussing on the generalised interaction with grasped objects. This very common task of the arm is non-trivial and provides good opportunities for improvement of the user experience. Especially approaching and grasping objects are reoccurring sub-tasks which could benefit greatly from situation-aware DoFs that may, for example, lead directly to objects. In addition, such objects implicitly introduce distinct goals for the task, whereas a task such as *Scratching oneself* is structurally undefined to a degree where shared-control assistance is futile.

In contrast to the neural network-generated paths (see Figure 15 of Section 5.1), the behaviour combination based on sigma points allows the representation of directed DoFs. This is possible due to the incorporation of the distribution's expected value. For the neural network, the sign of paths (Y) were lost due to the outer product operation of the covariance calculation, as well as not being taken into account with the Mahalanobis loss. In Figure 26, the fictitious *Simple Forward* and *Approach Home* behaviours encode such directed (signed) DoFs by having Q_b not centred at zero. The *Look Around*-behaviour in the same figure however solely displays a single undirected (unsigned) DoF.

With these object-related tasks in mind, 7 behaviours were introduced; 6 of which are fairly simple and model some fundamentals of robotic interaction, whereas the last is most dependent on the camera input and provides object-associated individual *Approach*-behaviours. An overview of the baseline behaviours is given below. For a detailed description, including behaviour-specific sigma point definitions, see [V]:

- *Forward*: As the most basic interaction, users often aim to command the robot to continue its trajectory, i.e. follow the heading currently defined by the gripper's orientation, either forwards or backwards. This constant behaviour therefore provides specifically that DoF, however slightly weighted towards moving forwards.

- *Look Around*: In alignment with the previous behaviour, this represents a user's desire for a rotation around the world's vertical axis. Without a preference of direction, the generated DoF allows to orient the gripper roughly towards objects, thereby implicitly paving the way for *Forward* to be applied, while also moving task-relevant objects into the camera's field of view.

- *Rotate Upright*: During regular operation and task execution, the robot's end effector will often reach unexpected kinematic-induced orientations. This can cause confusion in users and often hinders further operation ('Why does it point there?'). The *Rotate Upright* behaviour therefore encompasses a directed DoF which provides an incentive for the end effector to either remain upright or return to this state. Its likeliness is therefore scaled with the inclination angle.

- *Approach Home*: Similar to remaining upright, users often find their robotic arm outstretched or in unusable positions. Here, a return to a predefined home-position generally helps both the user and the robot to more easily continue their task. In addition, the home position might also be a valid retrieval pose for grasped objects, e.g. when the home pose is above the wheelchair's table. For such situations, the *Approach Home* behaviour provides two directed DoFs, one for the translational approach to the home pose and one for orienting towards. The likeliness of either are scaled dependent on the distance and orientation to the goal.
- *Grasp*: As the only baseline behaviour to utilise external sensory input, the *Grasp* behaviour accesses the wrist-mounted depth camera and provides a finger closing-DoF. The likeliness of this behaviour is scaled by the number of pixels indicating graspable objects between the fingers, effectively mapping the likeliness of closing the fingers with the likeliness of a graspable object in place.
- *Liftoff*: Conceptually close to an extension to the *Grasp*-behaviour, this behaviour describes an increased probability to vertically lift the gripper after any grasping action has taken place. This embeds the natural lifting of a recently grasped object to avoid scraping, as well as retrieval after releasing it later on.

The *Approach Object* behaviour differs from the baseline behaviours by extensively processing the camera data for graspable objects and basically creating child-behaviours for the approach of each detected object instance over time. To do this, a *Segment Anything Model* [35, 62] is used, which isolates generic object segments in arbitrary colour images. In order not to lose generality, the model was applied off-the-shelf without retraining it to specific scenarios or objects. Instead, the generated image segments were used in conjunction with the camera's depth data to calculate an object's pose and physical extent. This, coupled with information about the object's direct physical environment, was used to filter graspable objects. In short, segments are considered as graspable objects, if they are within reach, have physical extents not too small or too large for the gripper, and have sufficient chasms to its side that allow fingertips to embrace the object. Further, the likeliness of the user attempting to reach each such object is estimated based on the relative angle of the gripper to the object and its distance.

As such, the behaviour has no instance detection or tracking. Instead, it continuously accesses the video stream and generates new targets to be persistently stored in a global map. By decreasing the probabilities of stored targets over time, this setup allows for a global perception of the environment (e.g. remembering an object no longer directly in the line of sight) with directly embedded uncertainties of outdated data.

In alignment with the *Approaching Home* behaviour, the generic *Approach Object* behaviour also generates 2 isolated DoFs for each object: translational approach and reorientation towards an object. This follows the assumption that users would rather sequentially orient towards an object and draw near than control the interweaving operation of approaching with superimposed rotation. The likeliness of each object and associated DoFs is calculated on the fly based on its remaining certainty in the persistent map, and distance and orientation of the gripper towards the target. This generates a selection dynamic for the user.

Behaviour Synergy in Shared Control In application, the adaptive DoFs generated from probabilistically fusing the behaviours (see Figure 26), behaves just as described in Section 4. As long as no graspable objects are detected or their relative likeliness is too low, the baseline behaviours take command. Working together, they assure the user to be able to scan the environment (*Look Around*) and approach interesting areas (*Forward*), whilst either avoiding cumbersome poses or being able to swiftly unravel them (*Rotate Upright*, *Approach Home*). The latter is automatically infused into the adaptive DoFs based on their situational relevance, i.e. likeliness.

Once graspable objects come into the mix and are sufficiently likely, the *Approach Object* behaviour will be a major factor of the mixture distribution. For example, if the gripper is relatively far away from targets and/or the targets are evenly spaced in view (e.g. t_0 in Figure 12 on page 18), the fusion of *Forward* and the translational component of the *Approach Object*'s sub-behaviours, will produce a DoF in the general direction of the set of targets. This only changes once the robot is closer to the targets (t_1 in Figure 12), as the necessity to select a target will then outweigh the general translational approach. Now, either the rotational selection component or orthogonal translational-distinction will define the adaptive DoFs. Once a selection is made (either by closing the distance or rotating towards a distinct object), the associated sub-behaviour will rise in probability, thereby provide a clear DoF to the target (t_2 in Figure 12).

The more the object extends between the fingers, the more will the *Grasp* behaviour rise in likelihood for the user to choose. Finally, the *Liftoff* behaviour resolves the situation after grasping by removing the grasped object from the environment, possibly fused with either retrieving towards home (*Approach Home*) or reaching a stable state (e.g. with *Rotate Upright*).

A video of the interaction synergy can be found online[29] focussing on user interaction and the *Approach Object* behaviour. In this, the interaction of object segmentation, graspability analysis, and UI is shown with sub-videos.

5.2.3. Evaluations

The adaptive DoF control utilizing probabilistically combined behaviours was evaluated in 2 separate studies: A preliminary technical user study (Study 8) with $N_8 = 18$ able-bodied participants in a lab environment [V], followed by a more extensive evaluation (Study 3) with $N_3 = 24$ individuals from the target group conducted at a trade fair [VI].

Figure 27: Study Apparatus at the Trade Fair of Study 3, illustrating the placement of user, table, and shelf, as well as the UI visualised on the smart glasses (top left). Reused from [VI] © 2024 IEEE

The ADL *Shopping for Groceries* (see Section 3.1) was selected as example scenario for both studies, as it is highly relevant to users and sufficiently complex to expose advantages and limitations of the control. In addition, fixed object poses allowing the task to be neatly defined and thus repeatable allowed for a clean study procedure, whereas the absence of physical human-robot contact minimised risks to participants, especially during the trade fair-based study.

Figure 27 provides an over-the-shoulder view of the study environment at the trade fair. As the destined participants (wheelchair users with limited upper limb mobility) came with their own wheelchairs, a

[29]Video of behaviour-based adaptive DoFs interaction available at `https://www.informatik.uni-bremen.de/agebv/DoF-Adaptiv#Behaviour_Based_Control`, last visited 10th December 2024

transfer into the demonstrator was not feasible. This, coupled with safety precautions, led to the robot being mounted on a small table, similar to the setup in Section 3. During the study, the participants sat on one side of the table, with a shelf on the other. The task at hand involved retrieving two objects from the shelf (a book and a package of cereals in Figure 27) and placing them in a basket on the table. The preliminary lab study followed a very similar setup only with the able-bodied participants sitting on a regular chair.

The user interface is build around the smart glasses worn by the participants (see Section 4.1). Here, the glasses visualise the currently active adaptive DoFs as simplified arrows in their overlay (see top left add-in in Figure 27), with the currently selected direction in the centre and two alternatives to the sides. Once the system suggests a more likely alternative, this is emphasised in the UI with a yellow border and the keyword *New* (see right icon in the image). As discussed in Section 4.2, the user can, at any time, choose to either stick to their currently selected DoF (i.e. the icon at the centre) or switch to an alternative. The actual control input is aligned with the pre-existing *munevo DRIVE* wheelchair control such that tilting the head right and left steers the robot along the selected DoF (positive and negative), whereas sudden head-nods are used to navigate the menu.

The complete procedure of the study, including the scenario, user head motions, and UI-overlay can be seen in the media attachment video recorded in the lab-environment[30].

Preliminary Technical Study The preliminary study (Study 8) compared the adaptive DoF control with a classic control using the same interface (i.e. the smart glasses with 7 predefined and unchanging cardinal DoFs). Following a within-subjects design, each user performed the task with both controls in a random order, followed by a short qualitative questionnaire. The results showed a slight decrease of task completion times when using the adaptive control, coupled with a very clear reduction in the number of mode switches.

Subjectively, 14 participants preferred the adaptive control, while only 4 selected the classic alternative. This was partially ascribed to the adaptive control being perceived easier, though requiring familiarization, and more compact due to fewer options in the UI at most times. In addition, participants explicitly praised diagonal DoFs, which provided motion options to directly approach a target. On the other side, some users struggled with understanding the arrow-representation of DoFs, therefore deeming the fixed directions as a clear advantage of the classic control. In this case, they were able to learn the order of operation by heart and simply operate the arm from memory alone. This aligns with similar feedback given for preference of the 3D mouse in Section 3. For more details, see the associated publication at [V].

Trade Fair Study with Target Group This technical success led to the follow-up study with the target population, which was conducted during the *REHACARE* trade fair[31]. The experiences and success of the trade fair-located Study 1 (see Section 3) showed that such a location offers an extremely realistic environment and large target population, thus allowing an evaluation to be very close to an in-the-wild-study, even while restricting participants entirely to people with limited upper limb mobility. This user group is usually very difficult to sample, in part due to full schedules and limited mobility. In this case however, the trade fair offered unique opportunities by providing a large intrinsic motivation for the target group to visit, thus resulting in a largely increased localised population to sample from.

As the control's general viability was shown in the previous study, this second evaluation focussed on verifying results with the actual target group, as well as assessing the generalisability of the control to different interfaces. For this, the study followed a between-subject design, recording 81 trials over the

[30]Video of the behaviour-based adaptive control in the lab study setting is available at `https://doi.org/10.1145/3652037.3652071`, last visited 10th December 2024

[31]*REHACARE* trade fair. `https://www.rehacare.de/`, last visited 10th December 2024

$N_3 = 24$ participants with varying motor impairments using 3 input devices. The latter are shown in Figure 28 and included the smart glasses (a), a custom-built joystick (b), and assistive mobile buttons (c) that could be positioned on arbitrary positions to adapt to user capabilities. These devices were selected to compare the highly specialised smart glasses with more established joystick and button interfaces, as well as to examine the impact of non-continuous control inputs with the buttons.

(a) (b) (c)

Figure 28: Input Devices used at the Trade Fair of Study 3: (a) Smart Glasses [42], (b) custom-built Joystick, and (c) Assistive Buttons. Reused from [VI] © 2024 IEEE

Conducted in the previously described environment shown in Figure 27, the users were given a trial-based coaching with assistance from the study administrator with either the glasses or the joystick in order to prepare for 1 to 4 measured trials with this device. Based on personal capabilities, 16 subjects also followed this up with a single trial using the assistive buttons for comparison. At the end, each user was interviewed regarding their experiences.

In terms of task completion times and mode switches, the results showed no significant differences between the input devices. Yet, all participants successfully completed the task, with most of them remarking positively on the adaptive control. This shows that the use of an adaptive DoFs-control is not only limited to the specialised case of smart glasses, but can actually improve various input devices [VI].

A thematic content analysis of user comments and interviews revealed high acceptance after a short period of acclimatisation. Here, most users initially did not find the control to be intuitive. However, this quickly changed to becoming easier and more successful once 'the concept was understood' [VI]. In part, these initial complications were due to the unaccustomed use of smart glasses. Two participants with prior experience with smart glasses had the least start-up difficulties. Similarly, participants with prior experiences controlling an assistive robot arm with classic interfaces had to, in part, unlearn coping mechanisms. For example, these users initially controlled one axis after another to reach a target, instead of using suggested directed diagonal DoFs.

Performing such a research-oriented study of a prototypical assistive device in a trade fair environment is highly unusual and came with various unforeseen situations and complications. However, the easy access to a relatively wide range of participants from this diverse user group greatly improved their recruitment and consequently reliability of the results. This was further discussed as an in depth analysis and was published separately [XIII], thus providing the research community with the associated special requirements and limitations of this unconventional approach, accompanied by an overview of lessons learned.

Overall, all users expressed high hopes for the adaptive control, each of them assuming to becoming only faster with more time. In addition, they also explicitly praised the implemented degree of automation, as it balanced suggestions with manual control [VI].

6. Conclusion

This thesis delved deep into user-controlled personal assistive robotics. With the purpose of empowering people living with limited upper limb mobility, it steered away from unnecessarily complex and artificially theoretical constructs, but instead focussed on concepts and innovations achievable by the industry within a reasonable time frame. As such, this thesis does not introduce new mechanical or electrical designs, but mostly extends pre-existing machines (e.g. the Jaco) with intelligent software and context-bridging interfaces. During this process, each associated technological step successfully followed a user-centric participatory design, keeping care-receivers in the loop and developments closely aligned with the nursing community. Hopefully, this paves the way for the technology to swiftly reach affected users.

Especially novel interaction designs have the opportunity to vastly improve the users' lives by providing them with more autonomy and a possibly more self-determined life. However, a fine line needs to be considered in this setting, as developers tend to over-achieve and aim too high in automating operations. While this minimises the burden laid upon users, it also reduces their opportunities to make decisions themselves. As shown in the analysis of requested autonomy, no on-size-fits-all system can be installed. In most cases, assistance is highly valued, however always linked to the customisable fallback operation of manual control.

However, even in this seemingly most basic interaction variant of manual control, the analysis of contemporary control interfaces clearly showed the users swaying from the seemingly simpler manufacturer-provided joystick to more complex devices. Here, users expressed their eagerness to train with a more difficult, but also more capable system and gather experience over time. More skilful interfaces could therefore bridge a huge gap by reducing unnecessarily patronizing simplifications and vastly expanding user-robot capabilities. This should be understood as a shout-out to the industry to provide users with technological opportunities to grow alongside their assistive devices.

A realisation of such an opportunity is introduced as the *Adaptive DoF Control*, which itself could be made market-ready in a fairly short amount of time, as its underlying control synergy completely avoids possibly ambiguous and unsafe autonomous operation and merely extends user-robot capabilities, thereby maintaining sustainable fallback options. Hereby, the concept keeps the users in complete control while enabling them to operate intricate situations with adaptively changing DoFs.

As the ability to learn operations by heart was a major selling point for participants during the analysis of contemporary control interfaces, the lack of it is possibly the largest limitation of this situation-adaptive low-DoFs control. However, this ability directly correlates to the number of input DoFs to be skilfully controlled by the user, as only high-dimensional input devices allow this form of interaction by creating an immediate mapping of input DoFs to output DoFs. In turn, they demand a level of dexterity and physical capabilities not met by all users. This is where the adaptive DoF control explicitly shines for low-dimensional input devices by allowing complex manoeuvring with simple control interfaces.

While only a simplified proof of concept of an end-to-end trained generation of adaptive DoF was achieved, the behaviour-based implementation was capable to be evaluated in multiple studies, including extensive testing by the target study in a semi-unstructured environment. This showed remarkable success, with a clear reduction of mode switches and subjective task load of the users, when compared to a classic control. Possibly even more important, affected people explicitly expressed contentment with this control, with bright prospects of a more independent self-determined life.

6.1. Future Work

At this point, two distinct paths of further developments are possible: One, probably preferred by the nursing community, sees a rapid system integration and introduction to the market of presented technologies. For this, simple and standardised interfaces to independently connect (custom) input devices can pave the way for more self-determined manual interactions. Just like the rest of society, users of assistive robotics grow more and more accustomed to computers and machines. These systems should reflect this by allowing users to choose their interaction preferences.

For the adaptive DoF control, the shortest path to a market introduction probably utilises the behaviour-based approach. This was already shown to be functional even with minimal set of behaviours and is easy to expand. Also, as the control concept itself induces basically no additional risks, a public roll-out as add-on to a robot should require minimal effort. However, for a future-proof product, the set of behaviours needs to be expandable, both by manufacturers and users. Overall, user-piloted assistive robotics could benefit greatly from a public marketplace (app-store-esque) to share such and similar innovations. If not limited to be used by manufacturers, this could also be a share-point for the community, which often reports on developing customised solutions for activities of daily living. We have this for games on our smartphones, why not for life-changing devices with so much room for improvement?

The second path of future work deals with further improvements of the adaptive DoF control and the associated scientifically evaluation. Here, new behaviours can be devised which further extend the capabilities. For example, the presented *Approach Object* behaviour could be extended or replaced by building upon the recently introduced Grasp Anything system [43]. Also, reincorporating a deep learning component could be useful to allow online learning. This could adjust behaviours user-dependently over time and learn preferences and habits. Further, it might be fruitful to assess combining the control concepts of adaptive DoFs with that of latent action spaces [39].

The largest scientific value could however be gained by a series of long-term testing. Even though this thesis evaluated the adaptive control both in lab setting and the semi-unstructured environment of a trade fair, an extensive study in the real world is still missing. Preferably iteratively, i.e. allowing for the expansion of capabilities adjusting to user requirements at given points in time, this could be used to gain major insights into the control's actual usability and limitations in everyday life.

Acknowledgements

This thesis was supported by three projects, each of which was funded by the German Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung; BMBF). In chronological order, these were

- MobILe: Physische Mensch-Roboter-Interaktion für ein selbstbestimmtes Leben (FKZ: 16SV7867) July 2017 until June 2021,
- AdaMeKoR: Ein adaptives Mehrkomponenten-Robotersystem für die Pflege (FKZ: 16SV8534) March 2020 until September 2023, and
- DoF-Adaptiv: Adaptive Freiheitsgradeinbettung als kooperatives Userinterface für einen Assistenzroboter (FKZ: 16SV8563) February 2021 until April 2024.

Publications

Peer-Reviewed

[I] Annalies Baumeister, Elizaveta Gardo, Patrizia Tolle, Barbara Klein, Max Pascher, Jens Gerken, **Felix Goldau**, Yashaswini Shivashankar and Udo Frese. 'The Importance of Participatory Design for the Development of Assistive Robotic Arms. Initial Approaches and Experiences in the Research Projects MobILe and DoF-Adaptiv'. en. In: Connected Living: international and interdisciplinary conference (2021). Frankfurt University of Applied Sciences. Frankfurt am Main, 2021. 10.48718/8p7x-cw14. URL: `https://nbn-resolving.org/urn:nbn:de:bsz:946-opus4-62882` (visited on 2024/08/21).
My Contribution: 5 %
This work discusses the participatory research design of the two projects MobILe and DoF-Adaptiv. I was part of both projects and conducted (mostly technical) studies with the extended target group.

[II] Annalies Baumeister, Max Pascher, Yashaswini Shivashankar, **Felix Ferdinand Goldau**, Udo Frese, Jens Gerken, Elizaveta Gardo, Barbara Klein and Patrizia Tolle. 'AI for Simplifying the Use of an Assistive Robotic Arm for People with severe Body Impairments'. In: *Gerontechnology* 21 (Oct. 2022). 10.4017/gt.2022.21.s.578.5.sp7. URL: `https://doi.org/10.4017/gt.2022.21.s.578.5.sp7` (visited on 2024/08/21).
My Contribution: 5 %
This conference presentation and published abstract discussed further results of the DoF-Adaptiv project, including first technical findings. I developed the project's simulation environment and was involved in the technical studies.

[III] **Felix Goldau**, Yashaswini Shivashankar, Annalies Baumeister, Lennart Drescher, Patrizia Tolle and Udo Frese. 'DORMADL - Dataset of Human-Operated Robot Arm Motion in Activities of Daily Living'. In: *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. Piscataway, NJ: IEEE Press, 2023, pp. 11396–11403. 10.1109/IROS55552.2023.10341459. URL: `http://doi.org/10.1109/IROS55552.2023.10341459` (visited on 2024/08/21).
My Contribution: 60 %
This publicly available dataset provides pose and movement information for an assistive robot arm in simulation and reality. I developed the underlying robot control in both simulation and reality, created the simulation environment itself, and cleaned, processed and prepared the data.

[IV] **Felix Ferdinand Goldau** and Udo Frese. 'Learning to Map Degrees of Freedom for Assistive User Control: Towards an Adaptive DoF-Mapping Control for Assistive Robots'. In: *The 14th PErvasive Technologies Related to Assistive Environments Conference*. PETRA 2021. Corfu, Greece: Association for Computing Machinery, 2021, pp. 132–139. ISBN: 9781450387927. 10.1145/3453892.3453895. URL: `https://doi.org/10.1145/3453892.3453895` (visited on 2024/08/21).
My Contribution: 85 %
In this work, we initially introduced the adaptive control in 2D. I created the simulation environment, gathered training data, developed and trained the neural network, and conducted and evaluated the user study.

[V] **Felix Ferdinand Goldau** and Udo Frese. 'Probabilistic Combination of Heuristic Behaviors for Shared Assistive Robot Control'. In: *Proceedings of the 17th International Conference on PErvasive Technologies Related to Assistive Environments.* PETRA '24. Crete, Greece: Association for Computing Machinery, 2024, pp. 147–155. ISBN: 9798400717604. 10.1145/3652037.3652071. URL: `https://doi.org/10.1145/3652037.3652071` (visited on 2024/08/21).
Best Technical Paper Award
My Contribution: 85 %
As an alternative to a machine-learned adaptive control, this system provides similar functionality based on probabilistic combinations of heuristic behaviors. I implemented the robot control, including the behaviors and mathematical combination. I also conducted and evaluated the study.

[VI] **Felix Ferdinand Goldau**, Max Pascher, Annalies Baumeister, Patrizia Tolle, Jens Gerken and Udo Frese. 'Adaptive Control in Assistive Application - A Study Evaluating Shared Control by Users with Limited Upper Limb Mobility'. In: *2024 33rd IEEE International Conference on Robot and Human Interactive Communication (ROMAN).* Pasadena, CA: IEEE, 2024. 10.1109/RO-MAN60168.2024.10731229. URL: `https://ieeexplore.ieee.org/document/10731229`.
My Contribution: 30 %
This work presents the successful final evaluation of the adaptive control with users from the target group at a trade fair. I developed the underlying control and was part of the team that planned, conducted, and evaluated the study.

[VII] **Felix Ferdinand Goldau**, Tejas Kumar Shastha, Maria Kyrarini and Axel Gräser. 'Autonomous Multi-Sensory Robotic Assistant for a Drinking Task'. In: *2019 IEEE 16th International Conference on Rehabilitation Robotics (ICORR).* 2019, pp. 210–216. 10.1109/ICORR.2019.8779521. URL: `http://doi.org/10.1109/ICORR.2019.8779521` (visited on 2024/08/21).
My Contribution: 65 %
In this robotic solution to assistive drinking without a straw or any elements physically attached to the user, I designed the general interaction design and focussed on the task of serving the drink. I also conducted and evaluated the primary experimental study (not part of this thesis), and assisted on the second.

[VIII] Kirill Kronhardt, Stephan Rübner, Max Pascher, **Felix Goldau**, Udo Frese and Jens Gerken. 'Adapt or Perish? Exploring the Effectiveness of Adaptive DoF Control Interaction Methods for Assistive Robot Arms'. In: *Technologies* 10.1 (14th Feb. 2022), p. 30. ISSN: 2227-7080. 10.3390/technologies10010030. URL: `https://www.mdpi.com/2227-7080/10/1/30` (visited on 2024/08/21).
My Contribution: 10 %
This study initially evaluating adaptive control served as our wizard-of-oz-esque proof of concept in virtual 3D. I developed most of the simulation environment and provided expertise for the control method.

[IX] Max Pascher, **Felix Ferdinand Goldau**, Kirill Kronhardt, Udo Frese and Jens Gerken. 'AdaptiX - A Transitional XR Framework for Development and Evaluation of Shared Control Applications in Assistive Robotics'. In: *Proc. ACM Hum.-Comput. Interact.* 8.EICS (June 2024). 10.1145/3660243. URL: `https://dl.acm.org/doi/10.1145/3660243` (visited on 2024/08/21).
Best Paper Award
My Contribution: 35 %
This framework is designed to test and evaluate control concepts for assistive robotics in various stages from VR to reality. I developed most of the VR simulation and assisted on the integration of the real robot.

[X] Max Pascher, Kirill Kronhardt, **Felix Ferdinand Goldau**, Udo Frese and Jens Gerken. 'In Time and Space: Towards Usable Adaptive Control for Assistive Robotic Arms'. In: *2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. Piscataway, NJ: Institute of Electrical and Electronics Engineers (IEEE), 2023, pp. 2300–2307. 10.1109/RO-MAN57019.2023.10309381. URL: `http://doi.org/10.1109/RO-MAN57019.2023.10309381` (visited on 2024/08/21).
My Contribution: 10 %
Analysing the Robot-Human interaction, this work compares two means of communicating and adjusting DoFs of an adaptive control. I provided most of the simulation environment and provided expertise regarding the possible interaction variants.

In Review

[XI] Jonathan Behrens, Jonas Schulz, **Felix Goldau**, Christian Kowalski, Daniel Dorniok, Sandra Helmers, Stephanie Raudies, Rebekka Stellfeldt, Yashaswini Shivashankar, Niels Will, Serge Autexier, Cletus Brauer, Thomas Breisig, Andreas Hein and Manfred Hülsken-Giesler. 'Interdisciplinary Development and Reflection of a Robotic System for Nursing'. In: *International Journal of Social Robotics* (Special Issue on Robotic Systems for Nursing Care 2024). In Review.
My Contribution: 17.5 %
This large interdisciplinary work combines findings of multiple partners of the AdaMeKoR project and presents the overarching results. Among assisting to mold the collaboration to a whole, I did most of the work on the comparison of control methods. This included implementations, as well as study design, execution, and evaluation.

Not Peer-Reviewed

[XII] Annalies Baumeister, **Felix Goldau**, Yashaswini Shivashankar, Barbara Klein, Patrizia Tolle and Udo Frese. 'Die Perspektive potenzieller Nutzer*innen auf assistive Roboterarme in ambulanten Settings'. In: *Zeitschrift für Gerontologie und Geriatrie*. Vol. 55. SUPPL 1. Springer Heidelberg Tiergartenstr. 17, D-69121 Heidelberg, Germany. 2022, pp. 107–108. 10.1007/s00391-022-02095-7. URL: `https://doi.org/10.1007/s00391-022-02095-7` (visited on 2024/08/21).
My Contribution: 10 %
This conference presentation and published abstract discussed intermediate participatory results of the DoF-Adaptiv project. I was involved in the participatory workshops and subsequent development of application scenarios.

[XIII] Annalies Baumeister, **Felix Ferdinand Goldau**, Max Pascher, Jens Gerken, Udo Frese and Patrizia Tolle. *Evaluating Assistive Technologies on a Trade Fair. Methodological Overview and Lessons Learned*. arXiv. Reviewed submission planned. 2024. 10.48550/arXiv.2408.10933. URL: `https://arxiv.org/abs/2408.10933` (visited on 2024/08/21).
My Contribution: 27.5 %
Based on the studies in [XI, VI], this work discusses the opportunities gained by conducting evaluation studies on a trade fair as well as their implications. I was heavily involved in organising the studies, conducting both of them, and evaluating the lessons learned.

References

[1] Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra, Jorge Luis Reyes-Ortiz et al. 'A public domain dataset for human activity recognition using smartphones'. In: *Proceedings of the 21th International European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*. Vol. 3. Bruges, 2013, pp. 437–442. ISBN: 978-2-87419-081-0. URL: `http://hdl.handle.net/2117/20897`.

[2] Serge Autexier, Christoph Lüth and Rolf Drechsler. 'Das Bremen Ambient Assisted Living Lab und darüber hinaus – Intelligente Umgebungen, smarte Services und Künstliche Intelligenz in der Medizin für den Menschen'. In: *Künstliche Intelligenz im Gesundheitswesen: Entwicklungen, Beispiele und Perspektiven*. Ed. by Mario A. Pfannstiel. Wiesbaden: Springer Fachmedien Wiesbaden, 2022, pp. 835–850. ISBN: 978-3-658-33597-7. 10.1007/978-3-658-33597-7˙40.

[3] Anna Brinkmann, Conrad Fifelski-von Böhlen, Christian Kowalski, Sandra Lau, Ole Meyer, Rebecca Diekmann and Andreas Hein. 'Providing physical relief for nurses by collaborative robotics'. In: *Scientific Reports* 12.1 (May 2022). ISSN: 2045-2322. 10.1038/s41598-022-12632-4.

[4] Connor Brooks and Daniel Szafir. 'Visualization of Intended Assistance for Acceptance of Shared Control'. In: *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. Piscataway, NJ, USA: IEEE Press, 2020, pp. 11425–11430. 10.1109/IROS45743.2020.9340964.

[5] Rodney A. Brooks. 'Intelligence without representation'. In: *Artificial Intelligence* 47.1 (1991), pp. 139–159. ISSN: 0004-3702. 10.1016/0004-3702(91)90053-M. URL: `https://www.sciencedirect.com/science/article/pii/000437029190053M`.

[6] Barbara Bruno, Fulvio Mastrogiovanni, Antonio Sgorbissa, Tullio Vernazza and Renato Zaccaria. 'Analysis of human behavior recognition algorithms based on acceleration data'. In: *2013 IEEE International Conference on Robotics and Automation*. IEEE. 2013, pp. 1602–1607. 10.1109/ICRA.2013.6630784.

[7] Berk Calli, Arjun Singh, Aaron Walsman, Siddhartha Srinivasa, Pieter Abbeel and Aaron M Dollar. 'The YCB object and Model set: Towards common benchmarks for manipulation research'. In: *2015 international conference on advanced robotics (ICAR)*. IEEE. 2015, pp. 510–517. 10.1109/ICAR.2015.7251504. URL: `https://ieeexplore.ieee.org/abstract/document/7251504`.

[8] Gerard Canal, Guillem Alenyà and Carme Torras. 'Personalization Framework for Adaptive Robotic Feeding Assistance'. In: *Social Robotics*. Ed. by Arvin "Agah, John-John Cabibihan, Ayanna M. Howard, Miguel A. Salichs and Hongsheng" He. Vol. 9979. Cham: Springer International Publishing, 2016, pp. 22–31. ISBN: 978-3-319-47437-3. 10.1007/978-3-319-47437-3˙3.

[9] Cheng-Shiu Chung, Hongwu Wang and Rory A Cooper. 'Functional assessment and performance evaluation for assistive robotic manipulators: Literature review'. In: *The Journal of Spinal Cord Medicine* 36.4 (2013), pp. 273–289. 10.1179/2045772313Y.0000000132.

[10] B. J. F. Driessen, H. G. Evers and J. A. van Woerden. 'MANUS—a wheelchair-mounted rehabilitation robot'. In: *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine* 215.3 (2001), pp. 285–290. 10.1243/0954411011535876.

[11] BJF Driessen, TK Ten Kate, F Liefhebber, AHG Versluis and JA Van Woerden. 'Collaborative Control of the Manus Manipulator'. In: *Universal Access in the Information Society* 4.2 (2005), pp. 165–173. 10.1007/s10209-005-0121-6.

[12] F. Flemisch, D. A. Abbink, M. Itoh, M.-P. Pacaux-Lemoine and G. Weßel. 'Joining the blunt and the pointy end of the spear: towards a common framework of joint action, human–machine cooperation, cooperative guidance and control, shared, traded and supervisory control'. In: *Cognition, Technology & Work* 21.4 (Aug. 2019), pp. 555–568. 10.1007/s10111-019-00576-1.

[13] D. Gallenberger, T. Bhattacharjee, Y. Kim and S. S. Srinivasa. 'Transfer Depends on Acquisition: Analyzing Manipulation Strategies for Robotic Feeding'. In: *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 2019, pp. 267–276. 10.1109/HRI.2019.8673309.

[14] Stephan M Gerber, René M Müri, Urs P Mosimann, Tobias Nef and Prabitha Urwyler. 'Virtual reality for activities of daily living training in neurorehabilitation: a usability and feasibility study in healthy participants'. In: *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE. 2018, pp. 1–4. 10.1109/EMBC.2018.8513003.

[15] Corey Goldfeder, Matei Ciocarlie, Hao Dang and Peter K. Allen. 'The Columbia grasp database'. In: *2009 IEEE International Conference on Robotics and Automation*. 2009, pp. 1710–1716. 10.1109/ROBOT.2009.5152709.

[16] Deepak E. Gopinath and Brenna Dee Argall. 'Mode Switch Assistance To Maximize Human Intent Disambiguation'. In: *Proceedings of Robotics*. Ed. by Nancy Amato, Siddhartha Srinivasa, Nora Ayanian and Scott Kuindersma. 2017. ISBN: 978-0992374730. 10.15607/RSS.2017.XIII.046.

[17] Axel Gräser, Torsten Heyer, Leila Fotoohi, Uwe Lange, Henning Kampe, Bashar Enjarini, Stefan Heyer, Christos Fragkopoulos and Danijela Ristić-Durrant. 'A Supportive FRIEND at Work: Robotic Workplace Assistance for the Disabled'. In: *IEEE Robotics & Automation Magazine* 20.4 (2013), pp. 148–159. 10.1109/MRA.2013.2275695.

[18] Sorin M. Grigorescu, Thorsten Lüth, Christos Fragkopoulos, Marco Cyriacks and Axel Gräser. 'A BCI-controlled robotic assistant for quadriplegic people in domestic and professional life'. In: *Robotica* 30.3 (2012), pp. 419–431. 10.1017/S0263574711000737.

[19] Reem K. Al-Halimi and Medhat Moussa. 'Performing Complex Tasks by Users With Upper-Extremity Disabilities Using a 6-DOF Robotic Arm: A Study'. In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 25.6 (2017), pp. 686–693. 10.1109/TNSRE.2016.2603472.

[20] Laura V. Herlant, Rachel M. Holladay and Siddhartha S. Srinivasa. 'Assistive Teleoperation of Robot Arms via Automatic Time-Optimal Mode Switching'. In: *The Eleventh ACM/IEEE International Conference on Human Robot Interaction*. HRI '16. Christchurch, New Zealand: IEEE Press, 2016, pp. 35–42. ISBN: 9781467383707. 10.1109/HRI.2016.7451731. URL: `https://ieeexplore.ieee.org/abstract/document/7451731`.

[21] G. Hirzinger. 'ROTEX — The first space robot technology experiment'. In: *Experimental Robotics III*. Ed. by Tsuneo Yoshikawa and Fumio Miyazaki. Berlin, Heidelberg: Springer Berlin Heidelberg, 1994, pp. 579–598. ISBN: 978-3-540-39355-9. 10.1007/BFb0027622.

[22] Leigh R Hochberg, Daniel Bacher, Beata Jarosiewicz, Nicolas Y Masse, John D Simeral, Joern Vogel, Sami Haddadin, Jie Liu, Sydney S Cash, Patrick Van Der Smagt et al. 'Reach and grasp by people with tetraplegia using a neurally controlled robotic arm'. In: *Nature* 485.7398 (2012), pp. 372–375. 10.1038/nature11076.

[23] Leigh R. Hochberg, Mijail D. Serruya, Gerhard M. Friehs, Jon A. Mukand, Maryam Saleh, Abraham H. Caplan, Almut Branner, David Chen, Richard D. Penn and John P. Donoghue. 'Neuronal ensemble control of prosthetic devices by a human with tetraplegia'. In: *Nature* 442.7099 (July 2006), pp. 164–171. ISSN: 1476-4687. 10.1038/nature04970.

[24] Catherine Holloway. 'Disability Interaction (DIX): A Manifesto'. In: *Interactions* 26.2 (Feb. 2019), pp. 44–49. ISSN: 1072-5520. 10.1145/3310322.

[25] Ming-Kuei Hu. 'Visual pattern recognition by moment invariants'. In: *IRE Transactions on Information Theory* 8.2 (1962), pp. 179–187. 10.1109/TIT.1962.1057692.

[26] Manfred Hülsken-Giesler and Sabine Daxberger. 'Robotik in der Pflege aus pflegewissenschaftlicher Perspektive'. In: *Pflegeroboter*. Ed. by Oliver Bendel. Wiesbaden: Springer Fachmedien Wiesbaden, 2018, pp. 125–139. ISBN: 978-3-658-22698-5. 10.1007/978-3-658-22698-5˙7.

[27] Anja Jackowski, Marion Gebhard and Roland Thietje. 'Head Motion and Head Gesture-Based Robot Control: A Usability Study'. In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 26.1 (2018), pp. 161–170. 10.1109/TNSRE.2017.2765362.

[28] Siddarth Jain and Brenna Argall. 'Probabilistic Human Intent Recognition for Shared Autonomy in Assistive Robotics'. In: *J. Hum.-Robot Interact.* 9.1 (Dec. 2019). 10.1145/3359614.

[29] Siddarth Jain and Brenna Argall. 'Robot Learning to Switch Control Modes for Assistive Teleoperation'. In: *RSS 2016 Workshop on Planning for Human-Robot Interaction: Shared Autonomy and Collaborative Robotics*. 2016. URL: `https://people.csail.mit.edu/cdarpino/RSS2016WorkshopHRcolla/abstracts/RSS16WS_06_RobotLearningSwitchModes.pdf` (visited on 2024/07/02).

[30] Siddarth Jain, Ali Farshchiansadegh, Alexander Broad, Farnaz Abdollahi, Ferdinando Mussa-Ivaldi and Brenna Argall. 'Assistive robotic manipulation through shared autonomy and a Body-Machine Interface'. In: *2015 IEEE International Conference on Rehabilitation Robotics (ICORR)*. Piscataway, NJ, USA: IEEE Press, 2015, pp. 526–531. 10.1109/ICORR.2015.7281253.

[31] D. Kim, R. Hazlett-Knudsen, H. Culver-Godfrey, G. Rucks, T. Cunningham, D. Portee, J. Bricout, Z. Wang and A. Behal. 'How Autonomy Impacts Performance and Satisfaction: Results From a Study With Spinal Cord Injured Subjects Using an Assistive Robot'. In: *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans* 42.1 (2012), pp. 2–14. 10.1109/TSMCA.2011.2159589.

[32] Dae-Jin Kim, Ryan Lovelett and Aman Behal. 'An empirical study with simulated ADL tasks using a vision-guided assistive robot arm'. In: *2009 IEEE International Conference on Rehabilitation Robotics*. 2009, pp. 504–509. 10.1109/ICORR.2009.5209527.

[33] Dae-Jin Kim, Ryan Lovelett and Aman Behal. 'Eye-in-hand stereo visual servoing of an assistive robot arm in unstructured environments'. In: *2009 IEEE International Conference on Robotics and Automation*. 2009, pp. 2326–2331. 10.1109/ROBOT.2009.5152821.

[34] Kinova. *Jaco assistive robot - User guide*. "EN-UG-007-r05-202111". 2021. URL: `https://assistive.kinovarobotics.com/uploads/EN-UG-007-Jaco-user-guide-R05.pdf` (visited on 2024/05/21).

[35] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, Piotr Dollár and Ross Girshick. *Segment Anything*. 2023. arXiv: `2304.02643 [cs.CV]`.

[36] Maria Kyrarini, Adrian Leu, Danijela Ristić-Durrant, Axel Gräser, Anja Jackowski, Marion Gebhard, Jochen Nelles, Christina Bröhl, Christopher Brandl, Alexander Mertens et al. 'Human-Robot Synergy for cooperative robots'. In: *Facta Universitatis, Series: Automatic Control and Robotics* 15.3 (2016), pp. 187–204. 10.22190/FUACR1603187K.

[37] Maria Kyrarini, Fotios Lygerakis, Akilesh Rajavenkatanarayanan, Christos Sevastopoulos, Harish Ram Nambiappan, Kodur Krishna Chaitanya, Ashwin Ramesh Babu, Joanne Mathew and Fillia Makedon. 'A Survey of Robots in Healthcare'. In: *Technologies* 1 (2021). ISSN: 2227-7080. 10.3390/technologies9010008. URL: https://www.mdpi.com/2227-7080/9/1/8.

[38] Ian Lenz, Honglak Lee and Ashutosh Saxena. 'Deep Learning for Detecting Robotic Grasps'. In: *The International Journal of Robotics Researc* 34.4-5 (Apr. 2015), pp. 705–724. ISSN: 0278-3649. 10.1177/0278364914549607.

[39] Dylan P. Losey, Krishnan Srinivasan, Ajay Mandlekar, Animesh Garg and Dorsa Sadigh. 'Controlling Assistive Robots with Learned Latent Actions'. In: *2020 IEEE International Conference on Robotics and Automation (ICRA)*. Piscataway, NJ: IEEE Press, 2020, pp. 378–384. 10.1109/ICRA40945.2020.9197197.

[40] Veronique Maheu, Philippe S Archambault, Julie Frappier and François Routhier. 'Evaluation of the JACO robotic arm: Clinico-economic study for powered wheelchair users with upper-extremity disabilities'. In: *2011 IEEE International Conference on Rehabilitation Robotics*. IEEE. 2011, pp. 1–5. 10.1109/ICORR.2011.5975397.

[41] Bente Martinsen, Ingegerd Harder and Fin Biering-Sorensen. 'The meaning of assisted feeding for people living with spinal cord injury: a phenomenological study'. In: *Journal of Advanced Nursing* 62.5 (May 2008), pp. 533–540. 10.1111/j.1365-2648.2008.04637.x. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-2648.2008.04637.x.

[42] Mikepanhu. *Google Glass with frame.jpg*. Adapted under CC BY-SA 3.0 license. Wikimedia Commons. 2014. URL: https://upload.wikimedia.org/wikipedia/commons/b/be/Google_Glass_with_frame.jpg (visited on 2024/07/21).

[43] Malte Mosbach and Sven Behnke. 'Grasp Anything: Combining Teacher-Augmented Policy Gradient Learning with Instance Segmentation to Grasp Arbitrary Objects'. In: *2024 IEEE International Conference on Robotics and Automation (ICRA)*. 2024. 10.48550/arXiv.2403.10187. URL: https://www.ais.uni-bonn.de/papers/ICRA_2024_Mosbach.pdf.

[44] Max Pascher, Uwe Gruenefeld, Stefan Schneegass and Jens Gerken. 'How to Communicate Robot Motion Intent: A Scoping Review'. In: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems - CHI '23*. 2023. 10.1145/3544548.3580857.

[45] Horst Penkert, Julius C. Baron, Konstantin Madaus, Wolfgang Huber and Achim Berthele. 'Assessment of a novel, smartglass-based control device for electrically powered wheelchairs'. In: *Disability and Rehabilitation: Assistive Technology* 16.2 (2021), pp. 172–176. 10.1080/17483107.2019.1646817.

[46] Laura Petrich, Jun Jin, Masood Dehghan and Martin Jagersand. 'A Quantitative Analysis of Activities of Daily Living: Insights into Improving Functional Independence with Assistive Robotics'. In: *2022 International Conference on Robotics and Automation (ICRA)*. 2022, pp. 6999–7006. 10.1109/ICRA46639.2022.9811960.

[47] Anita Pollak, Mateusz Paliga, Matias M. Pulopulos, Barbara Kozusznik and Malgorzata W. Kozusznik. 'Stress in manual and autonomous modes of collaboration with a cobot'. In: *Computers in Human Behavior* 112 (2020), p. 8. ISSN: 0747-5632. 10.1016/j.chb.2020.106469. URL: https://www.sciencedirect.com/science/article/pii/S0747563220302211.

[48] Diego A. Rozas Llontop, José Cornejo, Ricardo Palomares and Jorge A. Cornejo-Aguilar. 'Mechatronics Design and Simulation of Anthropomorphic Robotic Arm mounted on Wheelchair for Supporting Patients with Spastic Cerebral Palsy'. In: *2020 IEEE International Conference on Engineering Veracruz (ICEV)*. 2020, pp. 1–5. 10.1109/ICEV50249.2020.9289665.

[49] Marco Ruzzon, Alessandro Carfì, Takahiro Ishikawa, Fulvio Mastrogiovanni and Toshiyuki Murakami. 'A multi-sensory dataset for the activities of daily living'. In: *Data in Brief* 32 (2020). ISSN: 2352-3409. 10.1016/j.dib.2020.106122. URL: `https://www.sciencedirect.com/science/article/pii/S2352340920310167`.

[50] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov and Liang-Chieh Chen. 'MobileNetV2: Inverted Residuals and Linear Bottlenecks'. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).* June 2018, pp. 4510–4520. 10.48550/arXiv.1801.04381.

[51] Jeroen Schäfer and Marion Gebhard. 'Feasibility analysis of sensor modalities to control a robot with eye and head movements for assistive tasks'. In: *Proceedings of the 12th ACM International Conference on PErvasive Technologies Related to Assistive Environments.* PETRA '19. Rhodes, Greece: Association for Computing Machinery, 2019, pp. 482–488. ISBN: 9781450362320. 10.1145/3316782.3322774.

[52] Sebastian Schröer, Ingo Killmann, Barbara Frank, Martin Völker, Lukas Fiederer, Tonio Ball and Wolfram Burgard. 'An autonomous robotic assistant for drinking'. In: *2015 IEEE International Conference on Robotics and Automation (ICRA).* 2015, pp. 6482–6487. 10.1109/ICRA.2015.7140110.

[53] Statistisches Bundesamt (Destatis). *Statistik der schwerbehinderten Menschen, Kurzbericht 2021.* Genesis-Online Datenlizenz by-2-0: `https://www.govdata.de/dl-de/by-2-0`. Sept. 2022. URL: `https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Gesundheit/Behinderte-Menschen/Publikationen/Downloads-Behinderte-Menschen/sozial-schwerbehinderte-kb-5227101219004.pdf` (visited on 2024/05/21).

[54] Katherine M Tsui, Dae-Jin Kim, Aman Behal, David Kontak and Holly A Yanco. '"I want that": Human-in-the-loop control of a wheelchair-mounted robotic arm'. In: *Applied Bionics and Biomechanics* 8.1 (2011), pp. 127–147. 10.3233/ABB-2011-0004.

[55] Philippe Turgeon, Michaël Dubé, Thierry Laliberté, Philippe S. Archambault, Véronique H. Flamand, François Routhier and Alexandre Campeau-Lecours. 'Mechanical design of a new device to assist eating in people with movement disorders'. In: *Assistive Technology* (2020). PMID: 32105199, pp. 1–8. 10.1080/10400435.2020.1734111.

[56] Sumukha Udupa, Vineet R. Kamat and Carol C. Menassa. 'Shared autonomy in assistive mobile robots: a review'. In: *Disability and Rehabilitation: Assistive Technology* 18.6 (June 2021), pp. 827–848. ISSN: 1748-3115. 10.1080/17483107.2021.1928778.

[57] Rudolph Van Der Merwe and Eric Wan. 'Sigma-Point Kalman Filters for Probabilistic Inference in Dynamic State-Space Models'. AAI3129163. PhD thesis. Oregon Health & Science University, 2004, p. 377. URL: `https://www.gatsby.ucl.ac.uk/~byron/nlds/merwe2003a.pdf` (visited on 2024/07/02).

[58] Jörn Vogel, Katharina Hertkorn, Rohit U. Menon and Máximo A. Roa. 'Flexible, semi-autonomous grasping for assistive robotics'. In: *2016 IEEE International Conference on Robotics and Automation (ICRA).* Piscataway, NJ: IEEE Press, 2016, pp. 4872–4879. 10.1109/ICRA.2016.7487692.

[59] Yujin Wakita, Eimei Oyama, Woo-keun Yoon, Hideyuki Tanaka, Yoshio Matsumoto, Arthur Blom and Harry Stuyt. 'User evaluation of service robotic arms based on ICF through interviews with people with upper-limb disability'. In: *2013 IEEE International Conference on Robotics and Biomimetics (ROBIO).* 2013, pp. 1282–1287. 10.1109/ROBIO.2013.6739641.

[60] W. Grey Walter. 'An Imitation Of Life'. In: *Scientific American* 182.5 (1950), pp. 42–45. ISSN: 00368733, 19467087. URL: `http://www.jstor.org/stable/24967456` (visited on 2024/01/11).

[61] Yu Xiang, Tanner Schmidt, Venkatraman Narayanan and Dieter Fox. *PoseCNN: A Convolutional Neural Network for 6D Object Pose Estimation in Cluttered Scenes*. Pittsburgh, Pennsylvania, June 2018. 10.15607/RSS.2018.XIV.019.

[62] Xu Zhao, Wenchao Ding, Yongqi An, Yinglong Du, Tao Yu, Min Li, Ming Tang and Jinqiao Wang. *Fast Segment Anything*. 2023. arXiv: 2306.12156 [cs.CV].

[63] Ming Zhong, Yanqiang Zhang, Xi Yang, Yufeng Yao, Junlong Guo, Yaping Wang and Yaxin Liu. 'Assistive Grasping Based on Laser-point Detection with Application to Wheelchair-mounted Robotic Arms'. In: *Sensors* 19.2 (2019), p. 303. 10.3390/s19020303. URL: https://www.mdpi.com/1424-8220/19/2/303.

[64] Jakub Złotowski, Kumar Yogeeswaran and Christoph Bartneck. 'Can we control it? Autonomous robots threaten human identity, uniqueness, safety, and resources'. In: *International Journal of Human-Computer Studies* 100 (2017), pp. 48–54. ISSN: 1071-5819. 10.1016/j.ijhcs.2016.12.008. URL: https://www.sciencedirect.com/science/article/pii/S1071581916301768.

[65] Dragiša Žunić and Joviša Žunić. 'Shape ellipticity from Hu moment invariants'. In: *Applied Mathematics and Computation* 226 (2014), pp. 406–414. ISSN: 0096-3003. 10.1016/j.amc.2013.10.062. URL: https://www.sciencedirect.com/science/article/pii/S0096300313011193.

Appendix A List of Mathematical Symbols

b		A single behaviour describing an elementary operation
B		Set of behaviours $(b_1, b_2, \dots)$
c	$\in \mathbb{R}^m$	User input
D	$\in \mathbb{R}^{n\times m}$	Robot control mode as set of DoFs
$\hat{D}$	$\in \mathbb{R}^{n\times n}$	Set of modes, Complete set of DoFs
e_i	$\in \mathbb{R}^n$	Normalised eigenvector ($\lVert e \rVert_2 = 1$) generated by the AI's PCA
E_f		Set of the first f predicted eigenvectors and eigenvalues
f	$\in \mathbb{N}$	Number of DoFs used in the PoM
G		RV describing which behaviour the user desires given the situation z
h_b		Distribution of a behaviour b's RV of desired motion V_b
k	$\in \mathbb{N}$	Number of sample paths Y generated by the deep learning backbone
m	$\in \mathbb{N}$	Number of DoFs of an input device
n	$\in \mathbb{N}$	Number of DoFs of a controlled robot
N_i	$\in \mathbb{N}$	Number of participants in Study i
Q_b		Set of sigma points in behaviour b
S		Function $S(Q, \mu)$ calculating the general covariance of a set of sigma points Q with defined reference point μ
t_i	$\in \mathbb{R}$	Point in time of Figure 12
u	$\in \mathbb{R}^n$	Control output sent to the robot
u_0	$\in \mathbb{R}^n$	Baseline control output sent to the robot if $c = 0$
U		RV representing the n-dimensional direction of robot motion desired by the user
v_i	$\in \mathbb{R}$	Eigenvector-associated eigenvalue generated by the deep learning-head's PCA, with $\sum_{i=1}^{n} v_i = 1$
V_b		RV describing the direction of motion required by the user to perform the behaviour b's underlying operation in the current situation given the desired behaviour
y_i	$\in \mathbb{R}^n$	One of k sample paths (Y) generated by the deep learning backbone
Y	$\in \mathbb{R}^{n\times k}$	Set of sample paths generated by the deep learning backbone
z		Sensor data representing the current situation
Z		RV describing the current situation as measured by the systems sensors
α_b	$\in \mathbb{R}$	Scaling factor of the mixture mode of behaviour b
μ_b	$\in \mathbb{R}^n$	Mean of a behaviour b's normal distribution h_b
$\mu_{\text{behaviour-based}}$	$\in \mathbb{R}^n$	Mean of $U \mid Z = z$ modelled as a mixture distribution
$\mu_{\text{end-to-end}}$	$\in \mathbb{R}^n$	Mean of $U \mid Z = z$ modelled as a normal distribution
Σ_b	$\in \mathbb{R}^{n\times n}$	Covariance of a behaviour b's normal distribution h_b
$\Sigma_{\text{behaviour-based}}$	$\in \mathbb{R}^{n\times n}$	Covariance of $U \mid Z = z$ modelled as a mixture distribution
$\Sigma_{\text{end-to-end}}$	$\in \mathbb{R}^{n\times n}$	Covariance of $U \mid Z = z$ modelled as a normal distribution

Appendix B Button Mappings and Control Overviews

This section contains cheat sheets of control interfaces used in the AdaMeKoR project presented in Section 3:

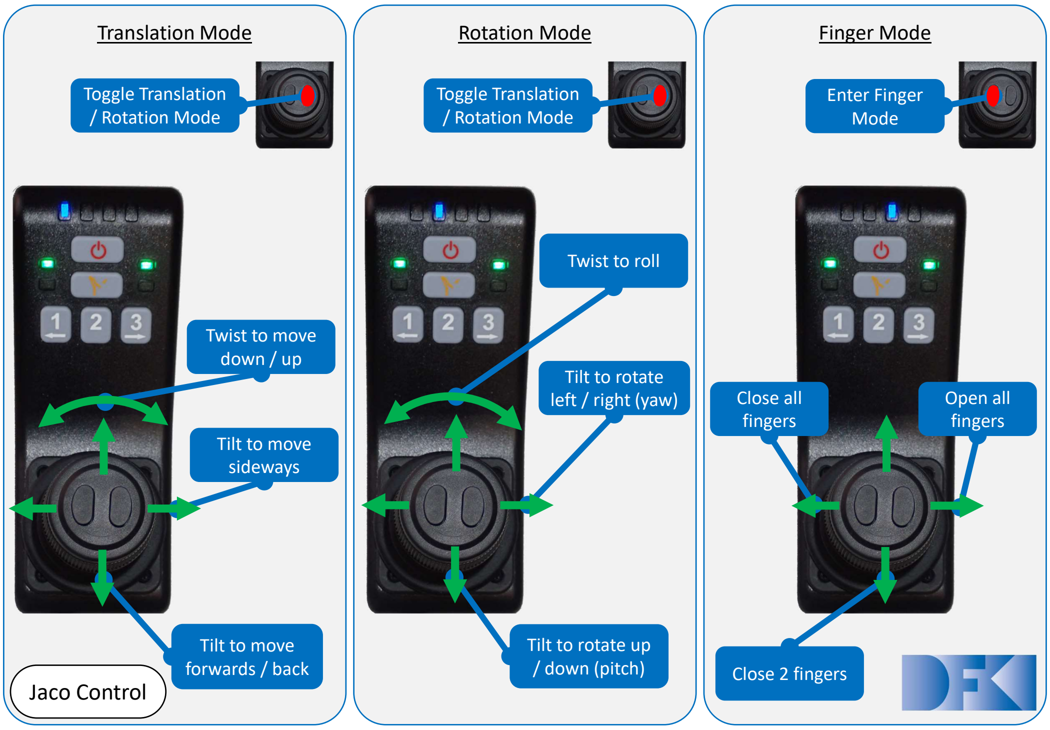


Figure 29: Control Interface Cheat Sheet: Kinova Joystick

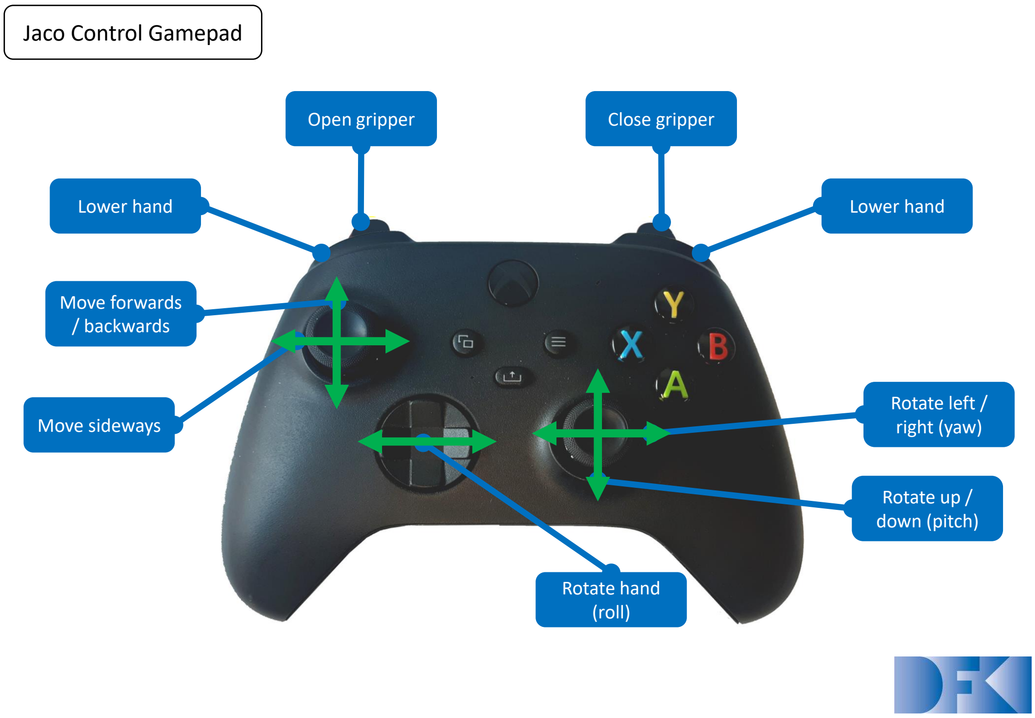


Figure 30: Control Interface Cheat Sheet: Gamepad

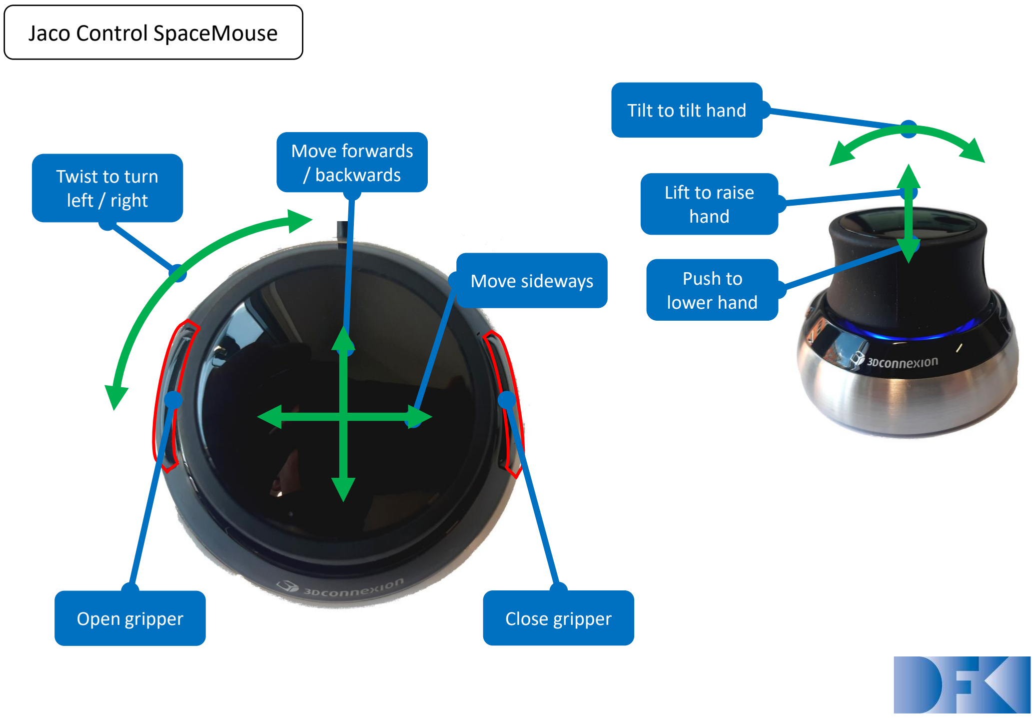


Figure 31: Control Interface Cheat Sheet: 3D Mouse

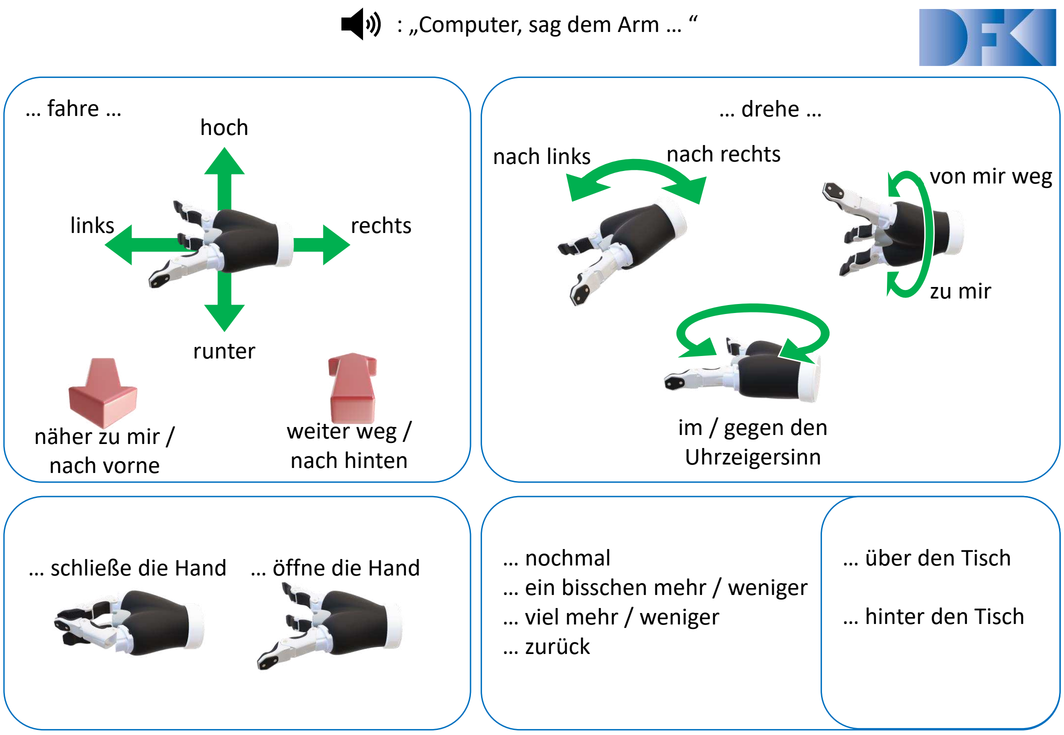


Figure 32: Control Interface Cheat Sheet: Voice Control Interface (German)

Appendix C Training Graphs for 3D

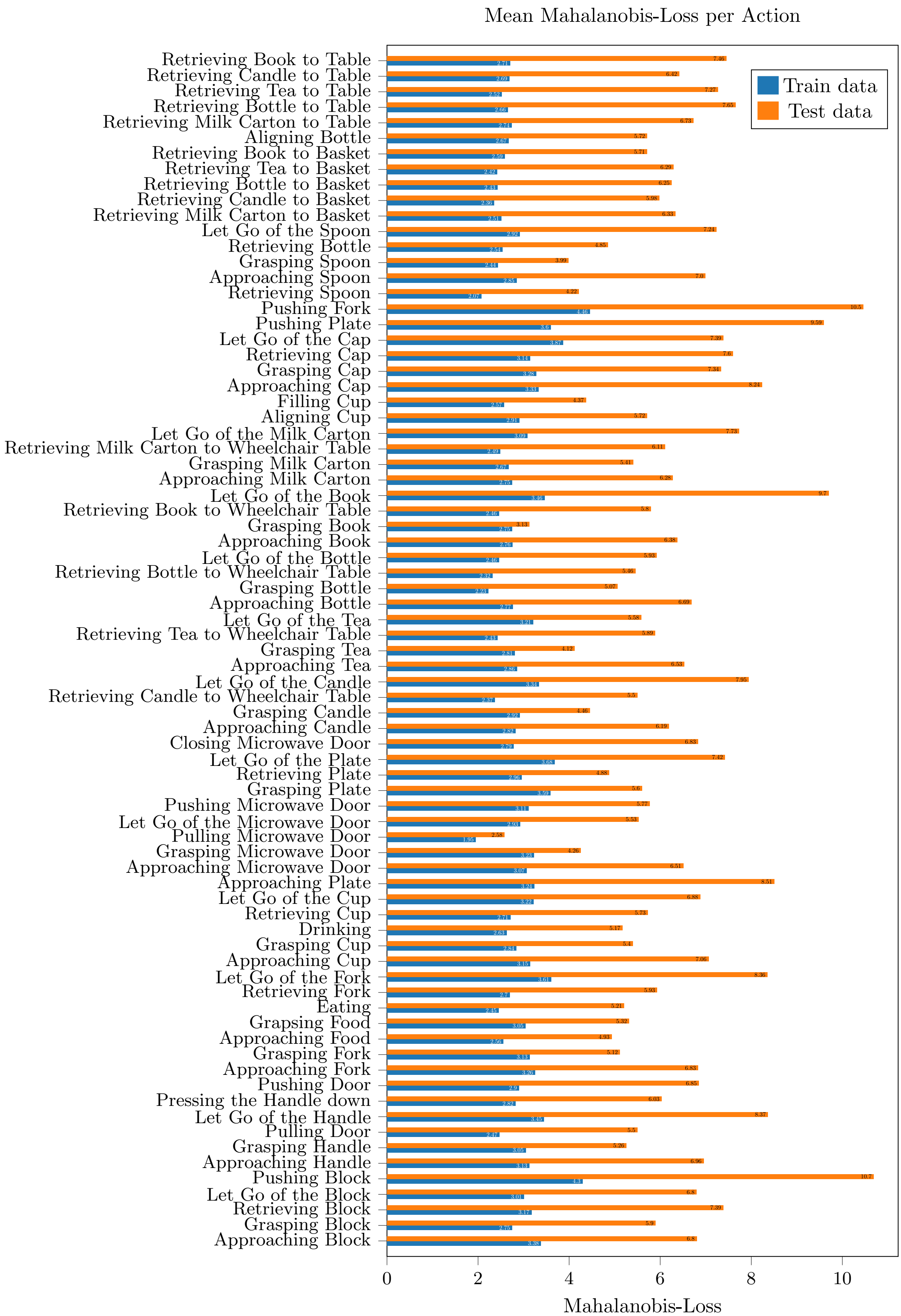


Figure 33: Mean Mahalanobis-Loss per Recording and Action for Test and Training Data

Appendix D Behaviour Control Loop

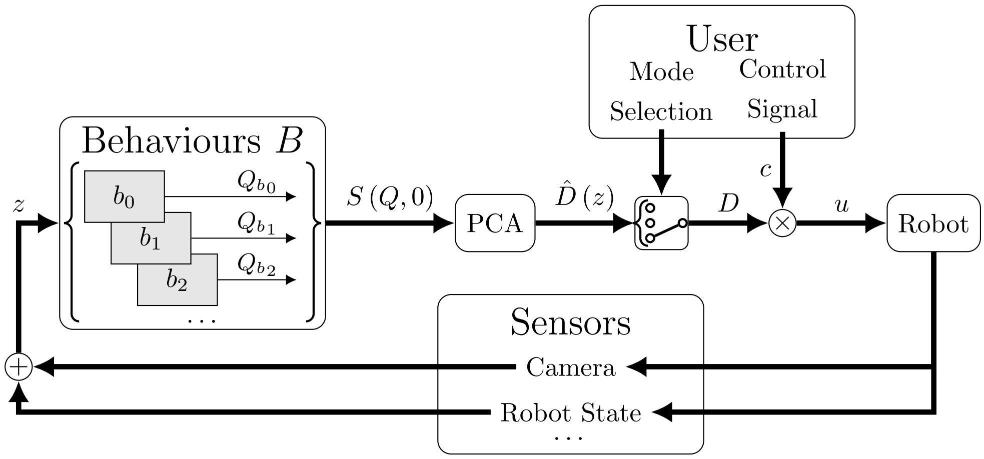


Figure 34: Behaviour-based Software Control Loop: Based on the current situation z, the set of behaviours B generates sigma points Q_b, accumulated to a probability distribution using $S(Q, \mu)$ (see Equation 16). The user controls the robot along axes of the latter's principal components. Adapted from [V]

Publication Documents

The following pages contain copies of the published documents, as listed in the Publications section, following the order of occurrence there. The respective reference for each file can be found at the top of each following page.

For most of the documents, the copies will be identical to the files available via the given URLs. Changes have only been made in regard to the documents available for [XII, II, XI], as the files online are not limited to the respective works but include the complete proceedings. In case of [II], this thesis only prints the relevant page, whereas the pages of [XII] (and [XI] in the electronic version) only reprint the abstract.

Copyright Disclaimer

[XI] The full paper is currently in review by the International Journal of Social Robotics: Special Issue on Robotic Systems for Nursing Care. The electronic version of this work therefore only includes the abstract.

The Importance of Participatory Design for the Development of Assistive Robotic Arms. Initial Approaches and Experiences in the Research Projects MobILe and DoF-Adaptiv

Annalies Baumeister, Elizaveta Gardo, Patrizia Tolle, Barbara Klein, Max Pascher, Jens Gerken, Felix Goldau, Yashaswini Shivashankar, Udo Frese

DOI: 10.48718/8p7x-cw14

Abstract

This Article introduces two research projects towards assistive robotic arms for people with severe body impairments. Both projects aim to develop new control and interaction designs to promote accessibility and a better performance for people with functional losses in all four extremities, e.g. due to quadriplegic or multiple sclerosis. The project MobILe concentrates on using a robotic arm as drinking aid and controlling it with smart glasses, eye-tracking and augmented reality. A user oriented development process with participatory methods were pursued which brought new knowledge about the life and care situation of the future target group and the requirements a robotic drinking aid needs to meet. As a consequence the new project DoF-Adaptiv follows an even more participatory approach, including the future target group, their family and professional caregivers from the beginning into decision making and development processes within the project. DoF-Adaptiv aims to simplify the control modalities of assistive robotic arms to enhance the usability of the robotic arm for activities of daily living. To decide on exemplary activities, like eating or open a door, the future target group, their family and professional caregivers are included in the decision making process. Furthermore all relevant stakeholders will be included in the investigation of ethical, legal and social implications as well as the identification of potential risks. This article will show the importance of the participatory design for the development and research process in MobILe and DoF-Adaptiv.

Keywords: Assistive robotics, Assisted Living Technologies, Participatory Design, Human-centered Design, User Acceptance, Risk Management

Assistive Robotic Arms for a Self-determined Life

At the end of 2019, the German Federal Statistical Office counted 7.9 million people with disa-bilities. Thereof 11.2% had functional losses in arms and/or legs and 10.4% in spine and torso. 7.6 million people with severe disabilities lived in private households with their families, part-ners or alone (Statistisches Bundesamt, 2020a + 2020b). As part of the German Spinal Cord Injury Survey (GerSCI) in 2017, 1,479 people with spinal cord injuries were questioned about their life situation, how they experiences living with a spinal cord injury and what kind of issues they were dealing with. Results towards daily activities and mobility showed that running an own household was viewed as extremely problematic by 40.7% of the respondents. Rated as very or extremely problematic were picking up small objects or opening containers (31.2%) and com-pleting everyday tasks (29.1%). Furthermore, 41% of the respondents said that they could open heavy doors only with some or great difficulties and 18.3% were unable to open heavy doors at all. The authors state that the descriptive statistical results will be followed by qualitative studies and in-depth analyses (Bökel et al., 2019, 22f.; 41).

Assistive robotic arms could be a solution to enable people with severe body impairment to perform activities of daily living independently. Currently, the online-Portal "rehadat-Hilfsmittel.de" lists three assistive robotic arms on the German market, JACO from Kinova inc., iArm from Assistive Innovations bv. and BATEO from EXXOMOVE UG. All robotic arms are mounted on an electric wheelchair and steered with the wheelchair controller. However, there are still some challenges. In case of functional losses in hands and arms, steering with a special controller is possible, but has its limits. Operating a robotic arm with a chin or mouth controller and thus eating or drinking, for example, are mutually exclusive. Another difficulty is the necessity to constantly switch between different modes of movement of the robotic arm, like up/down and left/right, while performing a task.

The research projects MobILe and DoF-Adaptiv presented in this article deal with new approaches to the above-mentioned challenges. Furthermore, ethical, social and legal implications (ELSI) and questions of risk and quality management are investigated. In the spirit of Responsible Research and Innovation (see Owen, Stilgoe et al., 2013), the participation of the future target group was or is being pursued in both projects. Against this backdrop, this article is intended to clarify the importance of participatory approaches or participation in the research process and technology development for research in the field of assistive robotic arms for people with severe physical limitations.

Investigating the Needs and Wishes of the User in MobILe

How could it be feasible to control and use a semi-autonomous robotic arm as a drinking aid, if using a joystick is not an option, due to functional losses in all four extremities? The aim of the MobILe project was to research and implement basic skills with and without direct physical contact between robots and humans. For robot control in a three-dimensional space, the use of head and eye movements as well as a combination of motion sensors and glasses with an eye tracker and electrooculography was investigated. For the interaction between the robot and the human, augmented reality

MobILe – Physical Human-Robot-Interaction	
Announcement BMBF:	Autonome Roboter für Assistenzfunktionen: Interaktive Grundfertigkeiten
Research funding:	VDI/VDE Innovation + Technik GmbH
Grand agreement ID:	16SV7868
Project coordination:	Prof. Dr. Marion Gebhard, Westfälische Hochschule
Project leadership Frankfurt UAS:	Prof. Dr. Barbara Klein
Joint project partners:	Friedrich Wilhelm Bessel Institut Forschungsgesellschaft m. b. H., Bremen Hidrex GmbH, Heiligenhaus Pi4 robotics GmbH, Berlin
Duration:	07.2017-06.2020, extended 06.2021
Website:	https://www.interaktive-technologien.de/projekte/mobile

Table 1: Fact sheet MobILe (authors figure)

(e.g., in the form of visual representations of intended actions by the robot) and a visual concept were developed. A safety system with redundancies ensures functional reliability. To ensure that the development of the control and interaction modalities meet the future users' needs and acceptance, a user-centred design strategy was implemented throughout the research and development process.

An Ethnographic/Small Data Analysis

At the beginning of the project an ethnographic/small data analysis with 15 persons of the future target group was conducted in Germany. Eleven male and four females took part in the survey. The diagnoses were spinal cord injury (10), locked-in-syndrome (2), multiple sclerosis (1), inclusion body myositis (1) and arthrogryposis (1). Participants were visited at home, where in-terviews and participatory observations of food and beverage intake took place.

The observations were documented with videos and photos. The interviews and videos were transcribed and the transcripts, videos and photos were analysed following the qualitative content analysis approach by Mayring (2015), the qualitative hermeneutical approach from social sciences by Reichertz & Englert (2011) and the photo analysis by Pilarczyk & Mietzner (2000). The aim was to gain a deeper insight into the life situation of the participants in order to work out aspects that promote acceptance and to develop user-centred recommendations for the technical development.

Demographic Data, Life and Care Situation

The age of the participants varied between 18 and 62 years and was distributed relatively evenly. The diagnoses or the accident events ranged from 2 to 35 years ago at the time of the survey. The only exception was in a case with congenital physical impairment. Around two-thirds of the participants have been living with a physical impairment for at least ten years or longer. However, no significant differences can be derived from the response behaviour in the interviews between participants who have been living with their impairment for decades and those who have obtained their physical impairment more recently. Experiences with assistive robots or technologies seem to be less dependent on the duration of the impairment than much more on one's own interests or accessibility.

All participants stated that they are active and like to go out as much as possible. Some have converted cars that the caregivers drive. Most of the participants use an electric wheelchair (ten participants). Others prefer active wheelchairs (three participants) or use push wheelchairs (three participants). The participants live in accessible houses (six participants) and in accessible or partially accessible apartments (nine participants). Seven participants live alone, three live with their parents, three with a wife/husband or partner (in one case with a child under 18 years) and one participant lives with one child over 18 years. Caregiving relatives exist in eight cases in which the participants live with their relatives, whereby the extent of the care provided varies. In seven of the eight cases, nursing services take on certain aspects of nursing (e.g., catheters or showers). In one case, assistants also support the participant and his relatives. Participants who are not cared for by relatives often use both nursing services and assistants (five participants). Only in two cases are participants supported exclusively by a nursing service or assistants.

The daily structure is characterized by fixed times of basic care (e.g., the morning toilet), relatively fixed/regular mealtimes, regular therapies and times of relief (pressure ulcer prophylaxis). Most of the time is spent in the bedroom (bed) and living room (wheelchair). The amount of time spent in the wheelchair varies between 8 and 14 hours a day. Participants said that they eat and drink in bed or while sitting in a wheelchair in the living room or in the kitchen at a dining table. Drinking has a special significance, as both the type of beverage (still water, tea) and the amount consumed are particularly important for health and general well-being. The participants often drink large quantities in a row and preferably from large glasses or drinking bottles. For all participants, drinking is linked to the aspects of autonomy and privacy. If they did not have to ask for help for every sip, but could drink independently, they would be able to spend several hours a day alone again. Always having to have someone around for assistance is perceived as exhausting.

The Users Perspective Towards an Assistive Robotic Arm as Drinking Aid

The analysis of the interviews and observations showed certain aspects that can influence the acceptance and use of a robotic arm as a drinking aid. Currently, the above mentioned assistive robotic arms can only be used while mounted on an electric wheelchair. However, some partic-ipants cannot or do not want to use an electric wheelchair. Participants would also like to use the robot while lying in their bed and would prefer a solution that enables them to use the robot independently. Most participants are cared for 24/7 by a mix of caregiving relatives, nursing services and assistants, which means, that different people are interacting directly or indirectly with the robot. Therefore, participants wish for an easy and intuitive control and interaction design that does not require too much explaining or a long training period. It should also be safe to use, even if other people are around and interacting with the participant. Furthermore, the robot should be robust, sturdy and show a high operating safety. Teething problems or safety issues would prevent participants to use the robot at all. How the robotic arm could be perceived by others is important to most participants. If it is too big and "showy" and/or looks too much like a medical device, participants do not feel comfortable and fear stigmatization. Instead, they wish for an unobtrusive and elegant design, like a lifestyle product and as space-saving as possible. If the robotic arm promotes independency and privacy, meaning that it enables the participants to drink without additional help from others, the participants stated that they would use it regularly and would like to spend more time on their own. On the other hand, participants would like to enjoy mealtimes with family and friends where everyone can eat and drink at the same time and communication is not disturbed while someone takes care of the participant. A robotic aid is seen as a possible solution as long as it is not the cause for new distraction. These findings lead to recommendations for the development of robotic drinking and eating aids, which are currently published at INTERACT 2021 (see Pascher/Baumeister 2021).

New Insights Through Including Caregivers

To discuss ethical and social implications of a robotic drinking aid, three workshops with all stakeholders were conducted in 2019. A total of 11 people with a disability, 2 caregiving parents, 4 assistants, 2 physiotherapists and 3 industry representatives took part. In addition, all project partners joined the first workshop, too. The MEESTAR model for the ethical evaluation of socio-technological arrangements was adapted and used to evaluate ethical issues. All workshops were transcribed and qualitatively analysed following the thematic analysis by Tuckett (2005, 75ff.).

Caregiving parents and assistants brought a new perspective towards the topics safety and independence. Both participating groups viewed an assistive robot as positive if they feel assured that the person they care for can be safely left alone to perform a task with it. If the robot promotes a higher independency, it would be a relief for caregivers. Assistants could use their time more efficiently and caregiving relatives would gain more time for themselves. However, it also became clear during the discussions that it is an issue to trust in the safety of an assistive robot and leave the person being cared for alone with it. Especially assistants said that they don't trust the robot and asked what would happen, when a problem occurs? They do not want to risk that the person being cared for comes to any harm or that they cannot help them (fast enough), if needed. Participants with a disability, on the other hand, were much more inclined to take risks, if only they could regain more privacy. Whilst participants with a disability say that safety standards in Germany are very high and they could call the caregivers with a mobile phone if necessary, the caregivers prefer them to be in another room/close by and quickly available when needed. So there is a profound conflict of interest between caregivers and the person being cared for about gaining more privacy and autonomy.

Lesson Learned for Further Research

During the visits of the participants, some asked why MobILe focused on drinking and how the decision was made. Although drinking was viewed as important, participants and caregivers mentioned further activities a robotic arm could help with. Participants mostly wished to pick up and manipulate objects or open doors. Again, in the context of the workshops, people with a disability, parents and assistants suggested more different activities, like support for dressing. It seems that people with severe body impairments do have several unmet needs that an assistive robotic arm could help with and that it is important to include the future target group into the decision-making before deciding what activity the assistive robotic arm should support. The inclusion of caregiving relatives and assistants in the workshop made clear that it is necessary to recognize the whole social system of people with severe body impairments and to include caregiving relatives, nurses and assistants into the research project as well. The conclusion was to pursue a more participatory approach in the succession project DoF-Adaptiv.

DoF-Adaptiv - Adaptive mapping of degrees of freedom as user interface for an assistive robot	
Announcement BMBF:	Adaptive Freiheitsgradeinbettung als kooperatives Userinterface für einen Assistenzroboter
Research funding:	VDI/VDE Innovation + Technik GmbH
Grand agreement ID:	16SV8564
Project coordination:	Prof. Dr. Udo Frese, DFKI, Bremen
Project leadership Frankfurt UAS:	Prof. Dr. Barbara Klein, Prof. Dr. Patrizia Tolle
Joint project partners:	Westfälische Hochschule, Gelsenkirchen Munevo GmbH, München
Duration:	02.2021-01.2024
Website:	https://www.interaktive-technologien.de/projekte/dof-adaptiv

Table 2: Fact sheet DoF-Adaptiv (authors figure)

Including All Stakeholders in the Research and Development Process of an Assistive Robotic Arm

The ongoing research project DoF-Adaptiv aims to simplify the use of assistive robotic arms, i.e. releasing the users from the necessity to switch between the various modes of degrees of freedom while performing a task. A combination of machine learning and artificial intelligence (AI) will be developed in order to improve the control system. In contrast to automatic control, humans remain in control. In a user-centred design process, possibilities of displaying feedback are explored using data glasses. In a participatory approach, the requirements are raised and the system is evaluated with those affected.

Participative Research, Participatory Approaches

DoF-Adaptiv is conducted as research with people, rather than as research about or for them. This course of action is a characteristic feature of participatory research (Bergold/Thomas, 2010, 333).
The DoF-Adaptiv project is guided by fundamental principles of participatory research. The following section focuses on two areas: the concept of „safe space“ for all participants during the research process and the roles of all participants in decision-making processes (von Unger, 2014, 39ff.; Bergold/Thomas, 2012, 6).

Participatory research can be understood as a research style in which professional researchers and co-researchers are equally involved in the process of knowledge construction (Bergold/Thomas, 2012, 2). In DoF-Adaptiv, two groups, the primary and secondary users of assistive technologies, form the group of co-researchers. People with disabilities (primary users), family caregivers, nurses, assistants (secondary users) collectively contribute their unique perspectives. Expressing and sharing their own opinions and experiences in institutional settings or with strangers requires a “safe space“ (Bergold/Thomas, 2012, 5). The “safe space“ allows, for example, to have and express different opinions or to resolve conflicts in a constructive manner (Bergold/Thomas, 2012, 7). In addition, a “safe space“ creates an opportunity for all members of the research group to experience that each opinion will be heard, but will not be judged or even devalued by others. The “safe space“ is also dynamic. It has to be reestablished over and over again via opening communicative spaces throughout the entire research process (Bergold/Thomas 2010, 338).

To determine whether a project fulfills the basic criterion for classification as participatory research, it is necessary to ask who is controlling the research in which phase of the project (Bergold/Thomas, 2012a, 9). These questions allow to determine which group of researchers (professional researcher or co-researcher) is involved in which decisions, whether the various actors are involved in decision-making-processes with equal rights in order to control and monitor the research and the course of the project (Bergold/Thomas, 2012, 10f.).
The research process in DoF-Adaptiv is designed in cycles. These cycles are based on the process of “Progressive Problem Solving with Action Research” (Riel quoted in Wright et al., 2013, 147). Each cycle describes the reflecting on practice, the actions taken, reflecting and taking further action (Riel n.d.)

In the following, the concept of “safe space“ and decision-making processes used to control and monitor the research process at DoF-Adaptiv are explained with a focus on the scenario development.

Including the Future Users Early on

Based on the experiences in MobILe, the future target group in DoF-Adaptiv includes not only people with a disability, but their caregiving relatives, nurses and assistants, too. Futhermore, the project aims for a more participatory approach that involves the future target group in the research and development process from the beginning and is iterative throughout the project. Starting with the development of application scenarios for the assistive robotic arm, workshops and interviews were held. It was decided which activities the robotic arm should support and which scenarios should be described great importance was attached to the fact that participants and researchers meet on an equal footing. Researches were in the role of listening and documenting the participant's thoughts and wishes, accepting them as experts in their own rights. Protocols, findings and work-in-process documents were shared with all participants throughout the development process. When deciding on application scenarios, the opinion of the participants was decisive for the decision-making. Furthermore, participants were given the opportunity to discuss the details of the scenarios design with all researchers during a project meeting. Those who couldn't attend the meeting were asked to give their view via e-mail or a one-to-one call with a researcher, making sure that every participant who wanted to be part of the development process and decision-making could participate. The transcripts, notes and protocols of the workshops, interviews and the project meeting are currently being analysed and the participative approach will be evaluated.

Impact of Participatory Approaches for Risk and Quality Management

In the field of medical technology, there is a separate standard for managing risks. DIN EN ISO 14971 describes requirements and possible procedures with a focus on risk analysis. Risks have to be identified and assessed to determine whether they are justifiable. Remaining residual risks must be set in relation to the benefits. In 2019, a revision of the risk management standard DIN EN ISO 14971 for medical devices was published: ISO 14971: 2019. The procedure has basically remained unchanged. However, some additions and explanations have been introduced. Among other things, the information for training option was discussed as a risk control measure. In addition, the concept of benefit for the patient was further specified. The standard has thus moved the issue of risk-benefit ratio even more into focus. In the area of risk management, the EU directives and EU regulations for medical devices such as the MDR (Medical Device Regulation (MDR)) have to be taken into account too. The risk analysis must factor in, inter alia, the effects of software errors, framework conditions and safety-relevant functions.

As part of a project on „Systemic risk management for the holistic consideration of entrepreneurial risks using the example of medical technology“ at RWTH Aachen University, some weaknesses in the current procedures and methods in the field of risk management in medical technology were discussed and described (SysRisk). Among other things, the current methods are seen insufficient in scope and depth to identify and assess residual risks as comprehensively as is necessary for risk-sensitive products. In order to minimise these weak points in the current project, risk management is being expanded to include methods and procedures that go beyond the mandatory standards and regulations.

For this reason, the current project (DoF-Adaptiv) puts a lot of emphasis on identifying as many risks as possible at an early stage and introducing appropriate measures in a timely manner. The following procedures and methods are used:

- Users are involved in the risk identification process from the beginning of the project.
- FMEA (Failure Mode and Effect Analysis) is used in the risk analysis to analyse the individual components of the product, in particular the resulting hazards.
- The scenario technique or scenario-based risk analysis is used for the investigation and evalua-tion of the influencing factors and simulated scenarios.
- Practical conclusions for the intended area of application of the product are derived from the action-oriented error taxonomy. The action-oriented error taxonomy is based on an error term that can be traced back to action-oriented error research (Freud, 1941, 25f.; Frese, Zapf, 1991, 11f.).

Conclusions

The early involvement of the future target group in the research project MobILe had a deep impact on the further development of human-robot interaction modalities. The insight into the life and care situation of people with severe body impairments brought findings towards aspects that promote a higher acceptance of assistive robotic arms, allowed a user-centred development and new input for further research was gained. In the ongoing project DoF-Adaptiv, the inclusion of the future target group into development and decision-making processes ensures that the research project meets the future target group needs and will increase their independence and autonomy in the future. In this way, both projects seek to contribute to a more independent living for people with severe body impairments in the future.

References

Bergold, J. & Thomas, S. (2012): Partizipative Forschungsmethoden: Ein methodischer Ansatz in Bewegung. In: Forum Qualitative Sozialforschung 13(1) 1-33, Online unter: https://www.qualitative-research.net/index.php/fqs/article/view/1801/3333. (accessed on: 15.06.2021).

Bergold, J. & Thomas, S. (2012a): Participatory Research Methods: A Methodological Approach in Motion. (13) 1, Forum: Qualitative Social Research 13(1) 1-31, Online unter: DOI: https://doi.org/10.17169/fqs-13.1.1801 . (accessed on: 03.07.2021).

Bergold, J. & Thomas, S. (2010): Partizipative Forschung. In: G. Mey & K. Mruck (Hrsg.): Handbuch Qualitative Forschung in der Psychologie. Wiesbaden: Springer, 333-344.

Bökel, A. / Blumenthal, M. / Egen, C. / et al (2019): Querschnittlähmung in Deutschland. Eine nationale Befragung (German Spinal Cord Injury Survey (GerSCI) Teilprojekt des Spinal Cord Injury Community Survey (InSCI)). DOI: 10.26068/mhhrpm/20191009-003. Online unter: https://www.mhh.de /fileadmin/mhh/rehabilitationsmedizin/downloads/GerSCI-Projektbericht_2019.pdf (accessed on: 07.07.2021)

Frese, M. & Zapf, D. (1991): Fehlersystematik und Fehlerentstehung: Eine theoretische Einführung Huber: Bern.

Freud S. (1941): Zur Psychopathologie des Alltagslebens. Imago: London.

Mayring, P. (2015): Qualitative Content Analysis: Basics and Techniques (nach “Qualitative Inhaltsan lyse: Grundlagen und Techniken“). Beltz: Weinheim.

Owen, R./ Stilgoe, J./ Macnaghten, P. et al. (2013): A Framework for Responsible Innovation. In: Owen, R./ Bessant, J./ Heintz, M. (Editors) Respnsible Innovation: Managing the Responsible Emergence of Sciences and Innovation in Society, Wiley: Chennai, 27-50. DOI:10.1002/9781118551424.ch2

Pascher, Max/ Baumeister, Annalies/ Schneegass, Stefan/ Klein, Barbara/ Gerken, Jens (2021): Recommendations for the Development of a Robotic Drinking and Eating Aid - An Ethnographic Study. In: Carmelo Ardito Rosa Lanzilotti, Alessio Malizia Helen Petrie Antonio Piccinno Giuseppe Desolda Kori Inkpen (Hrsg.): Human-Computer Interaction – INTERACT 2021, Springer, Cham, Forthcoming.

Pilarczyk, U./ Mietzner, U. (2000): Picture science methods in research on education and social science (nach „Bildwissenschaftliche Methoden in der erziehungs- und sozialwissenschaftlichen Forschung“). In: Zeitschrift für qualitative Bildungs-, Beratungs- und Sozialforschung 1(2), 343-364.

Reichertz, J.(2011): Introduction to Qualitative Video Analysis: A hermeneuticsociological Analy-sis Tool (nach „Einführung in die qualitative Videoanalyse. Eine hermeneutisch-wissenssoziologische Fallanalyse“). VS Verlag für Sozialwissenschaften: Wiesbaden.

Riel, M. (o.J.): Understanding Action Research. Center for Collaborative Action Research. Online unter: https://www.ccarweb.org/what-is-action-research. (accessed on: 02.07.2021).

Statistisches Bundesamt (Destatis) (2020a): Disability Facts and Figures - Brief Report 2019 (nach „Statistik der schwerbehinderten Menschen - Kurzbericht 2019“). https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Gesundheit/Behinderte-Menschen/_inhalt.html (accessed on: 07.07.2021).

Statistisches Bundesamt (Destatis) (2020b): Public social benefits – life situation of people with disability – Results oft he Microzensus 2019 (nach „Öffentliche Sozialleistungen - Lebenslagen der behinderten Menschen - Ergebnis des Mikrozensus 2019“) https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Gesundheit/Behinderte-Menschen/_inhalt.html (accessed on: 07.07.2021).

Tuckett A. G. (2005): Applying thematic analysis theory to practice: a researcher‘s experience. In: Contemporary Nurse, 19(1-2):75-87. DOI: 10.5172/conu.19.1-2.75.

von Unger, H. (2014): Partizipative Forschung. Einführung in die Forschungspraxis. Wiesbaden: VS Springer.

Wright, M.T.; Block, M.; Kilian, H.; Lemmens, K. (2013): Förderung von Qualitätsentwicklung durch Partizipative Gesundheitsforschung. In: Prävention und Gesundheitsförderung 8 (3), 147-154, Online unter: https://link.springer.com/article/10.1007/s11553-013-0396-z. (accessed on: 25.06.2021).

AI for simplifying the use of an assistive robotic arm for people with severe body impairments

A. Baumeister, M. Pascher, Y. Shivashankar, F. Goldau, U. Frese, J. Gerken, E. Gardó, B. Klein, P. Tolle

Purpose Assistive robotic arms, e.g., the Kinova JACO, aim to assist people with upper-body disabilities in everyday tasks and thus increase their autonomy (Brose et al. 2010; Beaudoin et al.2019). A long-term survey with seven JACO users showed that they were satisfied with the technology and that JACO had a positive psychosocial impact. Still, the users had some difficulties performing daily activities with the arm, e.g., it took them some time to finish a task (Beaudoin et al. 2019). Herlant et al. claim that the main problem for a user is that mode switching is time-consuming and tiring (Herlant et al. 2017). To tackle this issue, deep neural network(s) will be developed to facilitate the use of the robotic arm. A sensor-based situation recognition will be combined with an algorithm-based control to form an adaptive AI-based control system. The project focuses on three main aspects: 1) A neural network providing suggestions for movement options based on training data generated in virtual reality. 2) Exploring data glasses as a possibility for displaying feedback in a user-centered design process. 3) Elicitation of requirements, risks and ethical system evaluation using a participatory approach. **Method** In a first step, everyday scenarios that are relevant for the user, like eating and drinking (Pascher et al. 2021), were identified. Based on the iterative, cyclical process of action research by Riel (2020), two workshops and six interviews with people from our target groups were conducted to learn about their care situations and needs. Four scenarios were elaborated and decided on together. The simulation system consists of these scenarios along with detailed movements and manipulations (Kronhardt & Rübner et al. 2022). The user controls the simulated robot's hand via a VR controller. This makes it possible to record movements quickly, which is necessary to achieve a large training data set. Thus, the generated data is used for training a neural network to provide an adaptive set of controls. In the next step, a novel control method and possible visual cues for the DoF mappings were developed. The objective is to explore how the novel adaptive control method performs in a 3D environment compared to the standard mode-switch approach with cardinal DoF mappings and whether changes in the visual cues impact the performance of the adaptive control method. The participants repeatedly performed a simple pick-and-place task, controlling a virtual robot arm using the three control types. **Results and Discussion** The everyday scenarios that most correspond to the needs of the target group are: "eating and drinking", "open and close doors", "supermarket shelf/pick up", and "microwave". Simulation of these scenarios enables the user to control the robot akin to a normal hand, allowing more direct motions which are not influenced by the limitations of the input device and thus offer the possibility of quickly recording extensive data. Results show that the number of mode switches necessary to complete a simple pick-and-place task decreases significantly when using an adaptive control type.

References

Herlant, L. V., Holladay, R. M. & Srinivasa, S. S. (2016. Assistive teleoperation of robot arms via automatic time-optimal mode switching. In 2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI) (pp. 35-42). IEEE.

Beaudoin, M., Lettre, J., Routhier, F., Archambault, P. S., Lemay, M.& Gélinas, I. (2019). Long-term use of the JACO robotic arm: a case series. *Disability and Rehabilitation: Assistive Technology, 14*(3), 267-275

Brose, S. W., Weber, D. J., Salatin, B. A., Grindle, G. G., Wang, H., Vazquez, J. J. & Cooper, R. A. (2010): The Role of Assistive Robotics in the Lives of Persons with Disability. *American Journal of Physical Medicine & Rehabilitation 89* (6), 509-521

Riel, M. (2020). Understanding action research. CCARweb, accessed 17.03.2022, https://www.ccarweb.org/what-is-action-research

Pascher, M., Baumeister, A., Schneegass, S., Klein, B. & Gerken, J. (2021). Recommendations for the Development of a Robotic Drinking and Eating Aid - An Ethnographic Study. *Human-Computer Interaction – INTERACT 2021*, Springer, Cham

Kronhardt, K., Rübner, S., Pascher, M., Goldau, F. F., Frese, U. & Gerken, J. (2022). Adapt or Perish? Exploring the Effectiveness of Adaptive DoF Control Interaction Methods for Assistive Robot Arms. *Technologies 10*(1)

Keywords: adaptive AI-based control, degrees of freedom, Human-Robot Collaboration, assistive robotic arm, quality of live
Address: Science center FUTURE AGING, Frankfurt University of Applied Sciences, Frankfurt, Germany
Email: annalies.baumeister@fb4.fra-uas.de

Acknowledgement Supported by the German Federal Ministry of Education and Research

Figure 1 Scenario: User controls a robotic arm by head movements to eat and drink; Source: Own representation.

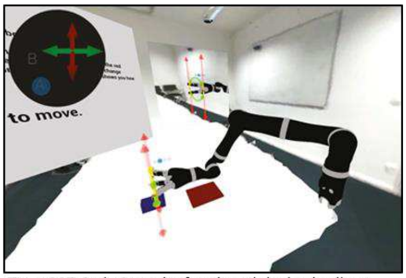


Figure 2 VR Study: Screenshot from the study in virtual reality to evaluate the DoF mapping; Source: Own research.

DORMADL - Dataset of human-operated Robot Arm Motion in Activities of Daily Living

Felix Goldau[1], Yashaswini Shivashankar[1], Annalies Baumeister[2], Lennart Drescher[1], Patrizia Tolle[2], Udo Frese[1]

***Abstract*— This work presents a dataset of human-operated robot motion to be used within the context of assistive robotics and assorted fields, such as learning from demonstrations, machine-learning based robot control, and activity recognition. The data consists of individual sequences of intentional robot motion performing a task in an environment of daily living. There are 2 973 sequences generated in a high-resolution simulation and 986 sequences performed in reality, totaling to 1.16 M datapoints. The data includes labels for the robot's pose, motion and activity. This paper also provides data augmentation methods and a detailed dataset analysis as well as simple models trained on the dataset as a baseline for future research.**

The dataset can be downloaded free-of-charge at `https://www.kaggle.com/f371xx/dormadl`.

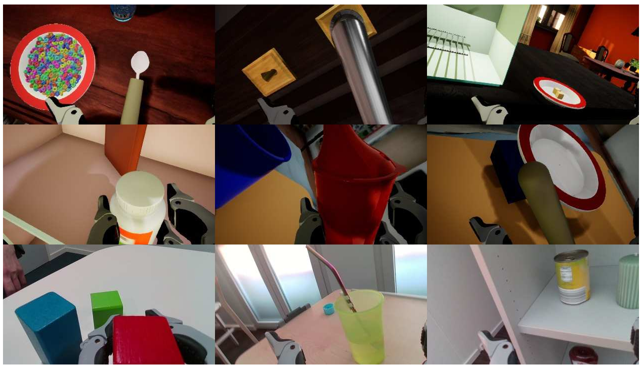


Fig. 1: Example images from the dataset

I. INTRODUCTION

The field of assistive robotics attempts to improve the lives of people who struggle with activities of daily living (ADLs), by using robotic assistance. The intended users often live with physical impairments which restrict their interaction with the environment. Apparently simple tasks, like drinking from a cup, moving a small object from one place to another, or even scratching oneself can become impossible or cumbersome to perform.

Research in this field shows a variety of individual solutions to a lot of tasks and is designed for diverse user groups. The common ground therein is the distinction from the typical application environment of robotics (i.e. industry) and focus on interaction with other people, their homes or private lives. These ADLs refer to the "basic tasks of everyday life, such as eating, bathing, toileting, and transferring" [1] and are well-represented in the literature. The applications range from custom eating utensils for users with spastics [2], simple fetch applications controlled by pointing with a laser pointer [3] or on a touch screen [4] for people with motion-impairments, up to partially autonomous systems to assist people with paraplegia with drinking using their remaining head motion [5] or brain-computer interfaces [6].

The market already provides wheelchair-mounted robotic arms (WMRAs) to be controlled directly by the person sitting in the chair. This creates a great opportunity since it allows for a mobile setup where users can interact with the environment, while also creating challenges, as most places are not designed for a robot arm. In addition to this, the interfaces used to control the robots limit the motion actually possible, as they generally offer less degrees of freedom (DoFs) than the robot is able to perform.

As this is a problem induced by the interface applied, different input modalities have been analyzed, often targeted to specific user groups. These concepts include both physical joystick-alternatives or additional sensors [7], as well as computer-aided control methods such as autonomously switching control modes [8]. More ambitious concepts apply shared control mechanisms, where the user-controlled action, defined by a pre-existing mode, is extended by fusing the result with autonomous solutions [9], possibly also combining this with custom input devices [10].

For modern data-driven approaches, it is necessary to have data representing the desired robot motion to perform a given task, be it for learning or evaluations. This data might also be interesting to the field of activity or intend recognition in order to detect patterns in the users' behavior. However, acquiring this data poses a chicken and egg problem, as the desired motion often cannot be controlled with standard interfaces: Opening a door, for example, requires pulling or pushing the door in an arc, whilst rotating the wrist to keep the alignment with the handle. Given a standard joystick distinguishing between translational and rotational modes, such a motion is simply impossible to perform.

A. *Contribution*

This work provides a novel dataset of user-controlled robot arm motion in activities of daily living. The main purpose of

*This work was supported by the German Federal Ministry of Education and Research BMBF (Bundesministerium für Bildung und Forschung)-funded projects *DOF-Adaptiv* and *AdaMeKoR* (FKZ 16SV8563, 16SV8534)

[1] The authors are with the German Research Center for Artificial Intelligence (DFKI)

[2] The authors are with the Frankfurt University of Applied Sciences
Corresponding Author: Felix Goldau, `felix.goldau@dfki.de`

the dataset is to learn the user's intended motion given the current situation, i.e. for a situation-adaptive user interface of an assistive robot.

In order to avoid the previously mentioned chicken and egg problem, part of the data is generated with a virtual robot arm controlled by 38 able-bodied participants in a simulation. As the robot arm precisely follows the human hand, no interface-induced motion restrictions apply. The virtual data is padded with a smaller sub-dataset created in our living-lab [11] with a real robot, that is controlled by 4 trained researchers using a 3D-mouse [12].

In short, we contribute a new dataset for assistive robotics of 3 959 recordings (1.16 M datapoints), which

- shows realistic and purposeful robot motion in eight simulated and one real scenario,
- provides aligned color and depth images for each datapoint (compare Fig. 1),
- provides poses, velocities, action-labels and the gripper status for machine-learning applications such as activity recognition, AI-based robot control or visual-servoing,
- is preprocessed and ready to use with a provided tensorflow-based [13] dataset loader,
- treats handedness by data-augmentation, and
- is available free-of-charge at `https://www.kaggle.com/f371xx/dormadl`.

II. STATE OF THE ART

To our knowledge, no dataset exists that includes activity-recognition data of an assistive robot arm or provides sequential pose data of such an arm during the execution of tasks in ADL. Both versions would hold the potential to support the development of shared control algorithms that focus on user intent.

In the field of activity recognition, various ADL-describing sequential datasets are in use: The Human Activity Recognition database [14] consists of recordings of activities such as walking or standing combined with inertial measurement unit (IMU) data of smartphones. The Dataset for ADL Recognition [15] relates wrist-worn accelerometer data to activities such as brushing one's teeth or eating soup. Other datasets have multiple IMUs, be it body worn [16] or partially attached to the environment [17] or use vision [18]. However, all of these are recorded by able-bodied participants and describe relatively broad tasks, where most activities are ADLs themselves.

In a rehabilitation or healthcare setting in particular fall or anomaly detection are interesting. The latter was examined by [19], who published a dataset generated in a simulated smart home environment for that express purpose. [20] shows that these virtual environments, though by far not perfect, are sufficiently realistic enough for neurorehabilitation.

From a more robotic perspective, the community prepared multiple datasets to be used in (assistive or service) robotics: The YCB [21] and YCB-Video [22] datasets link images of objects to their respective 6D poses, with YCB having readily available objects to be used as a benchmark. Knowing the poses of seen objects can be very beneficial for automatic grasping. The Cornell Grasping Dataset [23] expands on this idea by providing grasp rectangles and point clouds for objects in images. Even more specific, the Columbia Grasp Database [24] combines 3D object models with grasp poses for multiple variations of grippers.

In contrast to the datasets of activity recognition, the robotic image datasets mostly provide single-shot information and not video. The exception to this is the YCB-Video dataset; however, in this, the camera simply pans around the object without following any specific purpose (such as grasping the object). This leads to mostly (semi-) autonomous implementations in research of assistive robotics (e.g. visual servoing [25]), even though users prefer manual control [26].

To fill this gap, this work presents a dataset of detailed robotic arm motion in different ADLs. The dataset provides sequential image data of purposeful interactions during the activities and links these to pose and motion information, as well as human-readable action labels. This allows for applications to react to and analyze realistic situations.

III. RECORDING THE DATASET

The presented dataset is aimed to consist of *purposeful* motion for the current task at hand. We define the robot's motion to be such, if the task completion was successful and the robot hand behaved similarly to an able-bodied person using their own hand, given the robot's workspace restrictions.

In order to achieve the desired quality of motions, the data-generating users were instructed to perform only intentional, deliberate and clear motions during data generation. Each scenario was repeated multiple times per user, with recordings being stopped in between runs to prepare the next setup and a few initial runs to get acquainted with the environment. Thus, in this dataset, a *recording* will refer to a single attempt of a user to perform a task once.

During each recording, object poses, including the individual joints of the robot arm [27] were stored regularly to be later used to calculate camera poses and velocities. In addition to the poses, data from an RGBD camera [28] was gathered. This includes aligned color (RGB) and depth data, as well as unaligned data from infrared cameras used to calculate the depth data. In the simulation, this also included segmentation images.

A. Scenario Selection

In order to create a dataset with situations that are both realistic and relevant for the final users, we followed a participatory approach that included both primary users (i.e. people with disabilities who will actually control the arm) as well as secondary users (e.g. therapists or caregiving relatives who are indirectly affected). To develop these scenarios, we followed the action research model by Margaret Riel [29], [30] and the cycle-based participatory development process *Progressive Problem Solving with Action Research* [31], [32].

In two workshops, five primary users (one of them using a WMRA), two caregiving relatives, two physiotherapists and several researchers collectively discussed different scenarios in an open brainstorming session. The results were further evaluated with the participants (and one additional primary user) in individual one-to-one interviews to allow for more in-depth feedback. All workshops and interviews were transcribed and qualitatively analyzed following Tuckett's thematic analysis [33]. Protocols and interim results were distributed to everyone involved for heightened transparency and short feedback cycles.

During the workshops, the primary users stated that the most meaningful activities for them were eating and drinking prepared food and beverages. One primary user stated that "actually, eating is one of the major things (...) or rather being independent (of another person) during it".

Another activity that was important to all participants was to open and close doors independently. One primary user wished that they "could open the door, drive through it to eat and drink, in a way that (they) could then simply grab it (themselves)" to which a caregiver responded: "If (they) could really eat with it alone, that would be nice; That would be independence".

Though initially dismissed during the workshops because of constantly available human assistance, the scenarios of pick-up tasks, shopping, or microwave-usage, were later reintroduced in the interviews. Four primary users and two caregivers changed their minds and saw these scenarios as chances to gain more independence and relieve the caregivers. One primary user stated, "I now see it as rather interesting in my case (..). I'm home alone and my cell phone falls on the floor or I need something (..) important that is on the floor".

Further scenarios discussed in the workshops were activities towards one's personal hygiene, e.g. brushing teeth or using a sponge, which users regarded as unsafe or unrealistic.

Finally, four scenarios were decided upon. They are shown in Fig. 2 in reading order, starting at the top left. They are:

- **Eating and Drinking**: A prepared meal (e.g. a bowl of cereal or a set of small pre-cut pieces of food) is on the table. The robot arm grasps a fork or spoon, takes food and brings it to the mouth. For drinking, an open bottle or cup is grasped and brought to the mouth. Optionally, the liquid is poured from one container to the other.
- **Opening and Closing Doors**: The wheelchair is positioned close to the door. The door handle is pressed down with the gripper and the door is opened by pushing or pulling with the WMRA. The wheelchair is driven through the door and the door is closed using the robotic arm.
- **Microwave**: A microwave is placed on a table accessible to the robot arm. The gripper either pulls on the door or presses the button that opens the microwave. A plate with a prepared meal is grasped and placed into the microwave. The arm closes the microwave, activates it and retrieves the plate afterwards.
- **Supermarket Shelf / Pick Up from Floor**: This scenario is inspired by the setting of shopping for groceries. It includes a shelf with various objects, such as pasta packages and cans, on the lower levels and at least one additional object on the floor. The robot arm grasps the objects and places them on a table or in a basket.

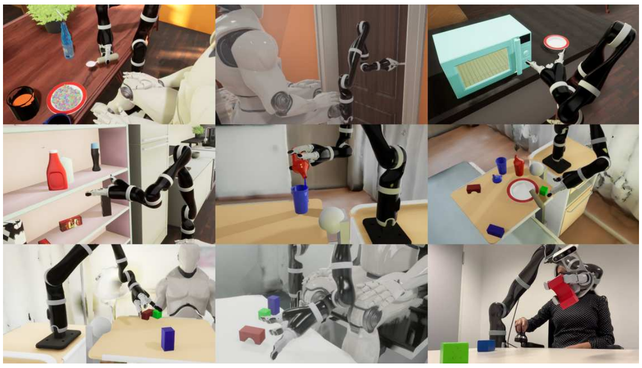


Fig. 2: Overview of the scenarios in the dataset. See media attachment for individual videos

All scenarios were implemented in simulation, with eating and drinking being combined into a single *meal* scenario. In these, the virtual robot arm is mounted to a stationary wheelchair. As the implicit motion of the wheelchair is not part of the dataset, the door scenario does not include the wheelchair moving, but instead only different positions from which to open and close the door.

Another project inspired a scenario where the user is sitting in a bed and the virtual robot arm is mounted to a sideboard with a table attached to one side. In this environment, we added the two scenarios **Fill Cup** and **Cleanup Table**, both based on a photogrammetry-scan of an existing room in our living lab (see [11]).

The *Fill Cup* scenario has two cups on the table with the robot arm grasping one of them to pour water into the other cup, whereas the *Cleanup Table* has various items (e.g. cups, plates, and cutlery) on the table and the robot rearranges them in an orderly fashion. These scenarios were chosen as they support mundane but complex tasks (e.g. pouring water, involving simultaneous rotation and translation), as well as cluttered tasks with various grasps and non-obvious sequential orders (e.g. rearranging objects), thus increasing the difficulty of the dataset.

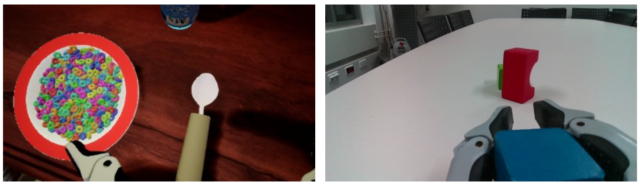
Finally, we added simple **Block** scenarios for both the bed and wheelchair settings. Here, the robot is used to re-position two blocks to a third block. These simple scenarios work as a baseline but also provide various actions of grasping and reorientation.

B. Recordings in Simulation

In order to create a sufficiently large dataset, part of the data was generated in a virtual reality (VR) simulation environment [34] based on a framework for shared control applications of assistive robots [35]. This includes a virtual version of the same robot arm used in reality. Simulation also allowed to customize the camera's pose or model after the actual recording sessions and automatically assured user anonymity, as only a virtual avatar is rendered.

To record data in the simulation, users were equipped with a VR Headset [36] and motion controllers. Here, they would perceive themselves, depending on the scenario, as either sitting in a wheelchair with the robot arm attached to its side, or sitting in a bed with the robot mounted to a sideboard. The end effector of the simulated robot arm is connected to the hand-held motion controller, thus enabling the users to basically complete the task using their own hand, only having to adjust for the limiting gripper functionality of the robot. This was hoped to generate human-like but still functional robot motion. The scenarios were developed such that the workspace limitations of the robot would not impede the user.

As this method of control required no initial training of the users, we gathered a variety of people to record data in order to allow different approaches to tasks and variations of motion in the data. An example image created in the simulation can be seen in Fig. 3a.



(a) (b)

Fig. 3: Example datapoints in simulation (a) and reality (b)

C. Recordings in Reality

For the recordings in reality, an assistive robot [27] was installed to a wheelchair and an RGBD camera [28] was mounted to the last joint of the robot arm. Unlike the setup in the simulation, the real robot cannot simply be moved by following the user's hand, as this would be visible in the image data. Instead, a 3D mouse [12] capable of controlling six DoFs (seven by adding two binary buttons) was used.

Controlling the real robot arm in sufficiently desired motions required training with the 3D mouse and was therefore limited to a selected group. This limited the number of recordings and variety generated in reality. As the setup and implementation of real scenarios takes a lot of time, these were also limited. The real scenarios include versions of Block, Fill Cup (including Drinking with a straw) and Supermarket Shelf / Pick Up from Floor. An example image generated in reality can be seen in Fig. 3b.

D. Data Labeling

As the recording system of the framework automatically stores image-pose pairs of the robot in every frame, no manual labeling of motions is necessary. On a higher level, however, we were able to add activity labels by manually assigning these to time ranges within each recording. As typical activities in the field of activity recognition are relatively broad and rather fit our definition of scenarios, we assign *Actions* instead. These cover shorter ranges of time and are more detailed. Mostly, these consist of a verb defined in reference to an object (e.g. approaching the cup).

Table I lists all components of the actions. Most action-verb combinations exist with only a few exceptions (e.g. the door and handle are an exclusive pair as they are conceptually connected). The table also lists verbs that only occur with a single object as singular verbs; as well as stand-alone actions without an object.

TABLE I: Overview of action components

verbs	Approach, Grasp, Let go, Push, Retrieve
objects	block, book, bottle, candle, cap, cup, door / handle, food, fork, microwave door, milk carton, plate, spoon, tea
singular verbs	Align *[cup]*, Close *[door]*, Fill *[cup]*, Press *[handle]*, Pull *[door]*
stand-alone actions	Discard, Drinking, Eating, Idle

As most actions are self-explanatory, we will describe only those with some ambiguity: *Retrieve* moves a held object to another position. For the shelf scenario, a suffix indicates the retrieval to the wheelchair table. *Align cup* positions a cup above another for pouring, which itself is described as *Filling cup*. *Eating* and *Drinking* both move a held item to the mouth, optionally tilting it, and partially retract afterwards.

Discard is a special label referring to sequences with recording issues. If recordings had Discard-labels at the beginning or end, they were trimmed accordingly or not included in the dataset altogether. *Idle* refers to the rest or pull-back motions of the user.

In addition, every action was given a binary success token, allowing for labeling of failed attempts. Unlike discarded-actions, failures do not refer to software issues, but indicate that the user was unable to complete their intended action (e.g. dropping an object).

E. Preprocessing / Dataset Cleaning

The recorded raw data was processed to cleanup the dataset and prepare it for easier use. For this, the initially measured pose data was smoothed and differentiated over time to generate motion information.

We define $T_{a\leftarrow b} := (\vec{p}, \vec{q}, g)$ as the transformation of a robot's coordinate frame b in reference to a and consisting of a 3-DoFs position $\vec{p}$, 3-DoFs quaternion-orientation $\vec{q}$, and 1-DoF gripper opening status g. In addition, we define $\vec{v}^{\,b} := (\vec{d}, \vec{r}, f)$ to be the relative velocity of the frame b consisting of a translational, rotational, and gripper-velocity respectively.

Let $\hat{T}_{\text{base}\leftarrow\text{EE}}$ be the raw data measured during recording. Due to the rotational component's dependencies, this data has to be treated as a manifold [37]. We utilized a version of the smoother on boxplus-manifolds proposed by [38] to smooth the raw pose data over time, remove outliers and handle data inconsistencies, thus creating a cleaner $T_{\text{base}\leftarrow\text{EE}}$. To further contextualize the pose with the camera data, the pose is transformed to $T_{\text{base}\leftarrow\text{cam}}$ with the camera frame being at the center of the color lens.

The dataset also provides the robot's velocity $\vec{v}^{\,\text{cam}}(t)$, which is approximated as the relative camera motion per timestep Δt_k, i.e.

$$\vec{v}^{\,\text{cam}}(t_k) = \frac{\Delta}{\Delta t_k} T_{\text{base}\leftarrow\text{cam}}(t_k) = \frac{T_{\text{cam}(t_{k-1})\leftarrow\text{cam}(t_{k+1})}}{t_{k+1} - t_{k-1}}. \quad (1)$$

For this, we define the division of a transformation T by a duration s to create the velocity $\vec{v}$, mostly following vector multiplication, except for orientation which needs to be converted to a rotation vector first, i.e

$$\vec{v} = T/s = (\vec{p}/s, \ln(\vec{q})/s, g/s) = (\vec{d}, \vec{r}, f). \qquad (2)$$

The preparation also includes datatype conversions between simulation and reality, outlier detection and removal, and handling of lost frames. In order to retain time-series information, invalid datapoints not at the beginning or end of recordings were kept in the data and marked as such.

1) Handedness: One special case of our dataset is handedness. Our robot arm has a non-symmetric 3-fingered gripper that is intended to be used on the right side of the wheelchair. However, a mirrored version of the same arm exists, that is to be mounted on the left side of a wheelchair. In practice, the chosen side generally aligns with the handedness of the user. As both the arm and users were right-handed during data generation, the dataset conforms to this practice in a way. The position of the robot generally affects from which side objects are approached, so data-points from both positions are needed if any machine learning shall work with both.

We propose to computationally augment the dataset with left-handed datapoints by mirroring the whole scene along the central plane of the robot base, which is parallel to the user's central plane. This means flipping the image and velocities relative to the camera around the YZ-plane (true also for an obliquely looking camera) and flipping poses along the central plane of the robot ($p_x = 0$). The specific formulas for the left-handed $T' := (\vec{p}\,', \vec{q}\,', g')$ and $\vec{v}\,' := (\vec{d}\,', \vec{r}\,', f')$ are simple but not trivial to derive:

$$\vec{p}\,' := \begin{pmatrix} -p_x \\ p_y \\ p_z \end{pmatrix}, \quad \vec{q}\,' := \begin{pmatrix} -q_x \\ q_y \\ q_z \\ -q_w \end{pmatrix}, \quad g' := g, \qquad (3)$$

$$\vec{d}\,' := \begin{pmatrix} -d_x \\ d_y \\ d_z \end{pmatrix}, \quad \vec{r}\,' := \begin{pmatrix} r_x \\ -r_y \\ -r_z \end{pmatrix}, \quad f' := f. \qquad (4)$$

IV. DATASET STRUCTURE

The dataset is split into subsets for training and testing and is structured in two index files with the labels. The image data is stored in a separate directory structure for each recording and referenced by the index files. The data split was performed on a per-user level, such that recordings of a single user are either in the test or training set in order to maintain more independence.

The recordings are processed and sampled at 10 Hz to generate clean time series data. All datapoints of the time series are listed in the csv files, with the recording number, a user number, the scenario, a timestamp within the recording, $T_{\text{base}\leftarrow\text{cam}}$, $\vec{v}^{\,\text{cam}}(t_k)$, the assigned action class and failure tag, the validity-flag, and file paths for the image data.

This multi-dimensional data structure allows for different settings and usages. The features can be a subset of the camera data, consisting of an RGB color image and an aligned depth image. In the simulation, a segmentation image is also generated. In addition, but uploaded separately[1] in order to reduce storage size, the infrared camera data used to generate the depth images can be loaded. The labels can be either the pose of the gripper or camera, their velocities, or the assigned actions. A python script is provided to assist loading default dataset configurations.

V. DATASET STATISTICS

This section gives a brief statistical overview. All numbers refer to the complete dataset, with the respective numbers for the subsets in brackets as *(training, test)*.

The dataset consists of 1.16 M (871 k, 290 k) datapoints from 3 959 (3 165, 794) recordings with an average runtime of 29.3 seconds, thus totaling to a length of approximately 29.3 (24.2, 8.1) hours. Thereof, 502 k datapoints from 986 sequences over 13.9 hours were created in reality. An example for the poses of a single recording are shown in Fig. 4.

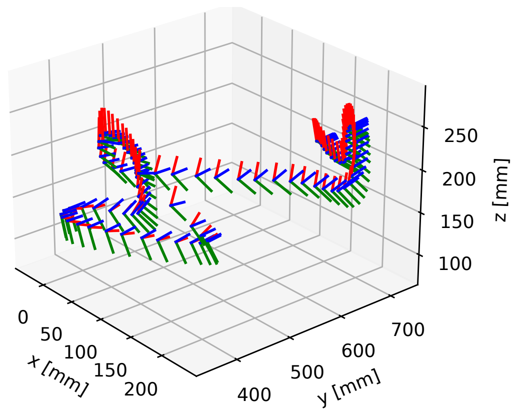


Fig. 4: Camera poses during a single example recording with (red, green, blue) being (right, up, forward)

The mean and a scaled representation of the sampled covariance of camera motion is shown in Fig. 5. The scaling is performed to be able to compare the translational, rotational, and gripper velocities with different units. For this, a rotational velocity of 180 °/s corresponds to a translational velocity of 1000 mm/s, and, respectively, opening the gripper once per second corresponds to 300 mm/s. To enhance visibility, the covariance's components are color-coded based on their absolute ratios to visualize relationships.

	unit	mean
Translation X	[mm/s]	-0.564
Translation Y	[mm/s]	5.629
Translation Z	[mm/s]	4.686
Rotation X	[°/s]	-0.177
Rotation Y	[°/s]	0.131
Rotation Z	[°/s]	0.331
Gripper G	[1/s]	0.010

covariance

	X	Y	Z	R_X	R_Y	R_Z	G
X	9319	-392	2091	1534	1951	-5619	-147
Y	-392	7026	528	-988	-703	412	66
Z	2091	528	17167	5367	554	-280	323
R_X	1534	-988	5367	9288	-57	-1105	83
R_Y	1951	-703	554	-57	13232	1217	120
R_Z	-5619	412	-280	-1105	1217	20442	78
G	-147	66	323	83	120	78	4788

Fig. 5: Mean and scaled covariance of camera motion

Fig. 6 shows the distribution of the labeled actions over the complete dataset. It can be clearly seen that objects are

[1]refer https://www.kaggle.com/f371xx/dormadl

represented more often which either have dedicated scenarios (block, cup, microwave door) or are reoccurring (plate, cup, bottle). The verbs show a focus on approaching and retrieving, as well as opening and closing of the gripper. This is expected as it is at the core of robot manipulation.

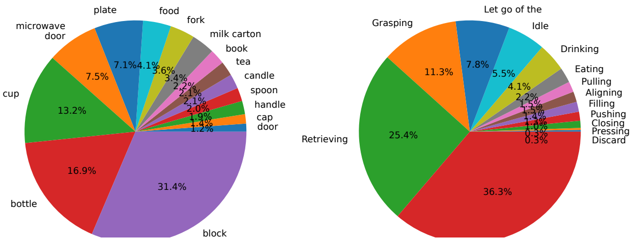


Fig. 6: Distribution of action labels in the dataset

VI. Baseline Model Statistics

As a baseline for comparison, we trained two simple networks on the dataset: One for activity (here action) recognition and one to predict the camera's motion direction for direct robot control, robot servoing, or approaches of shared control. Both can be thought of as predicting the user's *intent* from the current situation.

A. Baseline Action Recognition Model

MobileNetV2 [39] is used as a base model to predict the robot's action. The model backbone is extended with a 2D convolution layer with ReLU activation function, global average 2D-pooling, as well as a dropout and a dense layer with a softmax activation function. The model is trained to classify the images into one of the actions performed by the robot arm. A categorical cross-entropy loss function was used during the training. Table II shows the baseline model's results for both the test and training datasets for the loss and an accuracy metric.

TABLE II: Baseline results of action recognition model

	train	test
categorical cross-entropy	0.8554	0.9223
categorical accuracy	0.7066	0.6796

B. Baseline Motion Direction Prediction Model

For the motion direction prediction, we implemented a baseline model able to output a multi-dimensional Gaussian distribution of the motion, which allows for the use of statistical tools. Even if a single output is required, one can simply take the most likely value of the distribution.

The model is designed with a MobileNetV2 backbone [39] extended with a 2D-convolutional layer with a batch normalization, ReLU activation, as well as three fully-connected layers, the first two of which with a ReLU activation. For the final activation, a layer calculating a sample-based covariance (compare [40]) was used to generate the probability distribution. As this is intended as a simple baseline model, no further extensions, such as recurrences, were added. The only preprocessing was dimensional scaling to align translational, rotational, and gripper velocity (Section V). The model was trained using a Mahalanobis-loss [40].

To provide more intuitive values than the Mahalanobis-loss, we propose a new metric: Let $b = (\vec{b}_1, \ldots, \vec{b}_n)$ be the n-dimensional base spanned by the covariance's normalized principal components $\vec{b}_i$ (i.e. its eigenvectors), where $\vec{b}_1$ has the largest corresponding eigenvalue and $\vec{b}_n$ the smallest. We can now calculate the projection $\vec{p}$ of our labeled vector $\vec{v}$ onto a k-dimensional sub-base $s = (\vec{b}_1, \ldots, \vec{b}_k)$

$$\vec{p} = \sum_{i=1}^{k} \left\langle \vec{v} \cdot \vec{b}_i \right\rangle \cdot \vec{b}_i, \tag{5}$$

where $\langle \cdot \rangle$ is the scalar product of two vectors and $k < n$. Together with the projection $\vec{q}$ on the complementing sub-basis $b = (\vec{b}_{k+1}, \ldots, \vec{b}_n)$, $\vec{p}$ and $\vec{v}$ form a right triangle such that pythagorean theorem yields $|v|^2 = |p|^2 + |q|^2$. This allows us to define $|p|^2 / |v|^2$ as a metric: the percentage of the squared length of true motion that is represented by the k first principal components of the probability distribution. Note that the first k principal components maximize this metric among all choices of $\vec{b}_1, \ldots, \vec{b}_n$. This can also be seen as how well one could follow the true motion, while only moving along the axes $\vec{b}_1, \ldots, \vec{b}_k$.

Within a direction-prediction task, the percentage-of-motion-metric is to a Mahalanobis- or log-likelihood-loss what the accuracy-metric is to a cross-entropy-loss in a classification task: Both metrics reinterpret the evaluation to a more human-readable form by simplifying (reducing) the data. Here, accuracy assesses only the binary equivalence of the label and the most likely class while ignoring the actual probability value. Similarly, the percentage-of-motion-metric assesses only the projection of the true direction in the the sub-basis s while ignoring the remaining $n - k$ dimensions.

Table III shows the baseline model's results for the Mahalanobis-loss distance, negative log-likelihood and the percentage of motion metrics for $k = 1$ and $k = 2$, as well as the root-mean-squared-error between the label and the first principal component.

TABLE III: Baseline results of motion prediction model

	train	test
Mahalanobis-distance	4.6991	5.5219
negative log-likelihood	-11.1994	-10.4222
percentage of motion ($k = 1$)	0.5844	0.4975
percentage of motion ($k = 2$)	0.7538	0.6741
root mean squared error	0.5229	0.5268

VII. Limitations

The quality of the dataset is limited by two main factors: Issues in the methodology and the simulation reality gap. Even though the simulation is designed to be very close to reality, there are some aspects of robot interactions in reality that were not implemented for various reasons.

A. Methodology

As with typical learning-from-demonstration applications, the dataset can only be assumed to provide accurate information for situations similar to those recorded.

Apart from that, our method of robot control is based on the assumption that users control the robot arm similarly to a regular arm. This assumption might not be correct and therefore make the data partially invalid. The data generated in reality should improve this.

Only a small portion of the data is recorded in reality and on a small subset of scenarios. The dataset could be greatly improved by adding more real recordings.

Due to our participatory approach, we continued end-user interviews during data recording. This resulted in a request for the *Meal* scenario to be adjusted to eat cereals from a bowl instead of using a fork to eat fruit. We adjusted this and recorded additional data in a separate instance. This results in an imbalance in the number of scenarios and users.

B. Simulation-Reality-Gap: Camera Data

The simulated camera follows the real camera in terms of camera parameters and effects (compare [35]). In order to verify the simulated camera quality, we generated data with both the real and simulated cameras in environments as similar as possible (Fig. 7).

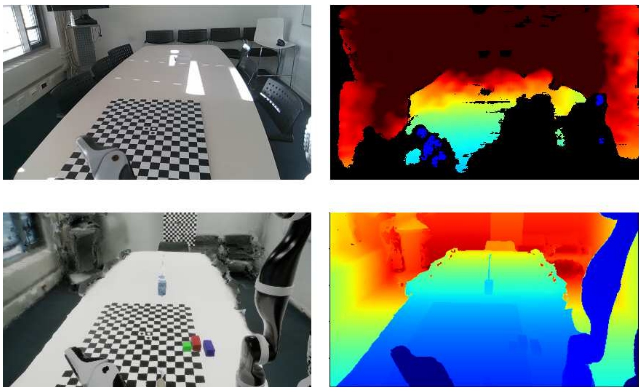


Fig. 7: Comparison of real (top) and simulated (bottom) camera data

It can be seen that the color data (left) is close to identical, whereas the depth data (right) shows vast differences, as the simulated camera is perfect while the real camera suffers from multiple image effects. These mostly stem from the stereo depth algorithm used by the real camera, which uses grayscale image data generated from two additional built-in cameras. We provide simulated versions of these images in a separate repository for users interested in calculating more realistic depth images.

There is no additional noise or image effects in the simulated data. If required, users can simply add these manually.

C. Simulation-Reality-Gap: Robot Arm

As mentioned in [35], the simulated robot arm is built from original robot data, including mechanical dimensions and meshes. However, in order to avoid movement limitations during data generation, the simulated robot is not controlled using an inverse kinematic and therefore not limited to the motion limits of the real robot. Instead, the simulated joints are spring-based and can, to some extent, move beyond the intended angle limits.

The simulated robot is controlled by moving the motion controller and having the end effector follow it. This results in each robot link in row being pulled along, such that the robot behaves similar to a rope in zero gravity. While this is a major difference to reality, this should not change the relative motion per datapoint which depends only on the camera pose.

D. Simulation-Reality-Gap: Grasping

The simulated grasping is not physics-based but instead a software solution, such that an object is considered attached to the robot hand if squeezed by opposite fingers. This can sometimes cause unrealistic behavior.

In our setup, however, objects are easily graspable with the robot. Poorly-graspable or very heavy objects would have more issues that are therefore avoided. Another factor weakening the effect is due to the use of human operators instead of a script. It can be assumed that humans instinctively prefer realistic grasps.

However, the two doors (room door and microwave) could not be designed as such, as they are not lifted but instead opened or closed. Custom interactions were designed for these, where the robotic hand would retain a relative pose to the handle whilst grasped.

VIII. CONCLUSIONS

In summary, we present a novel dataset applicable to multiple fields associated with assistive robotics. The dataset is easily accessible free of cost and can be used for both robot control as well as activity recognition tasks.

We provided detailed descriptions of the specific method applied to generate the dataset, using both a simulation environment and an associated setup in reality. The capabilities and limitations of the dataset were discussed in detail and metrics were presented as baselines for machine learning research.

Future work should focus on utilizing the provided data to gain insights into user behavior and optimize shared control implementations based on this knowledge. This can, for example, be achieved by analyzing the recorded motions and manually implementing specific interactions, possibly dependent on the current action or activity. Alternatively, data-driven machine-learning models could be trained to predict the user's intended motion in order to offer the most likely direction of control as part of a user interface.

REFERENCES

[1] J. M. Wiener, R. J. Hanley, R. Clark, and J. F. Van Nostrand, "Measuring the activities of daily living: Comparisons across national surveys," *Journal of gerontology*, vol. 45, no. 6, pp. S229–S237, 1990.

[2] P. Turgeon, M. Dubé, T. Laliberté, P. S. Archambault, V. H. Flamand, F. Routhier, and A. Campeau-Lecours, "Mechanical design of a new device to assist eating in people with movement disorders," *Assistive Technology*, vol. 0, no. 0, pp. 1–8, 2020, pMID: 32105199. [Online]. Available: https://doi.org/10.1080/10400435.2020.1734111

[3] M. Zhong, Y. Zhang, X. Yang, Y. Yao, J. Guo, Y. Wang, and Y. Liu, "Assistive grasping based on laser-point detection with application to wheelchair-mounted robotic arms," *Sensors*, vol. 19, no. 2, p. 303, 2019.

[4] K. M. Tsui, D.-J. Kim, A. Behal, D. Kontak, and H. A. Yanco, ""i want that": Human-in-the-loop control of a wheelchair-mounted robotic arm," *Applied Bionics and Biomechanics*, vol. 8, no. 1, pp. 127–147, 2011.

[5] F. F. Goldau, T. K. Shastha, M. Kyrarini, and A. Gräser, "Autonomous multi-sensory robotic assistant for a drinking task," in *2019 IEEE 16th International Conference on Rehabilitation Robotics (ICORR)*. IEEE, 2019, pp. 210–216.

[6] L. R. Hochberg, D. Bacher, B. Jarosiewicz, N. Y. Masse, J. D. Simeral, J. Vogel, S. Haddadin, J. Liu, S. S. Cash, P. Van Der Smagt *et al.*, "Reach and grasp by people with tetraplegia using a neurally controlled robotic arm," *Nature*, vol. 485, no. 7398, pp. 372–375, 2012.

[7] P. Try, S. Schöllmann, L. Wöhle, and M. Gebhard, "Visual sensor fusion based autonomous robotic system for assistive drinking," *Sensors*, vol. 21, no. 16, 2021. [Online]. Available: https://www.mdpi.com/1424-8220/21/16/5419

[8] S. Jain and B. Argall, "Robot learning to switch control modes for assistive teleoperation," in *RSS 2016 Workshop on Planning for Human-Robot Interaction: Shared Autonomy and Collaborative Robotics*, 2016.

[9] B. Driessen, T. T. Kate, F. Liefhebber, A. Versluis, and J. Van Woerden, "Collaborative control of the manus manipulator," *Universal Access in the Information Society*, vol. 4, no. 2, pp. 165–173, 2005.

[10] S. Jain, A. Farshchiansadegh, A. Broad, F. Abdollahi, F. Mussa-Ivaldi, and B. Argall, "Assistive robotic manipulation through shared autonomy and a body-machine interface," in *2015 IEEE international conference on rehabilitation robotics (ICORR)*. IEEE, 2015, pp. 526–531.

[11] S. Autexier, C. Lüth, and R. Drechsler, *Künstliche Intelligenz im Gesundheitswesen*. Springer Fachmedien Verlag Wiesbaden, march 2022, ch. Das Bremen Ambient Assisted Living Lab und darüber hinaus – Intelligente Umgebungen, smarte Services und Künstliche Intelligenz in der Medizin für den Menschen, pp. 835–850.

[12] 3Dconnexion. (2023) SpaceMouse Compact. [Online]. Available: https://3dconnexion.com/us/product/spacemouse-compact/

[13] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015, software available from tensorflow.org. [Online]. Available: https://www.tensorflow.org/

[14] D. Anguita, A. Ghio, L. Oneto, X. Parra, J. L. Reyes-Ortiz *et al.*, "A public domain dataset for human activity recognition using smartphones." in *Esann*, vol. 3, 2013, p. 3.

[15] B. Bruno, F. Mastrogiovanni, A. Sgorbissa, T. Vernazza, and R. Zaccaria, "Analysis of human behavior recognition algorithms based on acceleration data," in *2013 IEEE International Conference on Robotics and Automation*. IEEE, 2013, pp. 1602–1607.

[16] M. Ruzzon, A. Carfì, T. Ishikawa, F. Mastrogiovanni, and T. Murakami, "A multi-sensory dataset for the activities of daily living," *Data in Brief*, vol. 32, p. 106122, 2020.

[17] D. Roggen, A. Calatroni, M. Rossi, T. Holleczek, K. Förster, G. Tröster, P. Lukowicz, D. Bannach, G. Pirkl, A. Ferscha *et al.*, "Collecting complex activity datasets in highly rich networked sensor environments," in *2010 Seventh international conference on networked sensing systems (INSS)*. IEEE, 2010, pp. 233–240.

[18] P. Pareek and A. Thakkar, "A survey on video-based human action recognition: recent updates, datasets, challenges, and applications," *Artificial Intelligence Review*, vol. 54, pp. 2259–2322, 2021.

[19] T. Alshammari, N. Alshammari, M. Sedky, and C. Howard, "Simadl: Simulated activities of daily living dataset," *Data*, vol. 3, no. 2, 2018. [Online]. Available: https://www.mdpi.com/2306-5729/3/2/11

[20] S. M. Gerber, R. M. Müri, U. P. Mosimann, T. Nef, and P. Urwyler, "Virtual reality for activities of daily living training in neurorehabilitation: a usability and feasibility study in healthy participants," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2018, pp. 1–4.

[21] B. Calli, A. Singh, A. Walsman, S. Srinivasa, P. Abbeel, and A. M. Dollar, "The ycb object and model set: Towards common benchmarks for manipulation research," in *2015 international conference on advanced robotics (ICAR)*. IEEE, 2015, pp. 510–517.

[22] Y. Xiang, T. Schmidt, V. Narayanan, and D. Fox, "Posecnn: A convolutional neural network for 6d object pose estimation in cluttered scenes," 2017. [Online]. Available: https://arxiv.org/abs/1711.00199

[23] I. Lenz, H. Lee, and A. Saxena, "Deep learning for detecting robotic grasps," *Int. J. Rob. Res.*, vol. 34, no. 4–5, p. 705–724, apr 2015. [Online]. Available: https://doi.org/10.1177/0278364914549607

[24] C. Goldfeder, M. Ciocarlie, H. Dang, and P. K. Allen, "The columbia grasp database," in *2009 IEEE International Conference on Robotics and Automation*, 2009, pp. 1710–1716.

[25] D.-J. Kim, R. Lovelett, and A. Behal, "Eye-in-hand stereo visual servoing of an assistive robot arm in unstructured environments," in *2009 IEEE International Conference on Robotics and Automation*. IEEE, 2009, pp. 2326–2331.

[26] D. Kim, R. Hazlett-Knudsen, H. Culver-Godfrey, G. Rucks, T. Cunningham, D. Portee, J. Bricout, Z. Wang, and A. Behal, "How autonomy impacts performance and satisfaction: Results from a study with spinal cord injured subjects using an assistive robot," *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 42, no. 1, pp. 2–14, 2012.

[27] Kinova inc., "Jaco assistive robot - user guide," (last visited on 13.02.2023), 2021, "EN-UG-007-r05-202111". [Online]. Available: https://assistive.kinovarobotics.com/uploads/EN-UG-007-Jaco-user-guide-R05.pdf

[28] Intel Corporation. (2022, november) Intel realsense - product family d400 series. (last visited on 13.02.2023). [Online]. Available: https://www.intelrealsense.com/wp-content/uploads/2022/11/Intel-RealSense-D400-Series-Datasheet-November-2022.pdf

[29] M. Riel. (2020) Understanding action research. online. [Online]. Available: https://www.ccarweb.org/what-is-action-research

[30] A. Baumeister, M. Pascher, Y. Shivashankar, F. F. Goldau, U. Frese, J. Gerken, E. Gardo, B. Klein, and P. Tolle, "Ai for simplifying the use of an assistive robotic arm for people with severe body impairments," *Gerontechnology*, vol. 21, 10 2022. [Online]. Available: https://doi.org/10.4017/gt.2022.21.s.578.5.sp7

[31] M. T. Wright, M. Block, H. Kilian, and K. Lemmen, "Förderung von qualitätsentwicklung durch partizipative gesundheitsforschung," *Prävention und Gesundheitsförderung*, vol. 3, no. 8, pp. 147–154, 2013.

[32] A. Baumeister, E. Gardo, P. Tolle, B. Klein, M. Pascher, J. Gerken, F. Goldau, Y. Shivashankar, and U. Frese, "The importance of participatory design for the development of assistive robotic arms: initial approaches and experiences in the research projects mobile and dof-adaptiv," in *Connected Living: international and interdisciplinary conference (2021), Frankfurt am Main*, 2021.

[33] A. G. Tuckett, "Applying thematic analysis theory to practice: A researcher's experience," *Contemporary nurse*, vol. 19, no. 1-2, pp. 75–87, 2005.

[34] Epic Games. (2023) Unreal engine. (last visited on 13.02.2023). Version 4.27.2. [Online]. Available: https://www.unrealengine.com

[35] M. Pascher, F. F. Goldau, K. Kronhardt, U. Frese, and J. Gerken, "AdaptiX – A Transitional XR Framework for Development and Evaluation of Shared Control Applications in Assistive Robotics," in review.

[36] HTC Corporation, "Vive pro 2 headset," 2021. [Online]. Available: https://www.vive.com/us/product/vive-pro2/overview/

[37] C. Hertzberg, R. Wagner, U. Frese, and L. Schröder, "Integrating generic sensor fusion algorithms with sound state representations through encapsulation of manifolds," *Information Fusion*, vol. 14, no. 1, pp. 57–77, 2013.

[38] T. L. Koller and U. Frese, "The interacting multiple model filter and smoother on boxplus-manifolds," *Sensors*, vol. 21, no. 12, p. 4164, 2021.

[39] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4510–4520.

[40] F. F. Goldau and U. Frese, "Learning to map degrees of freedom for assistive user control: Towards an adaptive dof-mapping control for assistive robots," in *The 14th PErvasive Technologies Related to Assistive Environments Conference*, 2021, pp. 132–139.

Learning to Map Degrees of Freedom for Assistive User Control

Towards an Adaptive DoF-Mapping Control for Assistive Robots

Felix Ferdinand Goldau
German Research Center for Artificial Intelligence (DFKI)
Bremen, Germany
felix.goldau@dfki.de

Udo Frese
University of Bremen
Bremen, Germany
ufrese@uni-bremen.de

ABSTRACT

This paper presents a novel approach to shared control for an assistive robot by adaptively mapping the degrees of freedom (DoFs) for the user to control with a low-dimensional input device. For this, a convolutional neural network interprets camera data of the current situation and outputs a probabilistic description of possible robot motion the user might command.

Applying a novel representation of control modes, the network's output is used to generate individual degrees of freedom of robot motion to be controlled by single DoF of the user's input device. These DoFs are not necessarily equal to the cardinal DoFs of the robot but are instead superimpositions of those, thus allowing motions like diagonal directions or orbiting around a point. This enables the user to perform robot motions previously impossible with such a low-dimensional input device.

The shared control is implemented for a proof-of-concept 2D simulation and evaluated with an initial user study by comparing it to a standard control approach. The results show a functional control which is both subjectively and objectively significantly faster, but subjectively more complex.

CCS CONCEPTS

• **Computer systems organization** → **Robotic control**; Neural networks; • **Human-centered computing** → *Interaction devices*; *Interaction techniques*; • **Social and professional topics** → People with disabilities; **Assistive technologies**.

KEYWORDS

Assistive Robotics, Convolutional Neural Network (CNN), Deep Learning (DL), Human Machine Interface (HMI), Human Robot Interface (HRI), Shared User Control

ACM Reference Format:
Felix Ferdinand Goldau and Udo Frese. 2021. Learning to Map Degrees of Freedom for Assistive User Control: Towards an Adaptive DoF-Mapping Control for Assistive Robots. In *The 14th PErvasive Technologies Related to Assistive Environments Conference (PETRA 2021), June 29-July 2, 2021, Corfu, Greece.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3453892.3453895

PETRA 2021, June 29-July 2, 2021, Corfu, Greece

ACM ISBN 978-1-4503-8792-7/21/06…$15.00
https://doi.org/10.1145/3453892.3453895

1 INTRODUCTION

The general understanding of autonomy and technical systems is something akin to using a computer program to independently control the actuators of a machine or robot to solve a given task. While this might be appropriate for the default industrial scenario, it stands in vast contrast to applications of assistive robots, such as the Kinova Jaco [17, 20], which aim to (re-)enable a person to perform activities of daily living themselves, instead of having them performed by another person or program. However, the manual control of such devices can be very exhausting and taxing for the user due to the complexity of the system and the user's impairments, thus generating a necessity for easier and more accessible methods of control [5].

Some previous work has been done with the aim to automate or ease specific activities of daily living [6, 8, 24]. However, a study investigating the performance and satisfaction of spinal cord injured users of a wheelchair-mounted robotic arm with regards to manual and autonomous control modes showed a higher satisfaction for manual mode users, even though the autonomous mode required less effort [16]. The resulting call for more flexible interfaces coincides with findings by [21], who show the users' requirement to personalise their interaction such that personal standards and social norms are met. A situation with robotic assistance should be as similar as possible to a respective situation without impairments. Therefore, one should be very careful when applying fully automated solutions to such assistive scenarios.

The alternative to a system being controlled by a computer is usually to have it directly or indirectly controlled by a human using a form of Human Computer Interface (HCI) with a keyboard, joystick or similar input device. However, very few devices have sufficient Degrees of Freedom (DoFs) to directly control a robot like the Jaco and those that fulfil this specification require a significant dexterity from the user. For most users of assistive robots, this poses an impossible challenge due to their sicknesses or disabilities. In order to use the remaining mobilities of a user, specific HCIs have been developed [13, 19, 23, 23] which, due to the specifications and limitations, mostly cannot compare to the default control interfaces when it comes to their output DoFs. For example, the Jaco requires at least seven DoFs (three for positioning, three for rotation, and one for grasping), whereas input devices such as Eye-Trackers [23], Chin- or Tongue-Mouses [10] only provide two. Even the robot's joystick only provides a maximum of three DoF to be controlled at once, with buttons allowing to switch between different control modes (cf. [1, 13, 18]). An extensive literature review regarding functionality and performance of assistive robots concluded in a call to "develop a two-way user interface between higher dexterity [robots] that could be operated by fewer [DoFs] from end-users", whilst keeping the users in control, as desired [3].

Various forms of shared user control exist, where the systems utilise a combination of input from the user and the output of a computer program. For example, [25] initially lets the user control only the translational DoFs of a robot arm, whilst automatically handling rotation. Close to a defined target, the system starts blending the user input with an automated grasp approach based on the user-defined position, until finally applying a fully automated grasp action. Based on a literature study on multiple systems using shared control, [2] identifies the detection of user intent as one of the largest problems within this area and calls out for more Machine Learning (ML) in shared control approaches. Following this call, [7] presents a shared control approach for an electric wheelchair passing small doorways, where the user can activate a blend of their commands with a pre-trained ML-generated control.

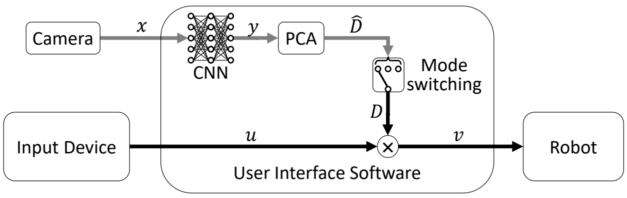


Figure 1: Control pipeline for a user-controlled assistive robot

This paper presents a proof of concept for a novel variation of shared control, where a Deep Learning (DL) based setup evaluates the current situation and adaptively proposes a set of high-dimensional DoFs of robot motion to be controlled by the user's low-DoF input device. Figure 1 shows the corresponding control pipeline: Usually, the user-generated input u is directly mapped to the robot-controlling input v (i.e. D is static), which enables the user to control a single cardinal DoF of the robot (i.e. x-axis, y-axis, z-axis, roll, pitch, yaw) with each DoF of their input device. In cases where the input device has fewer DoFs than the robot control, the user generally has the option to switch between pre-defined modes, thus changing the mapping from input device DoF to robot control DoF (i.e. exchange D). We break this static connection by using a Convolutional Neural Network (CNN) to describe the probabilistic distribution y of intended robot motion v given the camera data x (i.e. the current situation). A Principal Component Analysis (PCA) is applied to calculate a matrix D that adaptively maps the user-generated input to the robot motion, thus portraying modes of control.

The user stays in control; in particular a zero user command u always results in no motion. This eliminates much of the safety concerns of machine learning.

The presented approach enables the DL system not only to suggest the set of cardinal DoFs but also superimpositions of those, thus allowing motions previously impossible with a limited set of input DoFs, such as diagonal paths, orbiting around a point in space or approaching a goal at an angle (cf. Fig. 2). For this paper, the proof-of-concept scenario is limited to a simulated 2D environment with a robot defined by four cardinal DoFs (two positional, one rotational, and grasping). Figure 2 shows the robot with cardinal and adaptive DoFs, both represented by arrows.[1]

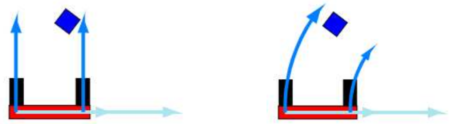


Figure 2: The simulated robot with two out of the four cardinal DoFs (left) and two adaptive DoFs (right)

The paper is organised as follows: After a review of previous research to handling the discrepancy of input to output DoFs in Section 2, Section 3 describes our approach in detail, with the simulation environment being described in Section 4. An initial user study is presented in Section 5, with Section 6 discussing the resulting implications and directions for future work.

This paper provides a proof of concept for adaptive DoF mapping in a 2D simulation environment.[2] Its contributions are

- the idea of a novel DL approach to shared control for an assistive robot arm,
- a general representation for DoF-based user control, optionally with modes,
- a 2D simulation environment for proof-of-concept of such methods, and
- an initial user study regarding the usability of such an approach to shared control.

2 RELATED WORK

The default method to controlling a high-DoF device using a low-DoF input device (e.g. controlling an assistive robot arm using a joystick) is mode switching. A single DoF of the input device controls a single cardinal DoF of the robot. Switching the selected mode changes this mapping, such that the same user input now controls a different cardinal DoF of the robot. To the best knowledge of the authors, no shared user control exists that allows the user to control a device along arbitrary online-defined DoFs. However, there are different approaches to mapping user input from a low-DoF input device to a high-DoF system, as well as ML setups that learn autonomous behaviours in a high dimensional environment.

For this paper we use *cardinal DoFs* to describe the set of DoFs defined by, and axis-aligned to, the Cartesian coordinate system of the robot, plus an additional DoF to handle closing the gripper. For a robot with at least six DoFs in 3D space, like the Kinova Jaco, this would be [X-Axis, Y-Axis, Z-Axis, Roll, Pitch, Yaw, Gripper].

Based on their method of user inclusion, it is possible to differentiate control approaches into two categories [11]: In one the user indicates targets and the autonomous system executes the action mostly without user interaction (cf. [26]). The other integrates the user as a direct source of movement control. If a user functions as a direct source of control input, they often have an HCI with low-DoF input device and different control modes. In experiments

[1] Video available at: http://www.informatik.uni-bremen.de/agebv2/downloads/videos/GoldauPetra21.m4v

[2] Resources available at: https://github.com/f371xx/adaptive_dof_mapping_2d

by [11] using an HCI with a standard button-based mode switching setup, more than one-sixth of the total execution time was spent changing the currently selected mode. Within a deterministic simulation environment and a predefined goal, they showed that an automatic mode switching approach already leads to an increase in user satisfaction.

Many manipulation actions require precise positioning. Therefore, when controlling a device towards a goal (e.g. grasping a cup), slight corrections in direction or orientation need to be made. Depending on the environment and perception of the user, this can be a difficult task. For the task of grasping a cup, this would be the precise positioning to not accidentally approach the cup off-center or tip it over with the fingers. Also, if applying a mode switching approach, these small adjustments generally require multiple mode switches, all with very small actual movements of the device within a single mode. To avoid this, research has shown remarkable success with control blending [5], which arbitrates the user's control input with computer generated control, thus allowing the computer to assist the user by avoiding obstacles or supporting with the final approach [4]. However a study has shown that the level of assistance should be customisable by the user to allow for perfect adjustment to the user's needs and abilities, as well as increase user satisfaction [14].

With more complex scenarios and non-deterministic users, multiple goal states can be possible in a given situation (e.g. multiple cups available from which the user can choose which to grasp). For these scenarios, [9] presents a different approach to assistive mode switching: The system isolates possible user intentions and chooses the control mode whose actions will maximise the arbitration of possible user goals in order to assist the ML System in identifying the underlying intention. Once a threshold certainty about the user's intent is surpassed, control blending is applied to assist the user. While this does show promising results, the user's control options are still limited to the cardinal DoFs.

Controlling more complex movements with a low-DoF interface has been realised by predefining sequences within a complex task and using autonomous planners to execute the task. Instead of directly controlling each cardinal DoF of the manipulator, the user utilises their low-DoF interface to define the velocity of the automation and switch between the automated trajectories [15].

A more general option of controlling a robotic device with an HCI is introduced by [22], who propose a neural network to map the sensory readings of an input device to the control signals for a robot. However, within their work they aim to learn an intuitive constant mapping per user and task, therefore restricting the mapping to be static and not adaptive to the situation.

3 MAPPING DEGREES OF FREEDOM

We want to not only do intelligent mode switching but instead loose the system's predefined definitions of DoFs and allow the user to control the robot along DoFs that are regularly redefined based on the current environment and situation.

3.1 Definitions

A DoF d is therefore not limited to the predefined set of cardinal DoFs but instead a vector $d \in \mathbb{R}^n, \|d\|_2 = 1$ in the cardinal coordinate space. This allows for DoFs that are not necessarily axis-aligned to the cardinal coordinate frame, such as moving diagonally or orbiting around a point. A 1-dimensional user input device (e.g. a 1D joystick) could therefore control a high-DoF robot along such an arbitrary n-dimensional DoF.

In the general case, given $u \in \mathbb{R}^m$ as the output of an m-dimensional user input device and $v \in \mathbb{R}^n$ as the n-dimensional robot motion, a matrix $D \in \mathbb{R}^{nxm}, D = (d_0, d_1, \ldots, d_m)$ can be defined such that

$$v = D \cdot u, \tag{1}$$

where D linearly maps an individual robot motion DoF d_i to each DoF of the user input device (cf. Fig. 1).

As most input devices supply fewer DoFs than the system which they control ($m < n$), a form of mode switching is generally applied. In our notation, this would be equal to exchanging the DoF-mapping matrix D. As an example, Figure 3 shows the static DoF-mapping matrices of the three default control modes of the Kinova Jaco joystick, omitting Drinking mode and the two-finger grasp option.

$$\begin{array}{c}
\text{X-Axis}\\ \text{Y-Axis}\\ \text{Z-Axis}\\ \text{Roll}\\ \text{Pitch}\\ \text{Yaw}\\ \text{Gripper}
\end{array}
\overbrace{\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\\0&0&0\\0&0&0\\0&0&0\\0&0&0\end{pmatrix}}^{\text{Translational mode}}
\quad
\overbrace{\begin{pmatrix}0&0&0\\0&0&0\\0&0&0\\1&0&0\\0&1&0\\0&0&1\\0&0&0\end{pmatrix}}^{\text{Wrist mode}}
\quad
\overbrace{\begin{pmatrix}0&0&0\\0&0&0\\0&0&0\\0&0&0\\0&0&0\\0&0&0\\1&0&0\end{pmatrix}}^{\text{Finger mode}}$$

Figure 3: The DoF mapping of the default control modes on the Kinova Jaco joystick

Assuming the use of an input device with sufficient DoFs ($m = n$) and corresponding DoF-map $\hat{D} \in \mathbb{R}^{n \times n}, \hat{D} = (d_0, d_1, \ldots, d_n)$ with linearly independent DoFs d_i and therefore $\operatorname{rank}(\hat{D}) = n$, a user would have complete control of the system without the necessity of switching modes. We name such a mapping a complete DoF-set. If each DoF of an input device directly controls a single cardinal robot DoF using a *complete DoF-set*, $\hat{D}$ would be equal to the identity matrix. For an input device with $m < n$, the mapping for different modes can be generated based on a complete DoF-set by stacking m columns of $\hat{D}$, optionally using zero-padding if $m \nmid n$. This method ensures that the set of modes collectively gives the user the same complete control as an input device with $m = n$ if each column (i.e. DoF) of $\hat{D}$ is represented in at least one mode. For the Kinova Jaco joystick, the underlying identity matrix-shaped $\hat{D}$ can easily be seen in Figure 3.

3.2 Approach

Our approach is to adaptively calculate the mapping D for a low-DoF input device, such that the most likely direction of control is represented by the first DoF in D. We require that the DoFs are perpendicular to one another, such that each of the remaining columns represents the next most likely direction for arbitration. Assuming an optimal mapping, the first DoF should therefore enable the user to manoeuvre the manipulator to their desired position, with the second DoF allowing them to adjust according to personal preferences. Further options of arbitration exist with the remaining DoFs.

For clarification, please see the following example: A user wants to pour water from an open bottle into a cup. Whilst approaching the bottle, the first DoF initially offers a 3D path command towards the cup, with the second DoF offering an adjustment in the z-direction, thus allowing to grasp the bottle higher or lower. Once in grasping range, these DoFs automatically switch to grasping and rotation around the bottle.

We generate the mapping D from a complete DoF-set $\hat{D}$. If the user wants to perform an action not represented by the current mapping, simple mode switching is applied as a fallback option to give the user the remaining modes for complete control. This can, for example, be automated by switching after a defined idle time, thus allowing to control a complex high-DoF system with a very low-DoF input device. Regarding the update rate of the mapping, internal tests showed the best results when keeping D static while the user is performing any action and, therefore, only updating D when the user gives no input (i.e. zero-input).

3.3 Learning Degrees of Freedom

In order to learn a mapping of DoFs given a certain situation, training data of robot motion is required. As we aim to extend the possibilities of control that are possible with a specific low-DoF input device, it is necessary to take advantage of more complex methods of control (i.e. high-DoF input devices) for the demonstration sequences. Therefore the control pipeline of the deployed implementation in Figure 1 differs from the training setup.

During data generation, using an m-dimensional input device to command an n-dimensional robot with $m \geq n$ allows maximum flexibility and avoids control-based restrictions of robot motions. Applying such a setup, the user interface software requires no mode switching and a simple identity matrix-shaped DoF-mapping D. For data generation and training, the control pipeline is therefore a direct link between input commands u and robot motions v. For our scenario, a joystick-equipped gamepad with continuous user input is used.

This setup allows to intentionally use able-bodied subjects with a very different method of control to generate training data, making it much easier to collect the dataset. Based on this, the CNN can learn a distribution y of arbitrarily complex robot motions v for a specific situation as described by the camera image x. This means for a situation as perceived by the camera image x, the CNN predicts which robot motions v are likely and unlikely to follow, expressed as a distribution of robot motions y.

3.4 Probabilistic view

We view the training data as samples from everyday activities performed by a robot arm. For the probabilistic view discussed here, an outcome of the considered probability space models a snapshot of a random moment of a random everyday activity.

Let X, Y and V be random variables, where X represents the image provided by the camera and V the robot motion. We are interested in the training distribution of V given X ($V|X = x$), i.e. what DoF the user will most likely command in the specific situation evident in the camera data $X = x$. This distribution shall be the basis for selecting an optimal DoF-mapping D and hence the output of the CNN.

Accordingly, we assume $P(V|X = x)$ to exist and follow a multivariate normal distribution $\mathcal{N}_n(\mu, \Sigma)$ with the mean vector $\mu \in \mathbb{R}^n$ and the symmetric, positive definite covariance matrix $\Sigma \in \mathbb{R}^{n \times n}$. Y contains parameters describing μ, Σ and is therefore also a random variable, depending on X.

Treating the control commands in training sequences as samples of V, a feed-forward CNN is used to estimate Y given the camera input X. The link between Y describing the conditional distribution of V and the particular V in the training sample is made by a specific loss (see below), similar to a maximum likelihood loss. We moreover define $\mu = (0, \ldots, 0)^T$ to represent a zero-motion when having the respective zero-input from the user. The CNN therefore only needs to calculate the covariance matrix Σ.

Knowing the distribution of user commands in a given situation allows us to extract a representation of principal components and use these as DoFs for our mapping. We can therefore calculate a complete DoF-set $\hat{D}$ by generating a matrix where each column represents an eigenvector of Σ, sorted in descending order by their respected eigenvalues. Thus, the mode generated by taking the first m columns of $\hat{D}$ as D represents the smallest expected error between the expected (intended) robot motion V and what the user can command with the input device using u. This will be derived in the following.

3.5 Mathematical Derivation of Optimal D

Our DoF-mapping D in (1) has fewer rows n than columns m, hence not every v can be obtained by an appropriate u. However,

$$u = D^+ v, \tag{2}$$

with D^+ as the Moore-Penrose-inverse of D gives the input u that produces a robot motion Du as close to v as possible.

With this in mind, we want to obtain the best DoF-mapping $D \in \mathbb{R}^{n \times m}$ given that the intended user command V in this situation is distributed as $V \sim \mathcal{N}_n(0, \Sigma)$. We define *best* by the following requirements:

$$\|Du\|_2 \leq \|u\|_2 \quad \forall u \in \mathbb{R}^m \tag{3}$$

$$\text{minimize } \mathrm{E}\left(\|V - DD^+V\|_2^2\right) \tag{4}$$

$$\text{among (4)-optimal } D \text{ minimize } \mathrm{E}\left(\|D^+V\|_2^2\right) \tag{5}$$

Requirement (3) forbids too large amplification of the user input, which would make the system hard to control. It also avoids an infinite optimum for D in (5). Requirement (4) expresses our primary goal, namely to minimize the expected difference between the robot motion desired by the user V and the one DD^+V that can be commanded via the input device. In general, there are several optimal solutions and among these, we prefer the one that minimizes the command (5).

Note that (4) depends only on the subspace spanned by the columns of D (span D), while (5) depends on D itself.

D can be singular-value decomposed as $D = A\,\mathrm{diag}(\sigma_1, ..., \sigma_m)B^T$, $A \in \mathbb{R}^{n \times m}$, $B \in \mathbb{R}^{m \times m}$, with orthonormal A and B. Due to (3),

$\sigma_i \leq 1 \,\forall\, i$, (5) can be rewritten in terms of the σ_i as

$$\mathrm{E}\left(\|D^+V\|_2^2\right) = \mathrm{E}\left(\|B \operatorname{diag}(\sigma_1^{-1}, ..., \sigma_m^{-1}) A^T V\|_2^2\right) \tag{6}$$

$$= \mathrm{E}\left(\| \operatorname{diag}(\sigma_1^{-1}, ..., \sigma_m^{-1}) A^T V\|_2^2\right) \tag{7}$$

$$= \mathrm{E}\left(\sum_{j=1}^{m} \left(\sigma_j^{-1} A_{\bullet j}^T V\right)^2\right) \tag{8}$$

$$= \sum_{j=1}^{m} \sigma_j^{-2}\, \mathrm{E}\left(\left(A_{\bullet j}^T V\right)^2\right) \tag{9}$$

Now D can be replaced by $D' = AB^T$ (equivalently $\sigma'_j = 1$) which is orthonormal, still meets (3), has the same span as D and hence the same (4). It has at least as large singular values as D and hence an equal or smaller (9). Thus it improves (5).

In conclusion, we can restrict our search for the optimal (4) to orthonormal D, because among the solutions equally good in (4), there is always an orthonormal one at least as good in (5).

We know, that DD^+V is the closest approximation of V in span D. Hence, $V - DD^+V$ is orthogonal to span D and DD^+V. It follows by the Pythagorean theorem, that

$$\|V - DD^+V\|_2^2 = \|V\|_2^2 - \|DD^+V\|_2^2 \tag{10}$$

$$= \|V\|_2^2 - \|D^+V\|_2^2 = \|V\|_2^2 - \|D^TV\|_2^2, \tag{11}$$

where the last two equations are because D is orthonormal. So (4) is equivalent to

$$\text{maximize}_{D \text{ orthonormal}}\, \mathrm{E}\left(\|D^TV\|_2^2\right) = \operatorname{tr} \operatorname{Cov}(D^TV) \tag{12}$$

$$= \operatorname{tr} D\Sigma D^T. \tag{13}$$

This is a well studied problem in linear algebra and as [12, Corollary 4.3.39] states, the maximum is obtained when D is chosen as orthonormal eigenvectors to the m largest eigenvalues.

This is the mathematical justification of our approach. It can be readily generated by defining the eigenvectors of Σ sorted by descending eigenvalues as a full DoF-set $\hat{D}$. First, D consists of the first m columns of $\hat{D}$. Should the desired robot motion not be (well) covered by these DoFs, the user can switch to the next m columns.

3.6 Implementation

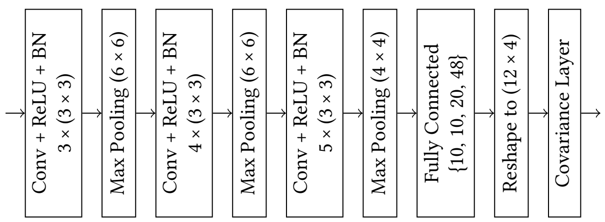


Figure 4: Neural Network

The structure of our CNN is shown in Figure 4. The image-shaped features are processed by Convolutional (Conv) layers with Rectifying Linear Units (ReLU), Batch Normalisation (BN) and max pooling such that fully connected layers can be applied on a flat feature vector. As the final layer, a sample-based method estimates the covariance matrix Σ, with

$$\hat{\Sigma} = \frac{1}{k} \sum_{i=1}^{k} \left(\frac{t_i}{\|t_i\|_2}\right) \left(\frac{t_i}{\|t_i\|_2}\right)^T, \tag{14}$$

$$\Sigma = \varepsilon\, I_n + \hat{\Sigma}, \tag{15}$$

where $\varepsilon > 0$, I_n is the n-dimensional identity matrix and $t_i \in \mathbb{R}^n$ are k samples generated by the previous layer. Each sample is normalised, such that

$$\operatorname{tr}(\hat{\Sigma}) = \sum_{i=1}^{n} \lambda_i = 1, \tag{16}$$

with $\lambda_i, i = 1 \ldots, n$ being the eigenvalues of $\hat{\Sigma}$. This method functions as a novel output layer for neural networks, allowing to learn conditioned covariance matrices, guaranteed to be positive definite with defined trace.

We trained our neural network using the loss function $l(v, \Sigma)$

$$l(v, \Sigma) = v^T \Sigma^{-1} v \tag{17}$$

based on maximum log-likelihood loss, to learn a distribution such that the probability of the robot motion $v \in \mathbb{R}^n$ is maximised. In comparison to the standard maximum log likelihood loss, we have omitted constant scaling factors and offsets, as well as the term $\ln |\Sigma|$. Conceptually, this term penalises the covariance matrix for growing too large. As we limit this already by defining the trace of the matrix and internal tests showed better training results without this term, we chose to omit it.

4 SIMULATION ENVIRONMENT

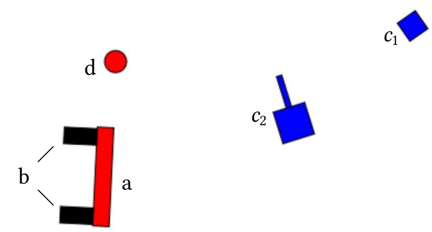


Figure 5: Element overview of the simulation environment

A simple 2D simulation environment was created to develop, test and evaluate the basic principles of adaptive DoF learning as a proof of concept. Figure 5 shows a section of the environment that includes all relevant features. To function as a minimal working example, the user-controlled device is a robotic manipulator (a) able to move forward and backward, sideways, rotate around its center, and close the gripper (b). This sums to a 4-dimensional setting, or 4 DoFs for the user to control. Two blue boxes (c_i) need to be grasped and moved towards the goal marker (d). The physics between the robot, gripper and boxes are handled by a Box2D JavaScript port[3], while the goal marker is solely visual and has no colliding component. At the start of an iteration, all components are positioned randomly. Optionally, the simulation can be toggled,

[3]https://github.com/hecht-software/box2dweb

such that the boxes have spikes on one side (cf. c_2), effectively adding an additional complexity to the scenario, as the gripper can now only grasp the boxes from the side opposite the spike.

Within this environment five options exist to control the robot:

(1) standard control using 8 binary buttons on the keyboard (2 per DoF, one positive and one negative) to control the robot along the cardinal DoFs, therefore allowing only a limited set of directions,
(2) standard control using 4 binary buttons and automated mode switching to cycle through all four cardinal DoFs,
(3) standard control using a joystick with multiple continuous inputs, thus fulfilling the requirement of a high-DoF input device in section 3,
(4) adaptive control using up to 4 binary buttons on the keyboard to steer the robot along up to two DoFs of the neural network-generated DoF-set, and
(5) adaptive control using a joystick with continuous input values based on the same DoF-set as 4.

Option 3 was used for data generation and options 2 and 4 for evaluation. Options 1 and 5 are used for testing and future work respectively.

A mode switching setup is used after five seconds without user input. The currently active DoFs are represented by colored arrows, showing the future state of the robot when following the respective DoF. Figure 2 shows an example situation, with the standard control shown on the left and the adaptive control on the right. When using adaptive control, a server evaluates the current state of the environment and generates the DoF-mapping matrix D for the simulation.

The simulation is implemented in JavaScript, therefore allowing quick and easy website deployment for user studies and evaluations. A variety of settings are customisable within a user interface and allow different deployment strategies for the changing DoFs, thus enabling us to evaluate how much DoF-variety, and therein complexity, users can handle. Internal tests showed the best results when not altering the DoF-set while the user enters any non-zero input and normalising the individual DoFs such that the largest component is always positive. While this prevents the neural network from constantly adjusting the DoFs to create smoother movements, it makes the motion more predictable for the user. The simulation can generate DoF-mappings either using rendered images for CNN-approaches or as an optional alternative using a slim eight-dimensional status vector.

5 USER STUDY

To evaluate the concept of adaptive DoF control, we ran an initial user study based on the 2D simulation system described above. The aim was to compare the standard control (i.e. a static identity matrix-shaped DoF-set) to our adaptive control.

Following the low-DoF HCIs of assistive systems, control option 2 was used for standard control and option 4 for adaptive control. The user input is therefore limited to four binary keyboard buttons to control two DoFs of the robot and having an automated mode switch after every five seconds without user input. The adaptive DoFs are redefined by the network whenever there is no user input, whereas the standard control is based on the cardinal DoFs.

The users were tasked with completing the scenario twelve times: Use the robot to grasp one box after the other and deliver each of them individually to the goal. After every three attempts, the control method switched between standard and adaptive control. After six attempts, spikes were activated for the boxes. To avoid preferences due to training effects, the initial control method was chosen randomly. Before the experiments, each user was shown an introductory video explaining the interface and control methods. During the experiment, the users were kept informed about the currently selected control method. Finally, each user was asked to anonymously evaluate their experience using a questionnaire.

To evaluate the impact of training, a small subset of users were given additional training of roughly ten minutes after their participation in the above-mentioned experiments. After this training, they repeated the adaptive sections of the experiment and gave their evaluation in a similar questionnaire.

5.1 Training

For the adaptive control we trained CNNs for both the scenario with and without spikes based on individual training sets, where the former dataset had spikes activated during data generation. In order to allow complete freedom of motion, the training data for both sets were generated with control option 3. For each training sequence, the simulation started with a random configuration and the users were tasked with grasping the boxes (on the non-spiked side if applicable) and delivering them to the target.

The dataset used for the scenario without spikes was generated by two people and consists of 392 sequences with a total of 29927 datapoints. The network converged in seven epochs.

The dataset used for the scenario with spikes was generated by three people and consists of 488 sequences with a total of 28075 datapoints. The network converged in eight epochs.

5.2 Results

The group of participants consisted of 23 people with a 8/13/1/1 gender split (female/male/diverse/no answer) with ages from 20 to 34 (25.96 ± 3.30). Of those, 2 male and 2 female, ages from 22 to 26, participated in the extended study after training. Regarding their previous experience with keyboard-based controls, the users responded between 1 and 10 (7.04 ± 3.10) on a scale from 1 (never used before) to 10 (usage on a daily basis).

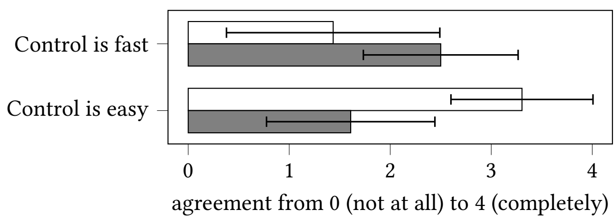


Figure 6: User evaluation of standard (white) and adaptive (grey) control

The users evaluated the speed and ease of both control methods in each scenario (square boxes and boxes with spikes) on a 5 point Likert scale. Figure 6 shows the results in a bar chart with the bar width representing the mean value and error bars showing the standard deviation.

We evaluated two hypotheses, H_1: *adaptive control is subjectively faster than standard control*, and H_2: *standard control is perceived easier than adaptive control* using dependant two-sampled one-sided t-tests. We were able to reject the null-hypotheses for both H_1 and H_2 and show the differences to be significant (cf. table 1).

On a scale from one to five, the users gave the standard control a rating of 3.17 ± 0.65 and the adaptive control 3.09 ± 1.00. Evaluating the suitability of the presented controls in more complex scenarios on a scale from one to ten, the users gave (4.87 ± 1.79) points for the standard control and 5.83 ± 1.99 for the adaptive control.

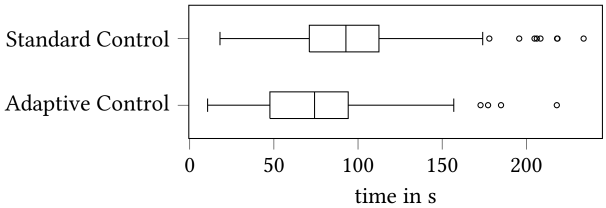


Figure 7: Sequence execution times

Figure 7 shows the distribution of sequence execution times using either standard or adaptive control. While the times vary strongly, it can be observed that the fastest sequences were always performed with the adaptive control, whereas the slowest used standard control. We evaluated hypothesis H_3: *adaptive control is faster than standard control* with a dependant two-sampled one-sided t-test and were able to reject the null-hypothesis and conclude the results to be significant (cf. table 1). This supports the subjective user responses regarding speed and shows that they were able to successfully utilise the subjectively more complex control to achieve lower execution times.

Table 1: T-test results

	M_D	SD_D	t	df	p
H_1	-1.07	1.46	-3.51	44	< 0.001
H_2	1.70	0.86	9.43	44	< 0.001
H_3	21.32	49.96	5.01	274	< 0.001

After additional training, the subset of users performing adaptive control a second time rated the adaptive control faster and easier than before training, while still not rating quite as easy as the standard control. The measured average execution times of the adaptive control sequences after training are lower than before, thus supporting their claim.

5.3 Limitations

The data obtained by this study has been generated entirely online and without any supervision. While this assures real anonymity and avoids personal bias, it cannot be assured that all users completely understood the control methods and the task itself. The participants of the study included a good gender diversity and variety of experience, but only a small age range.

In an optional comment field, some users expressed their desire for a more extensive training and the corresponding expectation that this would greatly benefit the adaptive approach. For the standard control, they also listed the mode switching delay as too long, with some requesting an additional button for switching. Users also complained about not using different subsets of cardinal DoFs (i.e. different definitions of modes). For the adaptive control, there were some complaints about too quick DoF changes, as well as occasional situations where the first and second DoF swapped among each other, therefore missing an opportunity to learn a button-to-action mapping for the user.

In addition to the data presented, five participants generated data, that was deemed flawed and omitted: One person left the simulation idle for several minutes, thus rendering the timings useless; three people seemingly did not follow the instructions by never actually grasping the boxes, and the data of one person was not transmitted completely.

6 CONCLUSION

In this work, we provided proof-of-concept of a novel method for shared control of an assistive robot and evaluated the idea within a 2D simulation environment. For this, we defined a new standardised representation of control modes and introduced a CNN structure to adaptively generate DoF-mappings based on camera data of the current situation and trained it using a specific output layer for conditioned covariance matrices.

The presented application is a simplified proof of concept with a larger scenario as perspective. Even though we expect the largest impact of adaptive DoF-learning in the more complex scenario, the results of our user study show a significant decrease in execution times even in the simple environment. We therefore conclude that adaptive DoF-mapping has the potential to provide a novel interface to assistive robot control and significantly lower task execution times. However, a big challenge for the robot arm application will be communicating the DoFs to the user.

6.1 Future Work

As this work is only a proof of concept in a low-DoF environment, the next steps will be integrating the CNN and concept of control in a more complex 3D environment. It will also be necessary to evaluate the control on more specific tasks of daily living, instead of simple 2D box manipulation.

By addressing more complex environments, an even more flexible interface is necessary. We will therefore evaluate the use of a joystick as an input device for our adaptive control. This will allow users to apply continuous commands, rather than binary button-outputs, to control the robot in the defined modes. This would enable the user to not only control directions of movement, but also control robot velocities.

ACKNOWLEDGMENTS

This work is partially funded by the German BMBF (Bundesministerium für Bildung und Forschung) project AdaMeKoR (FKZ 16SV8534).

REFERENCES

[1] Reem K Al-Halimi and Medhat Moussa. 2016. Performing complex tasks by users with upper-extremity disabilities using a 6-DOF robotic arm: A study. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 25, 6 (2016), 686–693.

[2] Brenna D Argall. 2013. Machine Learning for Shared Control with Assistive Machines. In *Proceedings of ICRA Workshop on Autonomous Learning: From Machine Learning to Learning in Real world Autonomous Systems, Karlsruhe, Germany.*

[3] Cheng-Shiu Chung, Hongwu Wang, and Rory A Cooper. 2013. Functional assessment and performance evaluation for assistive robotic manipulators: Literature review. *The journal of spinal cord medicine* 36, 4 (2013), 273–289.

[4] Anca D Dragan and Siddhartha S Srinivasa. 2012. *Formalizing assistive teleoperation.* MIT Press, July.

[5] BJF Driessen, TK Ten Kate, F Liefhebber, AHG Versluis, and JA Van Woerden. 2005. Collaborative control of the manus manipulator. *Universal Access in the Information Society* 4, 2 (2005), 165–173.

[6] D. Gallenberger, T. Bhattacharjee, Y. Kim, and S. S. Srinivasa. 2019. Transfer Depends on Acquisition: Analyzing Manipulation Strategies for Robotic Feeding. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI).* 267–276. https://doi.org/10.1109/HRI.2019.8673309

[7] Aditya Goil, Matthew Derry, and Brenna D Argall. 2013. Using machine learning to blend human and robot controls for assisted wheelchair navigation. In *2013 IEEE 13th International Conference on Rehabilitation Robotics (ICORR).* IEEE, 1–6.

[8] Felix Ferdinand Goldau, Tejas Kumar Shastha, Maria Kyrarini, and Axel Gräser. 2019. Autonomous Multi-Sensory Robotic Assistant for a Drinking Task. In *2019 IEEE 16th International Conference on Rehabilitation Robotics (ICORR).* IEEE, 210–216.

[9] Deepak Edakkattil Gopinath and Brenna Argall. 2017. Mode Switch Assistance To Maximize Human Intent Disambiguation.. In *Robotics: Science and Systems.*

[10] Axel Graser, Torsten Heyer, Leila Fotoohi, Uwe Lange, Henning Kampe, Bashar Enjarini, Stefan Heyer, Christos Fragkopoulos, and Danijela Ristic-Durrant. 2013. A supportive friend at work: Robotic workplace assistance for the disabled. *IEEE Robotics & Automation Magazine* 20, 4 (2013), 148–159.

[11] Laura V Herlant, Rachel M Holladay, and Siddhartha S Srinivasa. 2016. Assistive teleoperation of robot arms via automatic time-optimal mode switching. In *The Eleventh ACM/IEEE International Conference on Human Robot Interaction.* IEEE Press, 35–42.

[12] Roger A Horn and Charles R Johnson. 2012. *Matrix analysis.* Cambridge university press.

[13] Anja Jackowski, Marion Gebhard, and Roland Thietje. 2017. Head Motion and Head Gesture-Based Robot Control: A Usability Study. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 26, 1 (2017), 161–170.

[14] Siddarth Jain and Brenna Argall. 2019. Probabilistic human intent recognition for shared autonomy in assistive robotics. *ACM Transactions on Human-Robot Interaction (THRI)* 9, 1 (2019), 1–23.

[15] Siddarth Jain, Ali Farshchiansadegh, Alexander Broad, Farnaz Abdollahi, Ferdinando Mussa-Ivaldi, and Brenna Argall. 2015. Assistive robotic manipulation through shared autonomy and a body-machine interface. In *2015 IEEE international conference on rehabilitation robotics (ICORR).* IEEE, 526–531.

[16] D. Kim, R. Hazlett-Knudsen, H. Culver-Godfrey, G. Rucks, T. Cunningham, D. Portee, J. Bricout, Z. Wang, and A. Behal. 2012. How Autonomy Impacts Performance and Satisfaction: Results From a Study With Spinal Cord Injured Subjects Using an Assistive Robot. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans* 42, 1 (2012), 2–14. https://doi.org/10.1109/TSMCA.2011.2159589

[17] Kinova Inc. 2019. *KINOVA JACO Assistive robotic arm.* Retrieved 2019-08-19 from https://www.kinovarobotics.com/en/products/assistive-technologies/kinova-jaco-assistive-robotic-arm

[18] Kinova Inc. 2019. *User Guide Kinova Gen2 Ultra lightweight robot.*

[19] Maria Kyrarini, Adrian Leu, Danijela Ristić-Durrant, Axel Gräser, Anja Jackowski, Marion Gebhard, Jochen Nelles, Christina Bröhl, Christopher Brandl, Alexander Mertens, et al. 2016. Human-Robot Synergy for cooperative robots. *Facta Universitatis, Series: Automatic Control and Robotics* 15, 3 (2016), 187–204.

[20] Veronique Maheu, Philippe S Archambault, Julie Frappier, and François Routhier. 2011. Evaluation of the JACO robotic arm: Clinico-economic study for powered wheelchair users with upper-extremity disabilities. In *2011 IEEE International Conference on Rehabilitation Robotics.* IEEE, 1–5.

[21] Bente Martinsen, Ingegerd Harder, and Fin Biering-Sorensen. 2008. The meaning of assisted feeding for people living with spinal cord injury: a phenomenological study. *Journal of advanced nursing* 62, 5 (2008), 533–540.

[22] Jartuwat Rajruangrabin, Isura Ranatunga, and Dan O Popa. 2012. Adaptive interface mapping for intuitive teleoperation of multi-dof robots. In *Conference Towards Autonomous Robotic Systems.* Springer, 49–60.

[23] Jeroen Schäfer and Marion Gebhard. 2019. Feasibility analysis of sensor modalities to control a robot with eye and head movements for assistive tasks. In *Proceedings of the 12th ACM International Conference on PErvasive Technologies Related to Assistive Environments.* ACM, 482–488.

[24] Katherine M Tsui, Dae-Jin Kim, Aman Behal, David Kontak, and Holly A Yanco. 2011. "I want that": Human-in-the-loop control of a wheelchair-mounted robotic arm. *Applied Bionics and Biomechanics* 8, 1 (2011), 127–147.

[25] Jörn Vogel, Katharina Hertkorn, Rohit U Menon, and Máximo A Roa. 2016. Flexible, semi-autonomous grasping for assistive robotics. In *2016 IEEE International Conference on Robotics and Automation (ICRA).* IEEE, 4872–4879.

[26] Ming Zhong, Yanqiang Zhang, Xi Yang, Yufeng Yao, Junlong Guo, Yaping Wang, and Yaxin Liu. 2019. Assistive Grasping Based on Laser-point Detection with Application to Wheelchair-mounted Robotic Arms. *Sensors* 19, 2 (2019), 303.

Probabilistic Combination of Heuristic Behaviors for Shared Assistive Robot Control

Felix Goldau
felix.goldau@dfki.de
German Research Center for Artificial Intelligence (DFKI)
Bremen, Germany

Udo Frese
udo.frese@dfki.de
German Research Center for Artificial Intelligence (DFKI)
Bremen, Germany

Figure 1: User Perspective of the Adaptive DoF Control, including a robot arm with wrist-mounted camera (left) and the user interface as shown on the smart glasses (right)

ABSTRACT

Technology in general is developed to improve the lives of their users, often by allowing them to handle individual struggles. Assistive robotics takes this concept to its extreme by (re-) enabling users to physically interact with their environment in their daily life. For a successful utilization however, a user interface is required that allows for easy and quick interaction. Based on the promising concept of Adaptive Degree of Freedom (DoF) Control, this paper presents a novel heuristic implementation of the underlying principle, without relying on learning user actions.

To achieve this, a mixture distribution is obtained which expresses how likely the user wants which motion. Here, every mode of the mixture represents a heuristic behavior. Each such behavior defines its own distribution of motion, as well as a weight indicating how likely it is in the current situation. The best fitting DoF is obtained from this mixture and offered to the user with an interface.

This general-purpose control method has been tested in a small technical study, the results of which show its general viability, promising chances for a significant reduction of mode changes, as well as very good quantitative feedback by the users.

PETRA '24, June 26–28, 2024, Crete, Greece

ACM ISBN 979-8-4007-1760-4/24/06
https://doi.org/10.1145/3652037.3652071

CCS CONCEPTS

• **Human-centered computing** → *Accessibility technologies*; Empirical studies in accessibility; User centered design.

KEYWORDS

Adaptive DoF Control, Latent Action Space, Assistive Robotics, Social Robotics, Shared Robot Control

ACM Reference Format:
Felix Goldau and Udo Frese. 2024. Probabilistic Combination of Heuristic Behaviors for Shared Assistive Robot Control. In *The PErvasive Technologies Related to Assistive Environments Conference (PETRA '24), June 26–28, 2024, Crete, Greece.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3652037.3652071

1 INTRODUCTION

Assistive robot arms help users with physical impairments who cannot use their own arms and hands to perform activities of daily living (ADLs). However, robot arms are complex devices with usually six degrees of freedom (DoFs) plus at least one for the hand. This makes designing an accessible and effective user interface for the robot difficult. On one hand, special input devices with many degrees of freedom, e.g. a 3D mouse, exist, but they require considerable dexterity which most people of the target group cannot exercise. On the other hand, input devices specifically tailored for this application, e.g. a head motion based joystick, only offer to input one or two degrees of freedom. Overall the challenge is the mapping of these two input DoFs onto the seven output DoFs.

The standard solution is a menu to switch between different assignments of cardinal DoFs, so-called modes, e.g. X, Y, Z, roll, pitch, yaw, gripper. As the required motions in ADLs are usually not aligned with a cardinal coordinate system, many mode switches are required slowing down the robot's use. For example pouring a glass of water takes $\approx 500\ s$ with ≈ 50 mode switches[7]. Our research investigates the idea that artificial intelligence shall analyse the situation from images of a hand mounted RGBD-camera and suggest a DoF the user will probably want to use. Usually, this will not be a cardinal one. If the computer is right, time is saved and efficiency gained; if the computer is wrong, the user can still manually choose an alternative DoF.

This shared control scheme leaves the user in command and reduces requirements for the artificial intelligence that does not need to be as perfect as in pure autonomous operation. Conceptually this can be viewed as a probabilistic prediction of $p(U_t|Z_{1:t})$: How likely is it, that the user will want the motion U_t, given the situation $Z_{1:t}$ as apparent from sensors such as the camera? With the large success of deep learning, it is tempting to learn $p(U_t|Z_{1:t})$ from recorded data, such as the DORMADL dataset[4]. However, there are some challenges in that, which we will discuss in detail in Section 6. The most notable two are: The necessity to work on image sequences $Z_{1:t}$ instead of single images Z_t, as the hand camera often simply does not see enough of the scene; And the difficulty to incorporate the actually available geometric information from the depth camera and kinematics.

Hence, this paper explores a more analytical alternative. The approach is to define $p(U_t|Z_{1:t})$ as a mixture distribution, where each mixture mode represents an elementary behavior that is heuristically defined and reports its own relevance to the situation, which in turn is used as a weight in the mixture. The behaviors can access the current depth and kinematic data Z_t, as well as the position of tracked objects even if they are currently outside the camera's view. The latter is a way of aggregating information over time, i.e. computing $p(U_t|Z_{1:t})$ instead of $p(U_t|Z_t)$. It is not as general as the $|Z_{1:t}$ notation suggests, but it addresses the most important issue, namely to remember objects outside the field of view.

1.1 Contributions

The contributions of this paper are

- a set of basic behaviors of assistive robots in ADLs,
- in particular an approaching and grasping behavior based on perceived geometry and a generic object segmenter,
- a method for probabilistic combination of behaviors in a shared control setting,
- and a user study showing the validity of the approach as well as opportunities to improve user satisfaction by reducing mode switches.

The paper initially gives a short overview of relevant shared control concepts in assistive robotics, then presents the proposed method and finally reports on the user study.

2 SHARED CONTROL IN ASSISTIVE ROBOTICS

One of the most promising operation concepts for an assistive robot is shared control. Based on the duality of control inputs from a user and software, the concepts generally pairs user input with an automation or similar support by a computer system (e.g. [16] automatically handles end-effector rotation with the user only controlling translational DoFs). This aims to reduce mode switches, mental load, and execution errors by providing targeted assistance during control operations by the user.

As assistive robotics in general aims to (re-) enable users to perform tasks of daily living, the focus of any associated control methodology needs to be the user themselves, therefore requiring interfaces that keep the users in control whilst allowing them to operate the high-DoFs robots[3]. As the user's life should be self-determined and not automated, classic automation-based robot control systems cannot be applied here. Instead, a clear understanding of user intent is vital[1] and needs to be incorporated into the very base of any functional shared control concept. Assuming a known intention, [11] allows to control the arm using a latent action space, where the user's lower dimensional control input results in a high dimensional motion of the robot.

For this work, we will focus on the similar concept of *Adaptive DoF Control* as presented in [5], which follows the idea of adaptively adjusting the DoFs controllable by user interface, dependent on the current situation. Generally, a classic system has a defined set of modes (e.g. [12] with translational, rotational, grasping). The adaptive DoF control effectively reduces the number of modes and adjust them accordingly: By allowing more diverse DoFs, it allows the user to control the robot along directions more appropriate to the current situation, for example by moving the robot along a diagonal instead of the robot-typical jagged motion of going first left, then forward.

3 PROBABILISTIC COMBINATION OF HEURISTIC BEHAVIORS

The field of *Behavior-based Robotics* follows the concept of creating seemingly complex interactions based on minimal and very simple sensor-driven actions (compare [2]). Generally speaking, this is implemented by direct linking of sensor input to specific action, as for example rotating to the right if light was detected on this side. The most prominent examples of this are the tortoise robots by W. Grey Walter[17]. Such behavior-based robots are often intentionally compared to biological systems, as they can quickly react to new sensory input.

Instead of directly generating actions, as in behavior-based robotics, the robot control presented in this work follows a different, but correlated approach: Use statistical features of a set of preferably simple behaviors as the suggested DoF in an adaptive DoF Control (see Sec. 2).

The idea behind this control is not to select isolated behaviors, but to instead create a combination of options via DoFs. This allows for more natural and smoother motions, especially in areas of transition, as well as blending of directions.

The associated general control structure is shown in Figure 2 and will be explained in detail during this work.

3.1 Definition of a Behavior

In this work, a behavior b describes a simple action from a finite set of actions (such as *Lifting the End-Effector*). More specific, it

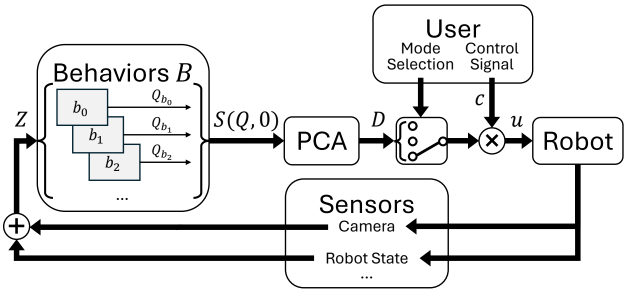


Figure 2: Software control loop. Based on (robot-associated) sensor input Z, the set of behaviors B generate sigma points Q_b, accumulated to a probability distribution S. The user controls the robot along an axis of the principal components of the latter.

is an adaptive multivariate probability distribution describing the likelihood of motion in an n-dimensional direction U_t relative to the end-effector given the situation as observed from the sensor readings $Z_{1:t}$.

$$p_b(U_t|Z_{1:t}) \sim b, \quad U_t \in \mathbb{R}^n \tag{1}$$

The n-dimensional vector U_t describes a velocity, which can be joint angle or Cartesian and include components for the gripper. In our case it consists of a stacked vector with translational, rotational and gripper velocities in end-effector coordinates. Since U_t contains rotational velocity not orientation, there is no problem with singularities or need for quaternions.

In order to ease further processing, each behavior's distribution is defined as a Gaussian $\mathcal{N}(\mu_b, \mathrm{S}_b)$ with mean μ_b and covariance S_b. It is represented by a set Q_b of sigma points, similar to usage in unscented Kalman filters[15] (see Figure 3).

$$\mu_b = \frac{1}{|Q_b|} \cdot \sum_{q \in Q_b} q \tag{2}$$

$$\mathrm{S}_b = \mathrm{S}(Q_b, \mu_b) \tag{3}$$

where $\mathrm{S}(Q, \mu)$ is the covariance of sigma points Q with reference point μ:

$$\mathrm{S}(Q, \mu) = \frac{1}{|Q|} \cdot \sum_{q \in Q} (q - \mu)(q - \mu)^T \tag{4}$$

In this context, each point represents a direction of control which originates at the end-effector's tool center point. The set of points therefore build a distribution with expected value E and covariance Cov conditioned on the behavior's underlying action and the current situation. In addition, each behavior provides a weight ω to represent the likelihood $p(b|s)$ of the underlying action in the current situation.

Using this setup, the simplest behaviors consist of a single sigma point $q_0 \in \mathbb{R}^n$ describing a point distribution ($\mu_b = q_0, \mathrm{S}_b = 0$) of a motion with the direction according to this sigma point. For example, a strict *Forwards* behaviour moves in the direction the end-effector is pointing.

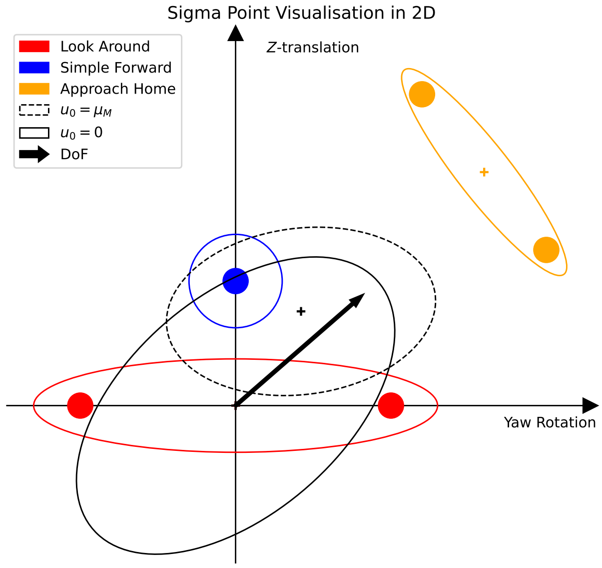


Figure 3: Illustration of the representation of $p(U_t|Z_{1:t})$ as a mixture distribution. The dots are the sigma-points of three behaviors (red, blue, orange), the ellipses show the corresponding mean and covariance. The black dotted ellipse shows the overall mean and covariance of the mixture, the bold ellipse corresponds to S from equation (14), and the black arrow shows the single optimal DoF. Mixture weights are represented by the distance of the sigma points to the origin.

A slightly more complex behavior may, for example, have no preferred sign of the direction (which would be the case for a general DoF) and could therefore represent this by supplying two opposite points $Q_b = \{q_0, -q_0\}$ leading to $\mu_b = 0$ and $\mathrm{S}_b = q_0 q_0{}^T$. An example of this kind of single-DoF behavior is the *Look Around* behavior described in Section 4.1, which can yaw left or right without preference.

More complex behaviors can use arbitrary numbers of sigma points to represent uncertainty in multiple dimensions. These sigma points can also depend on the environment.

If the behavior involves rotation and translation, there are several options: Both can be combined into one sigma-point with a defined ratio. This expresses that the user likely wants a combined motion, e.g. an orbit around an object. Or there can be two sigma-points, one with the rotation and no translation and one with the translation and no rotation. This expresses, that the user likely wants a translation, or a rotation or a combination of both. In theory rotation and translation could also be two different behaviors, expressing, that the user likely wants no combination. However, by the way the following processing is done, this makes no difference.

3.2 Combination of Behaviors as Mixture Distribution

The different behaviors are treated as modes of a mixture distribution with weights, as returned by the behaviors. The rationale behind this view is that we assume the user wants to follow one of the behaviors, but we don't know which one. As the probability of a behavior becomes larger when its better suited to the situation, the mixture distribution has a high chance to reflect the users intent.

$$p(U_t|Z_{1:t}) = \sum_{b\in B} p(b|Z_{1:t}) \cdot p(U_t|Z_{1:t}, b) \tag{5}$$

$$= \frac{1}{\sum\limits_{b\in B}\omega_b} \sum_{b\in B} \omega_b \cdot p_b(U_t|Z_{1:t}) \tag{6}$$

$$= \frac{1}{\sum\limits_{b\in B}\omega_b} \sum_{b\in B} \omega_b \cdot \mathcal{N}(\mu_b, \mathrm{S}_b)(U_t) \tag{7}$$

Therefore, we can calculate the expected value E_M and covariance Cov_M of the resulting mixture distribution directly from the sigma points Q_b of all behaviors $b \in B$ as

$$\mathrm{E}_M = \frac{1}{\sum\limits_{b\in B}\omega_b} \cdot \sum_{b\in B} \omega_b \cdot \mu_b \tag{8}$$

$$\mathrm{Cov}_M = \frac{1}{\sum\limits_{b\in B}\omega_b} \cdot \sum_{b\in B} \omega_b \cdot \mathrm{S}(Q_b, \mathrm{E}_M) \tag{9}$$

This is shown in Figure 3 as a black dotted ellipse.

3.3 Choosing the most suitable DoFs

To be usable for the adaptive DoF control, a matrix $D \in \mathbb{R}^{nxm}$ of individual DoFs needs to be generated from the mixture distribution, such that an m-dimensional user control signal $c \in \mathbb{R}^m$ can be mapped to a final robot motion u:

$$u = D \cdot c + u_0 \tag{10}$$

For this, the user input axis c_j controls motion along the DoF represented by column j of matrix D.

Similar to [5], our DoF-mapping D has more rows n than columns m, hence not every u can be obtained by an appropriate c. However,

$$c = D^+(u - u_0), \tag{11}$$

with D^+ as the Moore-Penrose-inverse of D gives the input c that produces a robot motion $Dc + u_0$ as close to u as possible.

Substituting (11) into (10), we want to minimize the expected squared error $\mathrm{E}(\|e\|^2)$ of this control

$$e = u - D \cdot D^+(u - u_0) + u_0. \tag{12}$$

The problem can also be viewed as minimizing the expected squared distance from the distribution to the m-dimensional subspace $\{D \cdot c + u_0 | c \in \mathbf{R}^m\}$. Typically, it is solved using a Principal Component Analysis (PCA)[13] with the optimum $u_0 = \mu_M$ and D consisting column-wise of the m eigenvectors with the largest eigenvalues.

$$\hat{D} = \mathrm{eigen}_{1:m}(\mathrm{S}_M), \quad u_0 = \mu_M \tag{13}$$

If we were to apply this to an adaptive control, an input of zero (i.e. a non-action of the user) would, instead of a standstill, result in a motion of the robot according to the expected value $u_0 = \mu_M$.

This is neither user-friendly, nor safe, so we change the optimization problem to enforce $u_0 = 0$. For that $\mu_M {\mu_M}^T$ needs to be added to the covariance S_M before calculating the eigendecomposition.

$$D = \mathrm{eigen}_{1:m}(S), \quad u_0 = 0, \quad S = \mathrm{S}_M + \mu_M {\mu_M}^T \tag{14}$$

The resulting distribution is shown in Figure 3 by the black ellipse, with the black arrow representing the eigenvector with the largest associated eigenvalue.

3.4 Mathematical Derivation

This section derives (14). The PCA provides the optimal solution (13) for an optimally chosen u_0. An arbitrary centered m-dimensional subspace is fitted to the distribution. This is well known[13].

As we defined to enforce $u_0 = 0$ to ensure a stationary robot when no user input is given, we modify the derivation from [13] for a zero-centered subspace. Let U be an $\mathbb{R}^n$ random variable. In our system $\mathrm{E}(U) = \mu_M$ and $\mathrm{Cov}(U) = \mathrm{S}_m$, but we will keep the derivation general. Let D parametrize the desired m-dimensional subspace $\{Dc | c \in \mathbb{R}^m\}$, where the columns of $D \in \mathbb{R}^{n\times m}$ span the subspace and D is orthonormal ($D^T D = I$). In [13] there is an additional center θ, which is fixed to 0 in our case.

Since D is orthonormal, $DD^+U = DD^TU$ is the closest point to U on the subspace and BU with $B = I - DD^T$ is the corresponding error vector e. Our optimization problem is

$$J(D) = E\left(\|BU\|^2\right) = \mathrm{E}\left(U^T B^T BU\right) \tag{15}$$

$$D = \arg \min_{D\in\mathbb{R}^{n\times m}, D^T D = I} J(D). \tag{16}$$

The matrix B is symmetric and idempotent ($B^2 = B = B^T$), so

$$J(D) = \mathrm{E}\left(U^T BU\right) = \mathrm{E}\left(U^T (I - DD^T) U\right) \tag{17}$$

$$= \mathrm{E}\left(U^T U\right) - \mathrm{E}\left(U^T DD^T U\right) \tag{18}$$

$$= \mathrm{tr}\left(\mathrm{E}\left(UU^T\right)\right) - \mathrm{tr}\left(D^T \mathrm{E}\left(UU^T\right) D\right) \tag{19}$$

$$= \mathrm{tr}(S) - \mathrm{tr}\left(D^T SD\right), \quad S = \mathrm{Cov}(U) + \mathrm{E}(U)\mathrm{E}(U)^T. \tag{20}$$

This is the same expression as in [13, (19), $W = D, X = U$], except that there $S = \mathrm{Cov}(U)$ and here $S = \mathrm{Cov}(U) + \mathrm{E}(U)\mathrm{E}(U)^T$. To conclude, enforcing $u_0 = 0$ leads to an additional term $\mathrm{E}(U)\mathrm{E}(U)^T$ in S. The rest of the proof is the same, deriving that the optimal $D = \mathrm{eigen}_{1:m}(S)$.

4 BEHAVIORS

For our assistive robotic system, we focussed the set of behaviors on a generalized task of interacting with grasped objects as it is a very common use of the arm. In addition, especially the approach and grasping of objects holds good opportunities for assistive support, as grasping objects often requires detailed alignment from the user, graspable objects can be readily identified and defined, and the task has foreseeable goals, instead of, for example, the inverse *Placement* task, which can end nearly arbitrary.

In the following, we will present the seven developed behaviors, six of which as *basic behaviors* in Section 4.1, and one in more detail in Section 4.2. However, all behaviors presented are treated independently and equally, as they are mixed together (see Section 3.3) instead of being selected individually.

4.1 Basic Behaviors

The six basic behaviors follow a very minimalistic structure and are designed to represent some fundamentals of robotic interaction.

We will briefly present each behavior with its sigma points Q_b in end-effector coordinates EE.

In the following, the terms <Rot $(a)^f$> and <Trans $(a)^f$> are used to describe unit rotations or translations along an axis a of a frame f, whereas <Grasp> describes the unit motion vector of a directional movement leading to closing of the fingers. Also, the transformation matrix $T_{B \leftarrow A}$ maps from coordinate system A to B

Please note, that in this framework, motion is described by a vector with components x, y, z, yaw (around x), pitch (around y), roll (around z), gripper. These refer to the relative motion *velocity* in end-effector-coordinates, not to (relative) poses, so the question of avoiding singularities by matrices or quaternions does not arise.

Look Around. In order to reach a target position, it is often necessary to reorient the gripper towards it. This constant behavior describes the necessary rotation solely around the vertical axis, which aims to avoid possible spillage by not changing the end-effector's alignment with the horizon. In addition, this orientation change towards a target improves the situational awareness of the gripper-mounted camera. Without additional information, this behavior provides a DoF, by supplying two opposite sigma points:

$$Q^T_{LookAround} = \begin{pmatrix} \text{Rot}\left(T_{EE \leftarrow Base} \cdot vertical^{Base}\right)^{EE} \\ -\text{Rot}\left(T_{EE \leftarrow Base} \cdot vertical^{Base}\right)^{EE} \end{pmatrix}, \quad \omega = constant. \tag{21}$$

Forward. Once the end-effector is oriented towards the user's target position, a common motion is driving, relative to the end-effector, forwards. For example, when preparing to grasp, users will often initially align with the object and then move in a direct line to reach a grasping pose. This behavior provides a constant, but light-weighted DoF pointing outwards z from the tool center point of the end-effector. This represents the action of continuing motion in the direction the gripper is pointing. The sigma points form a DoF in both directions, with a tendency to go forwards:

$$Q^T_{Forward} = \begin{pmatrix} \text{Trans}\,(z)^{EE} \\ \text{Trans}\,(z)^{EE} \\ -\text{Trans}\,(z)^{EE} \end{pmatrix}, \quad \omega = constant \tag{22}$$

Grasp. Continuing the though process of aiming to grasp an object, this behavior supplies the actual motion of closing the fingers. This behavior is distinct from the others, as it is the only one affecting this dimension. This avoids the accidental opening or closing of the fingers that would otherwise be possible by multidimensional DoFs. Using the attached depth camera, the behavior reacts to the number of close pixels px in between the gripper that suggest the presence of an object.

If an object is already grasped, the behavior will instead suggest opening the gripper upon standstill, with the likelihood increasing based on the time t_s since last movement and the distance d_g travelled since grasping the object.

$$Q^T_{Grasp} = \begin{cases} \left(-\text{Grasp}\right), & \omega = \text{const} \cdot t_s \cdot d_g \quad \text{if object grasped} \\ \left(\text{Grasp}\right), & \omega = \text{const} \cdot px \quad \text{otherwise} \end{cases} \tag{23}$$

Rotate Upright. In most cases, it is desirable to keep the end-effector upright. This simplifies grasping, avoids dropping or spilling of grasped objects, and is often easier for the user to fathom. Based on the current orientation (roll angle r, pitch angle p) of the end-effector, this behavior provides a rotational motion to reorient the gripper to be upright. The yaw angle has no effect on this behavior.

$$Q^T_{RotateUpright} = \left(-\left(\text{Rot}\,(roll)^{EE} \cdot r + \text{Rot}\,(pitch)^{EE} \cdot p\right)\right), \quad \omega = |r| + |p| \tag{24}$$

Liftoff. After grasping or placing an object, there is generally a short phase of retrieval, where either the object is lifted and positioned, or the arm is retracted from the position of the object. In either case, it makes sense to (slightly) lift the end-effector, as well as retrieve it in the general direction of the robots base $v_{EE \leftarrow Base}$ or the user. This should be a safe direction to move in most cases, as it roughly aligns with the user's line of sight and the robot's joints. This behavior diminishes with the time t_g and distance d_g since grasping.

$$Q^T_{Liftoff} = \left(\text{Trans}\,(vertical)^{EE} + v_{EE \leftarrow Base}\right), \quad \omega = \begin{cases} -\text{const} \cdot t_g \cdot d_g & \text{if object grasped} \\ 0 & \text{otherwise} \end{cases} \tag{25}$$

Approach Home. As certain pre-definable positions have recurring meaning, this behavior provides a direct approach towards a home position, weighted by the distance and relative orientation of the gripper. It's individually comprised of a rotation $\lambda\,(home)$ aiming to align the end-effectors z-axis with the target position, and the translational displacement $v_{home \leftarrow EE}$ of the target pose to the end-effector.

For a wheelchair user, one home pose might be above their wheelchair table, so that they can easily retrieve objects to there. For a detailed explanation of λ and ω in this case, please see Sec. 4.2.

$$Q^T_{ApproachHome} = \begin{pmatrix} \lambda\,(home) \\ v_{home \leftarrow EE} \end{pmatrix} \tag{26}$$

4.2 Approach Object Behavior

The *Approach Object* behavior actively scans for graspable objects in range and provides individual behavior distributions for each of them. Individually, each such distribution represents the action of approaching the specific object. However, in combination they provide a distribution of possible directions that best allow to approach the group of objects. This way it actually encapsulates multiple similar behaviors (one for each object stored) which are aggregated in this section. For a detailed analysis of the resulting performance synergy, see Section 4.3.

In general, per detected object o the behavior provides two sigma points $Q_{ApproachObject(o)}$ and corresponding weights ω^r_o, ω^v_o. The sigma points separately describe translation to, and orientation towards, the goal, as it is assumed that users will perform each of them stepwise one after the other (initially rotate towards the object and approach only then):

$$\begin{aligned} Q^T_{ApproachObject(o)} &= \begin{pmatrix} \lambda(o) \\ v_{EE\leftarrow o} \end{pmatrix}, \\ \omega_o^\lambda &= \theta_\lambda\left(\left|v_{EE\leftarrow o}\right|, \left|\lambda(o)\right|, t_o\right) \\ \omega_o^v &= \theta_v\left(\left|v_{EE\leftarrow o}\right|, \left|\lambda(o)\right|, t_o\right), \end{aligned} \tag{27}$$

where the rotational component $\lambda(o)$ is the angle between the forward-pointing z-axis of the end-effector and the vector towards the object and is calculated as

$$\lambda(o) = \angle\left(\text{Trans}(z)^{EE}, v_{EE\leftarrow o}\right). \tag{28}$$

Perception. The sensory input of this behavior is structured around the *FastSAM*[18] implementation of the *Segment Anything Model* (SAM)[10]. This neural network model is designed to isolate segments in a color image and can basically be run on arbitrary images. We use the model as-is and actively refrain from doing any model adjustments, as we want the underlying generalistic behavior.

Using the color images generated by the wrist mounted RGBD camera[8] (see Figure 1), the SAM model provides image masks for a fairly large number of possible objects it detects. As not all the generated image segments are valid real-life objects, we post-process each image segment in order to interpret it as an object and generate a reasonable target.

Using the camera-proved depth data, the 3D physical extends of the objects are calculated based on a camera project matrix and the rotated minimal bounding box. In addition, the direct neighboring area around the objects is checked to verify, that the segment protrudes sufficiently from the background to be able to be grasped; In other words, it checks the sides of the objects for chasms that are deep enough for the fingertips. The depth data is also used to calculate the relative object pose.

Based on this data, objects are treated as targets if they are within reach, have grasp-appropriate physical extends, as well as having sufficient chasms to the objects sides for the gripper to protrude during grasping.

Each such object o is persistently stored, so it is remembered even when it falls outside the camera's field of view. The weights ω_o^λ and ω_o^v of an unseen object however decrease over the time t_o since the object was last seen, so to say *forgetting* the object.

Sigma Points. For every object o remembered, a sub-behavior is generated that handles direction orientation $\lambda(o)$ and approach $v_{EE\leftarrow o}$ towards the object.

For this, the weights are regularly updated using the functions $\theta_\lambda\left(\left|v_{EE\leftarrow o}\right|, \left|\lambda(o)\right|, t_o\right)$ and $\theta_v\left(\left|v_{EE\leftarrow o}\right|, \left|\lambda(o)\right|, t_o\right)$ respectively. These are simple functions calculating a relative estimate of the likelihood of the user aiming to grasp the object in this current situation. In our setup, they were designed to embrace objects that are close and aligned with the gripper, resulting in a selection dynamic (see Section 4.3). In detail, they are:

$$\begin{aligned} &\theta_v\left(\left|v_{EE\leftarrow o}\right|, \left|\lambda(o)\right|, t_o\right) = \\ &\quad t_o \cdot \left(\left(v_{\max} - \left|v_{EE\leftarrow o}\right|\right) \cdot \left(\frac{1}{2}\left(\left(\cos\left|\lambda(o)\right| + 1\right)\right)\right)^3\right) \end{aligned} \tag{29}$$

$$\begin{aligned} &\theta_\lambda\left(\left|v_{EE\leftarrow o}\right|, \left|\lambda(o)\right|, t_o\right) = \hat{f}_\rho \cdot \theta_v\left(\left|v_{EE\leftarrow o}\right|, \left|\lambda(o)\right|, t_o\right) \\ &\text{where } \hat{f}_\rho = \begin{cases} 1 & \text{if } \left|\lambda(o)\right| > \rho \\ f_\rho & \text{otherwise,} \end{cases} \end{aligned} \tag{30}$$

where $v_{\max}$ is the maximum distance reachable by the robot arm, and ρ is a threshold angle, below which the relative likelihood of the rotate-towards sigma point is reduced by a factor of f_ρ. The latter serves to favor the translational component in the final approach rather than optimizing orientation.

The *Approach Home* behavior (see Sec. 4.1) follows the same principle and ω-scaling.

4.3 Synergy in Shared Control

The presented behaviors can obviously not be used as a baseline for automation; The system's information is only very limited with respect to the environment and the intent of the user. In general, there is too much uncertainty for an automation task.

However, as discussed in Section 2, automation should not be the goal of assistive control strategies. Instead, the presented control was developed with the specific user requirements and relies heavily on interaction of the user, who not only selects the current motion to perform, but instead smoothly controls through them.

As each behavior evaluates its own likeliness and adjusts its weight ω accordingly, the weighted combination of the behaviors provides a quickly adapting set of DoFs. From the perspective of the user, the suggested DoFs barely make any decision between options, but instead provides the user with the means of selecting one themselves.

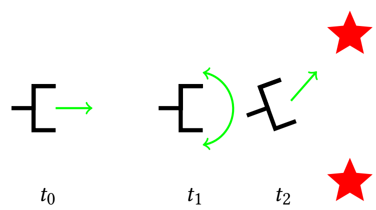


Figure 4: DoF Selection Example: Two star-shaped red targets and a 2D-robot at 3 points in time. For each point in time, a green arrow represents the suggested DoFs that are most likely given the situation.

A simple scenario showcasing these principles can be seen in Figure 4:

Initially, at t_0, the robot is far away from the targets. As the system has no information about which object the user prefers, it simply provides the direction towards the center of the targets, as this will get them closer to their actual goal. This is a direct result of the *Approach object* behavior for each target: The two translational components add up to a clear direction, whereas the rotational components balance out each other and only provide a less likely unsigned DoF.

Assuming the user follows the suggested motion until t_1, the robot has reached a point where the summed translational components have decreased sufficiently, such that they are now smaller than the balanced rotational element. In other words, the system assumes it's no longer effective to drive forward, but instead orient toward one of the targets. The user must not obey this, but can choose to select a new DoF at any point.

At t_2 the user has obviously continued controlling the robot in a translational manner and has only just started rotating slightly to the left. This orientation towards the leftmost target is however sufficient, such that the system assumes that they prefer this target. As a result, it provides a direct translational approach motion. This is a result of the ω-scaling of θ_λ and θ_v.

In summary, the system relies on the user to make their selection. In the setting described above, if the user had not made any decision to switch, the control would have missed the targets and passed through the center. The systems assist only once the user makes a clear enough decision. The same concept generalizes to an arbitrary number of targets, as well as effects from the other behaviors.

As long as the user input devices supports less input DoFs than the robot has ($m < n$), this method of control cannot be complete (i.e. there are poses the user cannot reach). To solve this, a user interface with a fallback option of controlling in the cardinal DoFs is suggested. This way, the adaptive DoF control only extends existing controls and does not restrict the user's options. See the media attachment for a video example.

Seemingly contradictory, this setup of behaviors does not influence the robot safety. The only but essential rule is that if there is no command from the user, the robot performs no motion. In this setting and with this type of robot, contact with the environment necessarily needs to be possible (be it for object interactions, or tasks like scratching), therefore no specific obstacle avoidance is implemented. This is left to the user.

5 TECHNICAL USER STUDY

To verify the technical usability of the control, we conducted a small user study in a laboratory environment using a Kinova Jaco Gen2 7DoF assistive robot arm[9], an Intel Realsense D435[8] and a ros-based software stack. As this was designed as a proof of technical concept, not an end-user compatibility scoring, we opted to test the control with able-bodied participants.

Based on the scenarios from [4], we selected a simplified supermarket shelf scenario. Two objects were placed on a shelf and, using the robot, the users were tasked to retrieve the objects to a basket on a nearby table, which was close to the stored *Home* position. Each user could select the order in which to retrieve the objects, so the system could not be adjusted to the specific setting. After each successful retrieval, the robot was reset to a starting pose to make the trials comparable.

We developed a user interface (UI) for our adaptive DoF control on the same principles as the munevo DRIVE[14] system for controlling an electric wheelchair with head gestures through a Google Glass[6]. This includes mode selection by flicking or nodding motions of the head, and control inputs by tilting the head. The UI can be seen in the top right of Figure 1. The DoF currently controlled is shown in the center, with the new suggestion being highlighted on the side. The cardinal DoFs can be reached by nodding. DoFs are represented with a set of simplified 3D arrows. The use of this interface aimed to provide a sense of a realistic interaction experience for our study participants.

During the study, each participant compared our control to the use of only the cardinal control, both using the same interface. The order of controls was switched for each user. In addition to extensive explanations, the users were also given time to get used to both control methods, and a test task (grasp a held bottle) for introduction. For the latter, they were verbally guided and assisted by the study administrator.

5.1 Results

Our user group of 18 people was aged between 20 and 34 years (25.8 ± 4.2). Of these, 8 reported their gender as female, 8 as male, and none as non-binary, with two choosing not to reply. All test subjects were able-bodied and reported no personal context to the field of care. Most of them (15) regularly used joysticks or keyboards, with 6 users working with robots on a weekly basis.

For the evaluation, we examined the time between the first motion of the robot and the start of opening the fingers to release the object at the target position. Figure 5 shows the execution times over all users with either control method, separated by execution order. For the first object (i.e. when there is very little experience with the control), it can be seen that they are slower with the adaptive control. Even though only slightly better than the classic approach, there is a noticeable improvement of the adaptive duration for the second object, when compared to the first. This is independent of the order the controls were presented in, although the effect is more prominent if classic was used first.

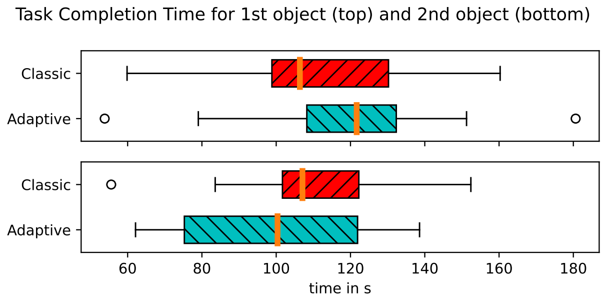


Figure 5: Task Completion Times of all users, separated by object

Another factor for the usability of an assistive control concept is the necessity of mode switches. Depending on the interface and the physical abilities of the user, mode switches can be difficult, exhausting, or time-consuming. It is therefore an important metric of the system. Figure 6 shows the number of mode switches for each control, again separated by execution order. This clearly shows a significant reduction in mode switches for the adaptive control for both objects.

A slight variation of this can be seen in Figure 7, which shows the number of user-interface selections necessary to reach the actual modes to control in. This differs, as the limited input options of the user interface requires the user to skim through various options to reach the desired control mode. The difference between classic and adaptive is even greater, showing reduced necessary skimming

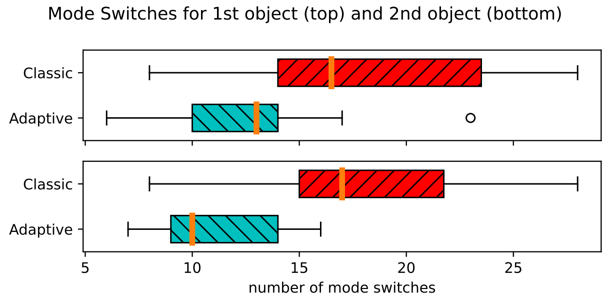


Figure 6: Number of Mode Switches of all users, separated by object

with the adaptive control. A clear training effect can be observed by the reduction of variability (broader range of interquartiles) from the first object to the second.

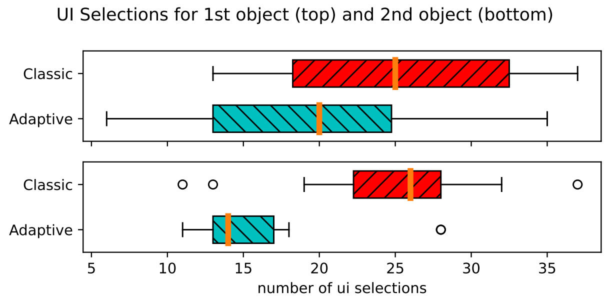


Figure 7: Number of UI Selection Gestures of all users, separated by object

The qualitative user responses had a clear preference: 14 chose the adaptive control and only 4 the classic. The users gave different reasons for this: Some users reported the adaptive control to be simpler, even if they require more familiarization, whereas others simply felt the visual user interface more compact (less mode options), therefore allowing for a better overview. The latter also caused for fewer necessary mode switches, which were sometimes physically exhausting for users. Multiple users explicitly praised the diagonal options of the adaptive control, assisting to move direction to a target, which in turn was an issue for a smaller subset, which had difficulty subconsciously grasping the directions of the more complex arrows.

For this user group, the classic side had a clear advantage, as one did not need to understand the arrows, but could instead learn the modes and positions by heart for this control. In addition, one user reported that the directions suggested by the adaptive control interface often did not fit his wishes.

The conducted NASA-TLX questionnaire was not conclusive: All categories showed very similarly distributed user responses. Based on the interviews, this can be traced back to differing perspectives. For example, for some users the mental demand of the adaptive control was higher because the arrows changed and one had to adapt, whereas others found this to be a reduction of mental load, as the more direct options removed intermediate steps and allowed for more streamlined executions. The physical demand was mostly identical, with some users reporting less strain on the adaptive control because of the reduction in mode changes. The results of the NASA-TLX can be seen in Figure 8.

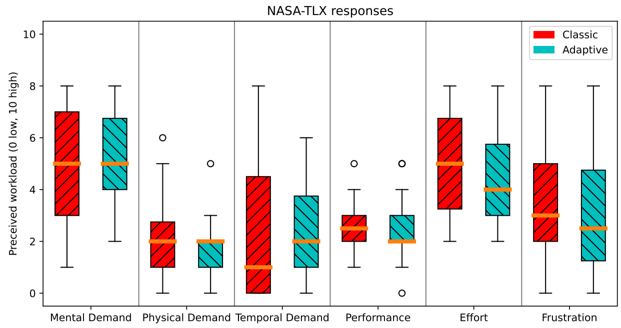


Figure 8: NASA-TLX User Responses

6 COMPARISON TO A HYPOTHETICAL END-TO-END LEARNING APPROACH

It is tempting to learn the desired $p(U_t|Z_{1:t})$ end-to-end from data. Indeed, in previous work we recorded the DORMADL dataset[4] exactly for this purpose. However, investigation of the data revealed considerable complications for such an approach. These motivated our investigation into a more classical engineering alternative reported in this paper and shall be discussed here.

First, as the camera is mounted slightly behind the hand (Fig. 1), it often sees only a small part of the scene. However, the motion is often motivated by something outside the view, e.g. when transitioning from one object to another on a table. This requires the system to memorize object positions to effectively predict motion. In our system this is done analytically with depth data and forward kinematics. An end-to-end learned system would have to learn this connection, which is not easy and of little value, since it is well described analytically. It would also require operating on image sequences not single images.

Second, we observed that users exhibit a large degree of arbitrariness in how they perform a certain motion. This makes it hard to predict. It is actually not necessary for the application to predict precisely which motion variant is desired as any useful variant is fine. However, this is not captured by typical least-squares or maximum likelihood losses.

Third, we observed that the presumably well predictable motions are reaching motions to grasp an object. As we showed, these are low-hanging fruits that do not need deep learning but are well realized geometrically.

Forth, an advantage of deep learning is that it would implicitly learn to understand the environment from images, which cannot be done well analytically. However, as shown in our system, this part can be covered by a network with a more specialized role (segment anything) that has actually been trained on much more data.

Lastly, as most often, the analytical approach is more transparent, because it is build of parts that have a human defined interface. It is also easier adaptable, because the heuristic definition of behaviors can be changed, while changing a learned one requires collecting a new dataset reflecting the desired changes in behavior.

Overall, while we don’t want to rule out an end-to-end learned approach, we claim that the proposed method is well suited for the considered task.

7 CONCLUSION

We have presented a system that suggests degrees of freedom to the user of an assistive robot arm. It models the distribution of what motion the user probably wants to do as a mixture of heuristically defined behaviors. Some of these, in particular the approaching and grasping behavior, incorporate sensor data, mostly in a geometric way. This has been shown as a viable alternative to an end-to-end learning idea.

The technical user study showed the viability of the system in a realistic scenario with a promising perspective on mode switch reduction. Also, the qualitative feedback of the users displayed clear preferences for the new control system, especially during the direct approach of objects.

Future work is to investigate, whether more general behaviors can be implemented this way and whether they are actually predictable to an extent that allows suggesting a DoF to the user. Also, the study needs to be repeated with the targeted user group.

ACKNOWLEDGMENTS

This work was supported by the project *DOF-Adaptiv*, funded by the German Federal Ministry of Education and Research BMBF (Bundesministerium für Bildung und Forschung), FKZ 16SV8563. It also received support from the *REXASI-PRO* H-EU project, call HORIZON-CL4-2021-HUMAN-01-01, under the Grant agreement ID 101070028.

REFERENCES

[1] Brenna D Argall. 2013. Machine learning for shared control with assistive machines. In *Proceedings of ICRA Workshop on Autonomous Learning: From Machine Learning to Learning in Real world Autonomous Systems, Karlsruhe, Germany*.

[2] Rodney A. Brooks. 1991. Intelligence without representation. *Artificial Intelligence* 47, 1 (1991), 139–159. https://doi.org/10.1016/0004-3702(91)90053-M

[3] Hongwu Wang Cheng-Shiu Chung and Rory A. Cooper. 2013. Functional assessment and performance evaluation for assistive robotic manipulators: Literature review. *The Journal of Spinal Cord Medicine* 36, 4 (2013), 273–289. https://doi.org/10.1179/2045772313Y.0000000132

[4] Felix Goldau, Yashaswini Shivashankar, Annalies Baumeister, Lennart Drescher, Patrizia Tolle, and Udo Frese. 2023. DORMADL - Dataset of Human-Operated Robot Arm Motion in Activities of Daily Living. In *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE Press, Piscataway, NJ, 11396–11403. https://doi.org/10.1109/IROS55552.2023.10341459

[5] Felix Ferdinand Goldau and Udo Frese. 2021. Learning to Map Degrees of Freedom for Assistive User Control: Towards an Adaptive DoF-Mapping Control for Assistive Robots. In *Proceedings of the 14th PErvasive Technologies Related to Assistive Environments Conference* (Corfu, Greece) *(PETRA ’21)*. Association for Computing Machinery, New York, NY, USA, 132–139. https://doi.org/10.1145/3453892.3453895

[6] Google. 2019. Glass Enterprise Edition 2.

[7] Laura V. Herlant, Rachel M. Holladay, and Siddhartha S. Srinivasa. 2016. Assistive teleoperation of robot arms via automatic time-optimal mode switching. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE Press, Piscataway, NJ, 35–42. https://doi.org/10.1109/HRI.2016.7451731

[8] Intel Corporation. 2022. Intel Realsense - Product Family D400 Series. (last visited on 13.02.2023). https://www.intelrealsense.com/wp-content/uploads/2022/11/Intel-RealSense-D400-Series-Datasheet-November-2022.pdf

[9] Kinova. 2021. Jaco assistive robot - User guide. (last visited on 13.02.2023). https://assistive.kinovarobotics.com/uploads/EN-UG-007-Jaco-user-guide-R05.pdf "EN-UG-007-r05-202111".

[10] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, Piotr Dollár, and Ross Girshick. 2023. Segment Anything. arXiv:2304.02643 [cs.CV]

[11] Dylan P. Losey, Krishnan Srinivasan, Ajay Mandlekar, Animesh Garg, and Dorsa Sadigh. 2020. Controlling Assistive Robots with Learned Latent Actions. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE Press, Piscataway, NJ, 378–384. https://doi.org/10.1109/ICRA40945.2020.9197197

[12] Veronique Maheu, Philippe S. Archambault, Julie Frappier, and François Routhier. 2011. Evaluation of the JACO robotic arm: Clinico-economic study for powered wheelchair users with upper-extremity disabilities. In *2011 IEEE International Conference on Rehabilitation Robotics*. IEEE Press, Piscataway, NJ, 1–5. https://doi.org/10.1109/ICORR.2011.5975397

[13] Abhilash Alexander Miranda, Yann-Aël Le Borgne, and Gianluca Bontempi. 2008. New routes from minimal approximation error to principal components. *Neural Processing Letters* 27 (2008), 197–207. https://doi.org/10.1007/s11063-007-9069-2

[14] Horst Penkert, Julius C. Baron, Konstantin Madaus, Wolfgang Huber, and Achim Berthele. 2021. Assessment of a novel, smartglass-based control device for electrically powered wheelchairs. *Disability and Rehabilitation: Assistive Technology* 16, 2 (2021), 172–176. https://doi.org/10.1080/17483107.2019.1646817

[15] Rudolph Van Der Merwe and Eric A. Wan. 2004. *Sigma-point kalman filters for probabilistic inference in dynamic state-space models.* Ph. D. Dissertation. Oregon Health & Science University. AAI3129163.

[16] Jörn Vogel, Katharina Hertkorn, Rohit U. Menon, and Máximo A. Roa. 2016. Flexible, semi-autonomous grasping for assistive robotics. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE Press, Piscataway, NJ, 4872–4879. https://doi.org/10.1109/ICRA.2016.7487692

[17] W. Grey Walter. 1950. An Imitation Of Life. *Scientific American* 182, 5 (1950), 42–45. http://www.jstor.org/stable/24967456

[18] Xu Zhao, Wenchao Ding, Yongqi An, Yinglong Du, Tao Yu, Min Li, Ming Tang, and Jinqiao Wang. 2023. Fast Segment Anything. arXiv:2306.12156 [cs.CV]

Adaptive Control in Assistive Application - A Study Evaluating Shared Control by Users with Limited Upper Limb Mobility

Felix Ferdinand Goldau[1,∘] and Max Pascher[2,3,∘] and Annalies Baumeister[4,∘] and Patrizia Tolle[4] and Jens Gerken[2] and Udo Frese[1]

***Abstract*— Shared control in assistive robotics blends human autonomy with computer assistance, thus simplifying complex tasks for individuals with physical impairments. This study assesses an adaptive Degrees of Freedom control method specifically tailored for individuals with upper limb impairments. It employs a between-subjects analysis with 24 participants, conducting 81 trials across three distinct input devices in a realistic everyday-task setting. Given the diverse capabilities of the vulnerable target demographic and the known challenges in statistical comparisons due to individual differences, the study focuses primarily on subjective qualitative data. The results reveal consistently high success rates in trial completions, irrespective of the input device used. Participants appreciated their involvement in the research process, displayed a positive outlook, and quick adaptability to the control system. Notably, each participant effectively managed the given task within a short time frame.**

I. Introduction

In 2023, the World Health Organization (WHO) estimated that approximately 15% of the global population lives with some form of disability [1], many of whom experience substantial, often permanent, reductions in limb usage. The resulting decreased mobility can severely restrict the ability to perform Activities of Daily Living (ADLs) without external assistance, necessitating the almost constant presence of caregivers [2]. However, constant caregiver presence is generally not desirable. Research by Pascher et al. demonstrated that individuals with physical disabilities strongly wish for personal space and alone-time [3], which might be facilitated through the use of dependable robotic assistance [3]. Similarly, a comprehensive review by Kyrarini et al. highlighted the beneficial effects of assistive robotic technologies — known as *cobots* — in aiding individuals with mobility issues [4]. Consequently, the decreased reliance on caregiver assistance supports the regaining of independence and addresses expressed wishes for self-determination.

However, introducing robots capable of (semi-) independent actions presents new challenges, potentially adding stress for the end-users if not properly considered during the design phase [5]. Pollak et al. [5] noted reduced sense of control felt by users when operating a cobot in autonomous mode, while switching to manual mode allowed participants to regain a sense of control and significantly lower stress levels. These findings are supported by Kim et al., who reported significantly higher satisfaction in the group using manual cobot control [6]. Unlike routine tasks in industrial settings, such as assembly jobs [7], the assistive care environment demands flexibility as cobots are tasked with a variety of support functions [8]. Managing robots in these scenarios remains demanding and requires continuous user involvement for efficient and safe system operation. A central issue arises from the types of robots employed, as multiple Degrees-of-Freedom (DoFs) either require complex multidimensional input devices or involve time-extensive mode switching (e.g., [9], [10]). The former option is often unmanageable for individuals with mobility impairments, while the latter leads to increased task completion times [11]. Consequently, these prevailing control strategies do not suit the needs of the intended audience.

In addressing this, adaptive DoF control merges semi-autonomous operations with manual flexibility, dynamically adjusting a robot's DoFs for simplified interactions based on the environment. Introduced by Goldau & Frese, this strategy enhances support for ADLs, outperforming traditional controls by using a Convolutional Neural Network (CNN) to select optimal DoFs from real-time environmental feeds [12]. Further research by Pascher et al. demonstrated a reduction in mode switching, indicating a notable improvement over standard controls [13], [14], [15] and explored different input devices for this adaptive control [16]. Goldau & Frese also confirmed the adaptive approach's advantages through heuristic behavior studies in a laboratory setting [17]. Nonetheless, the real-world applicability and impact of these advances, especially in user studies targeting specific groups, are yet to be fully examined. Building on these insights, the present study assesses the acceptance of adaptive control among the actual target group — people with limited upper limb functionality — through three select input devices.

Our contribution is two-fold: 1) we present a user study with the target group conducted at an international trade fair for rehabilitation and care, evaluating a novel shared control

∘ These authors contributed equally to this work.

* This research is supported by the *German Federal Ministry of Education and Research* (BMBF, FKZ: 16SV8563, 16SV8564, 16SV8565, and 01IW24001).

[1]Felix Ferdinand Goldau and Udo Frese are with the German Research Center for Artificial Intelligence (DFKI), 28359 Bremen, Germany felix.goldau@dfki.de, udo.frese@dfki.de

[2]Max Pascher and Jens Gerken are with the TU Dortmund University, Inclusive Human-Robot Interaction, 44227 Dortmund, Germany max.pascher@udo.edu, jens.gerken@udo.edu

[3]Max Pascher is also with the University of Duisburg-Essen, Human-Computer Interaction, 45127 Essen, Germany max.pascher@uni-due.de

[4]Annalies Baumeister and Patrizia Tolle are with the Frankfurt University of Applied Sciences, 60318 Frankfurt am Main, Germany annalies.baumeister@fb4.fra-uas.de, tolle@fb4.fra-uas.de

approach, and 2) provide an in-depth analysis of control performance data (average task completion time and average number of control switches) and subjective feedback (perceived workload, technology acquisition, and acceptance), highlighting the concepts' adaptability to various devices.

II. Assistive Robotics in Domestic Care

When designing assistive technologies for vulnerable user groups, such as people with disabilities, efficient human-robot collaboration becomes paramount. Assistive robotics have the potential to significantly enhance independence and improve care by assisting and supplementing caregivers, thereby enhancing the quality of life for those in need [6], [18], [19], [20]. Research attention has increasingly focused on how assistive robotic systems can assist individuals with motor impairments. Notably, projects like *Robots for Humanity* led by Chen et al. [21] and seminal studies like by Fattal et al. [8] explored the feasibility and user acceptance of these technologies. While the overarching aim is to fully integrate individuals with severe motor impairments into professional and social contexts, current assistive technologies predominantly target the performance of ADLs [22]. These activities range from basic tasks like eating and drinking to more complex ones, including grooming and leisure activities [23].

Continual research efforts are expanding the capabilities of cobots and enhancing task performance. For instance, Gallenberger et al. utilized camera systems and machine learning in an autonomous robotic feeding system [24], while Canal et al. introduced a learning-by-demonstration framework for feeding tasks [19]. Both methods demonstrate how robotic arms can execute (semi-)autonomous tasks with minimal user intervention, thus underscoring the potential benefits of assistive technology. Implementing safe, user-friendly robotic solutions can fundamentally improve the quality of life for individuals needing assistance while ensuring that the user retains control [25]. This increased independence is particularly vital for those with motor impairments, reflecting their desire for more privacy and prolonged alone time [26].

Drolshagen et al. found that individuals with disabilities readily adapt to working alongside cobots, even in close quarters [27]. Moreover, people with motor impairments tend to positively receive robotic assistance, especially when their specific needs are considered during the design process [28], and when sufficient oversight ensures a sense of security [29]. Thus, effective communication of the robot's motion intent emerges as a crucial factor in achieving high acceptance among end-users [30]. These findings align with Beaudoin et al.'s investigation into the long-term usage of the Kinova Jaco, a notable advancement in assistive technology [31].

A. Shared Robot Control Applications

The appropriate level of autonomy in assistive robots attracts attention in current research. Highly autonomous systems (e.g., [32]), which minimize user interaction to mere oversight, can induce stress [5] and feelings of distrust among users [33]. Conversely, for users with certain degrees of impairment, only minor adjustments to the users' otherwise manual control input [34] can pose significant challenges [21], [35]. Shared control provides a middle ground by integrating manual user operation through standard input devices with algorithmic software assistance to adjust the resulting motion [13]. This approach effectively mitigates concerns associated with purely autonomous systems and manual controls [36]. In shared control, there is a collaborative effort between the user and the robot, empowering individuals with motor impairments to actively participate in their care. By balancing autonomy and user involvement, shared control systems offer a more acceptable and comfortable experience for individuals relying on assistive technologies [37], [38], [39].

A distinct approach is the adaptive DoF control system proposed by Goldau & Frese [12]. This system isolates the most likely DoFs of a robotic arm based on the current situation and aligns them with a low-DoF input device. Effectively, this improves the classic mode-selection process by replacing the selectable modes with situation-adaptive directions of movement, allowing the user to easily control the arm. The process involves attaching a camera to the robotic arm's gripper and utilizing a CNN trained on ADLs performed by individuals without motor impairments [12], akin of the learning-by-demonstration method used in autonomous robots [19]. Furthermore, this CNN-based approach offers extensibility as it can be trained to distinguish between different situations, enhancing its practicality for everyday use. In their proof-of-concept study, which involved a 2D simulation environment featuring a robotic gripper representation and a target object, Goldau & Frese observed faster task execution with the proposed system than manual controls. However, users perceived the shared control approach as complex, expressing a preference for a more extensive training phase, even in this low-DoF environment. Their findings underscore the need for more intuitive and responsive interaction feedback when controlling the robot.

Pascher et al.'s *Adaptive DoF Mapping Control (ADMC)* concept draws inspiration from Goldau & Frese's approach but extends it to three dimensions [13]. This extension increases the potential DoFs, enabling a more precise realization of ADLs. In their case studies, they show the advantages of an adaptive against a non-adaptive control approach [15], [14] and explored different input devices for the ADMC concept [16]. Following the transition to 3D, Goldau & Frese expanded on their previous control by presenting a functional 3D prototype [17]. Here, instead of generating the DoFs with a CNN, they switched to a heuristic behavior-based approach. Using non-disabled participants in a laboratory environment, they showed the general viability of their control method, as well as the users' preferences for their novel approach.

However, as the adaptive DoF control is yet to be evaluated with the target group in a realistic real-world scenario, its general accessibility and user acceptance remains to be assessed. Due to the diverse limitations of the target demographic, this accessibility coincides with a generalizability to different input devices.

III. Technical Concept

In line with the adaptive control principles discussed in Section II-A, our study implemented a behavior-based heuristic control with a focus on assessing its applicability across various input devices and the acceptance within the targeted user group. The shared control approach adopted here is based on behaviors comparable to [17], albeit with a modification that incorporates known initial object poses. This adaptation aims to mitigate detection errors within the complex and cluttered environments typically encountered in trade fairs.

Unlike prior studies evaluating the usability of the proposed adaptive control concept [15], [14], our experimental setup differs by concurrently integrating multiple approach- and graspable objects, as opposed to a single defined target. Our implementation is designed to operate without a predetermined sequence of actions, allowing users flexibility in interaction. To facilitate a practical assessment, we modified the research-oriented *AdaptiX* [13] framework into a concise standalone Robot Operating System (ROS)-based system without a Mixed Reality (MR)-middleware. Similar to [17], our system uses a smart glass as a visualization interface, which we further augmented with options to use buttons or a joystick as input devices.

IV. Study Method and Materials

To assess the effectiveness of our adaptive control strategy, we conducted a supervised evaluation with a cohort of 24 participants. Our approach focused on qualitative data to gain individual insights into the broader implications of this diverse and challenging-to-generalize user group. Additionally, we supplemented this qualitative analysis by quantitative data derived from execution measurements and a NASA Raw-Task Load Index (Raw-TLX) questionnaire [40].

To achieve high external validity, we exclusively recruited participants from the target group and conducted the study in a relatively realistic environment, opting for a trade fair instead of an artificial laboratory setting. Participants used our adaptive control system with a designated input device to perform a simple task with a robot arm, after which they provided feedback on their experiences. The experiment primarily aimed to gather qualitative insights from the target group regarding the adaptive control strategy, supplemented by subjective questionnaires and performance data measurements.

A. Study Design

The study employed a between-subject design due to considerations of participant vulnerability (e.g., differing levels of fatigue) and diverse capabilities (e.g., only head-control being an option for some participants). Consequently, we used the input device as an independent variable, segmented into two distinct conditions: (1) *Head-Control* and (2) *Joystick*. Additionally, participants from both groups were asked to sample a third condition: (3) *Assistive Buttons*.

The evaluated input devices were selected to be sufficiently distinct from one-another to accommodate a wide range of users, with the *Joystick* and *Head-Control* requiring finger and head dexterity respectively, whereas the *Assistive Buttons* could be placed to be used with any body part. However, matching devices to participants' capabilities resulted in imbalanced data ($n_{\text{Head-Control}} = 16$, $n_{\text{Joystick}} = 8$, $n_{\text{Assistive Buttons}} = 16$). It is important to note that user familiarity with the devices varies greatly, as joysticks and buttons are more common than head-based controls.

To facilitate an in-depth analysis of immediate user perceptions, we recorded both audio and video during the study. Additionally, we evaluated the following dependent variables:

- **Average Task Completion Time:** The time to approach an object, pick it up, and position it at a designated target area was recorded (in seconds) for each trial.
- **Average Number of user interface (UI) Switches:** Within each trial, we documented instances of UI switching, i.e., selections within the UI independent of robot action, activated through a head-motion or button-press on the control device.
- **Average Number of Mode Switches:** We measured mode switching, i.e., successful UI switches followed by a user input to move the robot along a new DoF.
- **Perceived Workload:** Following the completion of each condition, we assessed the six dimensions of the Raw-TLX questionnaire [40] to gauge perceived workload.
- **Level of Autonomy:** Upon completing all trials, we asked participants to identify their preferred level of autonomy on a Likert-scale 1–10 (1: manual control, 10: full autonomy).

Following the practical part of the study, we engaged participants with several open-ended questions to explore their experiences, understanding of the control method, interpretation of directional cues, and any significant issues they encountered.

To extract participants' perceptions regarding the different control methods, the study's video and audio recordings were analyzed independently by three researchers through open coding. The resulting open codes were organized into affinity diagrams and further structured into themes, as detailed in Section V-B.

B. Hypotheses

Overall, we expect the adaptive control method to be well perceived by the target group, as long as the controls prove to be functional with the chosen input device. To assess this, we defined three hypotheses:

H1: After a short training, our target group of wheelchair-users with limited upper limb mobility is able to repeatedly use an adaptive DoF control for a grasp-and-retrieve task.

H2: Adaptive DoF control is perceived as promising and accessible by the target group to perform tasks of ADLs.

H3: The concept of adaptive control generalizes to different input devices.

Fig. 1: Study apparatus at the trade fair, illustrating the placement of user, table, and shelf, as well as the UI visualized on the smart glasses (top left)

(a) (b) (c)

Fig. 2: Input devices used in the study: **(a)** *Google Glass EE2* [47], **(b)** custom-built *Joystick*, and **(c)** *Assistive Buttons*

C. Apparatus

The system was assembled on a mobile nightstand, simulating a setup typically found in nursing homes or hospitals (cf. Figure 1). The central component was a *Kinova Jaco Gen 2* [41] 7-DoF assistive robotic arm with an *Intel Realsense D435* [42] color-and-depth-camera mounted to its end effector. As detailed in Section III, we evaluated multiple user input devices: A *Google Glass EE2* [43] with a customized *Munevo Drive* [44] software was used as the smart glasses, whereas an *Xbox Adaptive Controller* [45] with external *Assistive Buttons* and a custom-built *Joystick* served as hand-controlled input devices. The devices are depicted in Figure 2, with the UI-visualization shown in the top left of Figure 1. To minimize external influences in the busy trade fair environment, all devices communicated via wired connections to a ROS [46] interface of an embedded *Linux* computer. The only exception was the glasses, which were connected via a short-range *Bluetooth* connection.

D. Participants

We focused on a target demographic of wheelchair users with reduced upper-limb mobility and the capability of wearing and using smart glasses. Consequently, individuals were excluded if they had vision impairments that made the glasses inaccessible or if the glasses did not fit (e.g., due to custom headrests). In total, 24 individuals — 12 men and 12 women — with varying motor impairments participated in the study. The age range of participants was 19 to 68 years, with an average age of 43.75 years (SD = 14.68). All participants relied on wheelchairs and had diverse health diagnoses, including spinal muscular atrophy, ALS, DMD, mitochondrial disease, AMC, MS, ICP, stroke, GNE myopathy, Charcot-Marie-Tooth syndrome, MMN, spina bifida, and generalized dystonia. One participant reported fully functional arms and hands, 20 had limited arm and hand function, and three had complete loss of arm and hand function. Five participants had prior experience with an assistive robotic arm, another five had tested such an arm in the past, and 14 had no prior experience. Additionally, two individuals regularly used smart glasses, four had tried them before, and 18 had never used such glasses.

E. Procedure

The study was conducted at the *REHACARE*,[1] a leading international trade fair for rehabilitation and care in Düsseldorf, Germany. This location allowed for easy recruitment from the target group, as they are common visitors. The experiment setup was designed as part of a regular booth, with the participants facing a shelf to minimize visual distractions. Before the start, participants were thoroughly briefed about the research objectives and the to-be-completed tasks. Each participant provided explicit, informed consent to engage in the study and agreed to video/audio recording and documentation of all pertinent data.

The study administrator collected a socio-demographic questionnaire, monitored the experiment via a laptop, and provided instructions to participants on how to use the hardware and navigate the basic functions of the study environment. This followed the steps:

1) The participant engages in a training trial with one-by-one assistance from the study administrator.
2) 1–4 measurement trials (depending on individual capabilities) for the assigned condition are conducted.
3) Based on personal capabilities, a subset of participants tested the *Assistive Buttons* as an alternative input.
4) Finally, we conducted a Raw-TLX questionnaire [40] and a semi-structured interview.

The study concluded with a debriefing session, with a total average session duration of 60 minutes. Participants were compensated with a 10 EUR food voucher for their time and engagement, a detail that was not disclosed beforehand.

F. Experimental Setting and Task

A small basket was placed as a target drop zone on a table in front the participant, thus allowing for a design that does not specify the object's final orientation. From the user's perspective, four objects were placed inside a 2x2 shelf behind the table. The robotic arm, attached to the table, could reach both the shelf spaces and the basket. For each trial, the participants were tasked with guiding the robotic arm from its initial position to grasp an object from the shelf and put it into the basket. Upon successful placement, the trial concluded, and the object was removed from the basket. The robot, operated by the study administrator, returned to its starting point before starting a new trial for the

[1] *REHACARE* trade fair. `https://www.rehacare.de`, last retrieved June 7, 2024.

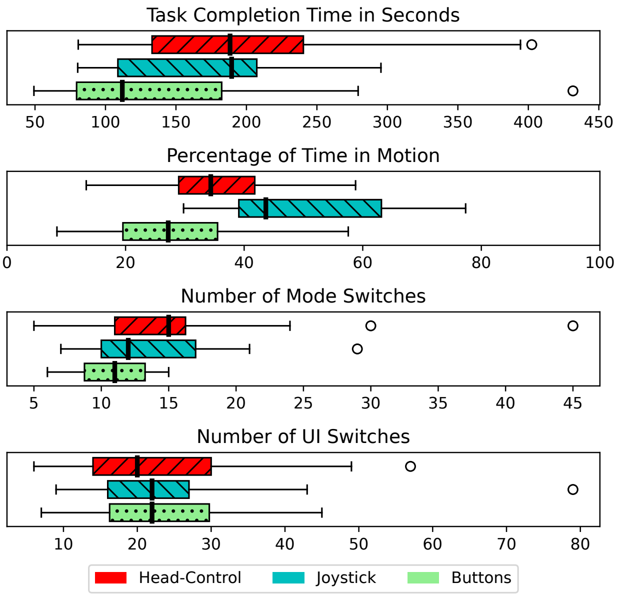


Fig. 3: Box plot of execution procedure measurements over all users with $N_{\text{Head-Control}} = 36$, $N_{\text{Joystick}} = 17$, $N_{\text{Buttons}} = 16$. The bold line represents the median

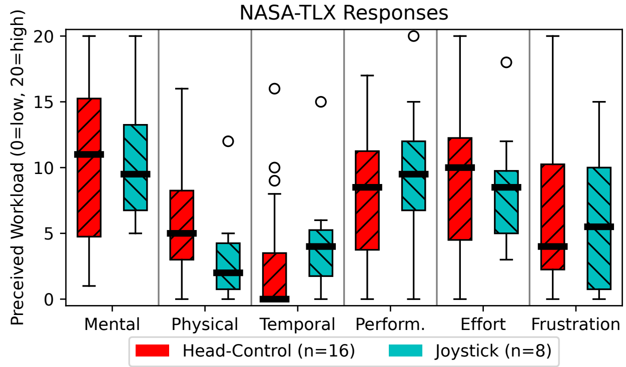


Fig. 4: Results of the NASA Raw-Task Load Index questionnaires

remaining objects on the shelf. Neutral box-shaped objects were selected to prevent bias and ensure consistency across trials.

V. Results

The study covers 81 measured trials (24 participants × 1–4 trials), with the training trials being excluded from analysis. To accommodate user capabilities, two thirds (16) of the participants evaluated the smart glass-based *Head-Control*, while eight used the *Joystick*. Additionally, 16 users evaluated *Assistive Buttons* as a second input method after their first trails.

A. User Procedure Analysis

As each trial begins at the same robot pose and involves only a single object with a pre-defined pose, we were able to analyze the user execution procedure for a singular task of approaching, grasping, and retrieving a single object. After excluding runs with external interruptions or major complications, we recorded 69 trials (36 *Head-Control*, 17 *Joystick*, 16 *Assistive Buttons*).

For each device, we recorded the overall task execution time, the percentage of time actually spend moving the robot, as well as the number of mode and UI switches. Figure 3 shows an overview of the collected data for all users and devices. Figure 4 presents the subjective Raw-TLX scores. Each dimension is displayed as a box plot, separated by the type of control (*Head-Control* or *Joystick*) initially employed by the user.

B. Thematic Content Analysis

Throughout the trials and interviews, the participants verbalized their experiences, including challenges, moments of success, and improvement suggestions from their perspective. The audio recorded during these sessions was transcribed for thematic content analysis. Since participants were specifically asked to reflect on their study trials, the analysis primarily focuses on their individual experiences and perception of the adaptive control.

1) Learning the Control: Participants experienced varying learning curves, with only a minority finding the adaptive control intuitive at the beginning. Instead, most found the control to be initially strenuous, before improving their opinion after a short training period. Once “the concept was understood” (P11), usage became easier and more successful.

The primary challenge referenced by all participants was confronting the unfamiliar technology, especially using head movements in conjunction with smart glasses to control a robotic arm via augmented reality. Difficulties were noted even by those who used a joystick for input and the glasses solely for visualization. Participants had to quickly learn new skills and adjust to the adaptive control system, leading some to report feeling mentally overstrained at times.

The two participants already familiar with smart glasses and head movements to control their wheelchairs experienced the least difficulty in learning to control the robotic arm. In contrast, participants accustomed to using a robotic arm with a traditional joystick struggled to transfer their previous experience to the new system, regardless of whether they controlled the robot with head movements or a joystick during the trial. This difficulty was partly attributed to the nature of the shared control system, which imposed adaptive motions rather than the traditional fixed cardinal motions.

Given the recurring theme of technology acquisition and learning, participants were asked to estimate the training time required to use the robotic arm with adaptive control at home proficiently. While all participants were confident they could achieve proficiency with time and practice, their estimated training times varied, ranging from a few hours to several days or even weeks, potentially including further coaching sessions.

2) Visualization: A central feature of the control system is the arrow-based visualization displayed on the smart glasses (see top left in Figure 1). Participants often felt uncertain about what action the robotic arm would take when following the arrow. As a result, they reported missing "the right one" (P8) and distrusting the system's suggestions. In these cases, frustration and uncertainty arose as the remaining task completion got more complicated and resulted in unexpected situations.

Overall, participants generally understood simple arrow indications. However, more complex movements were challenging as many had difficulty with rotations, curved movements (multidimensional paths), and distinguishing between forward and backward control directions. Some participants struggled to keep track of new suggestions and wanted the option for manual control in addition to the generated suggestions; an option that was not available during the study. Furthermore, participants also wished the system would indicate which object it was targeting.

3) Physical Devices: Participants using *Head-Control* repeatedly forgot which physical movement corresponded to each UI command, leading to periodic mix-ups. Conversely, those using the *Joystick* struggled with the differences between controlling a robotic arm and a wheelchair. However, after training, the *Joystick* users reported finding the adaptive control easy to use and an improvement over previously known controls.

Responses to the *Head-Control* varied among participants. While most users found the adaptive control easier with increased insight and practice, two users experienced stress and physical tension from the head movements. One participant attributed this to their neuro-psychological impairment, finding the movements tiring and challenging to focus on. Nevertheless, participants generally found the adaptive control to be an interesting new method of controlling a robotic arm. Many described it as enjoyable once they became accustomed to it, with some even finding the suggestions and control to be intuitive once they *got the hang of it* (P23).

Participants who also tested the *Assistive Buttons* often found them to be the more accessible and more comfortable solution. Only one out of five participants from the *Joystick* group who tested the *Assistive Buttons* preferred the *Joystick*. Among those who initially tested the *Head-Control*, 12 tried the *Assistive Buttons*, with only three preferring *Head-Control*. Participants who preferred the *Assistive Buttons* found them more familiar and easier to use. They also found the limited direction options (left, right, forward, backward) to be more accessible.

4) General User Remarks: The study occurred at a fair rather than in a controlled laboratory environment, which was noted by participants as contributing to nervousness. Additionally, the bright light at the trade fair caused difficulties in recognizing graphics on the transparent display of the smart glasses. Participants found it strenuous to shift focus between the real robotic arm and the display and expressed a desire for visual alignment. Despite these challenges, participants generally viewed the robotic system positively, appreciating the balance between suggestions and manual control.

5) Preferred Level of Automation: Given the frequent discussion surrounding the balance between automation and manual control in assistive robotics and shared control, participants were asked to express their own preferences regarding the level of automation on a scale ranging from 0 (i.e., no automation) to 10 (i.e., robotic system that functions completely autonomous).

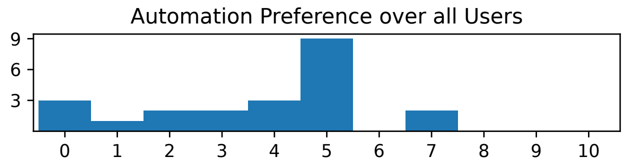


Fig. 5: Histogram of preferred level of (robotic) automation of all users (0 complete manual control, 10 complete automation)

The results, shown in Figure 5, display a peak at the midpoint (5) with overall relatively evenly distributed responses at lower levels of automation. Notably, none of the participants favored the highest levels of automation. Their reluctance towards complete automation stemmed from the significance they placed on maintaining independence from technical devices. However, there was a consensus among users that some degree of automation would be beneficial or even essential, considering their physical limitations. Overall, users expressed a preference for support in manually controlling the robot rather than full automation.

VI. Discussion

Previous research [14], [15], [16] has demonstrated the general functionality of adaptive control with predetermined input devices, conducted by non-disabled users. The user procedure data generated in the present study corroborates these findings by confirming a general functionality independent of the input device. The measured completion times and number of mode or UI switches showed no significant differences between input devices. Moreover, all participants completed the trials successfully and most evaluated the control concept positively. Consequently, these findings validate hypotheses H1 and H3 with a considerable degree of confidence.

In contrast, verifying H2 proves not as straightforward. Participants needed to learn the use of new technologies (glasses and adaptive control), which was mentally taxing and likely affected their perception of the adaptive control method. However, most participants did succeed after a brief period. Notably, they all anticipated that with more practice, usage would become easier, quicker, and more intuitive. This involves both learning the general concept of robot control and gaining a better understanding of the arrows and resulting robot motions. The latter, in particular, would lead to a clearer understanding of the robot's motion intent and encourage user trust in the suggestions and the overall system.

Overall, future users of an assistive system as used in this study could mitigate many — if not all — challenges experienced by the study participants. Among expected training effects, the glasses could be more personalized and calibrated more precisely to the individual user than is feasible in an experimental setup. Also, users are likely to select their preferred input device and become proficient with it. As shown by the preferred level of automation and user feedback, the presented concept fits requested the middle ground between automation and manual control. The results of the Raw-TLX indicate low physical and temporal demands when using adaptive DoF control with an assistive robotic arm, thus representing an added value compared to previous solutions.

A. Limitations

This study evaluated a novel research-based shared control concept specifically with the intended user group, in contrast to much of the existing literature, which often includes participants outside this demographic. Our approach allows us to draw conclusions directly relevant to the end users without relying on generalizations from non-disabled user cohorts. However, to achieve this, the study took place at an international trade fair for rehabilitation and care, resulting in certain environment-specific limitations.

Despite efforts to isolate trials and minimize external influences, the largely uncontrollable environment of the trade fair had a marked impact. However, even with the often audibly chaotic conditions, participants generally remaining focused during recordings, with only few getting noticeably distracted. Nevertheless, the noise and activity levels did affect the quality and options for recording quantitative data. As such, we focused more on the qualitative analysis of audible user comments during trials and their responses in the final interview. This approach yielded valuable insights, particularly because they came directly from the intended users themselves. Despite some distractions in the environment, they did not impact the qualitative data.

Finally, like most studies involving new control concepts, our participants only had a brief period to test the system. For comprehensive insights, the shared control approach requires extensive testing by target users in their everyday lives under assistive care settings.

VII. Conclusions

This study presents an evaluation of a novel but literature-known concept of shared assistive robot control within the context of ADLs through direct engagement of the intended user demographic in a realistic setting. Our findings demonstrate the successful implementation of the control mechanism across multiple input devices, thereby highlighting its versatility and broad applicability. As such, the proposed control mechanism extends beyond a standalone solution and offers a significant enhancement over current best practices.

Given that all study participants were representative of the target group, their quantitative feedback was particularly relevant and valuable. While some users initially encountered challenges with the system or found their assigned input device to be unfamiliar, all participants expressed confidence in being able to master the control with more time and practice. Notably, participants reported experiencing satisfaction in engaging with the presented control.

Overall, while our study does not conclusively show that adaptive control is straightforward to learn or intuitive, it does propose that the control method is indeed readily learnable within a short time frame, adaptable across different devices, and highly promising from an end-user perspective.

Acknowledgment

Our study was approved by the Ethics Committee of the *Faculty of Business Administration and Economics of the University of Duisburg-Essen*.

References

[1] World Health Organization (WHO), "Disability & health report," 7 March 2023. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/disability-and-health

[2] B. Martinsen, I. Harder, and F. Biering-Sorensen, "The meaning of assisted feeding for people living with spinal cord injury: a phenomenological study," *Journal of Advanced Nursing*, vol. 62, no. 5, pp. 533–540, May 2008. [Online]. Available: https://doi.org/10.1111/j.1365-2648.2008.04637.x

[3] M. Pascher, A. Baumeister, S. Schneegass, B. Klein, and J. Gerken, "Recommendations for the development of a robotic drinking and eating aid - an ethnographic study," in *Human-Computer Interaction – INTERACT 2021*, C. Ardito, R. Lanzilotti, A. Malizia, H. Petrie, A. Piccinno, G. Desolda, and K. Inkpen, Eds. Springer, Cham, 2021. [Online]. Available: https://doi.org/10.1007/978-3-030-85623-6_21

[4] M. Kyrarini, F. Lygerakis, A. Rajavenkatanarayanan, C. Sevastopoulos, H. R. Nambiappan, K. K. Chaitanya, A. R. Babu, J. Mathew, and F. Makedon, "A survey of robots in healthcare," *Technologies*, vol. 9, no. 1, p. 8, 2021.

[5] A. Pollak, M. Paliga, M. M. Pulopulos, B. Kozusznik, and M. W. Kozusznik, "Stress in manual and autonomous modes of collaboration with a cobot," *Computers in Human Behavior*, vol. 112, p. 106469, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0747563220302211

[6] D.-J. Kim, R. Hazlett-Knudsen, H. Culver-Godfrey, G. Rucks, T. Cunningham, D. Portee, J. Bricout, Z. Wang, and A. Behal, "How Autonomy Impacts Performance and Satisfaction: Results From a Study With Spinal Cord Injured Subjects Using an Assistive Robot," *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 42, no. 1, pp. 2–14, 2012.

[7] S. Bragança, E. Costa, I. Castellucci, and P. M. Arezes, *A Brief Overview of the Use of Collaborative Robots in Industry 4.0: Human Role and Safety*. Cham: Springer International Publishing, 2019, pp. 641–650. [Online]. Available: https://doi.org/10.1007/978-3-030-14730-3_68

[8] C. Fattal, V. Leynaert, I. Laffont, A. Baillet, M. Enjalbert, and C. Leroux, "SAM, an Assistive Robotic Device Dedicated to Helping Persons with Quadriplegia: Usability Study," *International Journal of Social Robotics*, vol. 11, no. 1, pp. 89–103, 2019.

[9] V. Maheu, P. S. Archambault, J. Frappier, and F. Routhier, "Evaluation of the jaco robotic arm: Clinico-economic study for powered wheelchair users with upper-extremity disabilities," in *2011 IEEE International Conference on Rehabilitation Robotics*, 2011, pp. 1–5.

[10] F. G. Pratticò and F. Lamberti, "Towards the adoption of virtual reality training systems for the self-tuition of industrial robot operators: A case study at kuka," *Computers in Industry*, vol. 129, p. 103446, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0166361521000531

[11] L. V. Herlant, R. M. Holladay, and S. S. Srinivasa, "Assistive teleoperation of robot arms via automatic time-optimal mode switching," in *The Eleventh ACM/IEEE International Conference on Human Robot Interaction*, ser. HRI '16. IEEE Press, 2016, p. 35–42.

[12] F. F. Goldau and U. Frese, "Learning to map degrees of freedom for assistive user control: Towards an adaptive dof-mapping control for assistive robots," in *The 14th PErvasive Technologies Related to Assistive Environments Conference*, ser. PETRA 2021. New York, NY, USA: Association for Computing Machinery, 2021, p. 132–139. [Online]. Available: https://doi.org/10.1145/3453892.3453895

[13] M. Pascher, F. F. Goldau, K. Kronhardt, U. Frese, and J. Gerken, "AdaptiX – A Transitional XR Framework for Development and Evaluation of Shared Control Applications in Assistive Robotics," *Proc. ACM Hum.-Comput. Interact.*, vol. 8, no. EICS, 2024, priprint on arXiv: https://arxiv.org/abs/2310.15887.

[14] M. Pascher, K. Kronhardt, F. F. Goldau, U. Frese, and J. Gerken, "In Time and Space: Towards Usable Adaptive Control for Assistive Robotic Arms," in *2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. Institute of Electrical and Electronics Engineers (IEEE), 2023, pp. 2300–2307.

[15] K. Kronhardt, S. Rübner, M. Pascher, F. Goldau, U. Frese, and J. Gerken, "Adapt or perish? exploring the effectiveness of adaptive dof control interaction methods for assistive robot arms," *Technologies*, vol. 10, no. 1, 2022.

[16] M. Pascher, K. Zinta, and J. Gerken, "Exploring of Discrete and Continuous Input Control for AI-enhanced Assistive Robotic Arms," in *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '24 Companion. New York, NY, USA: Association for Computing Machinery (ACM), 2024. [Online]. Available: https://doi.org/10.1145/3610978.3640626

[17] F. Goldau and U. Frese, "Probabilistic combination of heuristic behaviors for shared assistive robot control," in *The PErvasive Technologies Related to Assistive Environments (PETRA) conference (PETRA '24)*. New York, NY, USA: Association for Computing Machinery, 2024, p. 465–467. [Online]. Available: https://doi.org/10.1145/3652037.3652071

[18] R. Bemelmans, G. J. Gelderblom, P. Jonker, and L. de Witte, "Socially assistive robots in elderly care: a systematic review into effects and effectiveness," *Journal of the American Medical Directors Association*, vol. 13, no. 2, pp. 114–120, 2012.

[19] G. Canal, G. Alenyà, and C. Torras, "Personalization Framework for Adaptive Robotic Feeding Assistance," in *Social Robotics*. Cham: Springer International Publishing, 2016, vol. 9979, pp. 22–31.

[20] K. Hashimoto, F. Saito, T. Yamamoto, and K. Ikeda, "A Field Study of the Human Support Robot in the Home Environment," in *2013 IEEE Workshop on Advanced Robotics and Its Social Impacts (ARSO)*. Piscataway, NJ: IEEE, 2013, pp. 143–150.

[21] T. L. Chen, M. Ciocarlie, S. Cousins, P. M. Grice, K. Hawkins, K. Hsiao, C. C. Kemp, C.-H. King, D. A. Lazewatsky, A. E. Leeper, H. Nguyen, A. Paepcke, C. Pantofaru, W. D. Smart, and L. Takayama, "Robots for humanity: using assistive robotics to empower people with disabilities," *IEEE Robotics & Automation Magazine*, vol. 20, no. 1, pp. 30–39, 2013.

[22] L. Petrich, J. Jin, M. Dehghan, and M. Jagersand, "A quantitative analysis of activities of daily living: Insights into improving functional independence with assistive robotics," in *2022 International Conference on Robotics and Automation (ICRA)*, 2022, pp. 6999–7006.

[23] C.-S. Chung, H. Wang, and R. A. Cooper, "Functional assessment and performance evaluation for assistive robotic manipulators: Literature review," *The Journal of Spinal Cord Medicine*, vol. 36, no. 4, pp. 273–289, 2013.

[24] D. Gallenberger, T. Bhattacharjee, Y. Kim, and S. S. Srinivasa, "Transfer depends on acquisition: Analyzing manipulation strategies for robotic feeding," in *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2019, pp. 267–276.

[25] D. A. Rozas Llontop, J. Cornejo, R. Palomares, and J. A. Cornejo-Aguilar, "Mechatronics design and simulation of anthropomorphic robotic arm mounted on wheelchair for supporting patients with spastic cerebral palsy," in *2020 IEEE International Conference on Engineering Veracruz (ICEV)*, 2020, pp. 1–5.

[26] D. Park, Y. Hoshi, H. P. Mahajan, H. K. Kim, Z. Erickson, W. A. Rogers, and C. C. Kemp, "Active robot-assisted feeding with a general-purpose mobile manipulator: Design, evaluation, and lessons learned," *Robotics and Autonomous Systems*, vol. 124, p. 103344, Feb 2020.

[27] S. Drolshagen, M. Pfingsthorn, P. Gliesche, and A. Hein, "Acceptance of industrial collaborative robots by people with disabilities in sheltered workshops," *Frontiers in Robotics and AI*, p. 173, 2021.

[28] S. Federici, *Assistive technology assessment handbook*. Boca Raton, FL: CRC Press, Taylor & Francis Group, 2018.

[29] J. P. Boada, B. R. Maestre, and C. T. Genís, "The ethical issues of social assistive robotics: A critical literature review," *Technology in Society*, vol. 67, p. 101726, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0160791X21002013

[30] M. Pascher, U. Gruenefeld, S. Schneegass, and J. Gerken, "How to communicate robot motion intent: A scoping review," in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems - CHI '23*, 2023. [Online]. Available: https://doi.org/10.1145/3544548.3580857

[31] M. Beaudoin, J. Lettre, F. Routhier, P. S. Archambault, M. Lemay, and I. Gélinas, "Long-term use of the JACO robotic arm: a case series," *Disability and Rehabilitation: Assistive Technology*, vol. 14, no. 3, pp. 267–275, 1 2019.

[32] K. M. Tsui, D.-J. Kim, A. Behal, D. Kontak, and H. A. Yanco, ""i want that": Human-in-the-loop control of a wheelchair-mounted robotic arm," *Applied Bionics and Biomechanics*, vol. 8, no. 1, pp. 127–147, 2011.

[33] J. Złotowski, K. Yogeeswaran, and C. Bartneck, "Can we control it? Autonomous robots threaten human identity, uniqueness, safety, and resources," *International Journal of Human-Computer Studies*, vol. 100, pp. 48–54, 2017.

[34] J. Sijs, F. Liefhebber, and G. W. R. Romer, "Combined position & force control for a robotic manipulator," in *2007 IEEE 10th International Conference on Rehabilitation Robotics*. Institute of Electrical and Electronics Engineers (IEEE), 6 2007. [Online]. Available: https://doi.org/10.1109/icorr.2007.4428414

[35] K. Kallinen, "The effects of transparency and task type on trust, stress, quality of work, and co-worker preference during human-autonomous system collaborative work," in *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '17. New York, NY, USA: Association for Computing Machinery (ACM), 2017, p. 153–154. [Online]. Available: https://doi.org/10.1145/3029798.3038386

[36] D. A. Abbink, T. Carlson, M. Mulder, J. C. F. de Winter, F. Aminravan, T. L. Gibo, and E. R. Boer, "A topology of shared control systems—finding common ground in diversity," *IEEE Transactions on Human-Machine Systems*, vol. 48, no. 5, pp. 509–525, Oct. 2018. [Online]. Available: https://doi.org/10.1109/thms.2018.2791570

[37] S. Reddy, A. Dragan, and S. Levine, "Shared autonomy via deep reinforcement learning," in *Robotics: Science and Systems XIV*, ser. RSS2018. Robotics: Science and Systems Foundation, 6 2018. [Online]. Available: http://dx.doi.org/10.15607/RSS.2018.XIV.005

[38] D. Gopinath, S. Jain, and B. D. Argall, "Human-in-the-loop optimization of shared autonomy in assistive robotics," *IEEE Robotics and Automation Letters*, vol. 2, no. 1, pp. 247–254, 2017.

[39] S. Udupa, V. R. Kamat, and C. C. Menassa, "Shared autonomy in assistive mobile robots: a review," *Disability and Rehabilitation: Assistive Technology*, vol. 18, no. 6, p. 827–848, 6 2021. [Online]. Available: http://dx.doi.org/10.1080/17483107.2021.1928778

[40] S. G. Hart, "Nasa-Task Load Index (NASA-TLX); 20 Years Later," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 50, no. 9, pp. 904–908, 2006.

[41] Kinova inc., "Jaco assistive robot - user guide," Published online (last visited on June 7, 2024), 2021, eN-UG-007-r05-202111. [Online]. Available: https://assistive.kinovarobotics.com/uploads/EN-UG-007-Jaco-user-guide-R05.pdf

[42] Intel Corporation, "Intel realsense - product family d400 series," Published online (last visited on 13.02.2023), 11 2022, revision 014. [Online]. Available: https://www.intelrealsense.com/wp-content/uploads/2022/11/Intel-RealSense-D400-Series-Datasheet-November-2022.pdf

[43] Google, "Glass enterprise edition 2," 2019.

[44] H. Penkert, J. C. Baron, K. Madaus, W. Huber, and A. Berthele, "Assessment of a novel, smartglass-based control device for electrically powered wheelchairs," *Disability and Rehabilitation: Assistive Technology*, vol. 16, no. 2, pp. 172–176, 2021. [Online]. Available: https://doi.org/10.1080/17483107.2019.1646817

[45] Microsoft, "Xbox adaptive controller," Online: https://www.microsoft.com/en-us/d/xbox-adaptive-controller/8nsdbhz1n3d8 (last visited on June 7, 2024), 2024.

[46] Stanford Artificial Intelligence Laboratory et al., "Robotic operating system." [Online]. Available: https://www.ros.org

[47] Mikepanhu. (2014) Google glass with frame.jpg. Wikimedia Commons. Adapted under CC BY-SA 3.0 license. [Online]. Available: https://upload.wikimedia.org/wikipedia/commons/b/be/Google_Glass_with_frame.jpg

Autonomous Multi-Sensory Robotic Assistant for a Drinking Task

Felix Ferdinand Goldau
Institute of Automation
University of Bremen
Bremen, Germany
fgoldau@uni-bremen.de

Tejas Kumar Shastha
Institute of Automation
University of Bremen
Bremen, Germany
tejas@uni-bremen.de

Maria Kyrarini
Institute of Automation
University of Bremen
Bremen, Germany
mkyrar@iat.uni-bremen.de

Axel Gräser
Institute of Automation
University of Bremen
Bremen, Germany
ag@iat.uni-bremen.de

***Abstract*—Assistive robots have the potential to support people with disabilities in their Activities of Daily Life. The drinking task has a high priority and requires constant assistance by caregivers to be executed regularly. Due to incapacitating disabilities such as tetraplegia, which is the paralysis of all limbs, affected people cannot use classic control interfaces such as joysticks. This paper presents a robotic solution to enable independent, straw-less drinking using a smart cup and no physically attached elements on the user. The system's hardware and software components are presented and the overarching control scheme described. The cup approaches the mouth utilising a user-friendly and vision-based robot control based on head pose estimation. Once contact has been established, the user can drink by tilting the cup with a force sensor-based control setup. Two experimental studies have been conducted, where the participants (mostly able-bodied and one tetraplegic), could separately experience the cup's contactless approach and the contact-based sequence. First results show a high user acceptance rate and consistent positive feedback. The evaluation of internal data showed a high reliability of the safety-critical components with the test groups perceiving the system as intuitive and easy to use.**

Keywords—assistive robots, human-robot interaction, force control, head pose estimation

I. Introduction

People suffering from a severe disability, like tetraplegia, have difficulties performing Activities of Daily Living (ADLs). Tetraplegia is the paralysis of four limbs, thus limiting voluntary motor function of everything below the neck and, even though the treatment of paralysis has undergone great progress, people require the assistance of a caregiver to perform ADLs [1].

Assistive robotic manipulators have the potential to support individuals with tetraplegia to regain some of their independency in performing ADLs. One example is the wheelchair-mounted robotic manipulator FRIEND, which was used as a personal assistant for a tetraplegic end-user in performing ADLs and tasks in a working environment [2, 3]. A survey with potential end-users of robotic manipulators shows that drinking and eating are highly prioritized tasks [4].

This paper presents a robotic solution as an attempt to enable a person with tetraplegia to independently perform a drinking task using a cup, without external human aide and without any physically user-attached elements. The drinking task is executed without a straw or similar device and must therefore be accomplished with contact between the user and the robot-handled cup. The goal to be achieved, is to give the user a feeling of sovereignty over their own drinking and the perception of performing the drinking themselves, as opposed to being served a drink. First results and user feedback based on two small studies are presented.

The main contribution of the presented work is a consistent, user-friendly, and fully flexible concept of control with the human in the loop as opposed to a system using predefined positions. This allows a safe and intuitive human-robot interaction. The drinking is fully personalised with the user gaining control and comfort using natural head movements for the delivery of the cup to the mouth, and achieving adaptive control of the drinking process using cup-mounted force sensors.

The paper is organized as follows: In section II, related work is reviewed. Section III describes the proposed framework and Section IV presents the experimental results and the evaluation. Finally, Section V discusses the conclusion with directions for future work.

II. Related Work

An important topic in assistive robotics is Human-Machine Interfaces (HMIs), as users are often restricted in their movements and standard computer interfaces might not be usable. Based on a study using a vision-guided robot arm, [5] shows that many systems are too complicated for their respective end-users who have to follow tedious HMI-sequences. It also shows that systems should adopt higher velocities when not close to the user to minimise waiting times and user frustration.

Various systems have been developed within the field of assistive robotics, most of them based on a Wheelchair Mounted Robot Arm (WMRA). For example, the FRIEND systems I-IV [2] where FRIEND IV was capable of enabling a tetraplegic person to work as a librarian. The used HMI consists of a set of buttons operated by the user with head motions, a chin joystick, and a monitor for system feedback. This setup can be seen as a head operated computer mouse which allows the user to move and click with a cursor on the monitor to control the wheelchair and WMRA. Previous versions also addressed the drinking task, but were for safety reasons restricted to drinking with a straw [3].

Another approach to assistive drinking shown by [6] uses a Brain-Machine Interface (BMI) to control a robotic manipulator holding a cup for drinking without a straw. An externally mounted RGB-D camera is used to estimate the user's mouth by applying a golden-ratio approach on the detected face, thus bypassing issues of face occlusion by the cup and the robot. The user controls the scenario by giving GO-signals via the BMI and can thereby accelerate or stop the procedure in seven discrete predefined steps.

Moreover, hands-free human-robot interfaces could be used to assist individuals suffering from tetraplegia. Two interfaces using eye gestures [7] and Brain-Computer Interface (BCI) [8] have been developed to allow users with tetraplegia to control a 7 Degrees-of-Freedom (DoF) robotic arm and its gripper. A manipulation task was selected to evaluate both interfaces for the robot control. For research purposes the robotic degree of automation was set to zero, thus the user had to control the entire process step by step. It was shown that this kind of robot control burdens the user with high cognitive load due to the fast response needed during the drinking process.

III. Proposed Method

This section proposes a solution to the drinking task with contact between the user and the cup, maintaining user safety as an active element inside the loop and without using predefined steps. The drinking procedure is divided in two consecutive sequences which are individually described below:

- The 'Vision-based Robot Control to Serve a Drink' handles the delivery of a grasped cup to the user's mouth based on camera input, whereas
- the 'Robot Force Control for the Drinking Process' deals with the process of tilting the cup based on the force applied onto the cup by the user, thus enabling drinking.

A. Hardware Setup

The Kinova Jaco 2 [9] 7-DoF spherical ultra-lightweight robotic arm with a three-finger gripper attached as the end-effector (Fig. 1a) is used as the main assistive robotic manipulator. The Jaco 2 has been specifically designed for use as an assistive robot and has been thoroughly tested in a scientific context [10].

The robot arm grasps a smart cup (Fig. 1b), developed in previous work [11], consisting of a feeding cup with a beak, two force sensors, and a Bluetooth module which wirelessly transmits the force values to the operating computer. The force sensors are attached to the beak of the cup, just above and below the mouth piece. In the context of this work, the beak of the mouth piece is considered as the cup's origin.

a b

Fig. 1. The hardware components a) assistive robotic manipulator with vision sensor, b) smart cup with force sensors [11]

An Intel Realsense Depth Camera D435 [12] is selected as the vision sensor and is mounted on the robot's end-effector between the gripper and the last joint using a smooth 3D-printed attachment clamp. The vision sensor is USB powered and provides RGB-D (Red Green Blue - Depth) information of the scene ahead, with an RGB resolution of up to 1920x1080 pixels and a depth sensor range between 0.105 m and 10 m. The short distance between the camera and the robot's end-effector minimises the occlusion of the scene by the manipulator, as only the smart cup and the gripper's fingertips can be seen obstructing the scene. For this application, the camera is configured to allow minimal depth measurements, thus being able to detect the distances of objects right up to the tip of the cup.

A UNIX computer is used to combine all systems, interpret the sensor data and control the robot manipulator accordingly. The proposed system is developed on Robot Operatic System (ROS) [13].

B. Vision-based Robot Control to Serve a Drink

The concept of vision-based robot control, which is used to serve a drink, is shown in Fig. 2. The main idea is that the robot delivers the smart cup to the user's mouth using data gathered by the vision sensor. The information from the vision sensor is processed by the following modules: Vision-based User Face Detection and Tracking, Head Pose Estimation, and Mouth Pose Estimation. The task control module calculates the necessary robot action based on the mouth pose estimation and the robot pose, and controls the robot accordingly. Furthermore, the task control module uses the force data from the smart cup to ensure safe human-robot interaction. The modules in Fig. 2 are explained in detail as follows.

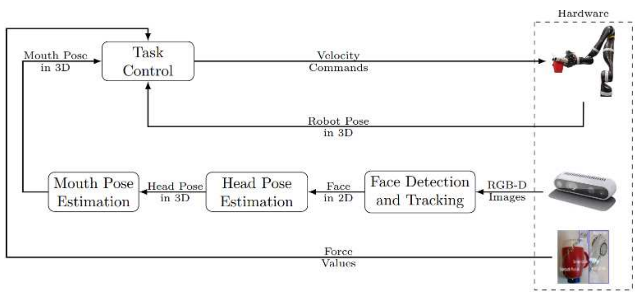


Fig. 2. Vision-based Robot Control to Serve a Drink

1. Face Detection and Tracking

The frontal-face detector of the dlib open source library [14] is used to search the 2D RGB-data of the vision sensor (camera) for a human face in order to detect the user. This detector is based on linear Support Vector Machine (SVM) using features of Histograms of Oriented Gradients (HOG) and has been trained on images of frontal faces from the Labeled Faces in the Wild database [15]. The output of this module is the 2D bounding box of the user's face.

In case the face is not detected, but information about a recent valid face detection from a previous iteration is available, a Discriminative Correlation Filter with Channel and Spatial Reliability (CSR-DCF) tracker [16] is used to estimate the bounding box of the face. This tracker compares the frequency domain of the relevant area and compares it to its surroundings to determine an object's movement in a sequence of frames.

2. Head Pose Estimation

Based on the detected user's face, a predefined 3D model of a human face is aligned with the data by applying an active shape model [17] based approach. The model is fitted by applying a cascade of regression trees, which have been previously trained on the 300-W dataset with 68 landmarks per face [18].

Using the landmark knowledge of the predefined 3D model, a PnP solution is calculated based on the RANSAC paradigm [19]. The result is a 3D position and orientation for the camera reference system (${}^{H}_{C}T$) in relation to the landmark coordinate system with the mouth at its origin. The head pose in relation to the world reference frame ${}^{W}_{H}T$ can be calculated by applying (1), with ${}^{W}_{E}T$ as the transformation of the robot's end-effector reference to the world reference system as result of the direct kinematics, and ${}^{E}_{C}T$ as the transformation of the camera reference system to the robot's end-effector system as defined by the rigid connection of the attachment clamp.

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

The final head pose is validated twice to ensure user safety. This is done by partially reversing the previous procedure, thus reprojecting the face landmarks of the final pose onto the image plane of the camera. This projection outputs new 2D landmarks which are compared to the equivalent landmark projection of the fitted active model. If the error is too large, a misdetection is assumed and the pose is omitted. The second validation is performed in case information about the user's pose is available from previous iterations, by comparing this prior information to the most recently calculated user pose. If either the orientation or the position shows major changes, the chances of an error are raised and the head pose is again omitted for a higher user safety.

Even though not intended by the user, it is possible that they move their head too much during two iterations, thus causing the second validation step to fail. This behaviour is intended for an increased user safety, as the situation is considered dangerous if too much user movement occurs and the possibility of false detections rises.

3. Mouth Pose Estimation

After the head pose has been estimated successfully, the pose ${}^{W}_{M}T$ of the mouth with respect to the world reference frame can be directly derived from the result as shown in (2), because both poses share the same reference system and the user's mouth as their origin.

$$ {}^{W}_{H}T = {}^{W}_{M}T \qquad (2) $$

When the robot is advanced close to the user (distance between the beak and the mouth is approx. 5cm), the head pose cannot be estimated anymore. Therefore, another method to estimate the mouth pose is developed in this work. The proposed mouth pose estimation calculates a result based on information gathered in previous iterations and predefined knowledge of the underlying path control logic, by comparing the previously tracked position of the mouth with an updated prediction. Two assumptions are considered:

1. If they want to drink, the user does not move, and
2. the robot moves the smart cup in a direct path towards the mouth.

The first assumption is only a minor constraint on the user's behalf as they would also refrain from moving if another person served them instead of a robot; and the second assumption is system-defined. If the user were allowed to move and this assumption were not made, the user's mouth position would be undefined until the transformation is available again. This would result in a safety-critical situation.

Using those two assumptions, an estimation of the mouth's position with respect to the camera can be calculated. Initially, the position obtained by previous head pose estimations is used and adjusted afterwards, while the position change of the camera is known by applying the robot's direct kinematics. The vector between this new camera pose and the previous head pose is then scaled, such that its length corresponds to a selection of distance values measured by the vision sensor.

The calculated position of the mouth is reprojected into the image plane. From the projected landmark points, a bounding box of the mouth-and-nose-region is calculated. In the meantime, a CSR-DCF tracker [16] trained during previous head pose estimations tracks the same region based entirely on RGB information and also introduces a bounding box. The resulting rectangles of both bounding boxes are compared with respect to their relative overlapping area. If the area exceeds a predefined threshold, it is assumed that the correct position of the mouth is known, and thus the position of the mouth is accepted.

The result of the head and mouth pose estimation is shown in Fig. 3 with the 3D-landmarks as red dots and the resulting pose as a color-coded coordinate system (red = x-axis, green = y-axis, blue = z-axis) at the user's mouth. A white rectangle represents the area used to calculate the safety-critical user distance.

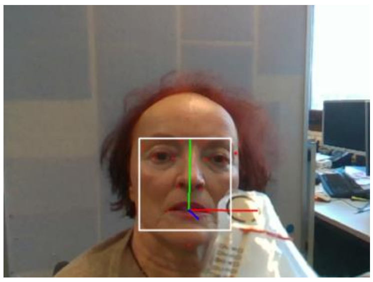


Fig. 3. Result of the Head and Mouth Pose Estimation with 3D-landmarks (red dots) and coordinate system (red = x-axis, green = y-axis, blue = z-axis)

In the event that the threshold is not met, it is implied that either the user has moved or another error has occurred. In either case, reliable information cannot be assumed and the calculated mouth pose is discarded. This also cancels future mouth pose estimations until a new head pose has been identified and the tracker is redefined.

4. Task Control

As defined by the control concept, the robot control is based on head pose information. The task control module is designed based on an analogy of a human assistant who serves a cup to the user and reacts to the user's head orientation. This results in the cup being brought to the user's mouth only if the head is oriented towards the cup. If this orientation is not given, the robot motion is stopped and, after a small delay, reversed towards a predefined home position of the robot.

When the head of the user is oriented towards the cup and the cup is within a predefined field of view with angle θ (in this work $\theta = 15°$) in either direction of the user's central z-axis, the robot moves the cup on a smooth and parabolic trajectory towards the mouth. The goal of this process can be described by (3) with ${}^{W}_{B}T$ as the pose of the cup's beak in relation to the world reference frame.

$$ {}^{W}_{M}T = {}^{W}_{B}T \qquad (3) $$

The pose of the beak ${}^{W}_{B}T$ can be calculated using the direct kinematics of the manipulator and structural information about the grasped cup as shown in (4).

$$ {}^{W}_{B}T = {}^{W}_{E}T \cdot {}^{E}_{B}T \qquad (4) $$

Fig. 4 shows an example path (purple) for the end-effector during the cup's approach towards a user with the respective coordinate systems of the user and the end-effector (red = x-axis, green = y-axis, blue = z-axis), and the θ-based field of view. Following a parabolic path, the cup initially converges towards the z-axis of the mouth before closing the distance. Once the cup has reached a distance of less than 5 cm to the mouth, the path is no longer defined by parabolic curves but moves the cup on a straight path directly to the user's mouth. The task control module sends velocity commands to the robot, in order to control it along the path.

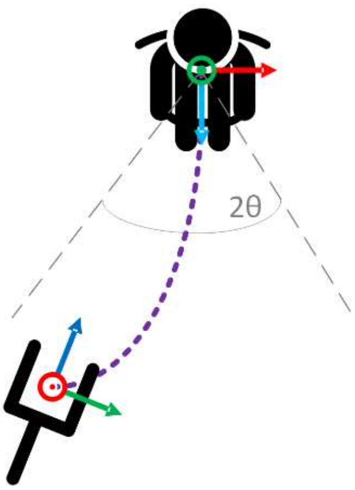


Fig. 4. Example path of the cup's approach

Moreover, in case the force sensors detect a contact at any point before the mouth is reached, the approach is stopped. If the contact has been initiated intentionally, the second sequence of the drinking process commences. (section III.C).

If, at any point, no head pose information is available, or if the user's orientation does not meet the requirements, the advance of the cup towards the user is stopped. If not updated within a short time period, the cup is retracted towards the home position, while the camera remains oriented towards the last known head position, thus enabling the user to regain control at any given point. The home pose is defined in a way that the user's head is seen by the camera in a standard scenario.

C. Robot Force Control for the Drinking Process

Once the smart cup, controlled by the vision-based robot control (Section III.B), gets close to the user's mouth, the force controller is initialised to enable the drinking process. The objective of this controller is to support the active drinking task wherein the user interacts with the smart cup by applying force to the force sensors (shown in Fig. 1b). Fig. 5 presents an overview of the force control for the drinking process. The inputs of this controller are the values read by the two force sensors on the cup and the current pose of the robot. The output is the Cartesian velocity command supplied to the robot. The controller drives the cup along a vertical plane which runs along the nose of the user, perpendicular to the face, to emulate a natural drinking motion.

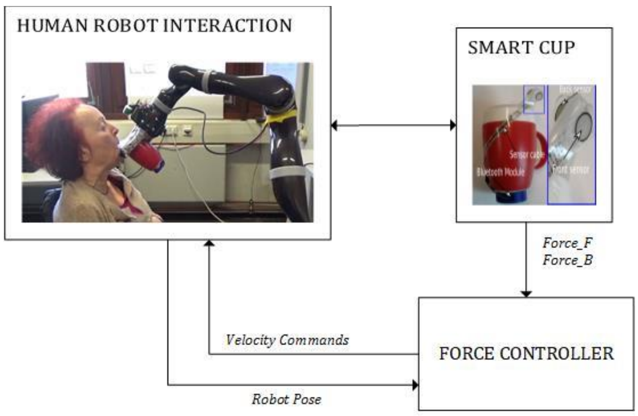


Fig. 5. Overview of force control for the drinking process

When the user applies force on the back sensor (Force_B), the cup's beak rotates down by a discrete angle (i.e.: feed motion). When the user applies force on the front sensor (Force_F), the cup's beak rotates up by a discrete angle (i.e.: non-feed motion). When forces are applied on both sensors simultaneously, the cup initiates a fall-back motion, moving laterally away from the user and stopping at the home position. The drinking task is completed.

To achieve this behaviour, three threshold values are defined for each of the force sensors:

- Trigger threshold - 0.5 N - Minimum amount of force to be applied to either sensors to trigger the respective action. Allows for slight pressure exerted when user swallows water.

- Fall-back threshold – 1.0 N – Simultaneous force required on both sensors to initiate fall-back.
- Safety threshold - 2.5 N - Maximum amount of force allowed to be exerted before an emergency halt is issued to the robot, stopping all motion.

This control schema is initially build based on intuition and available hardware. The next section explains how it was tested and proven to be effective.

IV. Experimental Results and Evaluation

Two small studies were conducted: the first to evaluate the safety and usability of vision-based robot control and the second to evaluate the general usability of the force-controlled feeding system.

At the end, each user was presented with a general feedback questionnaire based on a 5-point Likert scale. The questions are agreement based and the Likert scale ranges from 1 (“strongly disagree”) to 5 (“strongly agree”). All participants gave their informed signed consent to participate in this study.

A. Serving a Drink

The first study is designed to evaluate the safety and usability of the cup serving scenario with the proposed solution. 25 users, one of whom is tetraplegic, participated in the experiments. The average age was 31.08 ± 14.55 years with a gender distribution of 13 males and 12 females (11 able-bodied and one tetraplegic). The experiments were conducted individually and independently for each subject.

Seven tasks were performed by each user. In each task, the smart cup is already grasped by the robot. The user is seated on a chair (or a wheelchair for the tetraplegic user) at a pre-defined pose relative to the robot. The concept of control described in section III.B.1 is applied in each task. The tasks were as follows:

0. The user freely tested the system for one minute to get familiar with it.

1. The user constantly oriented their head towards the robot’s gripper, thus effectively commanding the robot to bring the cup to their mouth on a direct path.

2. The user performed so-called abort actions: The robot advanced the cup towards the user, but the user aborted the task by turning their head away. When the robot detected such an abort action, it stopped and withdrew towards its home position. The user was encouraged to compare the abort actions as reactions to motions to the top, bottom, left, and right.

3-6. The last four scenarios were performed with different starting positions for the robot and with the user imagining a standard drinking application, but also allowing themselves to become distracted once in a while to include an abort action.

Fig. 6 shows the robot’s end-effector paths of five different users during scenario 1 with the home position in the top right and the users’ mouths on the left. Though every user sat on the same chair with a defined position relative to the robot, the pose of the mouth, and with it, the final position of the end-effector, differs vastly. This is due to a variety of reasons including different heights and head orientations and it shows the necessity of an adaptive system as proposed in this paper.

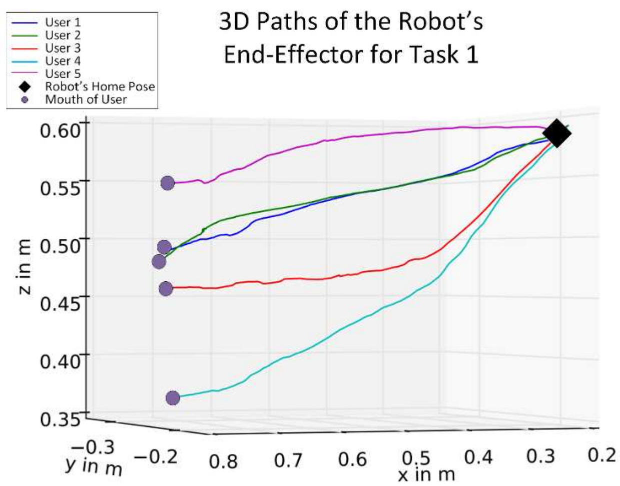


Fig. 6. 3D paths of the robot’s end-effector for 5 selected users in task 1

As the safety-critical elements during the cup serving sequence are mostly defined by the abort actions as well as the distance information obtained from the vision sensor and head pose estimation, these components are discussed in detail.

The outcome of abort actions during scenario 2 was manually sorted into the three categories: successful, delayed, and failed. Actions were considered delayed if the system did not react immediately and failed if an emergency stop was pressed or the user was forced to change the current abort motion in order to cause a reaction of the system. The users performed between 3 and 24 abort actions each, with 214 actions in total. Approximately 7 % of all actions were resolved with a delay and 4% of all actions failed, most of which occurred during upwards abort motions. The failing of upwards abort motions is caused by the fact that the cup already approached the user from above, which already requires an upwards tilting of the head. For the abort motion the users had to rotate their head very far upwards to use this specific abort action.

For redundancy purposes, the distance between the cup and the user is obtained using two separate methods. While one method uses the result of the previously mentioned head pose estimation, the other one is based on averaged values of the depth image in the region of the face. The robot control logic uses both methods and compares their values to lie within a defined offset for a single redundancy check. To evaluate the two methods during the experiments, the values of sequence 1 are compared to a reference value, which is calculated as the remaining distance of the robot’s end-effector to its final position. The results show an average error of 0.194 ± 0.038 m for the first method, and an average error of 0.142 ± 0.011 m for the second method respectively. The values show comparably high and constant systematic errors, which are mostly due to the offset of the final robot’s end-effector position to the mouth. This offset is a known factor and taken into account by the control setup. The remaining random errors are very small and considered as such in terms of safety critical actions.

The participants were able to test the system thoroughly in a realistic scenario during the final four sequences and reported their feedback on a questionnaire afterwards. The results of the Likert scale are shown in Fig. 7 as the mean values across subjects. The feedback is positive throughout with a rapid user familiarisation and high acceptability. The previously mentioned issues with the abort actions to the top are only slightly reflected by users' feedback. This is probably due to different home positions for the different tasks, thus not always approaching the user from above, but also from lower directions. The user feedback also shows a requirement to increase the velocity of the system.

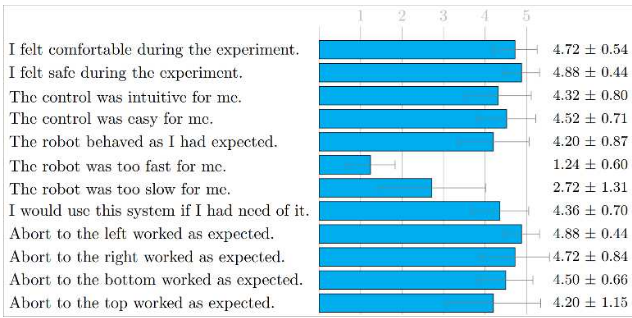


Fig. 7. Subjective user feedback for the first study

The feedback of the tetraplegic user aligns with the feedback of the other participants and is very positive in total, even verbally referring to the system as her "favourite one". As a very experienced user of assistive robotic systems, she did not express any concerns regarding safety or comfort, but instead listed the system as being rather too slow. Due to the user's restricted motion capabilities, abort motions to one side were not possible, but all others worked without any problems.

B. Drinking Process

The experiments for the drinking process were conducted with 16 participants (15 able-bodied and one tetraplegic). The average age was 26 ± 10 years. Out of 16 participants, seven (six able-bodied and one tetraplegic) were females. The users were allowed a brief acclimatization period followed by an attempt at one full successful run for the drinking process, terminating in a fall-back.

Fig. 8 shows the orientation of the end-effector in Euler angles (Fig. 8a) and the concurrent force sensor readings (Fig. 8b) during one of the trials of the drinking process. These values were collected as output values of the manipulator for the experiments of each user. Force_B (red curve) crossing the trigger line (yellow line) causes an increment in the roll angle (feed motion) and Force_F (blue curve) crossing the trigger line causes decrement in the roll angle (non-feed motion). Both sensors crossing the trigger line causes the smooth decrement in the roll angle, which is the fall-back. The initial increment in the roll angle seen before any force input is a pre-defined rotation for user convenience.

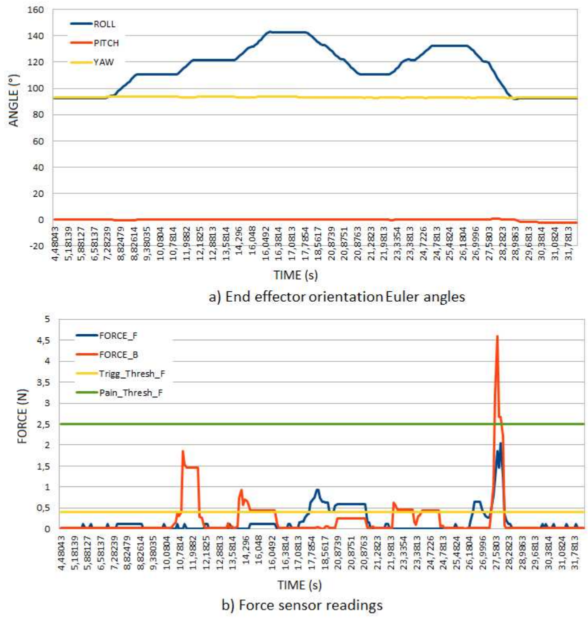


Fig. 8. Plots of feeding experiment of one of the trials

Since this is the second trial for the same user, the user is already familiar with the system, as evidenced by the fact that there are no overshoots above the pain threshold and that the user is able to manipulate the cup with only the minimum amount of force required.

Some of the feedback provided by the subjects is summarized in Fig. 9 as the mean feedback across the subjects. It can be safely concluded from the analyses of Fig. 8 and Fig. 9 that the force controller designed for the drinking process performs satisfactorily well. With regards to the drinking process, in alignment with the rest of the subjects, the tetraplegic user felt that the system was indeed comfortable and intuitive to use. She did have some critical feedback to provide, which forms the basis for some opportunities of further development:

- She preferred to use a normal cup without a beak
- She felt the need for a user-controlled emergency stop signal, such as a voice or eyes activated trigger.

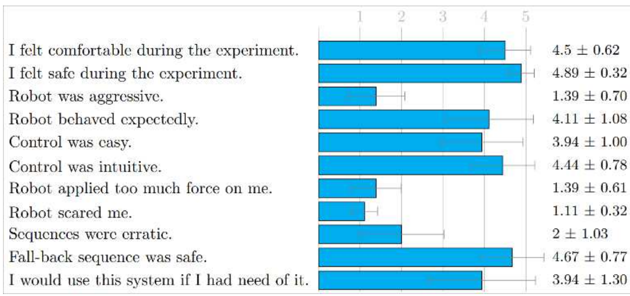


Fig. 9. User feedback

V. Conclusion and Future Work

The proposed solution functions as a user-friendly and safe concept of control for an assistive drinking task with the user in the loop. The mechanical setup requires no physical attachment on the user or in close proximity to them, thus allowing a flexible out-of-the-box usage with no preparations on the user's side.

The integration of all subsystems has been successful and first results could be obtained with potential users. The experiments show a high reliability of the safety-critical systems and quick responses on fall-back commands. The system has a high user acceptance and consistent positive feedback with an easy and intuitively perceived control scheme.

Future work will focus on replacing the current smart cup with a standard cup without the beak, as requested by the tetraplegic user, and increasing the overall velocity of the system in situations that are not safety-critical. To implement a system capable of executing the whole drinking operation, the autonomous filling and grasping of the cup will be necessary, as well as defining a comfortable and safe home position for the robot. The complete system will be tested within a larger user study with a higher ratio of potential end-users.

ACKNOWLEDGMENT

The research was supported by the German Federal Ministry of Education and Research (BMBF) as part of the project MobiLe (Physical Human-Robot-Interaction for Independent Living). The authors would like to thank all the persons involved, in particular the tetraplegic user, Mrs. Lena Kredel, who took part in both studies.

REFERENCES

[1] W. Strubreither, M. Neikes, D. Stirnimann, J. Eisenhuth, B. Schulz und P. Lude, Klinische Psychologie bei Querschnittlähmung, Vienna: Springer, 2015.

[2] A. Graser, T. Heyer, L. Fotoohi, U. Lange, H. Kampe, B. Enjarini, S. Heyer, C. Fragkopoulos und D. Ristic-Durrant, „A supportive friend at work: Robotic workplace assistance for the disabled,“ *IEEE Robotics & Automation Magazine 20.4,* pp. 148-159, 2013.

[3] I. Volosyak, O. Ivlev and A. Graser, "Rehabilitation robot FRIEND II-the general concept and current implementation," *9th International Conference on Rehabilitation Robotics (ICORR),* pp. 540-544, 2005.

[4] C.-S. Chung, H. Wang und R. A. Cooper, „Functional assessment and performance evaluation for assistive robotic manipulators: Literature review,“ *The journal of spinal cord medicine. Vol 36. No. 4,* pp. 273-289, 2013.

[5] D.-J. Kim, R. Lovelett und A. Behal, „An empirical study with simulated adl tasks using a vision-guided assistive robot arm,“ in *IEEE International Conference on Rehabilitation Robotics „(ICORR)*, 2009.

[6] S. Schröer, I. Killmann, B. Frank, M. Völker, L. Fiederer, T. Ball und W. Burgard, „An autonomous robotic assistant for drinking,“ *IEEE International Conference on Robotics and Automation (ICRA),* pp. 6482-6487, 2015.

[7] S. Alsharif, O. Kuzmicheva und A. Gräser, „Gaze Gesture-Based Human Robot Interface,“ *Technische Untersützungssysteme, die die Menschen wirklich wollen,* p. 339, 2016.

[8] B. Mindermann, *Untersuchung eines hybriden Brain-Computer Interfaces (BCIs) zur optimalen Auslegung als Mensch-Maschine-Schnittstelle,* University of Bremen, 2018.

[9] "KINOVA Gen2 Ultra lightweight robot," [Online]. Available: https://www.kinovarobotics.com/en/products/robotic-arms/kinova-gen2-ultra-lightweight-robot. [Accessed 1 Feb 2019].

[10] V. Maheu, P. S. Archambault, J. Frappier und F. Routhier, „Evaluation of the JACO robotic arm: Clinico-economic study for powered wheelchair users with upper-extremity disabilities,“ *IEEE International Conference on Rehabilitation Robotics (ICORR),* pp. 1-5, 2011.

[11] T. L. Koller, M. Kyrarini und A. Gräser, „Towards robotic drinking assistance: Low cost multi-sensor system to limit forces in Human-Robot-Interaction [accepted],“ in *12th ACM International Conference on PErvasive Technologies Related to Assistive Environments (PETRA)*, Rhodos, Greece, 2019.

[12] Intel Corporation, "Intel RealSense Depth Camera D435," Intel, [Online]. Available: https://click.intel.com/intelr-realsensetm-depth-camera-d435.html. [Accessed 1 February 2019].

[13] M. Quigley, K. Conley, B. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler und A. Y. Ng, „ROS: an open-source Robot Operating System,“ in *ICRA workshop on open source software. Vol. 3, No. 3.2*, Kobe, Japan, 2009, p. 5.

[14] D. E. King, „Dlib-ml: A Machine Learning Toolkit,“ *Journal of Machine Learning Research,* Bd. 10, pp. 1755-1758, 2009.

[15] G. B. Huang, M. Ramesh, T. Berg und E. Learned-Miller, *Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments,* Amherst: University of Massachusetts, 2007.

[16] A. Lukezic, T. Vojir, L. C. Zajc, J. Matas und M. Kristan, „Discriminative Correlation Filter With Channel and Spatial Reliability,“ *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* pp. 6309-6318 , 2017.

[17] T. F. Cootes, C. J. Taylor, D. H. Cooper und J. Graham, „Active shape models-their training and application,“ *Computer vision and image understanding,* Bd. 61, Nr. 1, pp. 38-59, 1995.

[18] C. Sagonas, E. Antonakos, G. Tzimiropoulos, S. Zafeiriou und M. Pantic, „300 faces in-the-wild challenge: Database and results,“ *Image and vision computing. Volume 47,* pp. 3-18, 2016.

[19] M. A. Fischler und R. C. Bolles, „Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography,“ *Commun. ACM Vol. 24 No. 6,* pp. 381-395, Juni 1981.

technologies

MDPI

Article

Adapt or Perish? Exploring the Effectiveness of Adaptive DoF Control Interaction Methods for Assistive Robot Arms

Kirill Kronhardt [1,*,†], Stephan Rübner [1,*,†], Max Pascher [1], Felix Ferdinand Goldau [2], Udo Frese [2] and Jens Gerken [1,*]

[1] Human-Computer Interaction Group, Department of Media Informatics and Communication, Westphalian University of Applied Sciences, 45897 Gelsenkirchen, Germany; max.pascher@w-hs.de

[2] German Research Center for Artificial Intelligence (DFKI) GmbH, 28359 Bremen, Germany; felix.goldau@dfki.de (F.F.G.); udo.frese@dfki.de (U.F.)

* Correspondence: kirill.kronhardt@studmail.w-hs.de (K.K.); stephan.ruebner@studmail.w-hs.de (S.R.); jens.gerken@w-hs.de (J.G.)

† These authors contributed equally to this work.

Abstract: Robot arms are one of many assistive technologies used by people with motor impairments. Assistive robot arms can allow people to perform activities of daily living (ADL) involving grasping and manipulating objects in their environment without the assistance of caregivers. Suitable input devices (e.g., joysticks) mostly have two Degrees of Freedom (DoF), while most assistive robot arms have six or more. This results in time-consuming and cognitively demanding mode switches to change the mapping of DoFs to control the robot. One option to decrease the difficulty of controlling a high-DoF assistive robot arm using a low-DoF input device is to assign different combinations of movement-DoFs to the device's input DoFs depending on the current situation (adaptive control). To explore this method of control, we designed two adaptive control methods for a realistic virtual 3D environment. We evaluated our methods against a commonly used non-adaptive control method that requires the user to switch controls manually. This was conducted in a simulated remote study that used Virtual Reality and involved 39 non-disabled participants. Our results show that the number of mode switches necessary to complete a simple pick-and-place task decreases significantly when using an adaptive control type. In contrast, the task completion time and workload stay the same. A thematic analysis of qualitative feedback of our participants suggests that a longer period of training could further improve the performance of adaptive control methods.

Keywords: assistive robotics; human–robot interaction (HRI); shared user control; augmented reality; virtual reality; visual cues

check for updates

Citation: Kronhardt, K.; Rübner, S.; Pascher, M.; Goldau, F.; Frese, U.; Gerken, J. Adapt or Perish? Exploring the Effectiveness of Adaptive DoF Control Interacion Methods for Assistive Robot Arms. *Technologies* **2022**, *10*, 30. https://doi.org/10.3390/technologies10010030

Academic Editor: Fillia Makedon

Received: 4 January 2022
Accepted: 10 February 2022
Published: 14 February 2022

1. Introduction

Robotic solutions are becoming increasingly prevalent in many areas of our professional and personal lives and have started to evolve into collaborators [1,2]. A non-negligible number of people live with motor impairments, ranging from slight limitations to severe paralysis [3]. While a near-complete integration into professional and social life is the final goal, current assistive robotic technologies focus on performing activities of daily living (ADLs). These include tasks ranging from essentials such as eating and drinking to more complex behaviors such as grooming and activities associated with leisure time [4].

A general problem with assistive robotic solutions is finding suitable methods and technologies for controlling such robots. Assistive robotic devices are often characterized as having a large number of Degrees of Freedom (high-DoF). For example, a robotic arm with a simple gripper can freely operate in 3D space and move along Cartesian space as well as yaw, pitch, and rotate. This typically results in five to seven DoFs. Standard input devices, such as joysticks, only cover two DoFs. To control a high-DoF device with a low-DoF input device, mode switching is used. This means that at any point in time, the user has to select a mode, which then maps the two DoFs of the input device to two of the total available DoFs of the robot while neglecting the others. While high-DoF input devices do exist, they are not often accessible for people with motor impairments.

Using a human–computer interface with a standard button-based mode switching setup, Herlant et al. discovered that more than one-sixth of the total execution time is spent changing the currently selected mode [5]. They showed that automatic mode switching leads to increased user satisfaction within a deterministic simulation environment and with a predefined goal.

Our latest research findings provide a proof-of-concept for a novel method of shared control of an assistive robot. We evaluated the idea within a 2D simulation environment [6]. The novel control method uses a Convolutional Neural Network (CNN) to adaptively generate DoF mappings based on camera data of the current situation. From a user perspective, this system can help the user choose an optimal mapping of available control DoFs for a low-DoF input device, either automatically or upon the user's request. In this paper, we build on this approach, focusing in particular on the user interface. Having an adaptive mapping of control DoFs to the input device can be challenging to understand and learn, which is why there is a need for visual feedback to convey that information to the user. The approach in our previous work included visual cues in the form of arrows. While the results are promising (see Section 2), the limitation of a 2D environment means that it is difficult to predict how this approach transfers to 3D. For example, certain DoF combinations might be more difficult to display with arrows in a 3D environment and lead to visual clutter.

The goal of this paper is to explore the proposed novel control method, as well as possible visual cues for the DoF mappings. In particular, we want to explore how the novel, adaptive control method performs in a 3D environment compared to the standard mode-switch approach with cardinal DoF mappings and whether changes in the visual cues have an impact on the performance of the adaptive control method.

We conducted a remote online study with 39 non-disabled participants, in which we compared three different control types with different DoF mapping behaviors and visual cues. These were *Classic* and *Double Arrow*, which used two arrows attached to the fingers as visual cues, and a visually reduced variant *Single Arrow*. *Single Arrow* only used one arrow through the middle of the gripper (see Section 3 for a detailed description of each control type).

The study was conducted inside a 3D Virtual Reality (VR) environment, utilizing Head-Mounted Displays (HMDs) for an immersive experience (see Section 4.3 for a complete description of the virtual environment). The participants repeatedly performed a simple pick-and-place task, controlling a virtual robot arm using the three control types (see Section 4.5 for a detailed description of the study design).

Due to the ongoing COVID-19 pandemic, we opted to recruit non-specific participants that had access to the required hardware (an *Oculus Quest* VR-HMD) to participate in our study. None of the recruited participants reported living with any motor impairments. We acknowledge this limitation and discuss how our findings can be transferred to the target group of people with motor impairments in Section 7.

As our main contribution, we present findings from our study, which compare our two adaptive control types with the standard mode-switch control type, explicitly focusing on task completion times, number of mode switches and workload. In addition, we contribute an extensive discussion of qualitative results from voice recordings of our participants, providing a deeper understanding of the benefits and challenges of each of the three control types.

2. Related Work

To assist people with physical or cognitive impairments, prior research often suggests possible solutions that use robots that automate specific tasks [7–10]. Assistive robots are found in a variety of designs. There are stationary robots specifically designed for meal-assistance [11], socially assistive robots for elderly people and people with cognitive impairments [12], navigational robots for blind people [13], and many more examples, both in research and commercially. Besides stationary robots (e.g., fixed to a table) [14], there are also moving robots attached to mobile platforms [15,16] or mounted to the user's wheelchair [9].

To help people with motor impairments, assistive robot arms are widely used, both within the workspace and in performing ADLs [17]. Their flexibility allows for many different applications, such as feeding assistance [18], fetch and pick-up tasks [15], and cataloging of books [7].

Robotic assistance is generally well-received by people with motor impairments. Drolshagen et al. found that people with disabilities quickly accept working with robots, even if the robots are in close proximity [19]. Regarding ADLs specifically, Pascher et al. conducted an ethnographic study with 15 participants with tetraplegia, multiple sclerosis, Locked-In Syndrome, and similar diseases [20]. They found that people with motor impairments would prefer to perform ADLs themselves with the help of a robotic aid as opposed to with the help of another person. People with motor impairments want to "live more independently" and "gain increased autonomy".

However, automating ADLs, as suggested in research, can have unintended consequences. Pollak et al. conducted a study comparing manual and autonomous modes of collaboration with a collaborative robot (cobot) [21]. They found that using the manual mode in which the cobot would perform tasks only upon interaction with the participants decreased stress significantly. The participants felt "more capable of coping with and controlling the situation" than in the autonomous mode.

Similarly, Kim et al. conducted a study with subjects with spinal cord injuries using an assistive robot arm in either a manual or an autonomous mode [22]. They found that overall task completion times for manual and autonomous usage for trained participants were similar, but user satisfaction was higher in manual mode. This is despite the fact that autonomous usage decreased the effort necessary to perform tasks significantly. The authors call for more flexible interfaces to control assistive robot arms.

When interacting with robots that carry out movements, a study by Cleaver et al. showed that users generally prefer to have a visual representation of the robot's future movements. However, having this visualization does not significantly affect the performance when executing tasks using the robot [23]. When using a visual representation of robot motion intent, the most prominent solution is to show the robot's movement using arrows [24–26]. In addition, most of these approaches rely on Augmented Reality to overlay the visual representation on the user's real environment.

Heeding the call for more flexible interfaces, we proposed in our recent work an adaptive control concept for assistive robot arms that promises to allow users to be in control at all times while still providing them with more assistance during ADLs than the standard mode switch control concept [6]. In this proposed concept, a CNN interprets the video feed of a camera attached to the robot arm and adaptively outputs the most likely movement DoFs.

With current control concepts, users with low-DoF input devices, such as simple joysticks, can only move the gripper of an assistive robot arm in cardinal directions (i.e., movement and rotation around Cartesian X-, Y-, and Z-Axes). The user has to switch and choose between the provided mappings of input DoFs to some of the robot's DoFs. This may include the pairings of different DoFs of the robot that are less than ideal for the given situation, resulting in many time-consuming and mentally demanding mode switches. Additionally, in any given mode, an input on an axis of a low-DoF device would move the gripper only in the cardinal direction currently assigned to this input DoF. Combinations of multiple output DoFs (such as orbiting an object, which is the combination of rotation and translation) require more than one input DoF (e.g., both the X- and Y-Axes of a joystick) to be engaged simultaneously in such systems.

To solve this problem, we proposed a representation of these assignments of input DoFs to output DoFs in the form of a matrix similar to the one seen in Figure 1 in our previous work. Each row in that matrix represents a cardinal output DoF, while each column represents the input DoFs of an input device. The values in a column determine which movement the robot's gripper will perform if the input DoF is engaged. For example, an identity matrix would yield a behavior identical to the cardinal mode switch approach, as each input DoF is only mapped to one cardinal output DoF.

$$\begin{array}{c} X-Axis \\ Y-Axis \\ Z-Axis \\ Roll \\ Pitch \\ Yaw \\ Gripper \end{array} \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad \begin{bmatrix} 0.5 & 0 & 0 & 0.5 & 0 & 0 & 0 \\ 0.5 & 0 & 0 & 0 & 0.5 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0 & 0.5 & 0 & 0 \\ 0 & 0.5 & 0 & 0 & 0 & 0.5 & 0 \\ 0 & 0 & 0.5 & 0 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Figure 1. Two different DoF mappings as matrices—(**Left**): classic control (one input DoF controls one cardinal output DoF); (**Right**): arbitrarily combined controls (one input DoF controls more than one cardinal output DoF at the same time).

This representation allows for combinations of multiple output DoFs for one input DoF. For example, if the first column contains a value of 0.5 in the first two rows, engaging the first input DoF would result in a diagonal movement along the XY plane of the robot's coordinate system (see the matrix on the right in Figure 1). According to the current situation, the proposed control concept adaptively fills this matrix to create the most useful combination of output DoFs.

We then conducted a small study with a 2D proof-of-concept simulation for our proposed control concept. A total of 23 participants used a "standard" and an "adaptive" control type for a simulated 2D robot that could drive forwards, sideways, rotate around its center, and close its fingers to move blue boxes to target red boxes (see Figure 2). This is the 2D equivalent of a simple pick-and-place task in 3D. Both control types switched modes after five seconds without user input.

The results of our study showed that, subjectively, the "adaptive" control was significantly faster but significantly more difficult than the "standard" control. "Adaptive" control also led to significantly shorter sequence execution times.

While these findings are promising, the concept requires further evaluation in 3D and in a more complex environment with devices that have more DoFs. We set out to do precisely that: evaluate the proposed concept of adaptive control in a more complex environment with a robot arm with seven DoFs.

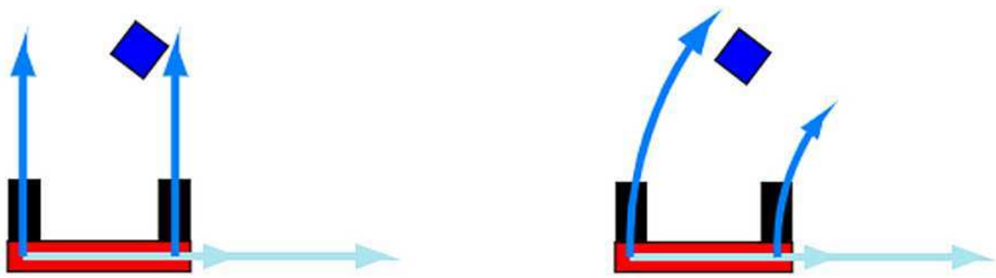
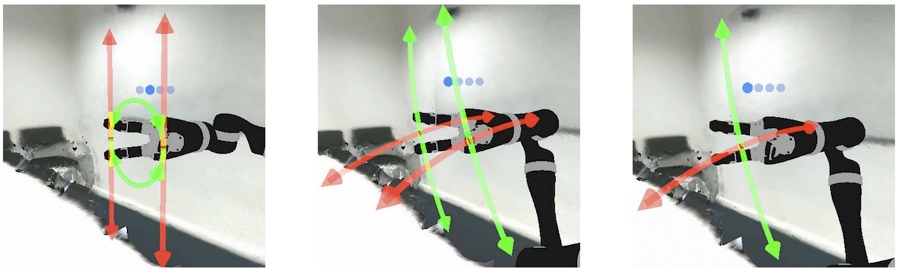


Figure 2. The simulated robot with two out of the four cardinal DoFs (**left**) and two adaptive DoFs (**right**) [6].

3. Control Types for a 3D Environment

To compare the standard control type of switching between cardinal modes to the adaptive approach, we implemented three control types (see Figure 3) in a simulated 3D environment (see Section 4.3). This simulated environment is meant to act as a proxy for a potential Augmented Reality (AR) implementation. There, users would control an assistive robot arm and see the visual feedback superimposed on the real world and robot via an AR-HMD device. Instead, in our 3D simulation, users wear an *Oculus Quest* VR-HMD, which superimposes the visual feedback directly in the computed 3D scene. An overview of the environment and the control types described in the following sections is provided as a video (see Video S1).

All three control types use arrows as visual cues. Specifically, the arrows show which direction the gripper will move if a user engages the corresponding input DoF. To allow the users to predict the robot's movement when engaging the input DoF with positive values (e.g., pressing the control stick up) and negative values (e.g., pressing the control stick down), the arrows have two heads. Each arrowhead points towards the corresponding movement direction.



(**a**) (**b**) (**c**)

Figure 3. Visualization for the different control types: (**a**) Classic; (**b**) Double Arrow; (**c**) Single Arrow.

Using visual cues in 3D as opposed to 2D often causes visual obstruction, e.g., if the gripper is close to the table, the active DoF would lower the gripper towards the table. In that case, the arrows would clip through the table, making them partially invisible to the user. It would also be common that the robot's gripper itself obstructs parts of the arrows, making them harder to see and interpret. To eliminate these problems, the arrows were made translucent and are always rendered above all other objects yet shown at the correct depth as if looking through whatever is blocking them. This behavior is similar to viewing the scene through Augmented Reality glasses, which would overlay the arrows onto a real scene as opposed to showing the arrows as part of the real world that can be blocked by other real-world objects.

To more easily communicate the currently active mode, all control types show a blue indicator above the robot gripper consisting of four spheres, each representing a mode (see Figure 3). The sphere representing the currently active mode is darker and less translucent than the inactive ones, indicating how many modes are left to switch through before returning to the first.

3.1. Manually Designed DoF-Calculations

The focus of this study was to evaluate how adaptively changing DoF mappings would impact the participant's experience in a more complex 3D environment. While we proposed a CNN to perform these calculations in our previous work [6], there are other ways of calculating these DoF mappings. We developed a manually scripted method of calculating these DoF mappings for the specific task used in the study instead of training a CNN. This method generates a matrix with the same rules described in our previous work (see Figure 1) to represent DoF mapping, thus providing the possibility of equal movements as generated by a CNN trained on camera data. Since our primary focus is the participant's experience with the adaptively changing DoF mappings, we assumed that this approach would significantly decrease the possibility of unpredictable behavior while having little impact on the applicability of our findings to a system using a CNN. A detailed description of the generated output values is presented in the description of the adaptive control types (see Sections 3.3 and 3.4).

This approach is akin to the widely used "Wizard of Oz" method, in which the output of a proposed system is instead provided by a human to test the user experience of that proposed system before finishing the implementation. In our case, we instead simulated the output of a complex CNN using a simpler system. As with "Wizard of Oz" experiments, our results should therefore be applied to the user experience with the system using a CNN, but the absolute performance measures may vary.

We developed three control types—*Classic*, *Double Arrow*, and *Single Arrow*—to function with different assistive robot arms and different input devices. To conduct the study, we decided to use the widely available stand-alone VR headset *Oculus Quest*. The *Oculus Quest* consists of the headset itself, and two motion controllers, one for each hand, with several buttons and a control stick each. Participants executed a simple pick-and-place task (see Section 4.6) in our VR environment using a virtual model of the *Kinova Jaco* robot arm using each of the control types (see Section 4.3 for a detailed description of the virtual environment and the VR setup).

3.2. Classic Control Type

The *Classic* control type implements the standard mode switch control type most commonly used to control assitive robot arms. This means that an input DoF always corresponds to a cardinal output DoF. Given the seven cardinal DoFs of the Jaco robot arm (X-Translation, Y-Translation, Z-Translation, Roll, Yaw, Pitch, Open/Close fingers) and two input DoFs (the X-Axis and Y-Axis on a motion controller's control stick) four modes are available to the users:

1. X-Translation + Y-Translation;
2. Z-Translation + Roll;
3. Yaw + Pitch;
4. Open/Close fingers + Nothing.

The last mode has no assigned output DoF for the X-Axis on the control stick to allow the users to learn an axis-to-action mapping.

Users can switch modes by pressing the A-Button on the right-hand motion controller. This allows them to perform the tasks at their own pace and assess the usefulness of a mode as long as they need to. Whenever the A-Button is pressed while the fourth mode is active, the first mode is selected again, allowing the users to cycle through modes at will.

Two arrows attached to the fingers of the gripper show the users which motion the gripper would perform, given a user's input on the respective input DoF. Red arrows represent the movement assigned to the Y-Axis of the control stick, and green arrows represent the movement assigned to the X-Axis of the control stick. As the motion controllers are also rendered in the virtual environment, we added a visual representation onto the control stick rendered in-game. A cross with one red axis and one green axis is shown on the motion controller to indicate which direction corresponds to which color. A blue sphere surrounds the A-Button to match it to the blue mode indicator (see Figure 4).

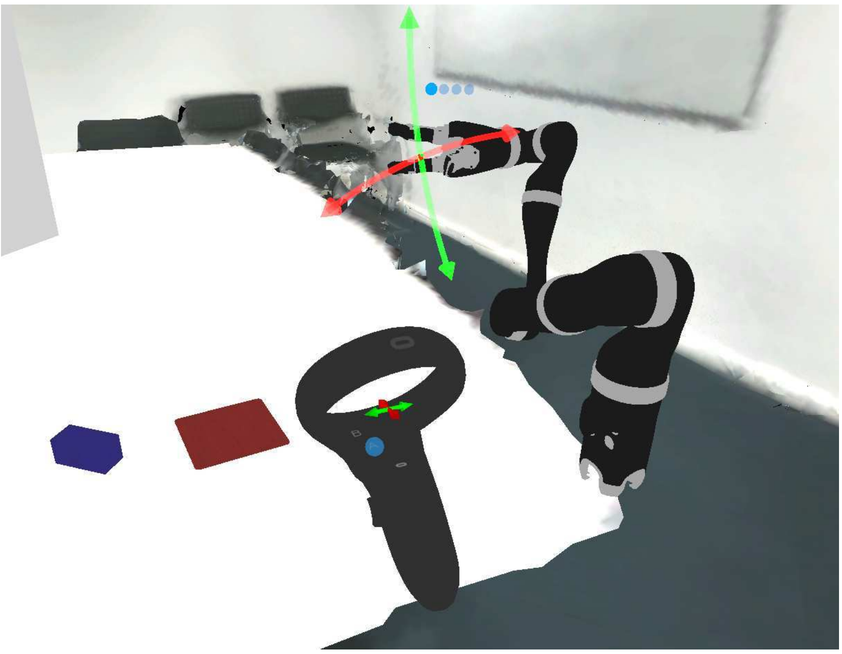


Figure 4. The virtual motion controller with directional indicators and the robot arm with matching arrows visualizing the currently selected mode.

3.3. Double Arrow Control Type

The *Double Arrow* control type implements the proposed adaptive control method using two arrows to show the position of the fingers if a user engages an input DoF. Therefore, each input DoF corresponds to a combination of cardinal DoFs determined based on the current situation. To ensure comparability with the *Classic* control type in regards to the number of mode switches necessary to return to the starting mode, four modes were developed. The modes are ordered by their complexity and usefulness to the users' goal of reaching the next target.

As in the *Classic* control type, two arrows attached to the fingers of the gripper show the users which motion the gripper would perform, given a user's input on the respective input DoF. Red arrows represent the movement assigned to the Y-Axis of the control stick, and green arrows represent the movement assigned to the X-Axis of the control stick.

The first mode assigns the Y-Axis of the control stick to a movement that both rotates and translates the gripper towards the next target simultaneously. More precisely, if the gripper is further than 10 cm away from the target, the movement is oriented towards a point 15 cm above the target. If the gripper is closer than 10 cm to the target, the movement is oriented towards the actual target. This ensures that the gripper tends to grasp and let go of objects from above, as opposed to trying to do so from the sides and thereby possibly crashing into the table. If the gripper is within reach of an object or target point where an object is supposed to be placed by the users, it also allows them to open and close the fingers. The X-Axis of the control stick in the first mode is assigned the same movement as the Y-Axis but rotated by 90° to allow for corrections perpendicular to the Y-Axis movement.

To provide users with more options, the second mode assigns the Y-Axis of the control stick to a linear translational movement towards the object and the X-Axis of the control stick to a rotational movement of the gripper towards the next target. Both of these assignments were chosen since only moving or only rotating are less likely to further the goal of the users. However, the individual movements themselves are still integral movements for coordinating the gripper orientation and some movement towards the goal. In the optimal case, this means that users would not need to use this mode, as both orientation and positioning would be taken care of simultaneously by the first mode.

The third mode assigns the Y-Axis of the control stick to the opening or closing the fingers, depending on whether an object was currently held or not. The X-Axis of the control stick has no assignment in this mode to ensure comparability with the *Classic* control type.

If users stop moving the gripper, they should always be able to move the same way they did before. To ensure this, the fourth mode always assigns the X- and Y-Axis of the control stick the same mappings that were last used to move the gripper. Otherwise, users who would want to assess if they had moved the robot far enough for their personal preference using a given mapping would have no possibility to correct their course.

The system calculates the next movement mappings whenever the users stop moving the robot. However, the system does not instantly assign the first mode to be active, as this would disrupt the users' flow of control (i.e., they might have stopped to asses the situation and then decided to continue with the DoF mapping they were using). Moreover, this would harm comparability to the *Classic* control type (as no automatic mode switches happen in that control type). This means that whenever the users stop moving, the blue mode indicator would show the fourth mode as being active, and a press on the A-Button would lead to the newly calculated first mode.

3.4. Control Type Single Arrow

During the development of *Classic* and *Double Arrow* we discovered that, while two arrows are a perfectly suitable visualization for a 2D environment, these arrows can result in a large amount of visual clutter during complex movement in 3D environments. We decided to develop a visualization that reduces visual clutter in a 3D environment and compare its usage to the *Double Arrow* control type.

Dubbed *Single Arrow*, the input-to-output DoF mappings are calculated in the exact same way as the mappings in *Double Arrow*. Switching between modes is also handled in the same way as in *Double Arrow*. However, the visualization changes from displaying two arrows at the tips of the fingers to displaying one arrow in the middle of the gripper, with a slight offset to allow certain movements to be displayed. This reduces visual clutter for all situations except when the fingers move.

4. Materials and Methods

We present a remote study with 39 participants to compare the proposed concept of adaptive control (in two variations) against the standard mode-switch control concept. In particular, we measured task completion times, the number of mode switches necessary to perform a task, the workload necessary to use the different control concepts via a NASA Raw-TLX (NASA Raw Task Load Index), and the participants' personal ranking of the three presented control types. Participants used their own *Oculus Quest* headset to perform a simple pick-and-place task using a virtual robot inside a realistic 3D environment.

4.1. Hypotheses

We propose the following hypotheses:

- Average Task Completion Time
  - *H1* Double Arrow *leads to lower task completion time than* Classic. The adaptive control of *Double Arrow* should significantly reduce the movements necessary to perform the task by combining different cardinal DoFs into one continuous movement, which otherwise would each have to be adjusted separately.
  - *H2* Single Arrow *leads to lower task completion time than* Double Arrow. Only using one arrow for each DoF mapping should reduce visual clutter. This should lead to a shorter processing time of the suggested movements, reducing the total time to execute a task.
- Average Number of Mode Switches

- *H3* Double Arrow *leads to fewer mode switches than* Classic*.* The adaptive control of *Double Arrow* should reduce the necessity to switch modes significantly. Since different DoFs are combined depending on the current situation, a change in position and rotation brings the robot arm closer to the target and can be performed without mode switches.
- *H4* Single Arrow *and* Double Arrow *need roughly an equal number of mode switches.* The behavior of the two adaptive control types is the same. Thus, while it might take participants longer to understand what movements they can perform with *Double Arrow* as opposed to *Single Arrow*, they should switch modes approximately as often in both control types.

- Workload
  - *H5* Double Arrow *leads to lower NASA TLX scores than* Classic*.* The adaptive control of *Double Arrow* calculates sensible movements to reach the next goal position and rotation. Thus, it should alleviate the participants from having to think of a sequence of movements to reach their goal, reducing workload. This is in contrast to the findings of our previous study, in which participants perceived the *Adaptive* control as more complex than the *Standard* control [6]. We expect the benefit of pre-calculated DoF combinations and the workload of developing a sequence of movements in cardinal DoFs to be higher in a 3D environment than in a 2D environment. Therefore, the workload for the adaptive control types should be lower than for *Classic* in 3D.
  - *H6* Single Arrow *leads to lower NASA TLX scores than* Double Arrow*.* Since we assume that reduced visual clutter leads to a shorter processing time for the suggested movements, the NASA TLX scores of *Single Arrow* should be lower.

4.2. Participants

In total, 39 people participated in our study (12 female, 26 male, 1 non-binary), which led to a data-set of 936 individual trials (8 per control type, 24 per participant). The age of participants ranged from $\leq$19 to 69, with 20 to 29 being the largest group with 22 participants. Four participants had prior experience with controlling an assistive robot arm, and no participants declared any motor impairments. All participants received EUR 10 as compensation unless they specifically denied the offer.

Due to the ongoing COVID-19 pandemic, we opted to perform a remote study using VR. We did not specifically search for participants with motor impairments because the potential target audience for people with VR setups at home that also have motor impairments appeared too small. There would not be enough time to gather enough participants in a realistic time frame. Instead, we searched for any participants that had access to the necessary equipment (an *Oculus Quest* headset, see Section 4.3) and were able to install our study software on their devices. With these non-specific participants, the performance measures for executing the tasks in our study with the different control types (see Section 4.6) can be compared relative to one another, even though they may not be representative of the intended target audience of such an assistive device. We acknowledge this limitation, which is further discussed in Section 7.

Participants were recruited via announcements in different social media communities relating to VR (e.g., r/OculusQuest: https://www.reddit.com/r/OculusQuest/, accessed on 3 January 2022), social media communities regarding assistive technologies (e.g., r/AssistiveTechnologies: https://www.reddit.com/r/AssistiveTechnology/, accessed on 3 January 2022), and platforms for acquiring participants specifically for XR studies (e.g., XRDRN: https://www.xrdrn.org/, accessed on 3 January 2022) among other more local announcements.

To ensure that VR sickness symptoms did not influence our results, the participants filled out the Virtual Reality Sickness Questionnaire (VRSQ) at the end of the study [27]. The VRSQ measures nine items on a four-point Likert scale and results in a value between 0 and 100, where 0 means no symptoms experienced and 100 means all symptoms were severe.

Reported values were low (Mean: 11.30, Std.-Dev.: 11.38), and none of the participants selected the "Severe" option for any of the items.

4.3. Apparatus

We designed a Virtual Reality environment based on a photogrammetry scan of a real room. The environment included a virtual model of the *Kinova Jaco* (Kinova Jaco robot arm: https://assistive.kinovarobotics.com/product/jaco-robotic-arm, accessed on 3 January 2022) robot arm attached to a table, a red target surface, a blue block, and two virtual screens—one for descriptions and questionnaires and one that would show example photos of the control types (see Figure 5). We decided to use a virtual model of a real robot arm (*Kinova Jaco*) to stay as close to a physical system as possible. Additionally, the *Kinova Jaco* robot arm is specifically designed and often used as an assistive device for people with motor impairments [5].

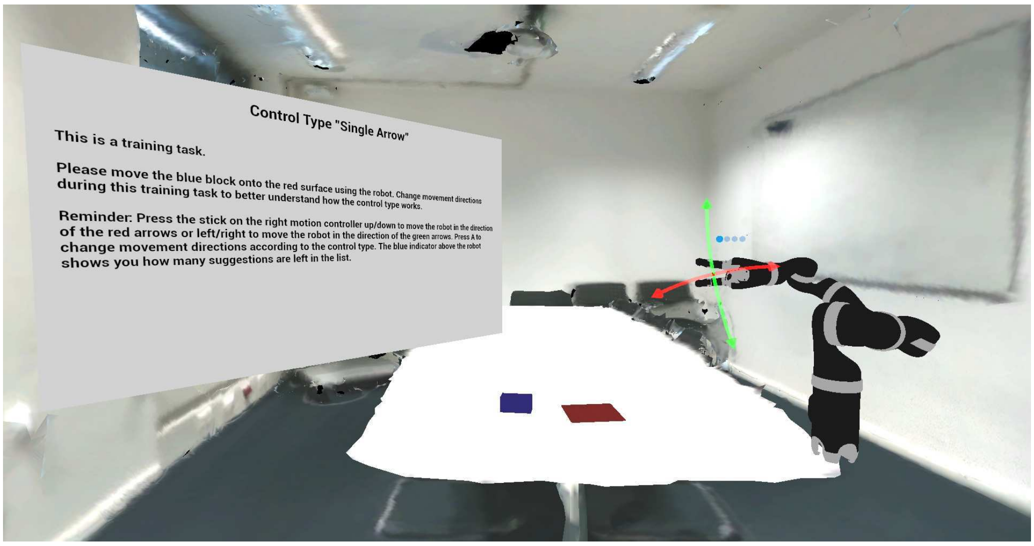


Figure 5. The virtual environment: description screen (**Left**); screen with example photos of the control types (not shown); *Kinova Jaco* with visualisation for control type *Single Arrow* (**Right**); table with blue block and red target (**Bottom**).

The virtual environment was created with the *Unreal Engine 4.26* and was developed to be deployed to the *Oculus Quest* VR headset. Participants had to either own or have access to such a headset and be able to install the study software on that headset using a computer (*Windows*, *macOS*, and *Linux* could be used). Although we tested our software on the original *Oculus Quest* hardware, we did not explicitly exclude the use of the newer and very similar *Oculus Quest 2* headset. The *Oculus Quest* consists of the VR headset and two motion controllers, one for each hand. Each motion controller has several buttons and a control stick. Participants controlled the robot using the right motion controller of the VR headset. In particular, the control stick of the motion controller moved the robot according to the currently active control type. This enabled the participants to control which DoFs were being used and how fast the robot would move. The A-Button was used to switch to the next mode cyclically, returning to the first mode when a mode switch was performed in the last mode.

To simulate the movement of the robot arm, the inputs did not move the joints of the robot as they would with a physical robot arm. Rather, the gripper of the virtual robot arm is moved in 3D space according to the inputs, and the arm of the robot is programmed to adopt a correct pose automatically. This was implemented using the physics system of the *Unreal Engine*.

4.4. Procedure

Participants were directed to a website with a brief introduction to the study, the duration of the study (around 30 to 45 min), the technical and non-technical prerequisites to participate in the study, and a description about what data would be collected during the study. Participants were informed that certain metrics and usage data, such as task completion times, will be recorded and sent to our servers during the study. They were also informed that they would need to fill out a short questionnaire after each condition of the study and that they would be able to record a short audio message after each condition. Lastly, participants were informed that cookies were being used on our website. Each participant gave informed consent by pressing a clearly labeled button to continue and start the study. After giving their consent, participants were instructed on how to install and open the study application and what to do when they were finished with the part of the study inside the VR headset. During the study, neither a video of the participants surroundings through the VR headsets external cameras and sensors nor a screen-recording was captured.

Next, the participants put on their VR headsets and opened our study application. They were greeted with a brief explanation of the study on a large virtual screen. Except for the questionnaires after each control type, any text that was available to read on that screen was also simultaneously read aloud as a prerecorded voice-over. The participants interacted with this screen via a common interaction method that was also used in the menus of the *Oculus Quest* headset: pointing a ray that originated from the motion controller towards the screen and using the trigger to confirm input.

After the study explanation, the participants were presented with a description of the first control type they would be using and the task they would be performing. This explanation was supplemented with an image on a second smaller virtual screen. The descriptions were written in a way that described how the gripper would move in relation to the current situation. We did not explicitly describe the intentions behind the different modes and their order in *Double Arrow* and *Single Arrow* (to provide ideally optimal mappings) to prevent possible biases. Otherwise, the participants might have been inclined to trust the adaptive mappings against their own judgment, thereby changing their behavior.

The explanation of each control type was followed by a series of trials of our pick-and-place task (see Section 4.6) the participants had to execute to progress through the study. For each control type, the task was performed once as a training trial and then eight more times for the same control type. During these eight trials, the task completion time and the number of mode switches performed was recorded.

After executing all trials for a control type, the participants were presented with the NASA Raw-TLX questionnaire to capture the participants' workload. Additionally, the participants could record a short audio message to point out additional things they felt were relevant during the execution of the trials. The recording of the audio message was optional. After filling out the questionnaire and optionally recording an audio message, the participants would continue with the next control type until they had executed all trials for all three control types.

Upon finishing the VR part of the study, participants received a unique code to be entered in a form on our website to complete the VRSQ [27] and our questionnaire. We asked the participants to report their demographic data and rank the control types presented in the VR section of the study. Lastly, participants left their contact information to receive the compensation.

4.5. Study Design

We used a within-subjects design with the control type as an independent variable with three levels: (1) *Classic*, (2) *Double Arrow*, and (3) *Single Arrow*. Each participant performed eight trials of a pick-and-place task for each of the three control types (see Section 4.6). Additionally, they performed one training trial for each control type to familiarize themselves with the control type. The order of control types shown to the participants was fully balanced.

We measured three dependent variables for each control type: *Average Task Completion Time*, *Average Number of Mode Switches*, and *Workload via a NASA Raw-TLX questionnaire*.

Average Task Completion Time in seconds While participants executed each trial with the robot arm, the time to complete the task was measured for each participant. Then, the average task completion time for each control type was calculated across all participants.
Average Number of Mode Switches While participants executed each trial with the robot arm, each mode switch executed by pressing a button on the input device was counted and stored as the number of mode switches. Then, the average number of mode switches for each control type was calculated across all participants.
Workload via a NASA Raw-TLX questionnaire After completing all trials within each control type, the participants were asked to fill out a NASA Raw-TLX questionnaire to obtain information about the participants' perceived workload. The questionnaire consists of the following six criteria, which participants would rate on a scale of 0 to 100 in steps of 5: mental demand, physical demand, temporal demand, performance, effort, and frustration [28].

In addition, the participants could record a short description of their experiences in the form of a voice message, although this was not mandatory. The recorded voice messages were transcribed and analyzed by multiple researchers to identify underlying themes and common impressions the participants had while using the virtual robot arm (see Section 5.2). Participants also provided a personal ranking of the three control types in a questionnaire at the end of the study.

4.6. Task

Participants were asked to repeatedly place a blue block onto a red target using the assistive robot. Participants performed this task eight times per control type. We did not use two blocks per trial to reduce variability in our results. We decided to use a simple pick-and-place task instead of a specific ADL (e.g., drinking from a glass) since pick-and-place tasks are part of many ADLs. Moreover, a specific ADL might have caused problems with participants' preconceived notions of that task (e.g., they would approach the glass in a particular way, while the adaptive system would approach it differently). This would have possibly distracted them from evaluating the control types as a whole, which we wanted to avoid.

In each of the eight trials per control type, the position of the blue block changed to one of eight predefined positions around the red target surface. The order in which the positions were used in the eight trials was randomized for each participant and control type.

5. Results

We recorded both quantitative and qualitative data from the participants during the trials. This section presents the results of each section from our data analysis.

5.1. Quantitative Results

The recorded quantitative data for each trial included *task completion time* (in seconds) and *the number of mode switches*. For each control type, the quantitative data included the *NASA Raw-TLX* results and the *Rank* given to the control type by the participants (lower rank numbers are better). The used abbreviations and symbols are:

- IQR: Interquartile Range;
- SD: Standard Deviation;
- SE: Standard Error;
- p: p-value as an expression of the level of statistical significance;
- N: Sample Size;
- $\chi^2(2)$: Chi-Squared with two degrees of freedom;

- F: F-Statistic for the Repeated-Measures ANOVA;
- M: Mean;
- df: Degrees of Freedom for the calculation of χ^2 for the Friedman Tests.

5.1.1. Task Completion Time

For each participant, we averaged the task completion times (see Table 1) of the trials for each control type. In an exploratory analysis, we removed outliers that had average task completion times $\geq 2.2 * \text{IQR}$ of the mean task completion time in at least one control type [29] (see Figure 6). Four outliers were excluded this way, leaving 35 participants for analysis of task completion times. An inspection of QQ-plots found the resulting data-set to follow a normal distribution.

Table 1. Statistics for average task completion times (in seconds, N = 35).

	Classic	Double Arrow	Single Arrow
Mean	47.41	42.62	44.04
Median	44.66	37.75	41.23
Std.-Dev.	12.55	19.28	22.24
IQR	14.03	24.33	31.68

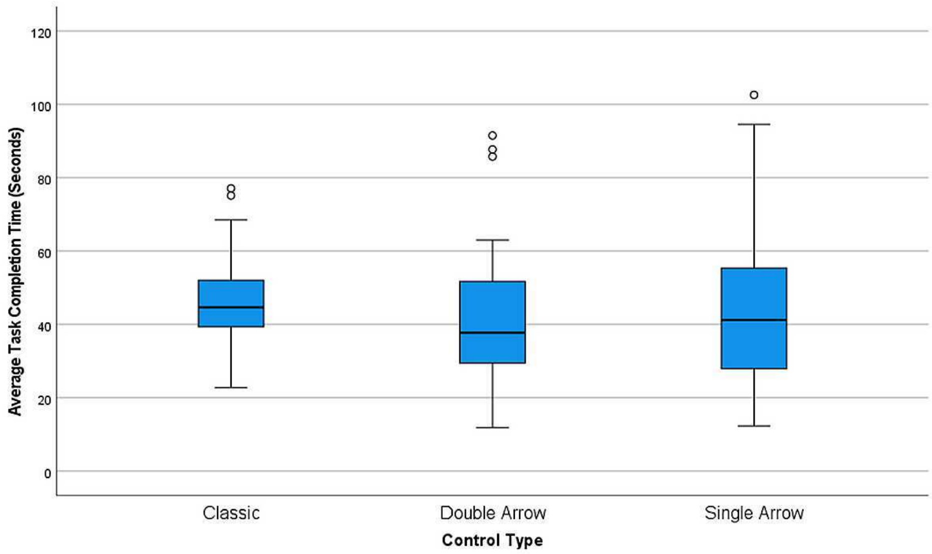


Figure 6. Boxplots for average task completion times.

To determine whether the control types had an effect on average task completion times, we performed a Repeated-Measures ANOVA (RM-ANOVA). However, we found no significant main effect ($F(2, 64) = 1.31$, $p = 0.28$).

In addition to the effect of control types, we examined whether the starting condition of a participant had an impact on task completion times. We included the starting condition as a between-subjects factor for the RM-ANOVA and discovered a significant interaction effect between the starting condition and the task completion times ($F(4, 64) = 8.86$, $p < 0.001$). Analyzing simple main effects, we discovered that the task completion times for *Classic* stayed roughly the same regardless of the starting condition. However, both adaptive control types heavily suffered when they were the starting condition (see Figure 7). A post hoc pairwise comparison (Estimated Marginal Means, Bonferroni adjusted) showed that task completion times for *Single Arrow* (M = 54.66 s, SE = 5.9) were significantly longer than those for *Double Arrow* (M = 33.74 s, SE = 4.6) if *Single Arrow* was the starting condition ($p = 0.001$). Conversely, task completion times for *Double Arrow* (M = 57.89 s, SE = 5) were significantly longer than those for *Single Arrow* (M = 37.82 s, SE = 6.41) if *Double Arrow* was the starting condition instead ($p = 0.002$). Another significant difference was found if *Single*

Arrow was the starting condition: *Classic* task completion times (M = 48.23 s, SE = 3.57) were longer than those of *Double Arrow* (M = 33.74 s, SE = 4.6) in that case (p = 0.013). The other comparisons yielded insignificant results.

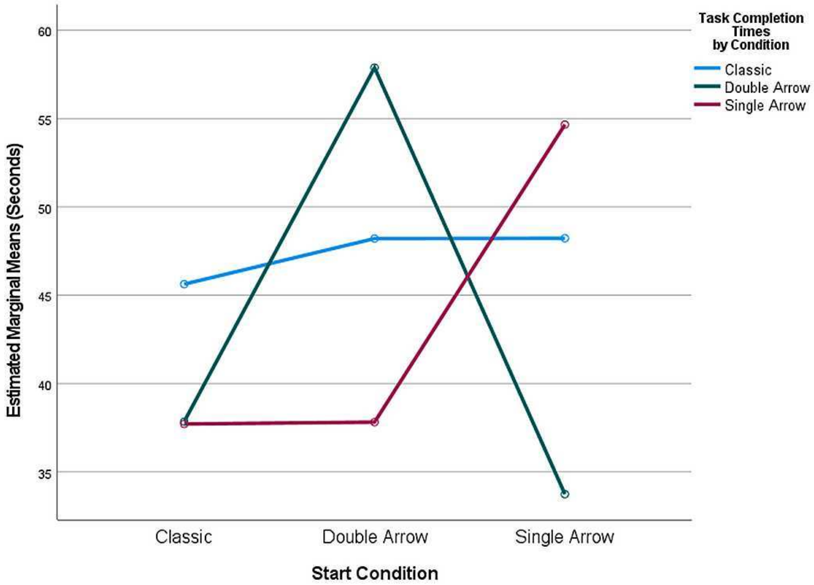


Figure 7. Estimated Marginal Means for average task completion times.

5.1.2. Mode Switches

To determine whether there were differences between the average number of mode switches between control types we used an RM-ANOVA. Due to a software error, mode switch data were only recorded correctly for 20 participants. We found a significant effect of control types on the average number of mode switches (F(2, 38) = 8.08, p = 0.001). Pairwise comparisons revealed that there were significant differences (p < 0.05) between the average number of mode switches for both adaptive control methods (*Double Arrow*: M = 12.93, SD = 3.91; *Single Arrow*: M = 14.23, SD = 5.15) and the *Classic* control method (M = 17.87, SD = 4.8). We found no significant difference between the average number of mode switches for *Single Arrow* compared to *Double Arrow* (p = 0.11, see Table 2 and Figure 8).

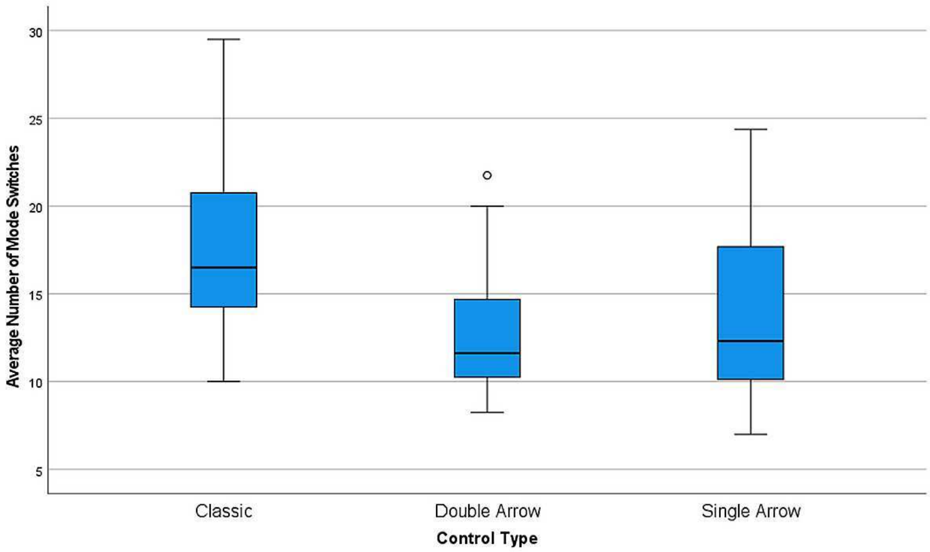


Figure 8. Boxplots for average number of mode switches.

Table 2. Statistics for average number of mode switches (N = 20).

	Classic	Double Arrow	Single Arrow
Mean	17.87	12.93	14.23
Median	16.50	11.63	12.31
Std.-Dev.	4.80	3.91	5.15
IQR	7.00	5.09	7.91

5.1.3. Workload and Rank

Each participant completed a NASA Raw-TLX questionnaire after completing the task with each control type, rating each dimension on a scale from 1 to 100. To evaluate whether there were any differences between the control types regarding workload, Friedman Tests were performed for both the overall NASA TLX value as well as the individual dimensions of the questionnaire. No significant differences were found for either the overall NASA TLX value ($\chi^2(2) = 5.33$, $p = 0.07$) or the individual dimensions (see Table 3).

We also evaluated whether the users preferred one control type over the others. To do so, the participants ranked the control types after completing all tasks. A lower number means the participant ranked that control type higher. No significant differences were found for the ranks ($\chi^2(2) = 0.97$, $p = 0.65$) (see Table 4).

Table 3. Statistics for individual NASA TLX Dimensions on a scale from 1 to 100 (df = 2, N = 39 for all Friedman Tests).

	Mental Demand	Physical Demand	Temporal Demand	Performance	Effort	Frustration
Classic (Mean)	53.33	30.26	36.92	32.05	48.59	41.41
Classic (Std.-Dev.)	24.64	21.67	21.07	20.48	24.84	24.52
Double Arrow (Mean)	56.28	28.21	40.38	38.97	52.82	43.08
Double Arrow (Std.-Dev.)	22.93	16.20	25.06	25.50	24.08	26.40
Single Arrow (Mean)	48.97	27.56	36.03	40.64	51.41	38.33
Single Arrow (Std.-Dev.)	24.69	22.94	20.56	26.61	23.25	26.34
Mean Ranks						
Classic	2.04	1.92	1.96	1.73	1.79	1.92
Double Arrow	2.21	2.17	2.18	2.15	2.18	2.17
Single Arrow	1.76	1.91	1.86	2.12	2.03	1.91
Friedman Tests						
χ^2	4.23	2.07	2.38	4.86	3.15	1.76
Exact Significance	0.12	0.37	0.31	0.09	0.21	0.43

5.2. Qualitative Results

Participants were asked to describe their experience with the control type they used in a voice message. They were asked to elaborate on the ease of controlling the robot, their understanding of movement directions, and the predictability of the next movement directions.

In total, 23 of the 39 participants recorded a message for all three control types. In addition, only four participants recorded voice messages for two of the three control types, and one participant just recorded a single voice message. This resulted in 26 voice messages for each control type.

Table 4. Statistics for NASA TLX on a scale from 1 to 100 and ranking on a scale from 1 to 3 (df = 2, N = 39 for all Friedman Tests).

	NASA TLX	Rank
Classic (Mean)	40.43	1.87
Classic (Std.-Dev.)	17.11	0.77
Double Arrow (Mean)	43.29	2.05
Double Arrow (Std.-Dev.)	15.32	0.86
Single Arrow (Mean)	40.49	2.08
Single Arrow (Std.-Dev.)	17.29	0.84
Mean Ranks		
Classic	1.85	1.87
Double Arrow	2.29	2.05
Single Arrow	1.86	2.08
Friedman Tests		
χ^2	5.33	0.97
Exact Significance	0.07	0.65

5.2.1. Thematic Analysis

The voice recordings were analyzed with the Thematic Analysis method described by Braun and Clarke [30]. This method was chosen because it has the flexibility to identify themes within the unstructured feedback from the recorded voice messages. Throughout the analysis, we identified themes related to our hypotheses, which gave us a better insight into how participants perceived their experience and success in executing the given tasks.

First, we transcribed the voice messages to be able to analyze them. Although most participants recorded their messages in English, a few recorded them in German. Some of the statements in the following chapters were therefore translated into English. Second, two of our researchers performed the Thematic Analysis using the six-phase method described by Braun and Clarke [30]. Each researcher read each transcribed voice message to become familiar with the participant's feedback. They then marked certain paragraphs and phrases to identify underlying topics related to our hypotheses that were relevant within multiple data-sets. Each marked phrase was assigned a short code describing its topic. We used the software *Obsidian* (Obsidian markdown note-taking software: https://obsidian.md, accessed on 3 January 2022) for managing and tagging the transcribed messages in a simple markdown text format with links and tags. Third, codes were organized and grouped into themes, and descriptive titles were assigned to each theme. For a visual representation, we developed visual thematic graphs; one of which is shown in Figure 9. Although some comments were related to several themes, we decided to sort them into the theme with the best fit. Fourth, themes were revised and evaluated by reading the related phrases and codes again to ensure that each theme was internally homogeneous. Fifth, both researchers worked together to refine the themes and compile them into a single thematic map presented in Figure 10. Sixth, a summary of the results was written based on the final thematic map.

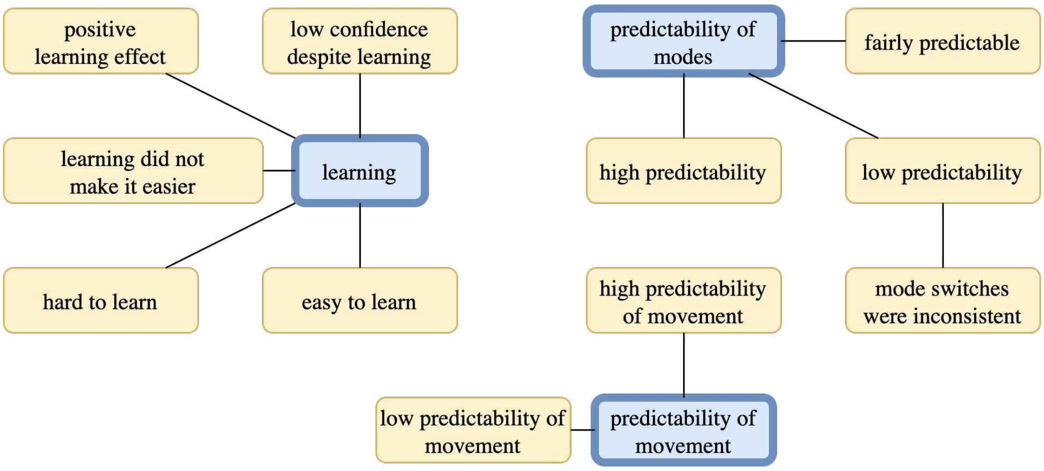


Figure 9. Early thematic map with codes shown in yellow and themes shown in blue.

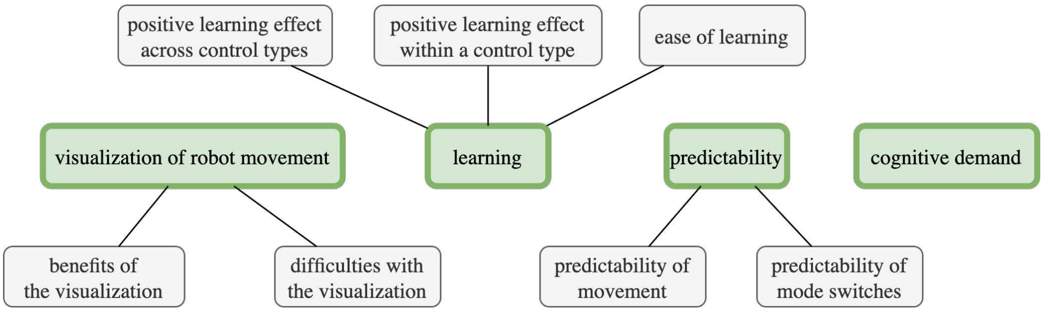


Figure 10. Final thematic map with themes shown in green and sub-themes shown in gray.

5.2.2. Results of the Thematic Analysis

We identified the following themes in the combined thematic map: *visualization of robot movement*, *cognitive demand*, *predictability of mode switching*, *predictability of movement* and *learning*. The excerpts from one participant's audio messages were marked with the participant's unique number (e.g., *P26* for the 26th participant out of the total 39 participants). Since participants often referenced the previous control types they used, we also added which control type they were referring to in brackets when citing them.

Visualization of robot movement: This theme comprises the difficulties and benefits of the visualization of the robot's movement. As expected when transferring over a visualization from a 2D environment to 3D, perspective was one source of errors across all three control types. *P5* stated, "Depending on the orientation of the robot arm, I could not see exactly which way the arrows were going." P4 added, "Sometimes moving the robot was a bit difficult because it just did not feel natural from different perspectives."

Regarding the control types *Double Arrow* and *Single Arrow*, many participants mentioned that the arrows are either hard to interpret or hard to see. Interestingly, the participants did not mention this problem with the *Classic* control type. Participants stated, "[...] the arrows that follow the change of the movement direction are a little more difficult and a little bit less intuitive to understand than the previous trial [control type *Classic*]" (*P9*), and "I think it is more difficult than the previous control type because it has more abstract movement [...]" (*P25*). Besides the curved arrows, many participants found it difficult to associate the differently colored arrows of the visualization with the different input DoFs across all three control types. *P31* made this clear after using the *Single Arrow* control type. They said, "The hardest part working with this method of motion was determining which direction pushing the analog stick would actually move the robot."

Across both adaptive control types, participants mentioned the helpfulness of the arrows. *P25* commented, "I think it was confusing at first, but those red and green arrows helped a lot to understand how the robot moved." After using the *Double Arrow* control type, *P8* mentioned, “Controlling the robot was better than before [control type *Single Arrow*], because one could tell more easily where the arm would go, based on the multiple arrows”. This suggests the possible benefits of having multiple arrows in the *Double Arrow* control type.

Cognitive demand: In this theme, we consolidated statements that describe a higher or lower cognitive demand while using a specific control type. Across all three control types, some participants mentioned a high cognitive demand. After using the *Single Arrow* control type, *P17* stated, “This one was more cognitively demanding than the previous one [control type *Classic*], maybe because this one did not have straight movement but a lot of rotational movements”. *P18* found it to be “a bit confusing, but okay.” Participants described the *Classic* control type as “confusing” (*P8*) and “counter intuitive” (*P18*). Using the *Double Arrow* control type, *P21* expressed the need to focus on the task and added, “I do not think you could do anything else while using this control method”.

While mentions of lower cognitive demand were equally frequent in total, many participants found the *Classic* control type to be “easy” or “easy to understand” (*P6*, *P9*, *P25*, *P27*, among others). After using the *Classic* control type, *P39* added, “Here it was best to intuitively remember where each function was”. This suggests a connection with the next two themes regarding predictability.

Predictability of mode switching: This theme describes the ability of the participants to anticipate the next set of movement combinations that the system provides when the participant executes a mode switch. Many of the difficulties participants had with the predictability were with the adaptive control types *Single Arrow* and *Double Arrow*. When using the *Double Arrow* control type, *P17* noted, “In this condition, I was not sure whether cycling through the different types of movements in there always were consistent. That was very confusing.” We also identified this statement as an expression of an increase in cognitive demand. For the same control type, *P21* added, “I did not know which combination would be next when I pressed A”. Using the *Double Arrow* control type, *P23* mentioned, “I could not predict the next movement, because I did not understand in which order the different movements are shown to me next.” We think this participant confused the ever-changing nature of the adaptive suggestions with the different modes. Only a few participants mentioned difficulties with predicting the next mode in the *Classic* control type. *P37* said, “Predictability was uncertain as well, until the later moves where I had enough training to do it effectively.” Additionally, many participants mentioned that they had to switch modes many times to find the proper movement they needed in a given situation, especially with the adaptive control types. Using the *Double Arrow* control type, *P3* stated, “So if I wanted it to go down I would have to switch through multiple modes [. . .]”. Furthermore, using the *Double Arrow* control type, *P5* mentioned, “I had to click through many modes to find the movement that I thought would bring me closer to the block”.

Mentions of good predictability were also spread across all three control types, although these were less common. For the *Classic* control type, *P39* stated, “It was very easy to understand and especially the predictability was the easiest here”. Using the *Single Arrow* control type, *P37* mentioned, “The ease of understanding the movement was a lot easier as well. With some of the movement directions being easier to understand and predict before they show up.” After executing the tasks with the *Double Arrow* control type, *P37* added, “It seemed more predictable and overall, a more optimum way of doing things”.

Predictability of movement: In contrast to the previous theme, this theme is about predicting how and where the robot arm will move when using the currently selected mode. As visualization plays a big part when predicting the robot’s movement, this theme is related to the first theme about visualization. Only a few participants mentioned the predictability of movement directly. After using the *Double Arrow* control type, *P4* said, “So I tried to do one thing and it would do a completely other thing. It felt really unnatural to try and get to the cube and even to pick it up”. For the *Classic* control type, *P10* stated, “Because of the immediate predictability [. . .], it was much easier to control the robot and to steer it into different vectors to approach the block in the different positions". Using the *Single Arrow* control type, *P10* added, "Therefore I could understand very well how it would move and how it would work out so I could reach the target".

Learning: This theme describes the participants' impression of their learning experience while using the different control types. Across all three control types, participants reported that they grew better at performing the tasks over time. For the *Classic* control type, *P26* stated, "Using this robot arm is pretty easy if you learn how to use them, [. . .]". After using the *Double Arrow* control type, *P25* mentioned, "The predictability of the next movement directions, I think, is easier as you practice with it, [. . .]". For the *Single Arrow* control type, *P39* said, "The more I practiced, the more confidence I got [. . .]".

As participants used the different control types, they noticed a learning effect even across the different control types. After finishing all trials of all control types, ending with the *Double Arrow* control type, *P25* said, "The predictability of the next movement directions, I think, is easier as you practice with it, [. . .]" After using the *Single Arrow* control type, *P33* stated, "Maybe I simply have more experience now, if I performed better in this task in any way".

Even though many participants felt that they needed more practice with the tasks so that they are easier to perform, some described that the process of learning felt relatively easy. When finishing the tasks with the *Classic* control type, *P16* stated, "It was quicker to get familiar with the system." *P33* expressed some difficulties with the *Double Arrow* control type but added, "At least it did not take long to notice a learning effect".

Additionally, we identified many instances where participants reported that they liked the second adaptive control type they used better than the one before, regardless of which control type came first and which came second. This also suggests that a learning effect is taking place. After using the *Double Arrow* and then the *Single Arrow* control types, *P27* stated, "I don't know what is the difference between double arrow and single arrow, but single arrow is much easier to control". For the *Double Arrow* control type, *P31* stated, "This method is a little bit easier to use than the second method [*Single Arrow* control type], but I think that was more a function of having a little bit more experience".

6. Discussion

Initially, our assumptions were that the overall task performance would be best when using the *Single Arrow* control type, followed by *Double Arrow*, and *Classic* would have the worst task performance. In comparison to the results of our previous study [6], the new results are not as pronounced in a realistic virtual 3D setting, at least not without considering the learning effects.

Regarding the task completion times, both Hypothesis 1 and Hypothesis 2 could not be substantiated. However, the interaction effect between the starting condition and task completion times suggests that, with time to learn, the adaptive control types could perform better than the *Classic* type. This is corroborated by participants' reports, as many participants said that their performance and understanding of the adaptive control types improved during the tasks. It is also worth noting that more participants experienced the second adaptive control type as "better" than the first, implying a learning effect not only for one control type but between control types.

Regarding mode switches, Hypothesis 3 and Hypothesis 4 could be substantiated by our results. From *Classic* to *Double Arrow*, we measured a significant reduction in the number of mode switches necessary to perform the task. In contrast, there was no significant difference between *Double Arrow* and *Single Arrow*. Interestingly, this contrasts the participants' opinions that they felt they had to switch many times to get to a mode that performed a movement they expected. However, this reduction in mode switches might be of higher benefit for people with motor impairments than for non-disabled people. Switching modes using a button requires a certain level of dexterity and causes the user to constantly divert their attention away from the original task, so more mode switches can cause more fatigue and time consumption, as explained by Herlant et al. [5]. The impact of this difference in the number of mode switches on people with motor impairments can thus only be evaluated in a future study with participants with motor impairments.

Regarding workload, Hypothesis 5 and Hypothesis 6 could not be substantiated. This could have multiple reasons. For example, the participants expressed that the predictability of the adaptive control types was low and that they did not necessarily know how the robot would move, even with the arrows. These impressions, combined with the statements regarding positive learning effects and overall high cognitive demand, could mean that with increased exposure to the adaptive control types, users could have a lower workload than with *Classic*.

According to some participants, using visual cues in a 3D environment caused problems with perspective. This made it difficult for them to predict how the robot would move, even with the visual cues provided by the arrows. To mitigate this problem, our concept might be combined with a "digital twin" of the robot arm, which demonstrates the movement virtually before the real robot performs it physically [31].

To improve the overall predictability of the system, both regarding the suggested modes and the movements of the robot, a training mode could be implemented. In this mode, the users would be able to teach the system the way they want specific tasks to be performed [32]. This should increase predictability, as the participants would know the proposed movements will be (partially) based on their own instructions. In addition, Spatial Augmented Reality can help the user's understanding of the robot's perception, e.g., which object the robot assumes the user wants to interact with [33]. In combination with the already implemented visual cues, this can help the users predict the robot's movement more accurately.

After further research and refinement of our proposed control methods, they might allow assistive robot arms to help with ADLs that currently require the help of caregivers or more complex robots, such as dressing [34] or bathing [35]. The fact that the users always stay in control of the robot while the robot performs more fluent, natural movements could also allow people with motor impairments to use the robot in social situations, e.g., at the workplace [36].

7. Limitations

Our study did not specifically involve or focus on people with motor impairments. Thus, we need to discuss how our results can be transferred to this target group. First, the absolute performance measures cannot be generalized to this target group. Individual differences are usually high within people with motor impairments due to varying degrees of physical limitations [37]. However, the study did not aim to provide absolute results in terms of performance but rather an insight into the relative performance of the three different control types. Since they all rely on the same physical interaction concept, we believe that the way motor impairments might affect performance should be comparable for all three control types. Second, Augmented Reality is necessary to provide the user with the type of visual feedback we implemented for our study. We are aware from our prior research that current-generation AR-HMDs are often not accessible to people with motor impairments. AR-HMDs such as the *Microsoft HoloLens* are too heavy and conflict too often with health-supporting systems [38]. We conducted this research with the firm belief that future AR hardware solutions will cope with requirements for people with motor impairments. We acknowledge, however, that this might make the visual feedback designs inapplicable for real-world systems at this point in time or the immediate future.

Additionally, our study involved the use of the *Oculus Quest* system and the *Oculus Quest* Motion Controller as the only input device. In the real world, however, assistive robot arms can be controlled with a wide range of input devices depending on the abilities and preferences of the person using them. We specifically only used the most basic functionality of the Motion Controller (the control stick and one button) to ensure that the results are also applicable when using a different input devices with two input axes. It is still possible that the use of different input devices might add more complexity to the overall usage of such a system.

Another limitation is the nature of our study being performed as a remote study. The level of control is limited for such a method, which means that the level of engagement of participants can vary. We addressed this limitation by keeping the duration of the study relatively short (30–45 min) and designing the task so that we could easily identify cases in which participants did not follow the study protocol. Our analysis further shows that only a few participants were identified as extreme outliers. In addition, the focus on one set of hardware devices made it possible to harmonize and control the kind of immersive experience that participants engaged with, further reducing potential biasing effects, such as low frame rates or other hardware-performance-related issues. Given the current COVID-19 pandemic, we believe that our study setup is sensible and still able to provide robust results. Still, we aim to replicate at least part of the study in a lab environment and with people with motor impairments in the future.

It is possible that our study does not provide insight into the quality of adaptive control through the means of a CNN. We simulated the adaptive control method to be able to have full control in the study. Otherwise, imperfect DoF mappings would have overshadowed the potential effects of the different visualizations, thus making it difficult to draw conclusions. As discussed, we believe that our approach significantly decreases the possibility of unpredictable behavior while having little impact on the applicability of our findings to a system using a CNN, as long as this CNN is able to perform at a high level of quality regarding the DoF mappings.

8. Conclusions

We conducted a study exploring and evaluating the user experience of an adaptive control concept for assistive robot arms in a realistic virtual 3D environment. Our results suggest a significant benefit of such an adaptive control concept regarding the necessary number of mode switches. However, task completion times and workload do not change when using an adaptive control concept without more intensive training.

By evaluating the interaction between the starting conditions and task completion times and applying a thematic analysis of qualitative data, we conclude that there could be a significant benefit of training that would reveal the potential of an adaptive control concept. Therefore, future work should consider longer training sessions before evaluating task completion times and workload. The targeted user group of assistive robot arms would use such devices not just once but daily and over extended periods and thus have more time to learn how to use the device. Therefore it is important to assess whether the adaptive control concept might have high cognitive demand in the beginning but is better than the *Classic* approach once the users are trained.

Our results seem to suggest that there is little to no difference between *Single Arrow* and *Double Arrow* regarding how well they convey the robots currently active DoF mapping to the users. However, an improved visualization could reduce the overall high cognitive demand users have experienced. Therefore, future work will also focus on different types of visualizations, which will not be restricted to MR-headsets and overlayed arrows but could (additionally) show the robot's future path using spatial Augmented Reality [39].

Future work should (whenever possible) include participants with motor impairments since their experience is vital in designing assistive technology [4]. The impact of a lower number of mode switches enabled by an adaptive control concept should be especially evaluated with people with motor impairments. This could significantly improve their execution of activities of daily living.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/technologies10010030/s1. Video S1: An Overview of the Environment and Control Types.

Author Contributions: Conceptualization, K.K., S.R., M.P., F.F.G., U.F. and J.G.; methodology, K.K., S.R., M.P. and J.G.; software, K.K. and S.R.; validation, M.P. and J.G.; formal analysis, K.K. and S.R.; investigation, K.K., S.R. and M.P.; resources, M.P., F.F.G. and J.G.; data curation, K.K. and S.R.; writing—original draft preparation, K.K. and S.R.; writing—review and editing, M.P., F.F.G., U.F. and J.G.; visualization, K.K. and S.R.; supervision, M.P., U.F. and J.G.; project administration, U.F. and J.G.; funding acquisition, U.F. and J.G.; original study, F.F.G. and U.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research is supported by the German Federal Ministry of Education and Research (BMBF, FKZ: 16SV8563 and 16SV8565).

Institutional Review Board Statement: Ethical reviews by an Ethics Committee are waived in Germany as they are not compulsory for fields outside medicine. Only the participant's consent was requested.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bauer, A.; Wollherr, D.; Buss, M. Human—Robot Collaboration: A Survey. *Int. J. Hum. Robot.* **2008**, *5*, 47–66. [CrossRef]
2. Fong, T.; Nourbakhsh, I.; Dautenhahn, K. A survey of socially interactive robots. *Robot. Auton. Syst.* **2003**, *42*, 143–166.
3. Statistisches Bundesamt (Destatis). Pflegestatistik—Pflege im Rahmen der Pflegeversicherung—Deutschlandergebnisse—2019. 2020. Available online: https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Gesundheit/Pflege/Publikationen/Downloads-Pflege/pflege-deutschlandergebnisse-5224001199004.pdf (accessed on 3 January 2022).
4. Pascher, M.; Baumeister, A.; Schneegass, S.; Klein, B.; Gerken, J. Recommendations for the Development of a Robotic Drinking and Eating Aid—An Ethnographic Study. In Proceedings of the Human-Computer Interaction—INTERACT 2021, Bari, Italy, 30 August–3 September 2021; Ardito, C., Lanzilotti, R., Malizia, A., Petrie, H., Piccinno, A., Desolda, G., Inkpen, K., Eds.; Springer: Cham, Switzerland, 2021. [CrossRef]
5. Herlant, L.V.; Holladay, R.M.; Srinivasa, S.S. Assistive Teleoperation of Robot Arms via Automatic Time-Optimal Mode Switching. In Proceedings of the Eleventh ACM/IEEE International Conference on Human Robot Interaction, Christchurch, New Zealand, 7–10 March 2016; pp. 35–42.
6. Goldau, F.F.; Frese, U. Learning to Map Degrees of Freedom for Assistive User Control: Towards an Adaptive DoF-Mapping Control for Assistive Robots. In Proceedings of the 14th PErvasive Technologies Related to Assistive Environments Conference, Corfu, Greece, 29 June–2 July 2021; Association for Computing Machinery: New York, NY, USA, 2021; PETRA 2021, pp. 132–139. [CrossRef]
7. Gräser, A.; Heyer, T.; Fotoohi, L.; Lange, U.; Kampe, H.; Enjarini, B.; Heyer, S.; Fragkopoulos, C.; Ristic-Durrant, D. A Supportive FRIEND at Work: Robotic Workplace Assistance for the Disabled. *IEEE Robot. Autom. Mag.* **2013**, *20*, 148–159. [CrossRef]
8. Foresi, G.; Freddi, A.; Monteriù, A.; Ortenzi, D.; Pagnotta, D.P. Improving mobility and autonomy of disabled users via cooperation of assistive robots. In Proceedings of the 2018 IEEE International Conference on Consumer Electronics (ICCE), Las Vegas, NV, USA, 2–14 January 2018; pp. 1–2. [CrossRef]
9. Tsui, K.M.; Kim, D.J.; Behal, A.; Kontak, D.; Yanco, H.A. "I want that": Human-in-the-loop control of a wheelchair-mounted robotic arm. *Appl. Bionics Biomech.* **2011**, *8*, 127–147. [CrossRef]
10. Goldau, F.F.; Shastha, T.K.; Kyrarini, M.; Gräser, A. Autonomous multi-sensory robotic assistant for a drinking task. In Proceedings of the 2019 IEEE 16th International Conference on Rehabilitation Robotics (ICORR), Toronto, ON, Canada, 24–28 June 2019; pp. 210–216.
11. Tanaka, K.; Mu, S.; Nakashima, S. Meal-assistance robot using ultrasonic motor with eye interface. *Int. J. Autom. Technol.* **2014**, *8*, 186–192. [CrossRef]
12. McColl, D.; Louie, W.Y.G.; Nejat, G. Brian 2.1: A socially assistive robot for the elderly and cognitively impaired. *IEEE Robot. Autom. Mag.* **2013**, *20*, 74–83. [CrossRef]
13. Guerreiro, J.; Sato, D.; Asakawa, S.; Dong, H.; Kitani, K.M.; Asakawa, C. Cabot: Designing and evaluating an autonomous navigation robot for blind people. In Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility, Pittsburgh, PA, USA, 28–30 October 2019; pp. 68–82.
14. Maheu, V.; Archambault, P.S.; Frappier, J.; Routhier, F. Evaluation of the JACO robotic arm: Clinico-economic study for powered wheelchair users with upper-extremity disabilities. In Proceedings of the 2011 IEEE International Conference on Rehabilitation Robotics, Zurich, Switzerland, 29 June–1 July 2011; pp. 1–5.
15. Hashimoto, K.; Saito, F.; Yamamoto, T.; Ikeda, K. A field study of the human support robot in the home environment. In Proceedings of the 2013 IEEE Workshop on Advanced Robotics and its Social Impacts, Tokyo, Japan, 7–9 November 2013; pp. 143–150.
16. Fattal, C.; Leynaert, V.; Laffont, I.; Baillet, A.; Enjalbert, M.; Leroux, C. Sam, an assistive robotic device dedicated to helping persons with quadriplegia: Usability study. *Int. J. Soc. Robot.* **2019**, *11*, 89–103. [CrossRef]

17. Kyrarini, M.; Lygerakis, F.; Rajavenkatanarayanan, A.; Sevastopoulos, C.; Nambiappan, H.R.; Chaitanya, K.K.; Babu, A.R.; Mathew, J.; Makedon, F. A survey of robots in healthcare. *Technologies* **2021**, *9*, 8. [CrossRef]
18. Canal, G.; Alenyà, G.; Torras, C. Personalization framework for adaptive robotic feeding assistance. In Proceedings of the International Conference on Social Robotics, Kansas City, MO, USA, 1–3 November 2016; Springer: Berlin/Heidelberg, Germany, 2016; pp. 22–31.
19. Drolshagen, S.; Pfingsthorn, M.; Gliesche, P.; Hein, A. Acceptance of industrial collaborative robots by people with disabilities in sheltered workshops. *Front. Robot. AI* **2021**, *7*, 541741. [CrossRef] [PubMed]
20. Pascher, M.; Baumeister, A.; Klein, B.; Schneegass, S.; Gerken, J. Little Helper: A Multi-Robot System in Home Health Care Environments. In Proceedings of the 2019 International Workshop on Human-Drone Interaction (iHDI) as Part of the ACM Conference on Human Factors in Computing Systems, Glasgow, UK, 5 May 2019.
21. Pollak, A.; Paliga, M.; Pulopulos, M.M.; Kozusznik, B.; Kozusznik, M.W. Stress in manual and autonomous modes of collaboration with a cobot. *Comput. Hum. Behav.* **2020**, *112*, 106469. [CrossRef]
22. Kim, D.J.; Hazlett-Knudsen, R.; Culver-Godfrey, H.; Rucks, G.; Cunningham, T.; Portee, D.; Bricout, J.; Wang, Z.; Behal, A. How Autonomy Impacts Performance and Satisfaction: Results From a Study With Spinal Cord Injured Subjects Using an Assistive Robot. *IEEE Trans. Syst. Man Cybern.—Part A Syst. Hum.* **2012**, *42*, 2–14. [CrossRef]
23. Cleaver, A.; Tang, D.V.; Chen, V.; Short, E.S.; Sinapov, J. Dynamic Path Visualization for Human-Robot Collaboration. In Proceedings of the Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction (HRI '21 Companion), Boulder, CO, USA, 8–11 March 2021; Association for Computing Machinery: New York, NY, USA, 2021; pp. 339–343. [CrossRef]
24. Walker, M.; Hedayati, H.; Lee, J.; Szafir, D. Communicating Robot Motion Intent with Augmented Reality. In Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction (HRI '18), Chicago, IL, USA, 5–8 March 2018; Association for Computing Machinery: New York, NY, USA, 2018; pp. 316–324. [CrossRef]
25. Shrestha, M.C.; Kobayashi, A.; Onishi, T.; Uno, E.; Yanagawa, H.; Yokoyama, Y.; Kamezaki, M.; Schmitz, A.; Sugano, S. Intent communication in navigation through the use of light and screen indicators. In Proceedings of the 2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI), Christchurch, New Zealand, 7–10 March 2016; pp. 523–524. [CrossRef]
26. Shindev, I.; Sun, Y.; Coovert, M.; Pavlova, J.; Lee, T. Exploration of Intention Expression for Robots. In Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction (HRI '12), Boston, MA, USA, 5–8 March 2012; Association for Computing Machinery: New York, NY, USA, 2012; pp. 247–248. [CrossRef]
27. Kim, H.K.; Park, J.; Choi, Y.; Choe, M. Virtual reality sickness questionnaire (VRSQ): Motion sickness measurement index in a virtual reality environment. *Appl. Ergon.* **2018**, *69*, 66–73. [CrossRef] [PubMed]
28. Hart, S.G. Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* **2006**, *50*, 904–908. [CrossRef]
29. Hoaglin, D.C.; Iglewicz, B. Fine-Tuning Some Resistant Rules for Outlier Labeling. *J. Am. Stat. Assoc.* **1987**, *82*, 1147–1149. [CrossRef]
30. Braun, V.; Clarke, V. Using thematic analysis in psychology. *Qual. Res. Psychol.* **2006**, *3*, 77–101. [CrossRef]
31. Park, K.B.; Choi, S.H.; Lee, J.Y.; Ghasemi, Y.; Mohammed, M.; Jeong, H. Hands-free human–robot interaction using multimodal gestures and deep learning in wearable mixed reality. *IEEE Access* **2021**, *9*, 55448–55464. [CrossRef]
32. Kyrarini, M.; Zheng, Q.; Haseeb, M.A.; Gräser, A. Robot learning of assistive manipulation tasks by demonstration via head gesture-based interface. In Proceedings of the 2019 IEEE 16th International Conference on Rehabilitation Robotics (ICORR), Toronto, ON, Canada, 24–28 June 2019; pp. 1139–1146.
33. Pascher, M.; Kronhardt, K.; Franzen, T.; Gruenefeld, U.; Schneegass, S.; Gerken, J. My Caregiver the Cobot: Comparing Visualization Techniques to Effectively Communicate Cobot Perception to People with Physical Impairments. *Sensors* **2022**, *22*, 755. [CrossRef]
34. Zhang, F.; Cully, A.; Demiris, Y. Probabilistic real-time user posture tracking for personalized robot-assisted dressing. *IEEE Trans. Robot.* **2019**, *35*, 873–888. [CrossRef]
35. Zlatintsi, A.; Dometios, A.; Kardaris, N.; Rodomagoulakis, I.; Koutras, P.; Papageorgiou, X.; Maragos, P.; Tzafestas, C.S.; Vartholomeos, P.; Hauer, K.; et al. I-Support: A robotic platform of an assistive bathing robot for the elderly population. *Robot. Auton. Syst.* **2020**, *126*, 103451. [CrossRef]
36. Arévalo Arboleda, S.; Pascher, M.; Lakhnati, Y.; Gerken, J. Understanding human-robot collaboration for people with mobility Impairments at the Workplace, a thematic analysis. In Proceedings of the 2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), Naples, Italy, 31 August–4 September 2020; pp. 561–566.
37. Arévalo Arboleda, S.; Becker, M.; Gerken, J. Does One Size Fit All? A Case Study to Discuss Findings of an Augmented Hands-Free Robot Teleoperation Concept for People with and without Motor Disabilities. *Technologies* **2022**, *10*, 4. [CrossRef]
38. Arévalo Arboleda, S.; Pascher, M.; Baumeister, A.; Klein, B.; Gerken, J. Reflecting upon Participatory Design in Human-Robot Collaboration for People with Motor Disabilities: Challenges and Lessons Learned from Three Multiyear Projects. In Proceedings of the 14th PErvasive Technologies Related to Assistive Environments Conference—PETRA 2021, Corfu, Greece, 29 June 2021–2 July 2021. [CrossRef]
39. Watanabe, A.; Ikeda, T.; Morales, Y.; Shinozawa, K.; Miyashita, T.; Hagita, N. Communicating robotic navigational intentions. In Proceedings of the 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Hamburg, Germany, 28 September–2 October 2015; pp. 5763–5769. [CrossRef]

AdaptiX – A Transitional XR Framework for Development and Evaluation of Shared Control Applications in Assistive Robotics

MAX PASCHER, TU Dortmund University, Germany and University of Duisburg-Essen, Germany
FELIX FERDINAND GOLDAU, German Research Center for Artificial Intelligence (DFKI), Germany
KIRILL KRONHARDT, TU Dortmund University, Germany
UDO FRESE, German Research Center for Artificial Intelligence (DFKI), Germany
JENS GERKEN, TU Dortmund University, Germany

(a) (b) (c)

Fig. 1. Setup with (**a**) a user's view in the Virtual Reality (VR) simulation environment, (**b**) setup of interaction with a physical robot, and (**c**) a combined view of physical robot and visual cues in Mixed Reality (MR).

With the ongoing efforts to empower people with mobility impairments and the increase in technological acceptance by the general public, assistive technologies, such as collaborative robotic arms, are gaining popularity. Yet, their widespread success is limited by usability issues, specifically the disparity between user input and software control along the autonomy continuum. To address this, shared control concepts provide opportunities to combine the targeted increase of user autonomy with a certain level of computer assistance. This paper presents the free and open-source *AdaptiX* XR framework for developing and evaluating shared control applications in a high-resolution simulation environment. The initial framework consists of a simulated robotic arm with an example scenario in Virtual Reality (VR), multiple standard control interfaces, and a specialized recording/replay system. *AdaptiX* can easily be extended for specific research needs, allowing Human-Robot Interaction (HRI) researchers to rapidly design and test novel interaction methods, intervention strategies, and multi-modal feedback techniques, without requiring an actual physical robotic arm during the early phases of ideation, prototyping, and evaluation. Also, a Robot Operating System (ROS) integration enables the controlling of a real robotic arm in a *PhysicalTwin* approach without any simulation-reality gap. Here, we review the capabilities and limitations of *AdaptiX* in detail and present three bodies of research based on the framework. *AdaptiX* can be accessed at https://adaptix.robot-research.de.

Authors' addresses: Max Pascher, TU Dortmund University, Emil-Figge-Str. 50. 44227 Dortmund, Germany, max.pascher@udo.edu; Felix Ferdinand Goldau, Enrique-Schmidt-Str. 5, 28359 Bremen, Germany, felix.goldau@dfki.de; Kirill Kronhardt, TU Dortmund University, Emil-Figge-Str. 50. 44227 Dortmund, Germany, kirill.kronhardt@udo.edu; Udo Frese, Enrique-Schmidt-Str. 5, 28359 Bremen, Germany, udo.frese@dfki.de; Jens Gerken, TU Dortmund University, Emil-Figge-Str. 50. 44227 Dortmund, Germany, jens.gerken@udo.edu.

ACM 2573-0142/2024/6-ART244
https://doi.org/10.1145/3660243

CCS Concepts: • **Computer systems organization** → **Robotic control**; • **Human-centered computing** → *Visualization techniques*; *Virtual reality.*

Additional Key Words and Phrases: assistive robotics, human–robot interaction, shared user control, augmented reality, virtual reality, mixed reality, visual cues

ACM Reference Format:
Max Pascher, Felix Ferdinand Goldau, Kirill Kronhardt, Udo Frese, and Jens Gerken. 2024. AdaptiX – A Transitional XR Framework for Development and Evaluation of Shared Control Applications in Assistive Robotics. *Proc. ACM Hum.-Comput. Interact.* 8, EICS, Article 244 (June 2024), 28 pages. https://doi.org/10.1145/3660243

1 INTRODUCTION

Robotic arms as assistive technologies are a powerful tool to increase self-sufficiency in people with limited mobility [33, 44], as they facilitate the performance of Activities of Daily Living (ADLs) – usually involving grasping and manipulating objects in their environment – without human assistance [50]. However, a frequent point of contention is the assistive robot's autonomy level. The reduction of user interaction to just oversight with purely autonomous systems elicits stress [51] and feelings of distrust in their users [67]. On the other side of the autonomy spectrum, manual controls can be challenging - or even impossible - to operate, depending on the significance and type of impairment. Shared control – a combination of manual user control through standard input devices plus algorithmic support through computer software adjusting the resulting motion – may have the potential to mitigate both concerns [1]. Here, both the user and the robot share a task on the operational level, enabling people with motor impairments to get involved in their assistance. As a result, such approaches can increase the feeling of independence while improving ease of use compared to manual controls [17].

A characteristic real-world scenario, motivated by our research, has an assistive robotic arm (e.g., a Kinova Jaco 2) attached to a wheelchair to support the user in ADLs. Here, the user is challenged with operating six or more Degrees-of-Freedom (DoFs), which requires complex input devices or time-consuming and confusing mode switches. This potentially results in increased task completion time and user frustration [21]. Addressing this, shared control systems can facilitate more straightforward and accessible robot operation. However, they may require well-designed communication of robot (motion) intent, so that the user retains awareness and understands the level of support they get from the system [45]. Also, different users might need distinct input devices or require multi-modal input to account for varying abilities.

Based on our experiences, we identified several challenges that currently influence and potentially impede the effective development of shared control approaches:

- Shared control systems for assistive technologies still pose open questions requiring considerable experimentation, tweaking and balancing between user and robot interaction [34].
- While much research explored robot motion intent, there is little insight into what works best in which situation and for which type of user. In assistive robotics, the visualization and feedback modality must be carefully adapted to the user's needs and abilities as there is no "one size fits all" solution [23].
- Similarly, suitable input devices may vary between users. Depending on individual preferences and capabilities, multi-modal input or the choice between different input modalities may be required [2].
- Bringing robots and humans physically together during research studies is difficult due to the laborious and costly transportation, safety concerns with robots and general availability of the user group [6].

Contribution. To allow researchers, designers and developers to address these challenges holistically and flexibly, we present *AdaptiX* – a free, open-source XR framework [1]. Aimed at Design and Development (D&D), *AdaptiX* combines a physical robot implementation with a 3D simulation environment. The simulation approach (analogous to simulations in industrial settings [37, 42, 59]) mitigates the assistive robotic arm's bulky, expensive, and complex nature. It also makes the integration of visualization feedback or different input modalities easier to explore and test, while a Robot Operating System (ROS) interface allows the direct transfer to the real robot. Testing new interaction and control options becomes much less time-consuming while simultaneously excluding potentially dangerous close-contact situations with users before glitches are managed [44]. In total, the framework facilitates the development and evaluation of assistive robot control applications *in-silico* and creates a practical and effective step between ideation, development, and evaluation, allowing HRI researchers more flexibility and facilitating efficient resource usage.

To summarize, the *AdaptiX* framework contributes the following:

- *AdaptiX* allows researchers to rapidly design and test novel visualization and interaction methods.
- The framework integrates an initial concept and implementation of a shared control approach.
- The integrated ROS interface facilitates connection to a non-simulated – physical – robotic arm to perform bidirectional interactions and data.
- The framework's concept enables a code-less trajectory programming by hand-guiding the simulated or physical assistive robotic arm to the specific location and saving the position and orientation of the Tool Center Point (TCP).
- Recording TCP data enables replaying user-controlled robot movements and results in a fully customizable system. Options include changing specific details during replaying, such as repositioning cameras or re-rendering background scenes.
- Finally, the entire continuum of Mixed Reality (MR) can be exploited in the *AdaptiX* environment. This allows applications in Virtual Reality (VR), pure screen space, Augmented Reality (AR), simultaneous simulation and reality, and pure reality (cf. the *virtuality continuum* of Milgram and Kishino [41]).

2 RELATED WORK

While robotic arms are a particularly useful and versatile subset of assistive technologies, their widespread success is limited by a number of design challenges concerning the interaction with their human user. In recent years, a growing body of research addressed these concerns and associated optimization options to increase their usability, e.g., [12, 20, 34]. During the *AdaptiX* development process, we aimed to include functionality to address the challenges of shared control optimization [19], intent communication [45], and attention guidance [48].

2.1 Shared Control for Assistive Robots

Current shared control systems operate along an autonomy continuum, respectively balancing user input and system adjustments. At one extreme, the systems tend to be heavily manual, with only minor adjustments to the user's input [56]. On the other end are systems where users primarily provide high-level commands for the robot to execute [60]. A number of different approaches – including time-optimal [21] and blended mode switching [16], shared-control-templates [52] and body-machine-interfaces [29] – are currently employed in various settings.

[1] *AdaptiX* framework. https://adaptix.robot-research.de, last retrieved May 20, 2024.

A fundamentally different approach is the shared control system proposed by Goldau and Frese [19]. Their concept combines a robotic arm's cardinal DoFs according to the current situation and maps them to a low-DoF input device. The mapping is accomplished by attaching a camera to the robotic arm's gripper and training a Convolutional Neural Network (CNN) by having people without motor impairments perform ADLs [19] – similar to the learning-by-demonstration approach for autonomous robots by Canal et al. [7]. The CNN returns a set of newly mapped DoFs, ranked by their assumed likeliness based on the CNN for the given situation, allowing users to access a variety of movements for each situation. In addition, the CNN-based approach allows the system to be easily extendable as the same system can be trained to discriminate between many different situations – making it a viable concept for day-to-day use. Goldau and Frese [19] conducted a proof-of-concept study comparing the control of a simulated 2D robot with manual or CNN-based controls. Task execution was faster with their proposed concept; however, users experienced it as more complex than manual controls [19].

Our framework *AdaptiX* is influenced by Goldau and Frese's approach, but extends it from 2D to 3D space. This increases the number of possible DoFs, which allows for an accurate representation of ADLs in the framework. By adding functionality, visualizations, and a ROS integration, *AdaptiX* can be used to develop and evaluate novel interaction control methods based on this approach for shared control, which we refer to as *Adaptive DoF Mapping Control (ADMC)*.

2.2 Robot Motion Intent

Regardless of the specific interaction details, it is necessary to effectively communicate the intended assistance provided by the (semi-)autonomous system [4]. Clear communication between robots and humans enhances the shared control system's predictability, avoids accidents, and increases user acceptance.

A crucial element of the D&D process of robotic devices is, therefore, the testing of intent communication methods. *Choreobot* – an interactive, online, and visual dashboard – proposed by van Deurzen et al. [61] supports researchers and developers to identify where and when adding intelligibility to the interface design of a robotic system improves the predictability, trust, safety, usability, and acceptance. Moreover, Pascher et al. [45] provide an extensive overview of the various types of visualization and modalities frequently used in communicating robot motion intent. These range from auditory [10] and haptic [9] modalities to anthropomorphizing the robot and using its gaze [38] or gestures [18]. Their findings are substantiated by Holthaus et al. [24], who used an ethnographic approach to derive a comprehensive communication typology.

While all these intent communication modalities are viable, visual representations of future movements are often quoted as less workload-intense for the end-user [13]. AR is, therefore, unsurprisingly a frequently used tool to convey detailed motion intent [8, 22, 53, 63, 65], allowing interactions to become more intuitive and natural to humans [36]. Suzuki et al. emphasize the benefits of AR-based visualizations for communicating movement trajectories or the internal state of the robot [58].

The visual feedback employed by *AdaptiX* mimics AR in a VR environment with directional cues registered in 3D space. This approach allows the user to understand different movement directions for the actual control and the suggested DoF combinations. To streamline understanding the control methods, one of our primary approaches is the usage of arrows – a straightforward and common visualization technique to communicate motion intent [54, 55, 63].

2.3 Feedback Modalities for User Attention Guidance

When creating systems using shared control, it is crucial to guide the user's focus to the assistance the robot is offering [49]. This guidance is particularly important if either party is moving the robot in a way that could lead to collisions or worsen the situation. To enhance the predictability of shared control systems, various feedback modalities have been proposed to guide user attention as a secondary feedback mechanism to AR. The goal is to provide a feedback solution that results in short reaction times, enabling users to quickly direct their focus to the information provided by the robot.

In the related discipline of autonomous driving systems, if the vehicle encounters a situation it was not programmed or trained to handle, it will issue a Take-Over-Request (TOR). This TOR prompts the driver to take manual control of the vehicle to prevent a collision or to drive in areas the vehicle cannot handle autonomously.

Auditory, visual, and tactile/haptic modalities are commonly used for TORs [64] – either as a single sensory input [49] or a combination of multiple variants [48]. Simulation studies, along with research on reaction times to different sensory stimuli, indicate that multi-modal feedback results in the lowest possible reaction times in shared control systems [5, 14, 31].

Implementing these feedback methods into existing assistive robot systems would be straightforward as the necessary output devices – like screens, speakers, or vibration motors – are commonly already present. To allow researchers to evaluate the benefits of the different modalities, *AdaptiX* includes three modes for attention guiding: visual, auditory, and tactile/haptic. Developers can either choose one modality or follow a multi-modal approach.

3 FRAMEWORK CONCEPT

The *AdaptiX* XR framework facilitates the development and evaluation of HRI shared control applications in an easy-to-use, high-resolution transitional MR environment. Equipped with a VR simulation environment containing a virtual *Kinova Jaco 2* and ample customization options, researchers can streamline their D&D process while simultaneously reducing overhead and boosting efficiency. Figure 2 provides an overview of the framework's architecture.

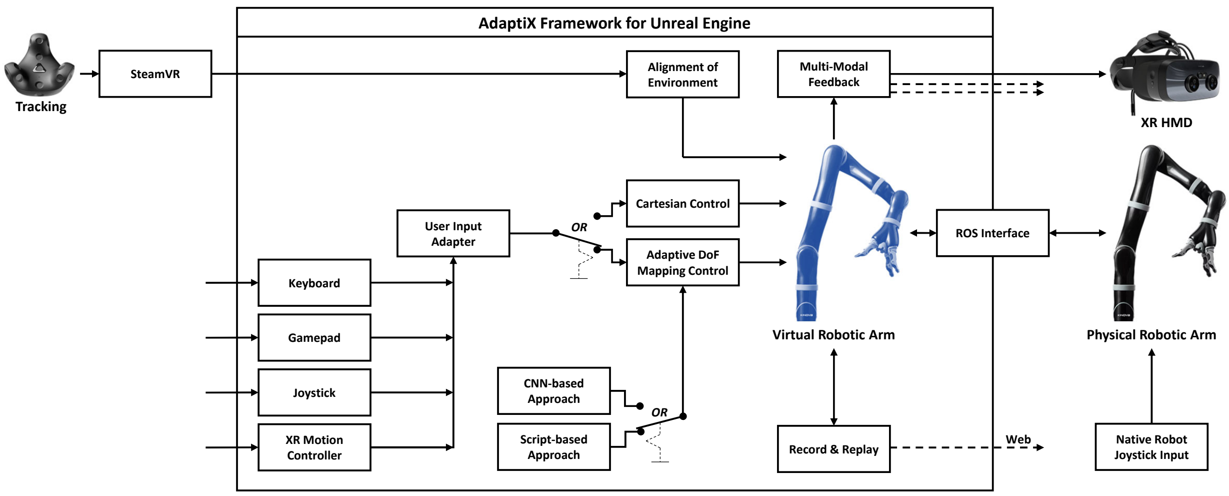


Fig. 2. Overview of *AdaptiX*' architecture, illustrating each component, their directional communication, and the crossover from and to the framework. The user input is either used for *Cartesian Control* or Adaptive DoF Mapping Control (ADMC). For ADMC, either a CNN-based or script-based rule engine can be selected.

In addition to an Cartesian robot control, we propose ADMC as an initial shared control approach, using suggestions by a rule engine (e.g., a CNN or script-based approach) to be controlled by the user. ADMC is implemented directly into the *Unreal Engine* to enable researchers and developers to fully customize the control methods, systems behavior, and feedback techniques by coding in *C++* or *Blueprints*.

AdaptiX supports several pre-implemented input devices and provides an adapter class for an easy development and implementation of further input devices. This supports researchers and developers to easy implement their ideas and concepts. The integrated ROS interface facilitates connection to a non-simulated – physical – robotic arm to perform bidirectional interactions and data exchange in a *DigitalTwin* and *PhysicalTwin* approach.

AdaptiX enables effortless trajectory programming by manually guiding the TCP of a simulated or physical robotic arm to a desired location and recording its position and orientation. Recorded data of user-controlled robot movements can be replayed. Offering the adjustment of specific details, such as camera positions and background scenes, results in a highly customizable system.

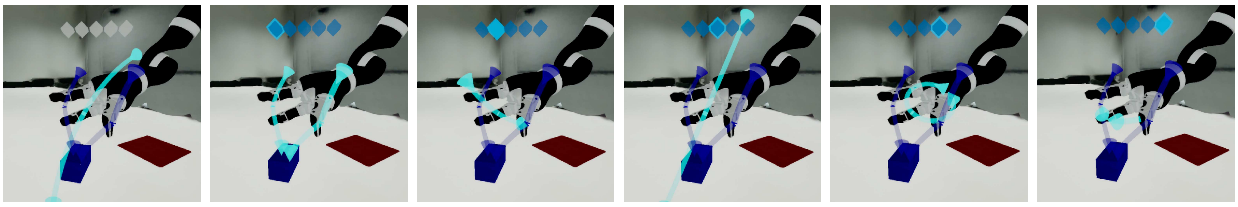
The aim is to provide a modular and extensible framework so that research teams do not need to start from scratch when implementing their shared control applications.

3.1 Adaptive DoF Mapping Control (ADMC)

For the adaptive DoF mapping – referred to as *ADMC* – of the robotic arm, the goal is to present a set of DoF mappings ordered based on their effectiveness in accomplishing the pick-and-place task used in the experiment. The concept of "usefulness" assumes that maximizing the cardinal DoFs of the robot assigned to an input-DoF while progressing towards the next goal is the most advantageous option.

This DoF mapping, referred to as the *optimal suggestion*, is assumed to be the best choice due to a significant reduction in the need for mode switches when multiple DoFs are combined into a single movement. The more DoFs are combined (assuming it is sensible for the given situation), the fewer mode switches are required. As a result, the DoF mappings are ordered based on the number of DoFs they combine.

In addition to the optimal suggestion, the second suggestion is a selection of an orthogonal variation of the first suggestion, which has the highest probability and most variation in spatial direction and keeps the number of combined DoFs unchanged. This secondary suggestion is likely useful to users as they can utilize it to adjust their position while maintaining a sensible orientation toward the next goal. The following DoF mappings were used (see Figure 3):



(a) (b) (c) (d) (e) (f)

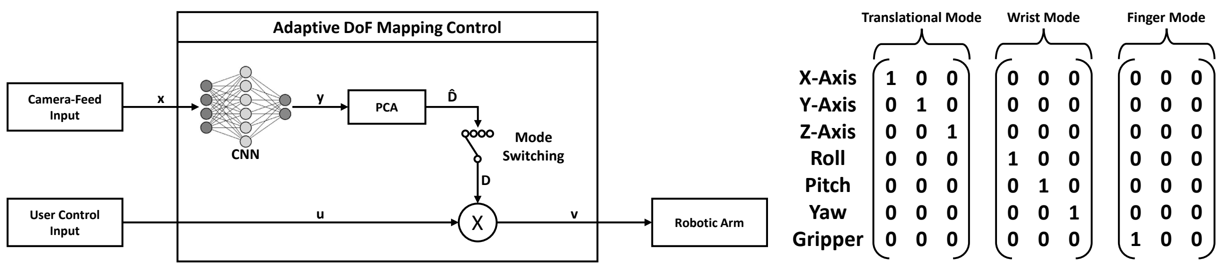
Fig. 3. Suggestions as Visualized in the ADMC, (**a**) Continue previous movement, (**b**) Optimal Suggestion, (**c**) Adjustment Suggestion, (**d**) Translation Suggestion, (**e**) Rotation Suggestion, (**f**) Gripper Suggestion. Colors: Bright cyan arrow: Currently active DoF mapping. Dark blue arrow: Next most likely DoF mapping.

(1) *Optimal Suggestion:* Combining translation, rotation, and finger movement [opening and closing] into one suggestion, causing the gripper to move towards the target, pick it up, or release it on the intended surface.

(2) *Adjustment Suggestion:* An orthogonal suggestion based on (1) but excluding the finger movement. Allows the users to adjust the gripper's position while still being correctly orientated.
(3) *Translation Suggestion:* A pure translation towards the next target, disregarding any rotation.
(4) *Rotation Suggestion:* A pure rotation towards the next target disregarding any translation.
(5) *Gripper Suggestion:* Opening or closing of the gripper's fingers.

3.1.1 CNN-based Approach. For the CNN approach, a color-and-depth camera is attached to the gripper of an assistive robotic arm. The live video feed is transmitted to a CNN, which is trained using data collected from non-impaired individuals performing ADLs using the robotic arm along with a high-DoF input device. The CNN does not need a model of the environment to provide these mappings. Principal Component Analysis (PCA) is employed to transform the CNN's output into a matrix $\hat{D}$, where each column represents a combination of cardinal DoFs along which the robotic arm can move.

Next, a subset of $\hat{D}$ is selected, containing as many columns as the number of DoFs provided by the input device. This selected subset is referred to as D, and it serves to map input-DoFs to output-DoFs. When an input-DoF is engaged, the robot's movements are determined by the values in the corresponding vector of D, which proportionally activate the robot's cardinal DoFs. A mode switch is defined as the exchange of D with a different subset of $\hat{D}$. This enables the system to switch between various mappings of input-DoFs to output-DoFs, adapting the robot's control according to the user's needs and preferences. A visual representation of this control pipeline is depicted in Figure 4a.



(a) (b)

Fig. 4. Concept of adaptive DoF mapping control. **(a)** Control pipeline for proposed adaptive shared control and **(b)** matrix representation of DoF mappings: Columns represent input-DoFs. Rows represent output-DoFs. Subsets represent modes. Two empty columns were added to represent zero movement mappings in *Finger Mode*.

$\hat{D}$ is a square matrix with dimensions based on the number of cardinal DoFs available on the robot to be controlled. In the case of the *Kinova Jaco 2* [30], this results in a 7×7 matrix. This matrix represents a mapping of input-DoFs to output-DoFs when the number of DoFs in both cases is equal. The values in each column, ranging from -1 to 1, indicate the proportion with which the specific cardinal DoF is utilized when engaging the corresponding input-DoF.

By defining $\hat{D}$ as an identity matrix, each input-DoF is mapped to a single output-DoF. Selecting an equal number of columns from $\hat{D}$ to form matrix D allows for manual control with mode switching along cardinal DoFs. Moreover, this representation enables the combination of multiple cardinal movements into arbitrary output DoF mappings. For example, a (transposed) column of *(0.5, 0.5, 0, 0, 0, 0, 0)* would result in diagonal movement along the X- and Y-Axes of the robot. Such combinations enable the offering of complex movements with different proportions depending on the situation, enhancing the control options available to users. The identity matrix for a *Kinova Jaco 2* with a 3-DoFs joystick is illustrated in Figure 4b.

3.1.2 Script-based Approach. As an alternative rule engine for our ADMC concept, we implemented a task-specific script. This approach eliminates potential biases that a more generic, but currently limited method like a CNN-based control might introduce. It is essential to note that our task-specific script is effective only in a controlled experimental environment.

The task-specific script assesses the end effector's current position, rotation, and finger position relative to a target, allowing it to adaptively calculate the matrix $\hat{D}$. This script recommends optimal movements to pick up an object and place it onto a target drop area, maximizing the combination of as many DoFs as possible. Additionally, it provides other DoF combinations that may be less beneficial to mimic the idea that each subsequent column in $\hat{D}$ has a decreasing likelihood of being useful. These additional DoF mappings are ordered by the number of combined DoFs in a decreasing manner.

To validate the effectiveness of this approach, we conducted pilot tests, comparing it to a *Wizard-of-Oz* method. In this scenario, a human "simulated a CNN" to explore user interaction with such a system.

3.1.3 Point of Time to Communicate the Suggestion. Our ADMC concept uses an adaptive DoF mapping system to recommend DoF mappings to the users depending on the current situation. The system visualizes the currently active DoF mapping as a bright cyan and the suggestion as a dark blue arrow (see Figure 3). This suggestion can be communicated – based on the the configuration – either continuously or only if the next most likely movement direction differs from the currently active DoF mapping by a certain threshold.

To calculate this threshold – the difference between the currently active and new most likely DoF mapping –, *cosine similarity* [57] is used, ranging from exact alignment [0%] to total opposite direction [100%]. The formula for cosine similarity of two n-dimensional vectors is defined as:

$$\text{cosine similarity} = \cos\left(\vec{a}, \vec{b}\right) = \frac{\vec{a}\vec{b}}{\|\vec{a}\|\|\vec{b}\|} = \frac{\sum_{i=1}^{n} a_i b_i}{\sqrt{\sum_{i=1}^{n} (a_i)^2}\sqrt{\sum_{i=1}^{n} (b_i)^2}} \tag{1}$$

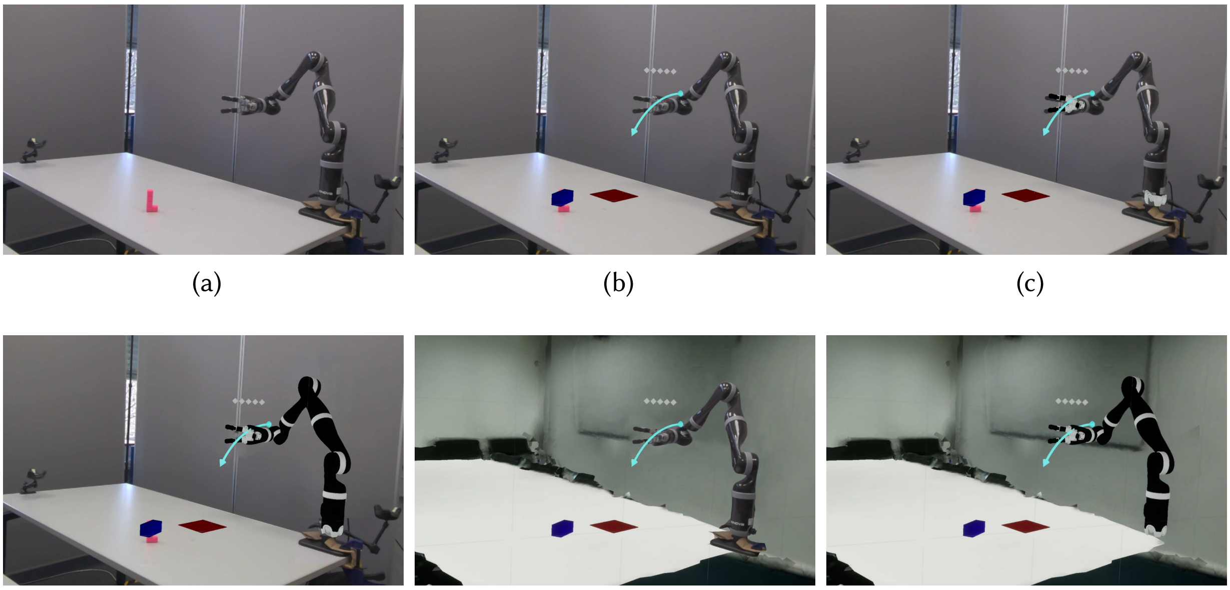
To implement a difference value, the cosine similarity needs to be transformed. As a cosine similarity of -1 indicates completely opposed vectors, the difference value needs to return 1 – i.e. the maximum possible difference – for a cosine similarity value of -1. A cosine similarity of 1, indicating exact similarity, should return a difference value of 0 – i.e. no difference. Perpendicular vectors with cosine similarity 0 should return a difference value of 0.5 – i.e. a 50% difference. To calculate the difference value **d**, the following formula is used:

$$\text{difference d} = 1 - \frac{\cos\left(\vec{a}, \vec{b}\right) + 1}{2} \tag{2}$$

This difference value represents the difference between two vectors. While the user moves the robot with an active DoF mapping, the adaptive DoF mapping system reevaluates the situation and calculates new suggested DoF mappings. The default difference value is set to 0.2 (20% difference between currently active and new most likely DoF mapping).

3.2 Full Mixed Reality Continuum

In our framework, we created an environment in which the entire continuum of MR is exploitable. This extends the use of *AdaptiX* to new scenarios and environments – including the real world. The



(a) (b) (c) (d) (e) (f)

Fig. 5. MR continuum with (**a**) only the real robotic arm in real environment, (**b**) augmenting of directional cues in the real environment with the real robotic arm, (**c**) additional visualizing the gripper and base of the virtual robotic arm in the real environment, (**d**) visualizing the simulated robotic arm in the real environment, (**e**) visualizing the real robotic arm in the virtual environment, and (**f**) visualizing the simulated robotic arm in the virtual environment.

virtual and real environments of the robotic arm are aligned, allowing researchers to seamlessly switch between the user controlling the real and virtual robot. The level of MR can be adjusted in various steps (cf. the *virtuality continuum* of Milgram and Kishino [41]).

The MR environment setups include:

(1) the completely real environment with the real robotic arm,
(2) the real environment extended with visual cues,
(3) the real environment into which the virtual robot is transferred and displayed (with and without visual cues),
(4) the virtual environment into which the real robot is transferred and displayed (with and without visual cues),
(5) the completely virtual environment with the virtual robotic arm.

A comparison of the user's view in reality and simulation can be seen in Figure 5. MR continuum level (1) is suitable for study baseline-condition, without any multi-modal feedback to the user. In level (2) an AR visualization technique is mimicked, showing the whole physical setup augmented by basic cues. Especially level (3) and (4) enable customizing either the robot itself or the environment to extent/exchange the physical setup but still not loosing the context. In (3) users can interact with a totally new or customized robot while being in a familiar environment. World's distractions can be excluded in (4) while the the original robot is presented. Finally, level (5) provides a VR environment that can be fully customized.

3.3 Interfaces

We designed *AdaptiX* to facilitate the comparison of different interaction designs, intervention strategies, and feedback techniques for shared robot control. The initial version of the framework includes interface types for extending user input, ROS integration, and multi-modal feedback. However, this baseline can easily be customized and extended by future development.

3.3.1 User Input. We provide a standard control approach where pressing a keyboard button moves the end effector along cardinal DoFs (x, y, z, roll, pitch, yaw, opening and closing the gripper). Using build-in functionalities, the designated keyboard input can easily be adjusted to other input devices like gamepads, joysticks, or customized assistive input appliances.

In contrast to tele-operating the robotic arm, a *follow-me* approach for any trackable object in 3D space – e.g., the user's handheld VR motion controller – was implemented. The robot's end effector directly follows the movement of the trackable object, which corresponds functionally to direct control. This can be used to generate high-dimensional input and record intended behavior quickly, providing an easy way of interacting and controlling the robot, especially for inexperienced users.

3.3.2 ROS Integration. The ROS integration allows for a bidirectional exchange of information between the simulation and a real robot, mirroring the robot's state *in-silico* and vice versa. Figure 6 shows the involved components: a ROS bridge facilitates the multi-device connection between the framework and the real robot while exchanging robot data. On the ROS side, the messages for the arm position and orientation control and the values for the angle-accurate control of the gripper fingers are read in via the ROS subscriber node. They are then processed, and the robot arm and gripper are controlled through our action client. In addition, the joint angles, the TCP, and the position of all three gripper fingers are published via ROS, which are then input by our *Unreal Engine* framework. The virtual and real robots are synchronized via ROS every 0.1 seconds.

Based on this, our framework provides – depending on the specific context – both a *DigitalTwin* and *PhysicalTwin* approach, allowing the control of either with the other.

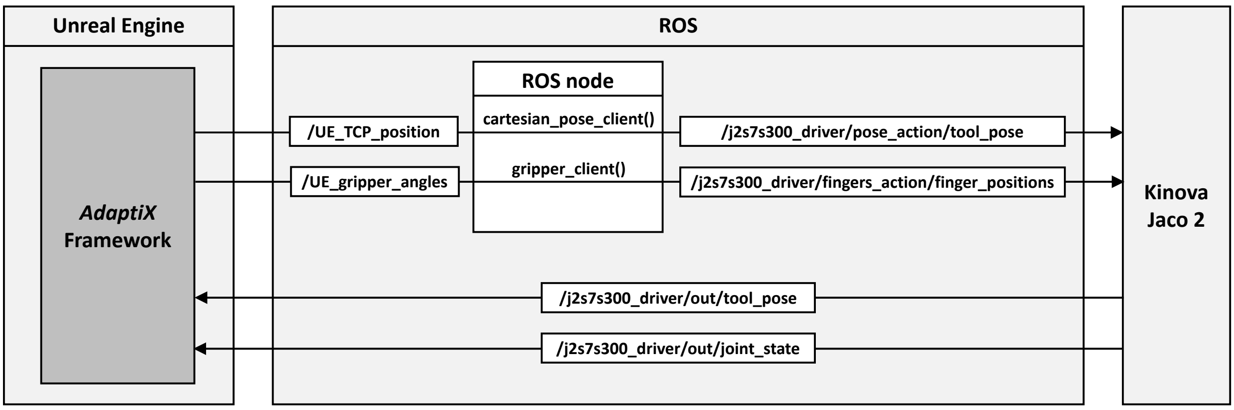


Fig. 6. Component connections of the ROS interface for mixed reality.

3.3.3 Multi-Modal Feedback. To communicate any combination of DoFs, our framework supports several visual cues to illustrate the intended movement trajectory and provides multi-modal feedback extensions via audio and haptic-tactile feedback. Visual feedback can be either provided dynamically attached to the virtual/physical robot's end effector, stationary in the world, or attached to the user's view.

AdaptiX aims to support the development of novel multi-modal interaction and feedback designs either in the pure VR simulation testbed environment or by interacting with a real robot in MR, which mimics an AR setting due to the stereoscopic video-feed. Moreover, it is also possible to show the real robot in our VR simulation environment instead of the simulated one.

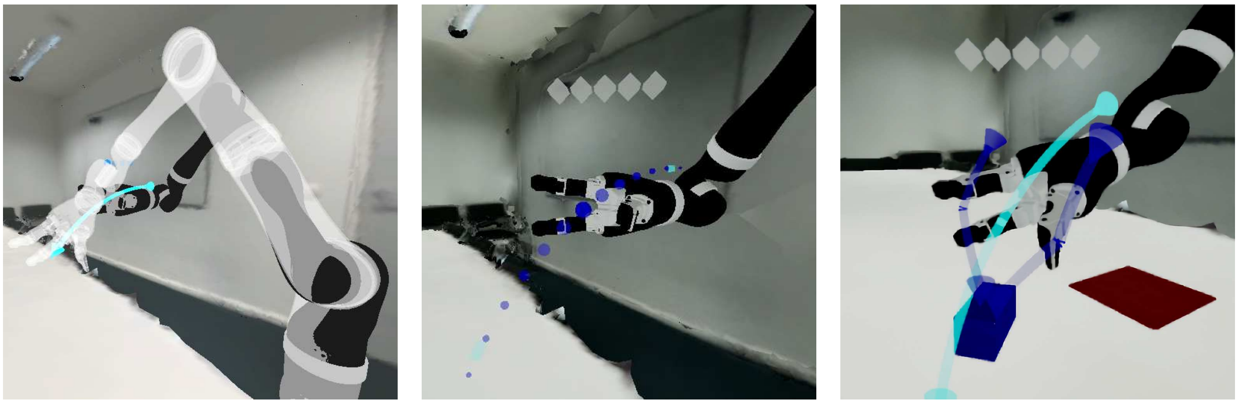
Figure 7 shows three exemplary AR-style visualizations provided by the framework, including (a) a robotic ghost overlay, (b) discrete waypoints in 3D, and (c) a variety of multidimensional arrows. Though varying in design, these visualizations can effectively communicate the robot's motion intent to the user.

Ghost: A visualization of robot motion intent by showing an additional version of the robot (or specific components) registered in 3D space, in another color and/or opacity. These visualizations communicate the exact position and orientation a robot at a given time, behaving precisely as though the real robot had been moved this way.

Waypoints: This visualization technique augments the position of a robot (or in our case, the gripper of the robotic arm) in 3D space at a certain point in the future. Usually, the robot navigates linearly between these *Waypoints*, which increases predictability.

Arrow: Among visualizations arguably the most basic but certainly also the most familiar (as seen in traffic navigation systems, road signs, and on keyboards). *Arrows* are found both in straight and curved varieties, where curved arrows indicate a rotation. Given the abundance of *Arrows* in daily life, it makes sense that many robot motion intent visualizations use them.

Classic: This visualization also uses *Arrows*, but in our prototype they are used as a baseline condition to evaluate adaptive and non-adaptive controls. Here, as with the standard input device *Kinova Jaco 2*, two axes can be controlled simultaneously and the user has to choose between different translations and rotations by mode-switching.



(a) Ghost (b) Waypoints (c) Arrows

Fig. 7. Visualization examples pre-implemented in the framework.

All interfaces are modular, enabling quick adaptations and switching between variations. This flexibility allows for studies with clean methodologies and easy comparisons without additional overhead. The community is invited to extend the implementations with any interfaces or control methods desired for their research.

3.4 Recording and Replay

AdaptiX contains an easy-to-use general-purpose system to record, store and replay simulation data, including detailed information about robot states, execution times, or the states of various objects in the environment. The recording system generates Comma-separated values (CSV) text files, which can be accessed with any data manipulation software (e.g., Python or MATLAB). The added output functionality differs significantly from the replaying system provided by the underlying *Unreal Engine*, which is mainly designed for visual replays and – among other things – does not support a CSV file format.

In addition, *AdaptiX*'s recording and replaying system is entirely customizable. Camera re-positioning or re-rendering background scene options are included in the initial version. By default, the recording system tracks the user's view, the robotic arm, and all moveable actors in the virtual environment. All other objects are assumed to be stationary, thus part of the level, and ignored as such. This approach allows for the randomization of background scenes by re-rendering.

The system stores the assigned virtual meshes, scales, possible segmentation tags for each tracked object, and the complete pose data per frame. During the replay process, all objects that were initially recorded in a specific level are swapped with the corresponding data stored in the loaded recording. However, if a different scene is being loaded, the objects from that scene are used instead. In every subsequent frame, all objects are positioned at their respective position until the loaded recording has finished. The system permits custom code to be run at the end of each loaded frame, thus enabling de-bugging and data rendering during replays.

Overall, *AdaptiX* facilitates the lightweight storage of recordings as CSV files with the option to render and store complex and large-scale data (e.g., images or videos) for subsequent evaluation. This lightweight approach is particularly useful when deploying experiments on external devices or recording extensive datasets.

4 FRAMEWORK IMPLEMENTATION

The *AdaptiX* simulation environment is based on the game engine *Unreal Engine 4.27* [15]. The advanced real-time 3D photoreal visuals and immersive experiences provide a suitable foundation for our framework, and assets for future extensions are readily available. *Unreal Engine 4.27* includes integrated options for various hardware setups, thus enabling the framework to be deployed on different operating systems while utilizing most currently available VR/MR/AR headsets, gamepads, and joysticks. At the time of writing, *Unreal Engine 4.27* is free to use, has a considerable user space, and allows unrestricted publications of non-revenue generating research products like the *AdaptiX* framework. Detailed implementation descriptions can be accessed in the *README* provided in the repository at https://adaptix.robot-research.de.

Fig. 8. Example scenario provided in *AdaptiX* including a table, a virtual *Kinova Jaco 2* robotic arm and colored blocks on the tabletop.

4.1 Simulation Environment

The *AdaptiX* default scenario centers on the photogrammetry scan of an actual room that contains a table with an attached virtual robotic arm (see Figure 8). A simulated camera is mounted on the arm's gripper. We added a toggle-off option to hide the camera from the user's view.

The framework includes a straightforward testbed scenario for pick-and-place operations, mimicking the basic principles of most ADLs. The simulation centers around a red surface as a drop target and a blue block as the to-be-manipulated object. Once the object has been successfully placed, the setup randomly re-positions the blue block on the table surface, and the task can be repeated.

We optimized the robotic arm simulation for operation via a VR motion controller with an analog stick, several playable buttons, and motion capture capabilities (e.g., *Meta Quest 2* [39]). These options provide a workable foundation to implement and test diverse interaction concepts, including adaptive concepts which can be configured to match the individual physical abilities of the intended user.

By incorporating the *Varjo XR-3* [62] – a particularly high-resolution XR-Head-Mounted Display (HMD) – we implemented a transitional MR environment. Using two *HTC VIVE* trackers [26], the virtual and real worlds are synchronized so that the robots' working areas are identical. By including the *HTC VIVE* motion controller [25], it is then possible to control the physical robot directly via the *PhysicalTwin* approach of *AdaptiX* (see Figure 1).

The virtual robotic arm is designed as a modular entity, allowing easy integration to new levels following the *Unreal Engine*'s *ActorBlueprint* class structure.

4.1.1 Simulated Robotic Arm. The commercially available *Kinova Jaco 2* assistive robotic arm [30] is specifically designed as an assistive device for people with motor impairments. It is frequently used by a) the target audience and b) researchers – e.g., [3, 21] – during HRI studies, hence the suitability for inclusion in *AdaptiX*.

We designed the simulated *Kinova Jaco 2* as close as possible to the actual product, using virtual meshes generated directly from computer-aided design (CAD) files provided by the manufacturer. Much like in reality, the virtual arm consists of a series of individual links connected by angular joints as shown in the annotated rendering of the assembled model Figure 9.

As *AdaptiX* – including the operation of its simulated robotic arm – is optimized for HRI studies, it focuses on user interaction rather than low-level robot control, whilst also able to incorporate those. Hence, rather than following the standard base-up control, the simulated arm moves in reverse: the user's input directly controls the end effector's motion; the connected joints are positioned to connect the end effector with the base.Each intermediate joint is modeled as a dampened spring with the links unaffected by gravity. This also resolves the redundancy, i.e., joint angle ambiguity a 7-jointed robot has.

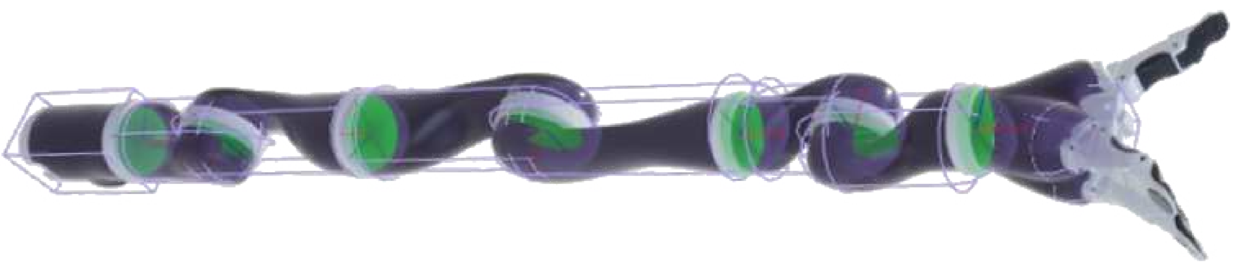


Fig. 9. Virtual Robotic Arm with Physics Constraints: purple capsules represent links, green discs represent angular constraints.

This approach allows for nearly arbitrary motion of the end effector and a semi-realistic interaction of the arm with the environment. As a beneficial side effect, developers can disconnect the end effector from the rest of the arm and allow the user to control a free-floating robot hand without any constraints. However, the internal physics engine to realistically handle collisions and interactions between the end effector and the environment is still active.

Likewise, we based the grasp concept on a custom interaction design for robotic grasping rather than physics. Physics-based grasping in a virtual environment is a challenging task [27] and would require substantial preparation and asset fine-tuning from future developers who use the framework. Instead, we defined a logic-based approach that we consider sufficiently realistic for shared control applications: an object is regarded as grasped once it has contact with two opposite fingers while closing the gripper until the fingers open again. The grasped object is rigidly attached to the end effector, keeping its relative position stable and moving alongside the end effector until released.

4.1.2 Simulated Camera System. Computer-aided robot control usually requires a camera system – or a comparable sensor – to measure context information about the current environment for the underlying software function. To provide a realistic equivalent in simulation, *AdaptiX* contains a virtual version of the commercially available *Intel Realsense D435* [28]. This camera system is commonly used in research applications [11, 66] and can deliver aligned color and depth images. The built-in color sensor generates depth data by applying a stereo-vision algorithm using grayscale image data of two built-in infrared (IR) imagers. To improve the texture information captured by the IR imagers, the camera also includes an IR projector, which projects a static pattern on the scene.

As with the simulated robotic arm, the virtual camera system is a modular actor that can be arbitrarily placed within the simulation environment. Its mesh and texture are derived directly from the manufacturer's CAD files to optimize authenticity. The virtual camera system includes all image sensors of the original, plus an optional virtual sensor generating a segmented image of the scene. We designed the virtual sensor parameters to be as close as possible to those of the actual sensors. They include – but are not limited to – sensor dimensions, lens structure, focal length, and aperture.

Because the framework can provide depth information directly from the 3D simulation, the virtual depth camera does not need to calculate its data using stereo-vision but instead yields perfect per-pixel depth information. If stereo-vision-generated depth data with realistic noise, errors, and other algorithm-specific effects is needed, the virtual system also delivers the IR images for a manual calculation.

Additionally, the simulated camera system supports the usage of the image data in-simulation and storing the data on disk for applications such as dataset generation or logging.

4.2 Adaptive DoF Mapping Control (ADMC)

The adaptive DoF mapping is implemented in the object *Axis Wizard*, which provides functions to calculate the optimal suggestion, as well as the other possible optimizations. The calculation relies solely on the virtual objects in the simulation environment instead of object recognition or camera data to enable development and evaluation without a physical robot setup. However, the camera feed for object recognition can be activated by developers to read positions and orientations. In addition to the positions and orientations of the *Gripper Mover* and the *Current Target* (which can be an object to pick up or a target surface to place the object on, depending on the context), two other parameters of *Axis Wizard* are important to ensure the correct calculations for the pick-and-place task – *Minimal Hover Distance* and *Hover Height*.

Disregarding the handling of edge cases, the calculation of the optimal suggestion is taken care of in three steps: 1) calculating *Translation*, 2) calculating *Rotation*, and 3) calculating the finger movement variable *Gripper*. The Blueprints for implementation details are provided in Appendix A.

4.2.1 Calculation of the Optimal Suggestion. Minimal Hover Distance represents the distance – projected on the XY-plane – between the *Gripper Mover* and the *Current Target*. When this distance is smaller than the *Minimal Hover Distance* (see Figure 12 in the appendix), the *Axis Wizard* uses a point above the *Current Target* for its calculations – referred to as the *Target Point*, instead of the *Current Target*'s position to prevent the robot from getting too close to the table, allowing for proper gripper rotation. Then, a vector from the *Gripper Mover*'s position towards the *Target Point* is calculated, normalized, and inversely rotated by the *Gripper Mover*'s rotation. This calculation returns a unit vector pointing from the *Gripper Mover* toward the *target point* in the *Gripper Mover*'s reference frame. This vector is then scaled by the *Vel Trans* value of the *Kinova Jaco 2* to get a translation of the size of the movement performed by the *Kinova Jaco 2* during one frame.

Hover Height determines the height of the aforementioned point above the *Current Target*. If the XY-projected distance between the *Gripper Mover* and the *Current Target* is smaller than the *Minimal Hover Distance*, the *Axis Wizard* directly uses the *Current Target*'s position for its calculations instead of the point above it.

To calculate the optimal suggestion's *Rotation*, the *Translation* – calculated in the first step – is used as input for the *Make Rot from X* node. This node returns a *rotator* representing the rotation required to make an object point toward the direction indicated by the input vector – *target point*. To mitigate an additional *roll* of *Gripper Mover*, the inverse value is added, keeping the *Gripper Mover's* orientation largely steady. Additionally, since only a small part of the rotation is performed during one frame, the *rotator* is scaled down. The calculation for the *Rotation*, excluding edge cases, is depicted in Figure 13 in the appendix.

4.2.2 Calculation of Gripper values. The *Gripper* value only depends on whether the target point is within reach of the robotic fingers, either with or without additional movement (i.e. if the fingers are almost close enough, there will be a movement towards the target point, otherwise the fingers will engage without moving the gripper) and whether or not an object is currently being grasped (i.e. if an object is grasped and the gripper is close to the target point, it suggests to open the fingers, otherwise close them).

4.2.3 Calculation of the Adjustment Suggestion. The adjustment suggestion is calculated by rotating the optimal suggestion's *Translation* by 90° around the Y-Axis, keeping the same *Rotation* and setting the *Gripper* value to 0. This results in a DoF mapping which moves roughly along the *Gripper Mover's* Z-Axis, or colloquially "up and down" between the fingers if the optimal suggestion is seen as "forward and backward". As *Rotation* is kept the same between the optimal and adjustment suggestions, the resulting movement keeps the fingers roughly facing the direction of the *Current Target*.

The translation, rotation, and gripper suggestions use much simpler calculations. The translation suggestion calculates a vector from the *Gripper Mover* towards the *Current Target*, inversely rotates it by the *Gripper Mover's* rotation to put it into the *Gripper Mover's* reference frame and uses that as the *Translation* value for the suggested *Adaptive Axis*. This vector is also what the rotation suggestion uses to calculate a *Rotator* representing a rotation towards the *Current Target*. The gripper suggestion checks whether an object is currently being grasped. If so, the suggestion is to open the fingers (*Gripper* = -1). Otherwise, the suggestion is to close the fingers (*Gripper* = 1).

4.2.4 Attention Guidance in Threshold*.* Both the *Continuous* and *Threshold* approaches share the same core calculation for DoF mappings. However, the *Threshold* approach has an additional task:

determining whether the optimal suggestion significantly differs from the currently active DoF mapping. This task is more related to visualization than the DoF mapping calculation itself and is managed by the *Gizmo* object.

The *Gizmo* object contains a *Realtime Threshold* variable, which represents the threshold as a value between 0 and 1. It also includes a function called *Adaptive Axes Nearly Equal*, which determines whether two *Adaptive Axes* are nearly equal by checking if their difference is below the *Realtime Threshold*. The threshold value is chosen to be between 0 and 1 to align with a percentage of difference (see Section 3.1.3), providing a more intuitive understanding of the amount of difference compared to the cosine similarity value used as the basis for the difference calculation.

As the *Unreal Engine* does not provide an arbitrarily sized vector structure, the calculations required needed to be programmed manually rather than with built-in vector operations. Therefore, two math expression nodes were defined, one calculating the dot product of two 7D vectors and the other calculating the magnitude of a 7D vector. Using these, the cosine similarity between two *Adaptive Axes* could be calculated in *Unreal Blueprints* (see Figure 14 in the appendix). To forego the transformation of the cosine similarity into a percentage difference, the *Unreal Engine's Nearly Equal* node was used to determine whether the cosine similarity was nearly equal to *1* – meaning the vectors align – with a threshold of *2 * Realtime Threshold*. The threshold needed to be multiplied by *2* as the range of the cosine similarity has a magnitude of 2. The result of this calculation is a boolean value that is true if the difference between the *Adaptive Axes* is below the threshold and false otherwise.

The resulting value is then used by the *Gizmo* to show the arrow corresponding to the optimal suggestion. It is also used to notify the *Game Mode* – an object representing the game, keeping track of study variables, etc. – that the threshold was broken. This triggers an event that causes a 1kHz sound to play and a haptic effect to occur on the motion controller. A reset variable is used to prevent the sound from constantly triggering. However, there appears to be a specific point during movement at which it is possible for users to stop their input and the software to get caught in a loop of firing the event and resetting it, causing a constant sound and vibration. If users continued their movement, the software stopped firing the event, seizing the sound and vibration. Unfortunately, this was only noticed during the experiment, which is why the problem persists in the current software version. Assuming *Threshold* is to be used in future research, a better solution for a single fire execution of the notification needs to be developed.

5 LIMITATIONS

In HRI research, the leading factor impacting user experience is usually the chosen method of (shared) control and the respective interfaces. Using frameworks like *AdaptiX* allows researchers to tweak these variables toward high user satisfaction through methodological studies and experiments.

However, like any simulation, *AdaptiX* only approximates reality and contains ingrained limitations when working with the system and evaluating generated results.

5.1 Scenario Selection

In the initial version, *AdaptiX* provides only a single level, as seen in all screenshots of this work. This scenario functions mainly as a model for simple tasks. As such, it lacks environment interactions or varying backgrounds and is not designed for a specific assistive task.

This single level might need to be revised to represent the complete application range of assistive shared control, which is why extensions are required. As such, *AdaptiX*'s modular design allows the community to generate custom levels for their specific research interests effortlessly.

5.2 Simulation Sickness caused by Head Mounted Display

HMDs are a popular tool to create immersive virtual environments, frequently used in research and industrial settings. However, using a HMD in HRI can create a significant displacement between the virtual object and the physical world through effects related to the resulting limited field of view, reduced depth perception, and distorted spatial cues.

For applications within the *AdaptiX* framework, these issues could result in users experiencing motion sickness, disorientation, discomfort, and potentially decreased performance when interacting with the simulated robotic arm or virtual objects. Researchers must consider these artifacts when designing experiments, especially when developing studies including qualitative questionnaires or when comparing different levels of MR continuum.

5.3 Simulation Environment

The simulation environment centers on the photogrammetry scan of an actual room that contains a table with an attached virtual robotic arm. Compared to a 3D modeling of a room, the photogrammetry does not provide a high resolution, leading to a partial blurred appearance.

AdaptiX does not provide a photo realistic virtual environment (yet). However, in our studies, the slightly blurred appearance never seemed to have had a negative effect. On the contrary, it has helped participants focus on the scene's relevant parts (i.e. the robot and objects). Researchers and developers are invited to create and evaluate a 3D modeled environment.

5.4 Simulated Robotic Arm

If controlled entirely in simulation, the robotic arm (as described in Section 4.1.1) does not move identically to an actual *Kinova Jaco 2* because of implementation decisions favoring physical interactions over accurate per-joint robot actions. In most other cases, the individual joints are in relatively realistic positions, even though they might not be identical to the underlying solution provided by an inverse kinematic of the real robot.

Especially in the *follow-me* approach (see Section 3.3.1), it is possible to reach outside of the mechanical range of the robotic arm. Due to the entirely physics-based connection, this results in partially disconnected joints. However, this is only an issue of visualizing the robotic arm in the simulation environment and does not affect the control or the TCP data recording.

Likewise, grasping simulated objects is based on a custom implementation, and grabbed objects are firmly attached to the end effector. Care must be taken for objects that are – in reality – too heavy for the gripper, have slippery surfaces, or have mechanical dimensions that make the object unstable when held. Theoretically, this "ideal kind of grasping" allows the virtual robot to move any arbitrarily large and heavy object. To address this, we added the object tag *Graspable* that allows developers to define permitted – and by omission – unpermitted objects.

5.5 Simulated Camera System

Although the simulated camera is based on manufacturer CAD files, comparison tests failed to deliver completely identical data to the actual recording system. These variances stem from environmental differences between simulation and reality, as light or dust/other particles in the air will cause effects in the produced image. However, these effects can be added in post-production or – if required – activated in the framework. By default, the respective settings are disabled as they would primarily introduce noise that not every developer might want.

On a technical level, the images generated by the virtual system differ slightly in terms of data types. The virtual grayscale IR images consist of three identical color channels instead of a single channel in reality. Also, the virtual IR and color images include an additional fourth alpha channel, which is not used in our framework. The generated depth data format also differs, as the actual camera system generates images as 16-bit unsigned integer, and the simulation provides them as 16-bit signed floats. The depth data generated by the framework is pixel-perfect, which ignores various camera system effects that occur in reality by the calculation of depth using stereo-vision.

All these technical differences are addressed within the framework through data transformation and should not noticeably affect the output of *AdaptiX*. However, researchers and developers should be aware of these adjustments for future developments and extension.

5.6 ROS Interface

The ROS interface connects the virtual with a real robot, each with its own environmentally-determined set of limitations. This results in some logical inconsistencies while using the interface. The obvious velocity limitations of the real system result in delayed execution if reality is to follow the simulation. Therefore, the maximum velocity of the virtual robotic arm is set automatically to the physical characteristics after enabling ROS. Also, as the virtual joints are not controlled by an inverse kinematics (IK) but instead based on physics, the interface sends only end effector poses to the real robot, omitting individual joint poses. This may result in differing robot configurations, with only the end effector point being aligned in some instances.

When sending pose data from the real robot to the virtual twin in simulation, most of these restrictions do not apply. The simulated robot can move arbitrarily fast, and its configuration aligns automatically with the real system. The only restriction is that, by default, no further information about the natural environment is available, resulting in a relatively empty virtual environment if relying purely on the ROS interface.

When designing expansions, developers also must be aware that ROS and *Unreal Engine* differ in handedness. ROS is based on a right-handed coordinate system, while the *Unreal Engine* uses a left-handed approach. *AdaptiX* internally does the necessary transformation for the robotic arm but will not automatically calculate this for other position and orientation data, e.g., obstacles. However, researchers can mitigate this by applying the provided coordinate transformation methods of the robotic arm to any further object.

6 FRAMEWORK EXAMPLE ADAPTIONS

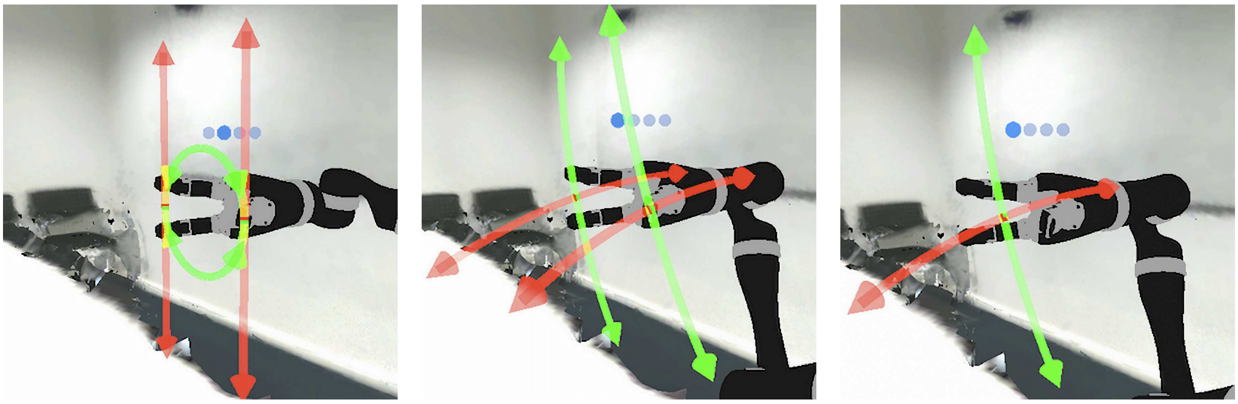
The *AdaptiX* framework has been successfully used and adapted in three case studies evaluating interaction concepts and multi-modal feedback with remote and laboratory-based focus groups.

6.1 Example Adaption 1: Adaptive Control of an Assistive Robot

In an initial study [32], the *AdaptiX* framework was used to explore the proposed ADMC control method with associated visual cues for various DoF mappings.

In particular, we analyzed how the novel adaptive control method – proposed by Goldau and Frese [19] – performs in a 3D environment compared to the standard mode-switch approach with cardinal DoF mappings. They also investigated whether changes in the visual cues' appearance impact the performance of the adaptive control method. Three different types of control with varying visual cues and methods of mapping DoFs were compared in a remote online study. These included the *Classic* visualization, one based on *Double Arrow* using two arrows attached to the gripper's fingers, and a visually reduced variant *Single Arrow*, using only one arrow through the middle of the gripper. See Figure 10 for a graphical comparison.

Due to the ongoing COVID-19 pandemic, the study was conducted entirely in a VR environment created by *AdaptiX*. Non-specific participants were recruited that had access to the required hardware (an *Oculus Quest* VR-HMD) for an immersive experience.



(a) Classic (b) Double Arrow (c) Single Arrow

Fig. 10. Evaluated interaction design and visualizations [32].

The participants repeatedly performed a simple pick-and-place task by controlling the virtual *Kinova Jaco 2* using one of the three control types. Comparative results established that adaptive controls require significantly fewer mode switches than the classic control methods. However, task completion time and workload did not improve. Study participants also mentioned concerns about the dynamically changing mapping of combined DoFs and the 2-DoF input device.

Framework contribution: *AdaptiX* demonstrated its effectiveness in this remote study to evaluate new interaction designs and feedback techniques. The innovative advantage is that the physical robotic device does not need to be present during these preliminary studies when testing and evaluating essential design elements. The *Record & Replay* functionality of *AdaptiX* allowed a remote analysis of participants data. This VR approach significantly increases the potential to include end-users in the research and design process while at the same time decreasing cost, time involvement, and accessibility concerns.

6.2 Example Adaption 2: Communicating Adaptive Control Recommendations

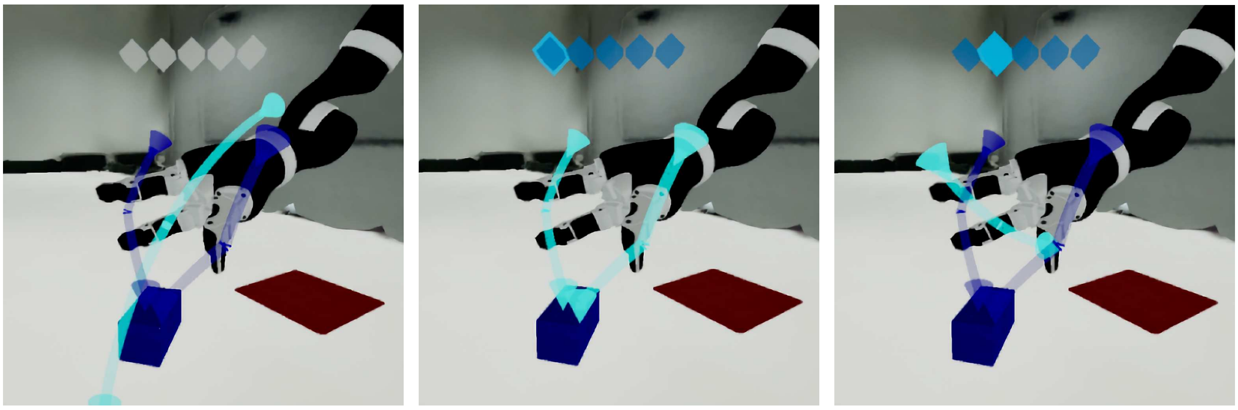
A follow-up study [46] evaluated two new adaptive control methods for an assistive robotic arm, one of which involves a multi-modal approach for attention guiding of the user.

We used *AdaptiX* in a laboratory study to cross-validate the initial study's findings on how participants interact with the environment. The adaptive system re-calculated the best combination of DoFs to complete the task during movement. These calculations were presented to the user as alternative control options for the current task. Users cycled through these suggestions – by pressing a button on the input device – to make a suitable selection or continue moving with the previous active DoFs (see Figure 11).

They contrasted the variants *Continuous* and *Threshold*, differing in the time at which suggestions are communicated to the user, against a non-adaptive *Classic* control method. Possible effects on task completion time, the number of necessary mode switches, perceived workload, and user opinions on each control method were compared. Further, we establish that *Continuous* and *Threshold* performed equally well in quantitative and qualitative insights. Consequently, both are promising approaches to communicating proposed directional cues effectively.

Framework contribution: The integrated multi-modal feedback is an integral feature of *AdaptiX*, capable of supporting the system's real-time suggestions by user attention guiding. Although some participants experienced the combined visual-auditory-haptic multi-modal feedback as "irritating" [46], it effectively communicated updated suggestions. One application of virtual frameworks like *AdaptiX* might be the differentiation between different modality types and corresponding user preferences in an easy-to-set-up study. Highlighting the advantage of our framework, we could evaluate our different visualizations and multi-modal feedback without implementing a VR environment [46].

Based on the successful implementation of *AdaptiX* in this laboratory study, we are confident that the framework performs well in remote and in-person studies.



(a) (b) (c)

Fig. 11. Suggested control alternatives in light blue, visualized as in case study 2: (**a**) Moving forward and downward towards the object, (**b**) Closing the fingers to grasp the object, and (**c**) Moving towards the target area.

6.3 Example Adaption 3: Comparing Input Devices for Controlling a Physical Robot in Mixed Reality

A third study [47] highlights the MR capability of the framework and the integration options with different input devices. This study used the *Varjo XR-3* XR-HMD to explore a similar interaction design and feedback technique to our *Threshold* approach [46]. By incorporating this XR-HMD, the prototype mimics an AR environment (see Section 3.2) to the user, seeing the physical setup augmented by visual cues. Instead of a virtual pick-and-place task as before, this study combined a physical object, a physical drop area, and a physical robotic arm with AR cues delivered via the headset.

Participants compared three assistive input techniques: 1) a head-based control by using the deflection of the head on the *pitch* axis for continuous input and on the *roll* axis for mode-switching, 2) a gamepad input by using the *Xbox Adaptive Controller* [40] extended with *Logitech Adaptive Gaming Kit* [35] buttons for a discrete input, and 3) the control-stick of a *Nintendo Joy-Con* [43] motion controller – as a baseline to our previous study [46].

Framework contribution: With its real-world setting augmented by virtual cues, the research moved closer to reality on the MR-continuum than the previous two case studies. *AdaptiX* successfully performed as an easy-to-use interface between the usage of a physical robot and virtual communication via a XR-HMD.

It also allowed the research team to quickly evaluate the efficiency of different input devices with the potential to control the robotic arm along the adaptive DoF mapping. The standardized *User Input Adapter* enables researchers to easily chose between different technologies – supporting continuous, discrete, and absolute user input – and further extend it to their needs by its modular nature.

7 CONCLUSION

Integrating *AdaptiX* into HRI research can streamline the development and evaluation of new interaction designs and feedback techniques for controlling assistive robotic arms. The framework is advantageous in remote and in-person studies as its usage negates the need for a physical robotic device during the initial ideation and prototyping stages, thus increasing flexibility, accessibility, and efficiency.

An initial shared control concept by adaptive DoF mapping is provided and implemented to support researchers and developers to either change, extend, or exchange methods with their ideas. In studies using a physical robot, the integration of ROS bridges the gap to reality, by enabling a bidirectional connection between virtual and physical robotic arm. ROS allows developers and users to choose between a *DigitalTwin* and *PhysicalTwin* approach while interacting with *AdaptiX*. Using *AdaptiX*, researchers benefit from the entire continuum of MR. As the simulated and real-world environments of the robotic arm are perfectly aligned, nearly seamless switching between controlling the real and virtual robot is possible. This functionality allows applications in pure screen space, VR, AR, simultaneous simulation/reality, and pure reality. *AdaptiX*'s 3D teach-in interface facilitates a code-less trajectory programming of an assistive robot by hand-guiding the simulated or real robot to the specific location and saving the position and orientation of the tool center point. These waypoints are interpolated to a combined movement trajectory. The framework's recording/replaying system is entirely customizable. It includes options to change details during replay, such as repositioning cameras or re-rendering background scenes. A fully integrated recording of participants interacting with the robot is possible, which can be analyzed afterward to evaluate the specific research variables.

Taken together, *AdaptiX* is a free and open-source tool that enables HRI researchers to test and evaluate their shared control concepts for assistive robotic devices in a high-resolution virtual environment. The cited case studies clearly demonstrate the benefits researchers and developers can draw from using the framework. The near-endless customization options allow users to tweak the initial version to their specific research needs, resulting in practically tailor-made environments.

7.1 Framework Extensions

We invite the community to extend the *AdaptiX* framework based on their requirements needs by creating custom levels/scenarios and integrating new interfaces. *AdaptiX* can be accessed free-of-charge at https://adaptix.robot-research.de. Refer to the *README* provided in the repository for a detailed description of how to implement experiments in *AdaptiX*.

ACKNOWLEDGMENTS

This research is supported by the *German Federal Ministry of Education and Research* (BMBF, FKZ: 16SV8563, 16SV8565).

REFERENCES

[1] David A. Abbink, Tom Carlson, Mark Mulder, Joost C. F. de Winter, Farzad Aminravan, Tricia L. Gibo, and Erwin R. Boer. 2018. A Topology of Shared Control Systems—Finding Common Ground in Diversity. *IEEE Transactions on Human-Machine Systems* 48, 5 (Oct. 2018), 509–525. https://doi.org/10.1109/thms.2018.2791570

[2] Stephanie Arévalo-Arboleda, Max Pascher, Annalies Baumeister, Barbara Klein, and Jens Gerken. 2021. Reflecting upon Participatory Design in Human-Robot Collaboration for People with Motor Disabilities: Challenges and Lessons Learned from Three Multiyear Projects. In *The 14th PErvasive Technologies Related to Assistive Environments Conference - PETRA 2021* (2021-06-29). ACM. https://doi.org/10.1145/3453892.3458044

[3] Maude Beaudoin, Josiane Lettre, François Routhier, Philippe S. Archambault, Martin Lemay, and Isabelle Gélinas. 2018. Impacts of robotic arm use on individuals with upper extremity disabilities: A scoping review. *Canadian Journal of Occupational Therapy* 85, 5 (2018), 397–407. https://doi.org/10.1177/0008417418820878 PMID: 30866682.

[4] Connor Brooks and Daniel Szafir. 2020. Visualization of Intended Assistance for Acceptance of Shared Control. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE Press, Piscataway, NJ, USA, 11425–11430. https://doi.org/10.1109/IROS45743.2020.9340964
[5] Jennifer L Burke, Matthew S Prewett, Ashley A Gray, Liuquin Yang, Frederick RB Stilson, Michael D Coovert, Linda R Elliot, and Elizabeth Redden. 2006. Comparing the effects of visual-auditory and visual-tactile feedback on user performance: a meta-analysis. In *Proceedings of the 8th international conference on Multimodal interfaces*. Association for Computing Machinery, New York, NY, USA, 108–117.
[6] Samuel Bustamante, Jan Peters, Bernhard Schölkopf, Moritz Grosse-Wentrup, and Vinay Jayaram. 2021. ArmSym: A Virtual Human–Robot Interaction Laboratory for Assistive Robotics. *IEEE Transactions on Human-Machine Systems* 51, 6 (2021), 568–577. https://doi.org/10.1109/THMS.2021.3106865
[7] Gerard Canal, Guillem Alenyà, and Carme Torras. 2016. Personalization Framework for Adaptive Robotic Feeding Assistance. In *Social Robotics*. Vol. 9979. Springer International Publishing, Cham, 22–31. https://doi.org/10.1007/978-3-319-47437-3_3
[8] Ravi Teja Chadalavada, Henrik Andreasson, Maike Schindler, Rainer Palm, and Achim J Lilienthal. 2020. Bi-directional navigation intent communication using spatial augmented reality and eye-tracking glasses for improved safety in human–robot interaction. *Robotics and Computer-Integrated Manufacturing* 61 (2020), 101830.
[9] Yuhang Che, Allison M. Okamura, and Dorsa Sadigh. 2020. Efficient and Trustworthy Social Navigation via Explicit and Implicit Robot–Human Communication. *IEEE Transactions on Robotics* 36, 3 (2020), 692–707. https://doi.org/10.1109/TRO.2020.2964824
[10] Tiffany L. Chen, Chih-Hung King, Andrea L. Thomaz, and Charles C. Kemp. 2011. Touched by a robot. In *Proceedings of the 6th international conference on Human-robot interaction - HRI '11*, Aude Billard, Peter Kahn, Julie A. Adams, and Greg Trafton (Eds.). ACM Press, New York, New York, USA, 457. https://doi.org/10.1145/1957656.1957818
[11] Marianna Ciccarelli, Alessandra Papetti, Cecilia Scoccia, Giacomo Menchi, Leonardo Mostarda, Giacomo Palmieri, and Michele Germani. 2022. A system to improve the physical ergonomics in Human-Robot Collaboration. *Procedia Computer Science* 200 (2022), 689–698. https://doi.org/10.1016/j.procs.2022.01.267
[12] Yann-Seing Law-Kam Cio, Maxime Raison, Cédric Leblond Ménard, and Sofiane Achiche. 2019. Proof of Concept of an Assistive Robotic Arm Control Using Artificial Stereovision and Eye-Tracking. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 27, 12 (2019), 2344–2352. https://doi.org/10.1109/TNSRE.2019.2950619
[13] Andre Cleaver, Darren Vincent Tang, Victoria Chen, Elaine Schaertl Short, and Jivko Sinapov. 2021. Dynamic Path Visualization for Human-Robot Collaboration. In *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction* (Boulder, CO, USA) *(HRI '21 Companion)*. Association for Computing Machinery, New York, NY, USA, 339–343. https://doi.org/10.1145/3434074.3447188
[14] Adele Diederich and Hans Colonius. 2004. Bimodal and trimodal multisensory enhancement: effects of stimulus onset and intensity on reaction time. *Perception & psychophysics* 66, 8 (2004), 1388–1404.
[15] Epic Games. 2024. Unreal Engine. Published online (last visited on May 20, 2024). https://www.unrealengine.com Version 4.27.2.
[16] Chinemelu Ezeh, Pete Trautman, Louise Devigne, Valentin Bureau, Marie Babel, and Tom Carlson. 2017. Probabilistic vs linear blending approaches to shared control for wheelchair driving. In *2017 International Conference on Rehabilitation Robotics (ICORR)* (London, United Kingdom). IEEE Press, Piscataway, NJ, USA, 835–840. https://doi.org/10.1109/ICORR.2017.8009352
[17] F. Flemisch, D. A. Abbink, M. Itoh, M.-P. Pacaux-Lemoine, and G. Weßel. 2019. Joining the blunt and the pointy end of the spear: towards a common framework of joint action, human–machine cooperation, cooperative guidance and control, shared, traded and supervisory control. *Cognition, Technology & Work* 21, 4 (Aug. 2019), 555–568. https://doi.org/10.1007/s10111-019-00576-1
[18] Michael J Gielniak and Andrea L Thomaz. 2011. Generating anticipation in robot motion. In *2011 RO-MAN*. IEEE Press, Piscataway, NJ, USA, 449–454.
[19] Felix Ferdinand Goldau and Udo Frese. 2021. Learning to Map Degrees of Freedom for Assistive User Control: Towards an Adaptive DoF-Mapping Control for Assistive Robots. In *The 14th PErvasive Technologies Related to Assistive Environments Conference* (Corfu, Greece) *(PETRA 2021)*. Association for Computing Machinery, New York, NY, USA, 132–139. https://doi.org/10.1145/3453892.3453895
[20] Muhammad Abdul Haseeb, Maria Kyrarini, Shuo Jiang, Danijela Ristic-Durrant, and Axel Gräser. 2018. Head Gesture-Based Control for Assistive Robots. In *Proceedings of the 11th PErvasive Technologies Related to Assistive Environments Conference* (Corfu, Greece) *(PETRA '18)*. Association for Computing Machinery, New York, NY, USA, 379–383. https://doi.org/10.1145/3197768.3201574
[21] Laura V. Herlant, Rachel M. Holladay, and Siddhartha S. Srinivasa. 2016. Assistive Teleoperation of Robot Arms via Automatic Time-Optimal Mode Switching. In *The Eleventh ACM/IEEE International Conference on Human Robot Interaction* (Christchurch, New Zealand) *(HRI '16)*. IEEE Press, Piscataway, NJ, USA, 35–42.

[22] Nicholas J. Hetherington, Elizabeth A. Croft, and H.F. Machiel Van der Loos. 2021. Hey Robot, Which Way Are You Going? Nonverbal Motion Legibility Cues for Human-Robot Spatial Interaction. *IEEE Robotics and Automation Letters* 6, 3 (jul 2021), 5010–5015. https://doi.org/10.1109/lra.2021.3068708

[23] Catherine Holloway. 2019. Disability Interaction (DIX): A Manifesto. *Interactions* 26, 2 (feb 2019), 44–49. https://doi.org/10.1145/3310322

[24] Patrick Holthaus, Trenton Schulz, Gabriella Lakatos, and Rebekka Soma. 2023. Communicative Robot Signals: Presenting a New Typology for Human-Robot Interaction. In *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction* (Stockholm, Sweden) *(HRI '23)*. Association for Computing Machinery, New York, NY, USA, 132–141. https://doi.org/10.1145/3568162.3578631

[25] HTC. 2024. VIVE Controller 2.0. Published online (last visited on May 20, 2024). https://www.vive.com/de/accessory/controller2018/

[26] HTC. 2024. VIVE Tracker 3.0. Published online (last visited on May 20, 2024). https://www.vive.com/de/accessory/tracker3/

[27] Markus Höll, Markus Oberweger, Clemens Arth, and Vincent Lepetit. 2018. Efficient Physics-Based Implementation for Realistic Hand-Object Interaction in Virtual Reality. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE Press, Piscataway, NJ, USA, 175–182. https://doi.org/10.1109/VR.2018.8448284

[28] Intel Corporation. 2022. Intel Realsense - Product Family D400 Series. Published online (last visited on May 20, 2024). https://www.intelrealsense.com/wp-content/uploads/2022/11/Intel-RealSense-D400-Series-Datasheet-November-2022.pdf Revision 014.

[29] Siddarth Jain, Ali Farshchiansadegh, Alexander Broad, Farnaz Abdollahi, Ferdinando Mussa-Ivaldi, and Brenna Argall. 2015. Assistive robotic manipulation through shared autonomy and a Body-Machine Interface. In *2015 IEEE International Conference on Rehabilitation Robotics (ICORR)*. IEEE Press, Piscataway, NJ, USA, 526–531. https://doi.org/10.1109/ICORR.2015.7281253

[30] Kinova inc. 2021. Jaco assistive robot - User guide. Published online (last visited on May 20, 2024). https://assistive.kinovarobotics.com/uploads/EN-UG-007-Jaco-user-guide-R05.pdf EN-UG-007-r05-202111.

[31] Robert J Kosinski. 2008. A literature review on reaction time. *Clemson University* 10, 1 (2008), 337–344.

[32] Kirill Kronhardt, Stephan Rübner, Max Pascher, Felix Goldau, Udo Frese, and Jens Gerken. 2022. Adapt or Perish? Exploring the Effectiveness of Adaptive DoF Control Interaction Methods for Assistive Robot Arms. *Technologies* 10, 1 (2022), 24 pages. https://doi.org/10.3390/technologies10010030

[33] Maria Kyrarini, Fotios Lygerakis, Akilesh Rajavenkatanarayanan, Christos Sevastopoulos, Harish Ram Nambiappan, Kodur Krishna Chaitanya, Ashwin Ramesh Babu, Joanne Mathew, and Fillia Makedon. 2021. A survey of robots in healthcare. *Technologies* 9, 1 (2021), 8.

[34] Audrey Lebrasseur, Josiane Lettre, François Routhier, Philippe S. Archambault, and Alexandre Campeau-Lecours. 2019. Assistive robotic arm: Evaluation of the performance of intelligent algorithms. *Assistive Technology* 33, 2 (May 2019), 95–104. https://doi.org/10.1080/10400435.2019.1601649

[35] Logitech. 2024. Adaptive Gaming Kit. Published online (last visited on May 20, 2024). https://www.logitechg.com/en-us/products/gamepads/adaptive-gaming-kit-accessories.943-000318.html

[36] Zhanat Makhataeva and Huseyin Atakan Varol. 2020. Augmented Reality for Robotics: A Review. *Robotics* 9, 2 (2020). https://doi.org/10.3390/robotics9020021

[37] Elias Matsas and George-Christopher Vosniakos. 2015. Design of a virtual reality training system for human–robot collaboration in manufacturing tasks. *International Journal on Interactive Design and Manufacturing (IJIDeM)* 11, 2 (Feb. 2015), 139–153. https://doi.org/10.1007/s12008-015-0259-2

[38] Alyxander David May, Christian Dondrup, and Marc Hanheide. 2015. Show me your moves! Conveying navigation intention of a mobile robot to humans. In *2015 European Conference on Mobile Robots (ECMR)*. 1–6. https://doi.org/10.1109/ECMR.2015.7324049

[39] Meta. 2024. Meta Quest 2. Published online (last visited on May 20, 2024). https://www.meta.com/de/quest/products/quest-2/

[40] Microsoft. 2024. Xbox Adaptive Controller. Published online (last visited on May 20, 2024). https://www.microsoft.com/en-us/d/xbox-adaptive-controller/8nsdbhz1n3d8

[41] Paul Milgram and Fumio Kishino. 1994. A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems* 77, 12 (1994), 1321–1329.

[42] Sarah Luisa Müller, Stefan Schröder, Sabina Jeschke, and Anja Richert. 2017. Design of a Robotic Workmate. In *Lecture Notes in Computer Science*. Springer International Publishing, 447–456. https://doi.org/10.1007/978-3-319-58463-8_37

[43] Nintendo. 2024. Nintendo Joy-Con. Published online (last visited on May 20, 2024). https://www.nintendo.de/Hardware/Zubehor/Zubehor-1183144.html

[44] Max Pascher, Annalies Baumeister, Stefan Schneegass, Barbara Klein, and Jens Gerken. 2021. Recommendations for the Development of a Robotic Drinking and Eating Aid - An Ethnographic Study. In *Human-Computer Interaction*

– *INTERACT 2021* (2021-09-01), Carmelo Ardito, Rosa Lanzilotti, Alessio Malizia, Helen Petrie, Antonio Piccinno, Giuseppe Desolda, and Kori Inkpen (Eds.). Springer, Cham. https://doi.org/10.1007/978-3-030-85623-6_21

[45] Max Pascher, Uwe Gruenefeld, Stefan Schneegass, and Jens Gerken. 2023. How to Communicate Robot Motion Intent: A Scoping Review. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems - CHI '23*. https://doi.org/10.1145/3544548.3580857

[46] Max Pascher, Kirill Kronhardt, Felix Ferdinand Goldau, Udo Frese, and Jens Gerken. 2023. In Time and Space: Towards Usable Adaptive Control for Assistive Robotic Arms. In *2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 2300–2307. https://doi.org/10.1109/RO-MAN57019.2023.10309381

[47] Max Pascher, Kevin Zinta, and Jens Gerken. 2024. Exploring of Discrete and Continuous Input Control for AI-enhanced Assistive Robotic Arms. In *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction* (Boulder, CO, USA) *(HRI '24 Companion)*. Association for Computing Machinery, New York, NY, USA, 5 pages. https://doi.org/10.1145/3610978.3640626

[48] Sebastiaan Petermeijer, Pavlo Bazilinskyy, Klaus Bengler, and Joost de Winter. 2017. Take-over again: Investigating multimodal and directional TORs to get the driver back into the loop. *Applied Ergonomics* 62 (2017), 204–215. https://doi.org/10.1016/j.apergo.2017.02.023

[49] Sebastiaan Petermeijer, Fabian Doubek, and Joost de Winter. 2017. Driver response times to auditory, visual, and tactile take-over requests: A simulator study with 101 participants. In *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. 1505–1510. https://doi.org/10.1109/SMC.2017.8122827

[50] Laura Petrich, Jun Jin, Masood Dehghan, and Martin Jagersand. 2022. A Quantitative Analysis of Activities of Daily Living: Insights into Improving Functional Independence with Assistive Robotics. In *2022 International Conference on Robotics and Automation (ICRA)*. 6999–7006. https://doi.org/10.1109/ICRA46639.2022.9811960

[51] Anita Pollak, Mateusz Paliga, Matias M. Pulopulos, Barbara Kozusznik, and Malgorzata W. Kozusznik. 2020. Stress in manual and autonomous modes of collaboration with a cobot. *Computers in Human Behavior* 112 (2020), 106469. https://doi.org/10.1016/j.chb.2020.106469

[52] Gabriel Quere, Annette Hagengruber, Maged Iskandar, Samuel Bustamante, Daniel Leidner, Freek Stulp, and Jörn Vogel. 2020. Shared Control Templates for Assistive Robotics. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE Press, Piscataway, NJ, USA, 1956–1962. https://doi.org/10.1109/ICRA40945.2020.9197041

[53] Emanuele Ruffaldi, Filippo Brizzi, Franco Tecchia, and Sandro Bacinelli. 2016. Third point of view augmented reality for robot intentions visualization. In *International Conference on Augmented Reality, Virtual Reality and Computer Graphics*. Springer, 471–478.

[54] Ivan Shindev, Yu Sun, Michael Coovert, Jenny Pavlova, and Tiffany Lee. 2012. Exploration of Intention Expression for Robots. In *Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction* (Boston, Massachusetts, USA) *(HRI '12)*. Association for Computing Machinery, New York, NY, USA, 247–248. https://doi.org/10.1145/2157689.2157778

[55] Moondeep C. Shrestha, Ayano Kobayashi, Tomoya Onishi, Erika Uno, Hayato Yanagawa, Yuta Yokoyama, Mitsuhiro Kamezaki, Alexander Schmitz, and Shigeki Sugano. 2016. Intent communication in navigation through the use of light and screen indicators. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 523–524. https://doi.org/10.1109/HRI.2016.7451837

[56] Joris Sijs, Freek Liefhebber, and Gert Willem R.B.E. Romer. 2007. Combined Position & Force Control for a robotic manipulator. In *2007 IEEE 10th International Conference on Rehabilitation Robotics*. IEEE. https://doi.org/10.1109/icorr.2007.4428414

[57] Amit Singhal. 2001. Modern information retrieval: A brief overview. *IEEE Data Eng. Bull.* 24, 4 (2001), 35–43.

[58] Ryo Suzuki, Adnan Karim, Tian Xia, Hooman Hedayati, and Nicolai Marquardt. 2022. Augmented Reality and Robotics: A Survey and Taxonomy for AR-Enhanced Human-Robot Interaction and Robotic Interfaces. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 553, 33 pages. https://doi.org/10.1145/3491102.3517719

[59] Emmanuele Tidoni, Mohammad Abu-Alqumsan, Daniele Leonardis, Christoph Kapeller, Gabriele Fusco, Cristoph Guger, Cristoph Hintermuller, Angelika Peer, Antonio Frisoli, Franco Tecchia, Massimo Bergamasco, and Salvatore Maria Aglioti. 2017. Local and Remote Cooperation With Virtual and Robotic Agents: A P300 BCI Study in Healthy and People Living With Spinal Cord Injury. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 25, 9 (Sept. 2017), 1622–1632. https://doi.org/10.1109/tnsre.2016.2626391

[60] Katherine M Tsui, Dae-Jin Kim, Aman Behal, David Kontak, and Holly A Yanco. 2011. "I want that": Human-in-the-loop control of a wheelchair-mounted robotic arm. *Applied Bionics and Biomechanics* 8, 1 (2011), 127–147.

[61] Bram van Deurzen, Herman Bruyninckx, and Kris Luyten. 2022. Choreobot: A Reference Framework and Online Visual Dashboard for Supporting the Design of Intelligible Robotic Systems. *Proc. ACM Hum.-Comput. Interact.* 6, EICS, Article 151 (jun 2022), 24 pages. https://doi.org/10.1145/3532201

[62] Varjo. 2024. Varjo XR-3. Published online (last visited on May 20, 2024). https://varjo.com/products/xr-3/

[63] Michael Walker, Hooman Hedayati, Jennifer Lee, and Daniel Szafir. 2018. Communicating Robot Motion Intent with Augmented Reality. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction* (Chicago, IL, USA) *(HRI '18)*. Association for Computing Machinery, New York, NY, USA, 316–324. https://doi.org/10.1145/3171221.3171253

[64] Hanna Yun and Ji Hyun Yang. 2020. Multimodal warning design for take-over request in conditionally automated driving. *European transport research review* 12, 1 (2020), 1–11.

[65] Mohammad Kassem Zein, Majd Al Aawar, Daniel Asmar, and Imad H Elhajj. 2021. Deep learning and mixed reality to autocomplete teleoperation. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 4523–4529.

[66] He Zhang and Cang Ye. 2019. Human-Robot Interaction for Assisted Wayfinding of a Robotic Navigation Aid for the Blind. In *2019 12th International Conference on Human System Interaction (HSI)*. IEEE. https://doi.org/10.1109/hsi47298.2019.8942612

[67] Jakub Złotowski, Kumar Yogeeswaran, and Christoph Bartneck. 2017. Can we control it? Autonomous robots threaten human identity, uniqueness, safety, and resources. *International Journal of Human-Computer Studies* 100 (2017), 48–54.

A BLUEPRINTS OF ADMC IMPLEMENTATION

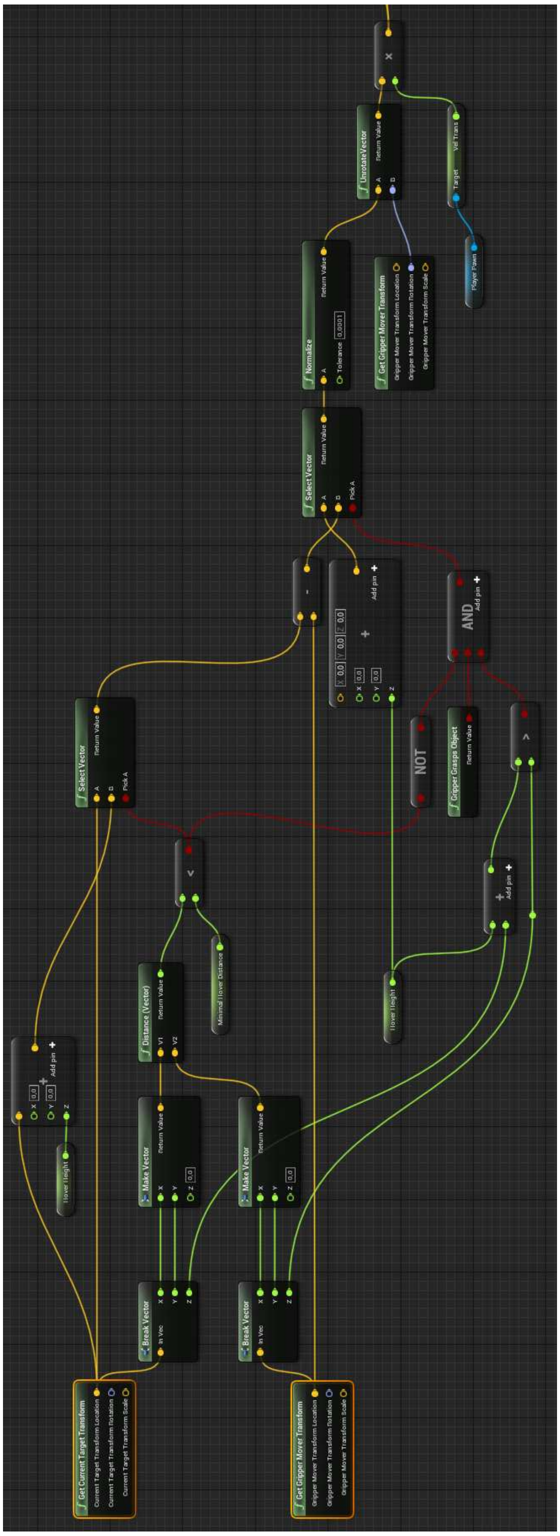


Fig. 12. Calculation of the translation for the *Optimal Suggestion*: Excerpt of *Blueprint* code calculating the *Translation* value of the *Adaptive Axis* for the *Optimal Suggestion*. Not pictured: Edge case handling for gripping an object.

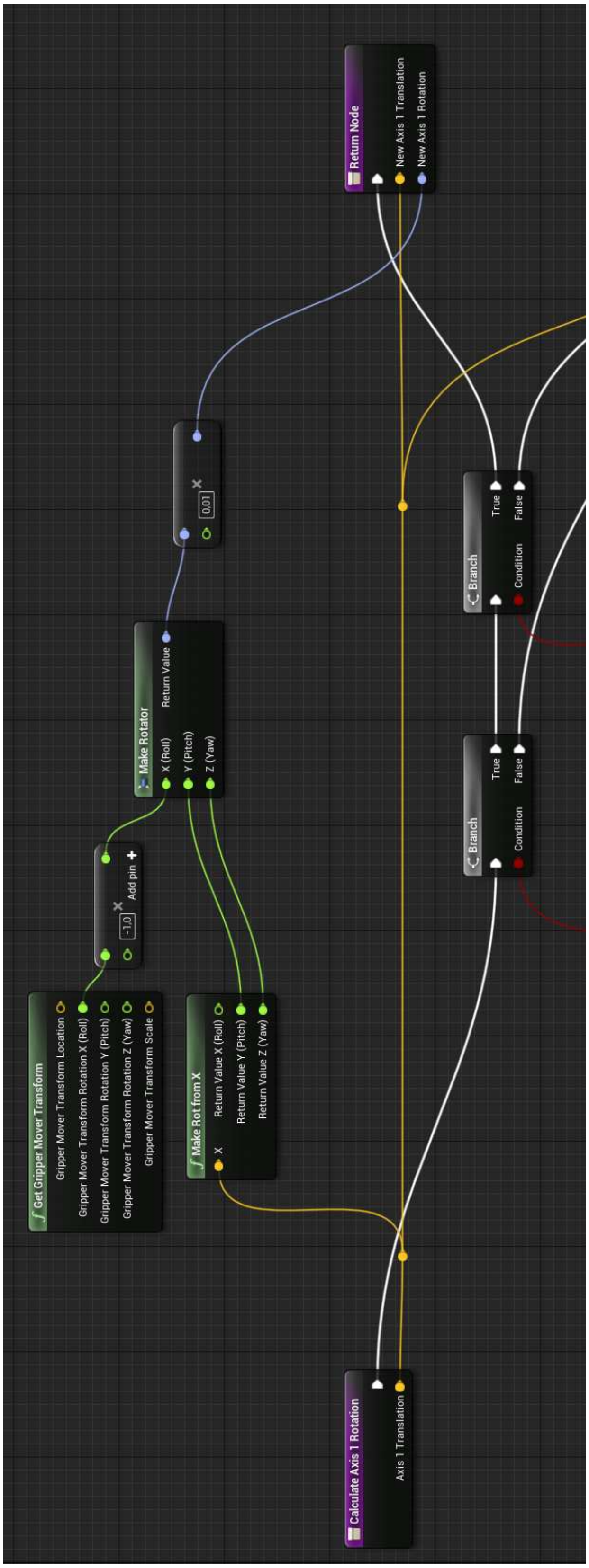


Fig. 13. Calculation of the Rotation for the *Optimal Suggestion*: Excerpt of *Blueprint* code calculating the *Rotation* value of the *Adaptive Axis* for the *Optimal Suggestion*. Not pictured: Edge case handling.

[IX]

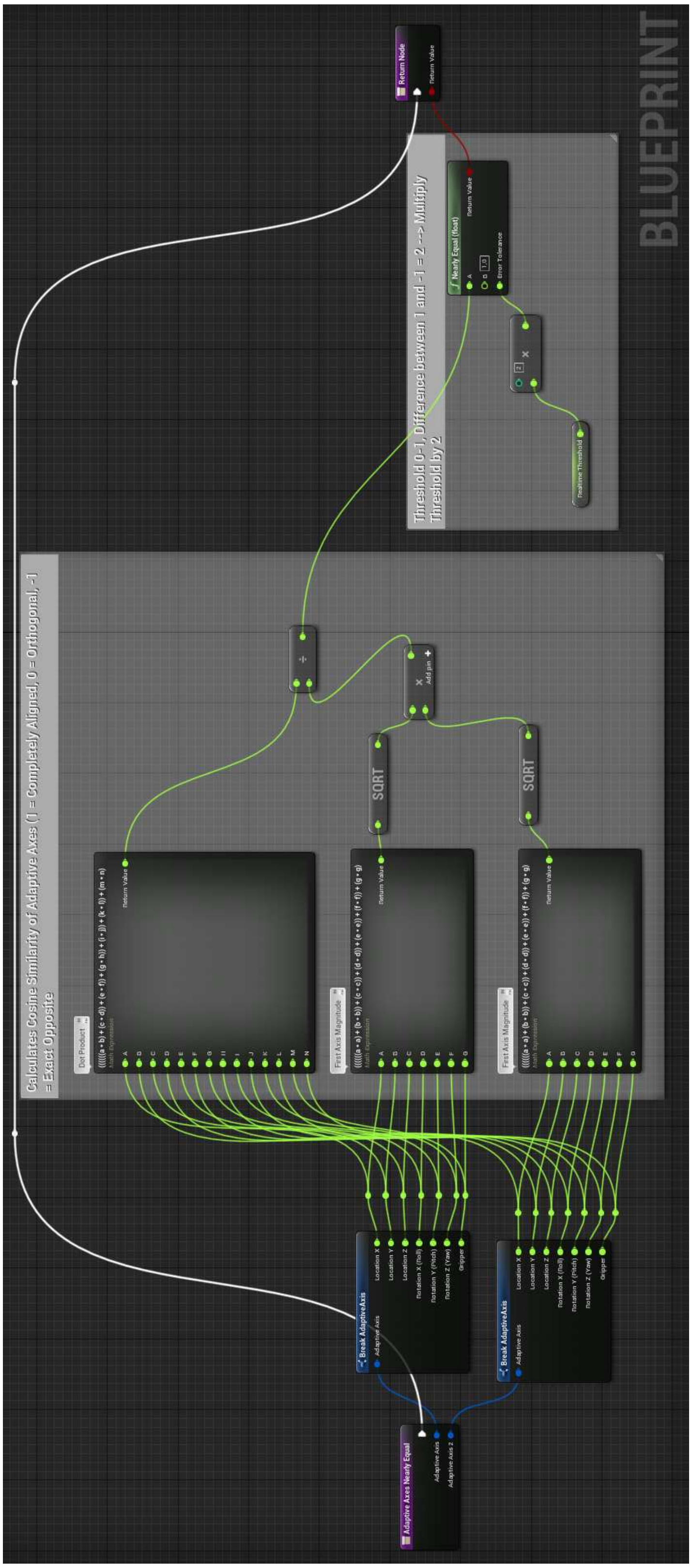


Fig. 14. *Adaptive Axes Nearly Equal* function to prepare the multi-modal attention guiding of the user.

First Submitted February 2023; Revised July 2023; Accepted October 2023

In Time and Space: Towards Usable Adaptive Control for Assistive Robotic Arms

Max Pascher[1,2] and Kirill Kronhardt[1] and Felix Ferdinand Goldau[3] and Udo Frese[3] and Jens Gerken[1]

***Abstract*— Robotic solutions, in particular robotic arms, are becoming more frequently deployed for close collaboration with humans, for example in manufacturing or domestic care environments. These robotic arms require the user to control several Degrees-of-Freedom (DoFs) to perform tasks, primarily involving grasping and manipulating objects. Standard input devices predominantly have two DoFs, requiring time-consuming and cognitively demanding mode switches to select individual DoFs. Contemporary Adaptive DoF Mapping Controls (ADMCs) have shown to decrease the necessary number of mode switches but were up to now not able to significantly reduce the perceived workload. Users still bear the mental workload of incorporating abstract mode switching into their workflow. We address this by providing feed-forward multimodal feedback using updated recommendations of ADMC, allowing users to visually compare the current and the suggested mapping in real-time. We contrast the effectiveness of two new approaches that a) *continuously* recommend updated DoF combinations or b) use discrete *thresholds* between current robot movements and new recommendations. Both are compared in a Virtual Reality (VR) in-person study against a *classic* control method. Significant results for lowered task completion time, fewer mode switches, and reduced perceived workload conclusively establish that in combination with feedforward, ADMC methods can indeed outperform classic mode switching. A lack of apparent quantitative differences between *Continuous* and *Threshold* reveals the importance of user-centered customization options. Including these implications in the development process will improve usability, which is essential for successfully implementing robotic technologies with high user acceptance.**

I. Introduction

While robotic devices have long been put behind fences for safety reasons, advances in the development of such (semi-) autonomous technologies have started to permeate almost all aspects of our personal and professional lives. These include increased close-quarter collaborations with robotic devices – from industry assembly lines [1] to mobility aides [2]. Assistive robotic arms are a particularly useful and versatile subset of collaborative technologies with varied applications in different fields, e.g., [3], [4].

Yet, new challenges arise when robots are tasked with (semi-) autonomous actions, resulting in additional stress for end-users if not correctly addressed during the design process [5]. Pollak et al. highlight the decreased feeling of control users experienced when using a robot's autonomous mode. Switching to manual mode allowed their study participants to regain control and decrease stress significantly. These findings are corroborated by Kim et al. whose comparative study of control methods resulted in markedly higher user satisfaction for the manual mode cohort [6].

A proposed solution from previous work [7] to these challenge are adaptive controls – referred as Adaptive DoF Mapping Controls (ADMCs) – which merge the advantages of (semi-) autonomous actions with the flexibility of manual controls. They combine multiple DoFs dynamically for a specific scenario to assist in controlling the robot. In our concept, a Convolutional Neural Network (CNN) interprets a camera's video feed of the environment and dynamically combines the most likely DoFs for a suggested movement. Building on this, we already showed that such ADMC combinations of the robot's DoFs can lead to a significantly lower number of mode switches compared to standard control methods [8]. However, our study could not show that this may also improve task completion time or reduce cognitive load. Also, challenges concerning the understanding of DoF mappings were raised during the study.

Based on these previous findings, the present study evaluates two novel ADMCs methods for an assistive robotic arm. We compare the variants *Continuous* and *Threshold*, differing in the time at which suggestions are communicated to the user, against a *classic* control method. In detail, we examine possible effects on task completion time, number of necessary mode switches, perceived workload, and subjective user experience. Our contribution is two-fold:

1) We demonstrate that both ADMC methods significantly reduce the task completion time, the average number of mode switches, and the perceived workload of the user.
2) Further, we establish that for *Continuous* and *Threshold*, each has specific advantages which some users may prefer over the other, raising the need for customizable configurations.

II. Related Work

Collaborative robotic solutions have received much attention in recent years. Previous work has generally focused on (a) different designs of robot motion intent and most recently (b) ADMCs for robots. The latter requires a critical yet seldom addressed topic in how collaborative robots can effectively communicate recommended movement directions to their user.

[1]Max Pascher, Kirill Kronhardt, and Jens Gerken are with the Westphalian University of Applied Sciences, Human-Computer Interaction, 45897 Gelsenkirchen, Germany max.pascher@w-hs.de, kirill.kronhardt@w-hs.de, jens.gerken@w-hs.de
[2]Max Pascher is also with the University of Duisburg-Essen, Human-Computer Interaction, 45127 Essen, Germany max.pascher@uni-due.de
[3]Felix Ferdinand Goldau and Udo Frese are with the University of Bremen, Mathematics & Computer Science, 28359 Bremen, Germany fgoldau@uni-bremen.de, ufrese@uni-bremen.de

A. Robot Motion Intent

Advance knowledge of the intended robot behavior and subsequent movements within the physical world are critical for effective collaboration when humans and robots occupy the same space and need to coordinate their actions [9]. In previous work, we analyzed existing techniques of communicating robot motion intent and identified different *intent types* as well as several intent properties, such as *location* and *information* or the placement of the technology [10]. Users generally prefer to have the robot's future movements represented visually [11]. To convey detailed robot motion intent, researchers often rely on Augmented Reality (AR) [12], [13], [14], as "with the help of AR, interaction can become more intuitive and natural to humans" [15].

Effective communication of robot motion intent is particularly relevant when using ADMCs for assistive robotic arms, as in such a shared or traded control environment each interaction needs to be precisely coordinated.

B. Adaptive DoF Mapping Controls

Traditionally, robot control methods include individual commands for each DoF, requiring frequent mode switches for controlling translations, rotations, and gripper functionality. Herlant et al. called into question the suitability of these standard control methods as task completion time markedly increases by using user-initiated compared to time-optimal mode switches [16].

To tackle this issue, we proposed in previous work the concept of ADMC – a dynamic combination of multiple DoFs, thus adjusted to specific scenarios or tasks [7]. This streamlining decreases the need for constant mode switching, resulting in faster and more efficient task fulfillment. In [7] we implemented a CNN as control unit to provide these dynamic DoF mappings and gave the user a triggering mechanism to request an update. In a 2D simulation study which had a 4-DoF robot control mapped to a 2-DoF input device, we found promising results.

We then extended this approach into a 3D VR simulation, thereby mapping a 7-DoF robot control to a 2-DoF input device [8]. We evaluated two ADMC methods – differing in their respective movement suggestion concept – against the baseline control method *Classic*. Simulating the effect of a CNN, our work relied on a task-specific script to provide DoF mappings based on the relative position and orientation between gripper and target. This removed the potentially confounding effect of a suboptimal CNN implementation. Results showed that the number of mode switches was significantly reduced compared to *Classic*, but task completion time was unaffected. Users reported high cognitive demand and difficulties understanding the mapping to 2 different input DoFs. In addition, the system felt difficult to predict and required trial and error [8].

III. Adaptive DoF Mapping Controls

Building on our previous work [8], we created a VR simulation of a Human-Robot Interaction (HRI) experimental setup to compare different ADMC methods to a non-adaptive baseline condition *Classic*. Like in previous work [8] we applied a task-specific script to explore our ADMC methods. We tackle previous issues by 1) visualizing not only the current but also the forthcoming DoF mapping suggestion (improving predictability) and 2) reducing the input to a single DoF (reducing cognitive demand). We propose two approaches as different trade-offs between control fidelity and cognitive demand.

The VR simulation includes a virtual model of the *Kinova Jaco 2*[1] – a commercially available assistive robotic arm frequently used in HRI studies, e.g., [4], [16]. Our proposed visual feedback mimics AR, with directional cues registered in 3D space. This allows the user to understand different movement directions for the actual control and the suggested DoF combinations. To simplify understanding, we use *arrows*, a straightforward and common visualization technique to communicate motion intent [9], [17], [18].

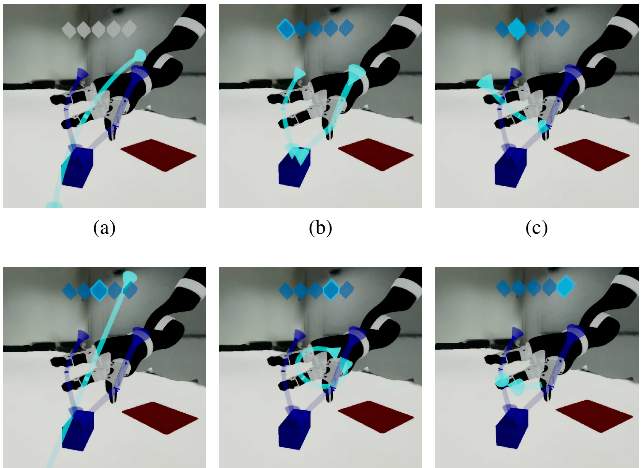
As a control method for the ADMCs, we implemented a task-specific script. This removed any potential bias that a more generic but currently still technically limited approach such as a CNN-based control method may introduce. Of course, our approach only works in a controlled experimental setting. The task-specific script evaluates the gripper's current position, rotation, and finger position relative to a target. The DoF mapping system then suggests five different movement options (referred in the following to as *modes*) – in order of assumed usefulness – to the user.

1) *Optimal Suggestion:* Combining translation, rotation, and finger movement [opening and closing] into one suggestion, causing the gripper to move towards the target, pick it up, or release it on the intended surface.
2) An orthogonal suggestion based on (1) but excluding the finger movement. Allows the users to adjust the gripper's position while still being correctly orientated.
3) A pure translation towards the next target, disregarding any rotation.
4) A pure rotation towards the next target without moving the gripper.
5) Opening or closing of the gripper's fingers.

During movement, the ADMC system re-calculates the best DoF combinations to fulfill the specific task, which are then presented as new suggestions. Users cycle through these modes – by pressing a button on the input device – to select a suitable one or continue moving with the previous active suggestion (see Figure 1). A suggestion indicator is visible above the gripper when users are not moving the robot to distinguish between the modes. Five slanted cubes represent the possible suggestions. The cubes appear gray if no suggestion is active and turn blue to indicate that a new suggestion is selected. The cube corresponding to the selected mode increases in size. In contrast to our previous work [8] and to the dual axis system of the baseline control method (see Figure 2), only one input axis is required to control the robotic arm. Consequently, the cognitive demand

[1]Kinova Robotic arm. `https://assistive.kinovarobotics.com/product/jaco-robotic-arm`, last retrieved October 16, 2024.

on the users is reduced as they can focus on evaluating one movement rather than two simultaneous suggestions.



(a) (b) (c) (d) (e) (f)

Fig. 1: Suggestions as visualized in the ADMC methods, (**a**) Continue previous movement, (**b**) Optimal Suggestion, (**c**) Adjustment Suggestion, (**d**) Pure Translation, (**e**) Pure Rotation, (**f**) Open / Close Fingers.

Continuous: This control method uses continuous feedback of robot motion intent to increase oversight of updated movement suggestions. Continuous feedback enables users to move in a direction and constantly evaluate the updated optimal suggestion by the ADMC system. If found fitting, users can switch to a new suggestion and move the robot in the updated path to fulfill the task. Here, two directional indicators are virtually attached to the robotic arm's gripper: a light blue and a dark blue arrow. The former represents the currently selected movement option (mode) mapped to the input axis. The forward movement of the input axis moves the gripper in the direction the arrow is pointing; engaging it backward moves the gripper in the arrow's reverse direction. The dark blue arrow represents the currently optimal suggestion at a given time. Users can only move the robot along the dark blue arrow if they switch to that suggestion first – which causes both arrows to overlap. While this approach increases transparency, users might be distracted by the constantly updating suggestions, potentially leading to more mode switches and perceived workload.

Threshold: In contrast to *Continuous*, *Threshold* uses time-discrete and multimodal feedback to indicate optimized movement suggestions. Again, a light blue arrow maps the selected movement option (mode) to the input axis. New suggestions are only shown to the users if the optimal mode differs – by a set degree – from the current movement. We followed Singhal et al. and used a cosine between-vector similarity measure to calculate this threshold [19], ranging from exact alignment [0%] to total opposite direction [100%]. In pretests, we determined a 20% difference between the current and optimal vector as the suggestion threshold. If exceeded, a short vibration pulse to the input device and a 1kHz sound inform the users of an updated suggestion. In addition, a dark blue arrow appears which visualizes the new suggested movement. Users can continue the active movement, switch to the new suggestion, or cycle through the other four modes before deciding on one. Unlike with *Continuous*, users can therefore entirely focus on the movement they are currently performing until explicitly notified and directed to a new suggestion. We expect *Threshold* to reduce perceived workload compared to *Continuous* as it does not require constant evaluation of the visual feedback. However, we expect task completion time to increase, as *Threshold* systematically interrupts the users' workflow. Additionally, *Threshold* might result in a perceived loss of control, potentially negatively influencing usability.

IV. Study Method and Materials

To explore the effectiveness of our ADMC methods, we conducted a supervised, controlled experiment as a VR simulation study with 24 participants. We compared our ADMC methods to *Classic*, which relies on mode switching to access and control all DoFs one after another. Approaches as *Classic* are well established (e.g., when driving a car) and are predictable and transparent for the user. Comparing ADMC methods to *Classic* allows HRI researchers to disentangle their respective advantages and disadvantages.

A. Study Design

We applied a within-participant design with *control method* as an independent variable with three conditions: (1) *Classic*, (2) *Continuous*, and (3) *Threshold*. Every participant performed eight training trials and 24 measured trials per condition, resulting in 72 measured and 24 training trials per participant and 1,728 measured trials in total. To counter learning and fatigue effects, the order of conditions was fully counter-balanced. We measured the following dependent variables:

1) **Average Task Completion Time** For each trial, we measured the time in seconds needed to pick an object and place it on the target surface.
2) **Average Number of Mode Switches** For each trial, we recorded every mode switch conducted by pressing a button on the input device.
3) **Perceived Workload** After completing each condition, we measured cognitive workload with the NASA Raw-Task Load Index (NASA Raw-TLX) questionnaire [20].
4) **Subjective Assessment** After completing each condition, we measured the five dimensions of the Questionnaire for the Evaluation of Physical Assistive Devices (QUEAD) [21]. After completing all trials, participants were further asked to rank the three conditions.

After each condition, participants were prompted with several open questions regarding their experience, their understanding of the control methods and the directional cues, plus any issue of interest they considered noteworthy. Additionally, participants were asked how they proceeded in situations when they could not solve the task at first.

Video and audio recordings of the interviews with the entire study cohort were assessed independently by two researchers. Open coding was applied to gather participants' opinions of the different control methods. We used Miro[2] – an online whiteboard [22] – to complete an affinity diagram of the open codes. Codes were then organized into themes (see Section V-F).

B. Hypotheses

Overall, we expected ADMC methods to reduce not just mode switches (as in prior work [8]) but – due to the advances in our designs – also improve on task completion time and workload.

H1: *Continuous* and *Threshold* lead to a lower task completion time compared to *Classic*. However, we expect *Continuous* to perform faster compared to *Threshold*, as the latter systematically interrupts the user during interaction.

H2: *Continuous* and *Threshold* result in fewer mode switches compared to *Classic*. We expect *Continuous* to require more mode switches than *Threshold*, as users have no clear guidance about when to switch modes. This may cause them to oversteer or accept new suggestions inefficiently.

H3: *Continuous* and *Threshold* cause lower perceived workload compared to *Classic*. However, we expect *Continuous* to cause a higher workload compared to *Threshold*, as it requires constant evaluation of the visual feedback while *Threshold* allows the user to relax until further notification.

C. Apparatus

Developing and testing new concepts for a robotic arm involves inherent challenges associated with a real robot's physical bulk and complexity. Quickly changing the experimental setup, adding feedback components, or providing information to the user further complicate testing regimes. We created a 3D testbed environment for HRI studies in VR to address these challenges. This testbed contains a simulated robotic arm (a virtual model of the *Kinova Jaco 2*) with multiple control mechanisms and a standardized pick-and-place task. Visual feedback mimics AR, with directional cues registered in 3D space. A *Meta Quest* motion controller is used as an input device to control the robotic arm.

Photogrammetry scans of an actual room were used to design the VR environment, which was created using the *Unreal Engine 4.27* and optimized for usage with a *Meta Quest* VR Head-Mounted Display (HMD) (see Figure 2). During the study, user behavior was recorded with appropriate software on a *Schenker XMG Key 17* laptop with *Windows 10 64-bit* and *Oculus Link* connected to the VR headset.

For our implementation of the baseline control method *Classic*, users cycled through four distinct modes to access all seven robot DoFs, as they are mapped on a two-DoF joystick, such as the control-stick on a *Meta Quest* motion controller:

1) X-Translation + Y-Translation
2) Z-Translation + Roll
3) Yaw + Pitch
4) Open/Close fingers

We illustrate the current mapping between the robot's DoFs and the input device through two arrows attached to the gripper. Light blue arrows indicate the robot's DoF assigned to the first, dark blue arrows to the second input axis. Looking at the joystick in VR, the same color-coded visualization is applied.

Users press a button on the input device – the A-Button of the *Meta Quest* motion controller – to switch between modes, cycling back to the first one at the end. Four blue spheres – in contrast to the slanted cubes used in our ADMC methods – above the robotic arm's gripper indicate the total number of available and the currently active mode when users are not moving the robot. The sphere representing the active mode is bigger and brighter than the spheres of inactive modes.

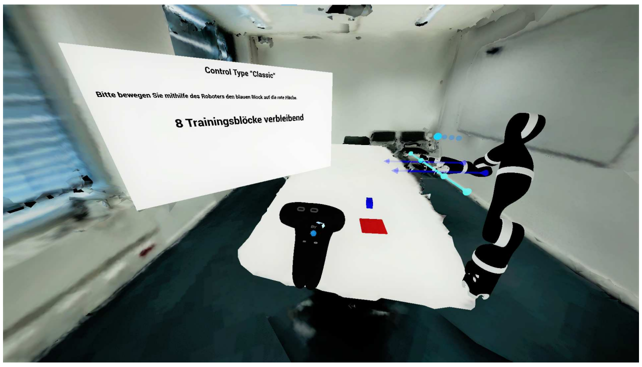


Fig. 2: Virtual environment consisting of (left to right): a virtual canvas, the motion controllers, a table with the blue object and red target, and a *Kinova JACO* with an arrow-based visualization

D. Participants

A total of 24 participants took part in our study (7 female, 17 male). The participants were aged 19 to 37, with a mean age of 26 years (SD = 4.85 years). No one declared any motor impairments that might influence reaction times. Five participants had prior experience with controlling a robotic arm. Participants were recruited from a university campus and an online appointment form.

E. Procedure

Utilizing the benefits of a standardized and portable VR simulation environment, the study was conducted in multiple comparable physical localities. Before commencing, participants were fully informed about the project objective and the various tasks they had to complete. Every participant gave their full and informed consent to partake in the study, have video and audio recordings taken, and have all the relevant data documented.

[2]Miro. `https://miro.com`, last retrieved October 16, 2024.

A study administrator observed the experiment on a laptop and briefed participants on using the hardware as well as the general functionalities of the study environment. Once set up, users followed command prompts embedded in the virtual simulation environment. For each of the three conditions, the following steps were performed:

1) Participants were given a written and standardized explanation of the control method used in the current condition.
2) Participants conducted eight training trials for familiarization with the respective control method.
3) Participants then conducted 24 measured trials.
4) Interview and questionnaires.

After completing all conditions, participants ranked the three control methods from most to least preferred and explained the reasoning behind their decision. The study concluded with a de-briefing. The average session lasted for 90 minutes and participants were compensated with 30 EUR.

F. *Experimental Task*

The experimental task is based on our previous work and resembles a common pick-and-place scenario [8]. A blue object appears on a table in front of the participant, which signals the start of a trial. The user has to control the robot from its starting position to pick the object and place it on a red target surface, also located on the table. To change the DoF mapping – for trial fulfillment – users could switch modes. Upon completion, the blue object disappears, and the robot automatically returns to the original starting position. A new blue object appears when this position is reached, and a new trial commences. For each trial, the position of the blue object is placed in one of eight possible locations spaced evenly around the red target surface. Each position occurred once during training and thrice during measured trials. However, the order of appearance was randomized. We used a neutral block shape rather than specific objects to avoid bias and ensure trial comparability.

V. Results

The study comprises 1,728 (24 participants $\times$ 3 control methods $\times$ 24 trials) measured trials. Training trials were excluded from the analysis.

We explored the distribution of the data through QQ-plots and either applied parametric Repeated Measures Analysis of Variance (RM-ANOVA) or non-parametric Friedman tests. For the latter, post-hoc pairwise comparisons using Wilcoxon signed-rank test with Bonferroni correction followed the omnibus test. Relevant effect sizes were calculated with r: >0.1 small, >0.3 medium, and >0.5 large effect.

A. *Task Completion Time*

Mean task completion time calculated per participant and control method (see Fig. 3) resulted in *Threshold* = 16.54s (SD = 4.09s); *Continuous* = 16.61s (SD = 4.77s); and *Classic* = 30.96s (SD = 4.89s). Outliers [N = 3] with average times $\geq 2.2 * \text{IQR}$ of the mean task completion time in at least one control method were excluded [23]. The QQ-plot of the remaining 21 participants followed a normal distribution.

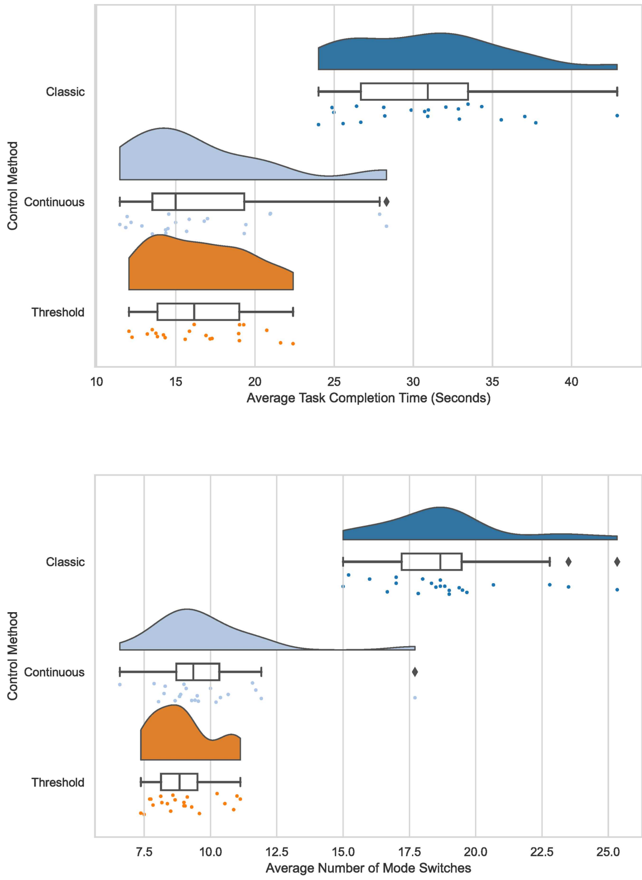


Fig. 3: Raincloud Plots for Average Task Completion Time and Mode Switches

A RM-ANOVA found a significant main effect (F(2, 36) = 130.92, p $\leq$0.001). A post-hoc pairwise comparison (Bonferroni corrected) showed a significant difference between *Continuous* and *Classic* (p $\leq$0.001) as well as between *Threshold* and *Classic* (p $\leq$0.001). No significant difference was found between *Continuous* and *Threshold* (p $\geq$0.999).

B. *Mode Switches*

We used a non-parametric Friedman test, as our data was not normally distributed, to determine differences between the average number of necessary mode switches between control methods. Two outliers – based on $\geq 2.2 * \text{IQR}$ of the mean value – were excluded prior to further analysis. This resulted in mean numbers of mode switches for *Threshold* = 9.28 (SD = 1.26); *Continuous* = 9.93 (SD = 1.47); and *Classic* = 19.55 (SD = 2.93) for N = 22. We found a significant main effect ($\chi^2(2)$ = 33.82, p $\leq$0.001, N = 22). Post-hoc pairwise comparisons showed a significant difference between *Continuous* and *Classic* (Z = -4.11, p $\leq$0.001, r = 0.62) as well as *Threshold* and *Classic* (Z = -4.11, p $\leq$0.001, r = 0.62). Again, we found no significant difference between the two ADMC methods (Z = -1.51, p = 0.131, r = 0.28) (see Fig. 3).

C. Perceived Workload

NASA Raw-TLX [20] scores [scale from 1 to 100] for all participants resulted in mean task load values of *Threshold* = 22.67 (SD = 13.86); *Continuous* = 23.23 (SD = 13.26); and *Classic* = 34.24 (SD = 14.65). We applied a Friedman test which revealed a significant main effect for perceived task load: ($\chi^2(2)$ = 9.87, p = 0.007, N = 24). Post-hoc pairwise comparisons show significant differences between *Continuous* and *Classic* (Z = −3.03, p = 0.002, r = 0.44), *Threshold* and *Classic* (Z = −2.76, p = 0.006, r = 0.40), but not between *Continuous* and *Threshold* (Z = −0.21, p = 0.830, r = 0.03).

D. Evaluation of Physical Assistive Devices

The QUEAD encompasses five individual scales (3 to 9 items each, 7-point Likert). Friedman tests for individual dimensions revealed significant main effects for *Perceived Usefulness (PU)*, *Perceived Ease of Use (PEU)*, *Emotions (E)*, and *Comfort (C)*, but not for *Attitude (A)*. Post-hoc pairwise comparisons indicate significant differences between *Continuous* and *Classic* for *PU*, *PEU*, and *C* as well as between *Threshold* and *Classic* for *PU* and *PEU* (refer to Table I for detailed scores).

TABLE I: Statistics for individual QUEAD dimensions: Perceived Usefulness (PU), Perceived Ease of Use (PEU), Emotions (E), Attitude (A), and Comfort (C).

	PU	PEU	E	A	C
Descriptive Statistics					
$M_{Classic}$	4.98	4.87	5.00	4.81	5.65
$SD_{Classic}$	1.39	1.20	1.71	1.75	1.71
$M_{Continuous}$	5.68	5.80	5.90	5.42	6.44
$SD_{Continuous}$	1.05	1.04	1.25	1.48	0.78
$M_{Threshold}$	5.77	5.90	5.68	5.44	6.13
$SD_{Threshold}$	1.02	0.97	1.43	1.58	1.14
Friedman Tests					
$\chi^2(2)$	7.49	15.22	7.20	1.76	6.39
p	0.022	$\leq$0.001	0.026	0.422	0.040
N	24	24	24	24	24
Pairwise Comparisons					
Classic vs. Continuous					
$\lvert Z \rvert$	2.32	2.47	1.85	—	2.29
p	0.021	0.014	0.064	—	0.022
r	0.33	0.36	0.27	—	0.33
Classic vs. Threshold					
$\lvert Z \rvert$	2.68	2.90	1.28	—	1.23
p	0.007	0.003	0.202	—	0.220
r	0.39	0.43	0.18	—	0.18
Continuous vs. Threshold					
$\lvert Z \rvert$	0.62	0.38	1.03	—	1.70
p	0.538	0.706	0.302	—	0.089
r	0.09	0.05	0.15	—	0.25

E. Individual Ranking

Participants ranked the control methods in order of preference from 1 = *favorite* to 3 = *least favorite*. Mean values in ascending order are *Continuous* = 1.67; *Threshold* = 2.04; and *Classic* = 2.29. A Friedman test revealed no significant main effect ($\chi^2(2)$ = 4.75, p = 0.100, N = 24).

F. Qualitative Insights

Overall, the open coding process led to the identification of five main themes, as discussed below.

1) Familiarization: While all three control methods included a training phase, comments suggest that in particular the ADMC methods required familiarization. Here, participants felt the controls were sometimes "inverted" (P3) and wanted to "move the stick in the direction the arrow was pointing at" (P6). They also reported that "it takes a while to get used to" (P24), but "routine set in fast" (P18).

2) Handling Adaptive DoF Mapping Suggestions: The study cohort showed a relatively uniform response to the two ADMC methods with clear distinctions between *Threshold* and *Continuous*. In *Threshold*, many participants "trusted the system" (P23) and switched to the new suggestion as soon as they perceived the multimodal indicator. They "did not have to think a lot" (P4) and "relied on what the suggestion says" (P7). This dependence on the system caused some to "draw a blank when something went wrong because [they] forgot they had other options" (P8). One participant even tried using the *Threshold* control method with eyes closed, which "worked surprisingly well" (P7).

In contrast, participants evaluated the suggestions in *Continuous* more thoroughly, as they had to decide when to switch without the help of threshold-based indicators. Some participants waited for suggestions with relatively simple direction cues, such as "straight arrows" (P6, P16) as an indication to switch modes, while others trusted their "gut feeling" (P23). Uncertainties of "How do I approach this?" (P23) were more frequent in this control method than *Threshold*. Participants dealt with problems in both ADMC conditions in one of two ways to find alternative suggestions that better align with their needs. They cycled through the further offered suggestions for an alternative option or reversed their current movement direction until a different suggestion was offered.

3) Visualization: Overall, participants understood the different visualizations. Yet, difficulties arose in all three conditions relating to depth perception and understanding if the gripper is positioned correctly to pick or place the object. Some participants suggested a "laser pointer" (P16) to indicate the gripper's position above the table for improved depth perception. This is a known problem for robot teleoperation. In the past, researchers have suggested and explored AR *Visual Cues* to counter that, which include similar approaches as the ones mentioned by our participants [24], [25].

Interestingly, some participants "manipulated" the second mode of *Classic* (X- and Y-Translation) to mimic this effect, as that mode shows straight up- and downward pointing arrows as directional cues along the y-axis.

4) Multimodal Feedback: As described above, most participants used *Threshold* as intended, switching to the next suggestion when they received the multimodal feedback.

However, some participants experienced the haptic and audio indicators as “irritating” (P20) or “weird and horrible” (P17). The poignant statement “If I had to do this for five more minutes, it would be too annoying.” (P7) reveals some participants’ strong reactions to this control method. As a possible mitigation, one participant suggested implementing multiple thresholds of varying intensity instead of a singular one that “instantly beeps loudly at me and says ’Do this now!’” (P24).

5) Control vs. Comfort: Participants reported substantial differences in the level of control and comfort between *Classic*, *Continuous*, and *Threshold*. By nature, *Classic* offers the highest control level but requires participants to decide individually on every task step. In contrast, *Threshold* allowed participants to perform tasks “entirely brainlessly” (P16) and only press “forward, then A, then forward, then A” (P17). Many participants expressed that they “felt too directed by [*Threshold*]” (P8), attesting *Continuous* a higher level of comfort or “freedom to experiment” (P24). Overall, participants described *Continuous* as a reasonable compromise or “the golden middle” (P14) between the comfortable execution in *Threshold* and the high level of control in *Classic*.

VI. Discussion

Adaptive DoF mapping controls have already been indicated to have benefits over classic methods [7], [8]. Yet, research is still limited, and analysis of *time-based dimensions* of directional cues is lacking. In this paper, we examined to what extent the two ADMC methods, *Continuous* and *Threshold*, differ from the *Classic* baseline – and each other – in terms of task completion time, necessary mode switches, perceived workload, and subjective assessment.

Significant results for all four metrics partially support our initial hypotheses. Most strikingly, ADMC methods reduced task completion time (*H1*) and mode switches (*H2*) by 50% respectively compared to *Classic*. As previously suggested by Kim et al., this establishes that ADMC methods lead to faster and less involved execution of pick-and-place tasks [6]. These findings are in line with previous work [7], underlining the benefits of ADMCs compared to *Classic* controls.

In contrast to previous results [8], our novel ADMC methods were able to significantly lower task completion time and perceived workload compared to the *Classic* method. The latter finding also partially supports *H3*. This highlights that ADMCs which communicate the suggested recommendation to the user – irrespective of timing – were able to increase usability. Notably, the decreased workload of ADMCs is particularly meaningful as the end goal should be the smooth integration of robotic devices into people’s lives and workflows, not to add stress.

Turning to the second part of our analysis – contrasting different time-based communication of feed-forward recommendations – we found no significant differences in the four metrics between *Continuous* and *Threshold*. The lack of measurable differences between *Continuous* and *Threshold* implies that both discrete and continuous communication of movement suggestions allows users to use ADMC methods efficiently. Insights gained by the results of the QUEAD and our qualitative interviews corroborate these findings, while the latter also helped to provide a more distinguished analysis.

Overall, participants expressed a positive stance regarding the ADMC methods. However, individual preferences vary greatly between *Continuous* and *Threshold*. While some participants preferred the higher level of control *Continuous* allowed, others favored the comfortable execution possible with *Threshold*. Consequently, future development of ADMC methods should – in accordance with Burkolter et al. – include individualization options to increase comfort and end-user acceptance [26]. Customizations would be particularly beneficial for *Threshold*-based controls as participants repeatedly criticized the multimodal feedback. Allowing users to adjust the modalities, the signal intensity, and even the threshold itself may improve usability while still offering the advantages of ADMC.

In contrast to expectations derived from our initial hypotheses, qualitative insights revealed that the *Classic* control method could still be a valuable addition in specific situations. Participants felt an apparent lack of control when the ADMC suggestions did not match their expectations. To improve usability, ADMC methods could incorporate static suggestions for certain situations. A potential way to address this could be combining ADMC and static suggestions using only the most common input-DoFs.

However, further experimental studies are needed to disentangle exactly which factors shape personal preferences and how customizations or crossover methods can deliver the best results.

A. Limitations

We explored the proposed ADMC methods in a VR simulation environment. While the usage of virtual simulations in industrial settings has been successfully established [27], [28], [29], future work should confirm if our promising findings can be replicated in the real world with a physical robot.

VII. Conclusions

Our ADMC methods *Continuous* and *Threshold* are promising approaches to communicate proposed directional cues effectively. We extend our previous work [8] by demonstrating that ADMCs significantly reduce task completion time (1), the average number of necessary mode switches (2), and the perceived workload of the user (3). Further, we establish that *Continuous* and *Threshold* perform equally well in quantitative measures while qualitative insights reveal individual preferences.

The observations of this study provide valuable implications for any HRI researcher involved in designing novel ADMC methods for human-robot collaborative settings. Future work should focus on disentangling quantitative and qualitative feedback of focus groups to develop optimal robot motion control methods, thus increasing usability, safety and – ultimately – end-user acceptance.

ACKNOWLEDGMENT

This research is supported by the *German Federal Ministry of Education and Research* (BMBF, FKZ: 16SV8563, 16SV8565). Our study is approved by the Ethics Committee of the *Faculty of Business Administration and Economics of the University of Duisburg-Essen* with the ID: 2202007.

REFERENCES

[1] S. Bragança, E. Costa, I. Castellucci, and P. M. Arezes, *A Brief Overview of the Use of Collaborative Robots in Industry 4.0: Human Role and Safety*. Cham: Springer International Publishing, 2019, pp. 641–650. [Online]. Available: https://doi.org/10.1007/978-3-030-14730-3_68

[2] C. Fattal, V. Leynaert, I. Laffont, A. Baillet, M. Enjalbert, and C. Leroux, "SAM, an Assistive Robotic Device Dedicated to Helping Persons with Quadriplegia: Usability Study," *International Journal of Social Robotics*, vol. 11, no. 1, pp. 89–103, 2019.

[3] T. B. Pulikottil, S. Pellegrinelli, and N. Pedrocchi, "A software tool for human-robot shared-workspace collaboration with task precedence constraints," *Robotics and Computer-Integrated Manufacturing*, vol. 67, p. 102051, Feb. 2021. [Online]. Available: https://doi.org/10.1016/j.rcim.2020.102051

[4] M. Beaudoin, J. Lettre, F. Routhier, P. S. Archambault, M. Lemay, and I. Gélinas, "Impacts of robotic arm use on individuals with upper extremity disabilities: A scoping review," *Canadian Journal of Occupational Therapy*, vol. 85, no. 5, pp. 397–407, 2018, pMID: 30866682.

[5] A. Pollak, M. Paliga, M. M. Pulopulos, B. Kozusznik, and M. W. Kozusznik, "Stress in manual and autonomous modes of collaboration with a cobot," *Computers in Human Behavior*, vol. 112, p. 106469, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0747563220302211

[6] D.-J. Kim, R. Hazlett-Knudsen, H. Culver-Godfrey, G. Rucks, T. Cunningham, D. Portee, J. Bricout, Z. Wang, and A. Behal, "How Autonomy Impacts Performance and Satisfaction: Results From a Study With Spinal Cord Injured Subjects Using an Assistive Robot," *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 42, no. 1, pp. 2–14, 2012.

[7] F. F. Goldau and U. Frese, "Learning to map degrees of freedom for assistive user control: Towards an adaptive dof-mapping control for assistive robots," in *The 14th PErvasive Technologies Related to Assistive Environments Conference*, ser. PETRA 2021. New York, NY, USA: Association for Computing Machinery, 2021, p. 132–139. [Online]. Available: https://doi.org/10.1145/3453892.3453895

[8] K. Kronhardt, S. Rübner, M. Pascher, F. Goldau, U. Frese, and J. Gerken, "Adapt or perish? exploring the effectiveness of adaptive dof control interaction methods for assistive robot arms," *Technologies*, vol. 10, no. 1, 2022.

[9] M. Walker, H. Hedayati, J. Lee, and D. Szafir, "Communicating robot motion intent with augmented reality," in *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 316–324. [Online]. Available: https://doi.org/10.1145/3171221.3171253

[10] M. Pascher, U. Gruenefeld, S. Schneegass, and J. Gerken, "How to communicate robot motion intent: A scoping review," in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems - CHI '23*, 2023. [Online]. Available: https://doi.org/10.1145/3544548.3580857

[11] A. Cleaver, D. V. Tang, V. Chen, E. S. Short, and J. Sinapov, "Dynamic path visualization for human-robot collaboration," in *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '21 Companion. New York, NY, USA: Association for Computing Machinery, 2021, p. 339–343. [Online]. Available: https://doi.org/10.1145/3434074.3447188

[12] N. J. Hetherington, E. A. Croft, and H. M. V. der Loos, "Hey robot, which way are you going? Nonverbal motion legibility cues for human-robot spatial interaction," *IEEE Robotics and Automation Letters*, vol. 6, no. 3, pp. 5010–5015, jul 2021. [Online]. Available: https://doi.org/10.1109%2Flra.2021.3068708

[13] R. T. Chadalavada, H. Andreasson, M. Schindler, R. Palm, and A. J. Lilienthal, "Bi-directional navigation intent communication using spatial augmented reality and eye-tracking glasses for improved safety in human–robot interaction," *Robotics and Computer-Integrated Manufacturing*, vol. 61, p. 101830, 2020.

[14] M. Pascher, K. Kronhardt, T. Franzen, U. Gruenefeld, S. Schneegass, and J. Gerken, "My caregiver the cobot: Comparing visualization techniques to effectively communicate cobot perception to people with physical impairments," *Sensors*, vol. 22, no. 3, p. 755, 2022. [Online]. Available: https://www.mdpi.com/1424-8220/22/3/755

[15] Z. Makhataeva and H. A. Varol, "Augmented reality for robotics: A review," *Robotics*, vol. 9, no. 2, 2020. [Online]. Available: https://www.mdpi.com/2218-6581/9/2/21

[16] L. V. Herlant, R. M. Holladay, and S. S. Srinivasa, "Assistive teleoperation of robot arms via automatic time-optimal mode switching," in *The Eleventh ACM/IEEE International Conference on Human Robot Interaction*, ser. HRI '16. IEEE Press, 2016, p. 35–42.

[17] M. C. Shrestha, A. Kobayashi, T. Onishi, E. Uno, H. Yanagawa, Y. Yokoyama, M. Kamezaki, A. Schmitz, and S. Sugano, "Intent communication in navigation through the use of light and screen indicators," in *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2016, pp. 523–524.

[18] I. Shindev, Y. Sun, M. Coovert, J. Pavlova, and T. Lee, "Exploration of intention expression for robots," in *Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '12. New York, NY, USA: Association for Computing Machinery, 2012, p. 247–248. [Online]. Available: https://doi.org/10.1145/2157689.2157778

[19] A. Singhal, "Modern information retrieval: A brief overview," *IEEE Data Eng. Bull.*, vol. 24, no. 4, pp. 35–43, 2001.

[20] S. G. Hart, "Nasa-Task Load Index (NASA-TLX); 20 Years Later," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 50, no. 9, pp. 904–908, 2006.

[21] J. Schmidtler, K. Bengler, F. Dimeas, and A. Campeau-Lecours, "A questionnaire for the evaluation of physical assistive devices (quead): Testing usability and acceptance in physical human-robot interaction," in *2017 ieee international conference on systems, man, and cybernetics (smc)*. IEEE, 2017, pp. 876–881.

[22] G. Harboe and E. M. Huang, "Real-world affinity diagramming practices: Bridging the paper-digital gap," in *Proc. 33rd Annual ACM Conf. Human Factors in Computing Systems*. ACM, 2015, pp. 95–104.

[23] D. C. Hoaglin and B. Iglewicz, "Fine-tuning some resistant rules for outlier labeling," *Journal of the American Statistical Association*, vol. 82, no. 400, pp. 1147–1149, 1987.

[24] S. Arévalo-Arboleda, F. Ruecker, T. Dierks, and J. Gerken, "Assisting manipulation and grasping in robot teleoperation with augmented reality visual cues," in *CHI Conference on Human Factors in Computing Systems (CHI '21)*. ACM, 2021. [Online]. Available: https://hci.w-hs.de/pub_VisualCuesCHI_compressed/

[25] S. Arévalo Arboleda, T. Dierks, F. Rücker, and J. Gerken, "Exploring the visual space to improve depth perception in robot teleoperation using augmented reality: The role of distance and target's pose in time, success, and certainty," in *Human-Computer Interaction – INTERACT 2021*, C. Ardito, R. Lanzilotti, A. Malizia, H. Petrie, A. Piccinno, G. Desolda, and K. Inkpen, Eds. Cham: Springer International Publishing, 2021, pp. 522–543.

[26] D. Burkolter, B. Weyers, A. Kluge, and W. Luther, "Customization of user interfaces to reduce errors and enhance user acceptance," *Applied Ergonomics*, vol. 45, no. 2, Part B, pp. 346–353, 2014. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0003687013000902

[27] E. Matsas and G.-C. Vosniakos, "Design of a virtual reality training system for human–robot collaboration in manufacturing tasks," *International Journal on Interactive Design and Manufacturing (IJIDeM)*, vol. 11, no. 2, pp. 139–153, Feb. 2015.

[28] S. L. Müller, S. Schröder, S. Jeschke, and A. Richert, "Design of a robotic workmate," in *Lecture Notes in Computer Science*. Springer International Publishing, 2017, pp. 447–456.

[29] E. Tidoni, M. Abu-Alqumsan, D. Leonardis, C. Kapeller, G. Fusco, C. Guger, C. Hintermuller, A. Peer, A. Frisoli, F. Tecchia, M. Bergamasco, and S. M. Aglioti, "Local and remote cooperation with virtual and robotic agents: A p300 BCI study in healthy and people living with spinal cord injury," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 9, pp. 1622–1632, Sept. 2017.

Abstract of the submitted paper 'Interdisciplinary Development and Reflection of a Robotic System for Nursing: An Overview of the AdaMeKoR Project'

The full paper is currently in review by the International Journal of Social Robotics: Special Issue on Robotic Systems for Nursing Care.

Jonathan Behrens[1*], Jonas Schulz[2], Felix Goldau[4], Christian Kowalski[3], Daniel Dorniok[2], Sandra Helmers[3], Stephanie Raudies[6], Rebekka Stellfeldt[4], Yashaswini Shivashankar[4], Niels Will[5], Serge Autexier[4], Cletus Brauer[6], Thomas Breisig[2], Andreas Hein[3], Manfred Hülsken-Giesler[1]

[1*]University of Osnabrück, Department of Nursing Science, Nelson-Mandela-Str. 13, Osnabrück, 49076, Germany.
[2]Carl von Ossietzky University of Oldenburg, Organisation und Personal, Ammerländer Heerstr. 140, Oldenburg, 26129, Germany.
[3]Carl von Ossietzky University of Oldenburg, Assistenzsysteme und Medizintechnik, Ammerländer Heerstr. 140, Oldenburg, 26129, Germany.
[4]Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI), Cyber Physical Systems, Enrique-Schmidt-Str. 5, Bremen, 28359, Germany.
[5]Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI), Robotics Innovation Center, Robert-Hooke-Str. 1, Bremen, 28359, Germany.
[6]Johanniter-Unfall-Hilfe e.V., Forschung und Entwicklung, An der Weinkaje 4, Elsfleth, 26931, Germany.

Abstract

The joint project An Adaptive Multi-Component Robot System for Nursing Care (AdaMeKoR) was dedicated to the development and research of robotic solutions to support nursing care practice from 2020 to 2023. In an interdisciplinary collaboration of technical, occupational science, nursing science and nursing practice perspectives, two robotic prototypes and conceptual designs to support care recipients and professional caregivers were developed and reflected. In addition to technical development, the empirical research of demonstrable relief potentials, profitable user interfaces as well as requirements and challenges for institutionally and socially successful implementation processes and ethical and nursing science legitimacy were the central areas of interest. The aim of this article is to provide a concise overview of the project work. Specifically, (1) empirical results of concrete evaluation studies and (2) overarching conclusions for the research and implementation of robotic systems for care will be presented. After a brief description of the areas of interest, an overview of the empirical research of the project is presented. Selected exemplary studies are then described individually in terms of their methodological approach. Subsequently, key results from the different scientific perspectives are presented. The results are then discussed against the background of the current discourse. Lastly, conclusions for overarching insights are outlined and the past three and a half years of the project are reflected.

Single-Abstract Extract of 'Abstracts des Gerontologie und Geriatrie Kongresses 2022 „Altern im Spannungsfeld von Resilienz und Vulnerabilität"' from page 107f.

S405–4

Die Perspektive potenzieller Nutzer*innen auf assistive Roboterarme in ambulanten Settings

A. Baumeister[1], F. Goldau[2], Y. Shivashankar[2], B. Klein[3], B.-P. Tolle[3], U. Frese[2]

[1]Forschungszentrum FUTURE AGING, Frankfurt University of Applied Sciences, Frankfurt a.M.; [2]Cyber-Physical Systems (CPS), Deutsches Forschungszentrum für Künstliche Intelligenz, Bremen; [3]Fachbereich Soziale Arbeit und Gesundheit, Frankfurt University of Apllied Sciences, Frankfurt a.M.

Assistive Roboterarme wie der JACO von Kinova inc. sollen Menschen mit Funktionsverlusten im Oberkörper bei täglichen Aktivitäten unterstützen und ihre Autonomie fördern (Brose et al. 2010). Dafür wird der JACO an einem E-Rollstuhl angebracht und über dessen Steuerung bedient (Baumeister et al. 2021). Allerdings ist die Ausübung alltäglicher Aktivitäten mit dem Roboterarm aufgrund einer zeitraubenden Steuerung mit Schwierigkeiten verbunden (Beaudoin et al. 2019; Herlant et al. 2017). In dem interdisziplinären BMBF Forschungsprojekt DoF-Adaptiv (Fördernr. 16SV8564) soll mit Hilfe von Künstlicher Intelligenz ein einfacheres Steuerungskonzept für den JACO entwickelt werden. Eine leichtere Steuerung könnte auch bei älteren Nutzer*innen die Akzeptanz von assistiven Roboterarmen erhöhen. Die Frankfurt UAS untersucht mit einem partizipativen methodischen Ansatz orientiert an der Aktionsforschungsspirale nach Riel (2020), ethische, rechtliche und soziale Implikationen (ELSI). In einem ersten Schritt wurden mit potenziellen Nutzer*innen, ihren pflegenden Angehörigen, Assistenz- und Pflegekräften in Workshops und Interviews Aktivitäten identifiziert, die im Alltag dieser Personengruppen von Bedeutung sind und mögliche Anwendungsszenarien für den JACO entwickelt. Neben den vier Szenarien „Essen und Trinken", „Türen öffnen und schließen", „Supermarktregal/Objekte greifen" und „Mikrowelle" konnten aus der Analyse der Workshops und Interviews auch nutzer*innenorientierte Anforderungen an assistive Roboterarme und deren Steuerung abgeleitet werden.

Literatur

1. Baumeister A et al (2021) The Importance of Participatory Design for the Development of Assistive Robotic Arms. Initial Approaches and Experiences in the Research Projects MobILe and DoF-Adaptiv. In: Günther, C; Klein, B; Schmidt, M; Weiland M (ed.), Connected Living. Conference Proceedings

2. Beaudoin M et al (2019) Long-term use of the JACO robotic arm: a case series. Disabil Rehabil Assist Technol 14(3):267-–275

3. Brose S et al (2010) The Role of Assistive Robotics in the Lives of Persons with Disability. Am J Phys Med Rehabil 89(6):509—521

4. Herlant L et al (2016) Assistive teleoperation of robot arms via automatic timeoptimal mode switching. In 2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI), 35–42. IEEE

5. Riel M (2020) Understanding action research. ccarweb.org/what-is-actionresearch, Stand 17. März 2022

Evaluating Assistive Technologies on a Trade Fair

Methodological Overview and Lessons Learned

ANNALIES BAUMEISTER*, Frankfurt University of Applied Sciences, Germany
FELIX GOLDAU*, German Research Center for Artificial Intelligence (DFKI), Germany
MAX PASCHER*, TU Dortmund University, Germany and University of Duisburg-Essen, Germany
JENS GERKEN, TU Dortmund University, Germany
UDO FRESE, German Research Center for Artificial Intelligence (DFKI), Germany
PATRIZIA TOLLE, Frankfurt University of Applied Sciences, Germany

Fig. 1. Impressions of a trade fair-based user evaluation

User-centered evaluations are a core requirement in the development of new user related technologies. However, it is often difficult to recruit sufficient participants, especially if the target population is small, particularly busy, or in some way restricted in their mobility. We bypassed these problems by conducting studies on trade fairs that were specifically designed for our target population (potentially care-receiving individuals in wheelchairs) and therefore provided our users with external incentive to attend our study. This paper presents our gathered experiences, including methodological specifications and lessons learned, and is aimed to guide other researchers with conducting similar studies. In addition, we also discuss chances generated by this unconventional study environment as well as its limitations.

CCS Concepts: • **Social and professional topics** → *People with disabilities*; • **Human-centered computing** → **HCI design and evaluation methods**; *Field studies*; Usability testing; *Empirical studies in HCI*; **Empirical studies in accessibility**.

Additional Key Words and Phrases: Assistive Robotics, User Study, Methodology, HRI, HCI, Action Research

*These authors – in alphabetical order – contributed equally to this research.

Authors' addresses: Annalies Baumeister, annalies.baumeister@fb4.fra-uas.de, Frankfurt University of Applied Sciences, Frankfurt, Germany; Felix Goldau, felix.goldau@dfki.de, German Research Center for Artificial Intelligence (DFKI), Bremen, Germany; Max Pascher, max.pascher@udo.edu, TU Dortmund University, Dortmund, Germany and University of Duisburg-Essen, Essen, Germany; Jens Gerken, jens.gerken@udo.edu, TU Dortmund University, Dortmund, Germany; Udo Frese, udo.frese@dfki.de, German Research Center for Artificial Intelligence (DFKI), Bremen, Germany; Patrizia Tolle, tolle@fb4.fra-uas.de, Frankfurt University of Applied Sciences, Frankfurt, Germany.

1 INTRODUCTION: WHY RUN USER EVALUATIONS ON A TRADE FAIR?

In 2023, the World Health Organization (WHO) constituted 15% of all people around the world to live with disabilities [39]. In Germany alone, 7.8 million people were identified as severely disabled by the end of 2021, with over half suffering from physical impairments, significantly impacting their mobility and leading to social and professional exclusion [31]. For many of these, who require consistent individual care-giving, assistive technologies can become an important tool to increase independence [19]. These range from basic aids to advanced robotics, offering independence and reducing caregiver dependency; thereby improving well-being and allowing individuals with disabilities to participate more fully in life. Furthermore, the aging population and their preference for aging in place amplify the demand for novel solutions [21].

The development of advanced assistive devices, such as robotic arms capable of performing daily tasks, presents a new frontier in support for those with motor impairments [5, 24]. However, challenges in the user's control and associated stress with autonomy need careful management. In these care-centered environments, flexibility and user-friendly controls are essential, as the technical proficiency of users varies widely and the complexity of such systems often poses barriers to effective use, especially for those with disabilities [33]. Current research in shared control and Artificial Intelligence (AI) aims to improve the usability and accessibility, emphasizing the importance of an intuitive operation and tailored user interfaces. This focus on enhancing human-robot interaction underscores the broader challenges in this field's research and development, including ethical and logistical hurdles, safety, recruitment, and the diverse needs of users.

Emphasizing user collaboration, research highlights the importance of involving users as active participants in the design process, leveraging their unique insights into individual needs and experiences [9, 22, 38]. This collaborative and interdisciplinary approach, supported by organizations like the WHO, underscores the value of viewing users as partners in the development and application of assistive technologies. Such involvement not only enhances functionality and accessibility, but also supports mental well-being and autonomy.

This is particularly relevant for those dependent on assistance for daily activities. The "Design for All" philosophy [32], integrating human-centered design with accessibility, advocates for incorporating user insights in the design process from the outset. Based on ethnographic studies and direct engagement with our target audience, our research builds on this foundation, identifying specific needs and challenges to inform the development of assistive technologies that address physical, social, and collaborative aspects for a more inclusive, empowering solution.

While this can, for some applications, be executed isolated in a controlled lab environment, most applications require interactions with users at some point during development; Some fields (e.g. Human-Computer Interaction (HCI), Human-Robot Interaction (HRI)) even have these at the very core of their research. However, sampling sufficient study participants to make reasonable claims is often difficult and not always a trivial task. This holds especially true for the field of assistive robotics, where the target population is limited in size and its members are often tightly scheduled and potentially vulnerable; be it physically, mentally, socially, or simply by introducing them to research-generated technologies that might help them, the production timelines of which are however too long to have any immediate use.

Nevertheless, as the field has high potential of improving the lives of people with technology, a lot of interesting and promising research is conducted and evaluated. Yet one can discuss the generalizability of various studies, as either only small shares of the study participants actually belong to the target population, or the total cohort of users is very small.

For example, Herlant et al. analyzed assistive robot control and compared the classic manual mode switching approach to one that is automatic and time-optimal. Their study shows interesting results, especially regarding the challenges associated with mode switches. However only their initial interviews were conducted with users from the target population ($N = 3$), whereas the rest of the evaluation was performed with able-bodied subjects [14]. Similarly, Jain et al. present an approach of assistive control using a body-machine interface and shared control. They tested it with 6 users, only one of which was a potential receiver of the technology [15]. Positive examples exist too, but often require specialized cooperation between partners: For example, Gopinath et al. proposed an assistive optimization framework with humans in-the-loop and conducted a pilot study with 17 subjects, 4 of which had spinal cord injuries. However, this group is located at the *Shirley Ryan AbilityLab*[1] and therefore profits from an established partnership with a rehabilitation hospital.

1.1 Contribution

In this work, we present an alternative approach, which utilizes the attracting effect of care-related trade fairs: By conducting studies in a booth of the fair, we were able to reach a considerably higher number of potential users when compared to a classical lab study and thus gain valuable insights for our research. As these studies required strategies specifically tailored to this environment, we aim to share our expertise with the community. In our case specifically, the studies focused on evaluating (shared) manual control of assistive robot arms for wheelchair users with limited upper limb mobility.

We therefore present our experiences regarding:

- opportunities, chances and advantages of trade fair-based studies,
- the special and/or unconventional requirements and preparations necessary to conduct such a study, and
- the limitations that arose in this unconventional environment.

2 RELATED WORK: PARTICIPATORY EVALUATIONS OF ASSISTIVE TECHNOLOGIES

Assistive technologies are increasingly recognized to be valuable tools in domestic care settings as they offer individuals with physical impairments the opportunity to regain a measure of independence by supplementing or reducing the need for ongoing human assistance [25]. Despite these benefits, the adoption and use of these technologies face challenges, including cases of non-acceptance and abandonment. Research by Klein [16] and Merkel and Kucharski [22] underscores the necessity of aligning these devices more closely with the specific needs and preferences of their intended users. Heeding this sentiment, Vines et al. [35] suggest involving potential users early in the development process to enhance device acceptance and utility.

In the social sciences, such participatory approaches of research are common ever since Kurt Lewin developed the *action research method* and gained increased importance in health care research with the WHO "Health for All Strategy". Here, the assigned goal is an active participation of those affected in the research process, thereby collaboratively gaining knowledge, reflecting, influencing, and thus changing the research process [3, 36].

[1] Shirley Ryan AbilityLab (formerly Rehabilitation Institute of Chicago) https://www.sralab.org/ last visited August 20, 2024

This synergy is also known in more technical fields, where they found the active participation of potential users in the design process of assistive technologies to be crucial but challenging [6, 17, 23]: Dalko et al. highlight the significant difficulties in patient involvement, particularly among those with long-term illnesses, due to hierarchical barriers in care-institutions and a lack of established patient groups. Nonetheless, this involvement is key to developing devices that meet the specific needs of users and facilitate better outcomes in terms of usability and acceptance [6]. Towards the end of the 1990s, action research was introduced to Information System research (IS) research, among others, by Baskerville and Wood-Harper and found its way into today's HCI research. Baskerville and Wood-Harper [2] and Hayes [13] both note that the participatory and collaborative approach of action research fits to methods and issues previously used by researchers in IS and HCI but extends their setup with an ethical framework.

In these technical fields, laboratory studies are very common and provide standardized and methodically controlled approaches, as well as being simpler and more economical by avoiding difficulties that could arise in the field. However, data generated in laboratories misses everyday conditions and consequently leads to discussions of data validity [3, 8]. In addition, these in-lab studies often face logistical challenges, especially when involving participants with mobility impairments. The difficulties in transporting individuals to and from study locations can significantly impact the feasibility and cost of research, suggesting a need for more accessible and inclusive research methodologies [20, 29, 30].

Field research on the other hand, refers to processes that are observed in real life's everyday settings, thus avoiding various issues inside laboratories. Downsides of field research however lie in the variation of conditions, the randomization of the perturbing conditions and therefore the method, as well as multiple (uncontrollable) effects that might limit internal validation of such studies [18]. Here, action research provides tools to methodically gain valid knowledge whilst collaborating with the target group [13].

One variation could be conducting studies within the homes of participants, as this allows for a more realistic understanding of how assistive technologies function in everyday settings while keeping the environment semi-controlled. However, these studies typically involve smaller sample sizes due to logistical and financial constraints, potentially limiting the generalizability of findings [1, 7].

Still, research conducted in real-world settings (in-the-wild studies) provides valuable insights into how assistive technologies are used in daily life. These studies can highlight issues of device acceptance and long-term use that may not be apparent in more controlled research settings [12, 17, 23].

In summary, the design and development of assistive technologies benefit significantly from involving the target user group at every stage, from ideation to final product testing. Identifying this requirement ensures that the resulting devices are not only technically sound but also tailored to the real-world needs and preferences of their users. Addressing the challenges associated with in-lab, in-home, and in-the-wild studies is essential for advancing our understanding of assistive technology use and improving outcomes for individuals with disabilities.

3 METHODOLOGY OF OUR EXAMPLE USER EVALUATIONS

A very important element of any user-centric evaluation is subject recruitment, the success of which depends primarily on one's location: It can be very struggling to sample a sufficient number of people from the target population if the study is location bound and the subjects are expected to travel to the research lab. As an alternative, we searched for places with a considerably higher-than-usual distribution of care-receiving individuals, finally landing on trade fairs for care and rehabilitation. Participants from previous studies suggested to look at these, as they are fixed annual events within the community.

Here, healthcare providers and (self-advocated) societies gather to exchange experience and inform the public, among other topics, about provided services, relevant regulations, and available federal social-care benefits. In addition, they also include a large marketplace for manufacturers to showcase their (new) designs and technologies to the target audience. This creates large incentives for those involved in care, as they can personally observe and experience a large number of products which might have the potential to improve their lives. The condensed experience is especially attractive for people who find traveling to be particularly strenuous, be it due to disabilities or other means.

Ultimately we decided on the *REHAB* trade fair[2] in Karlsruhe (Germany) and the *REHACARE* trade fair[3] in Düsseldorf (Germany), both internationally well known trade fairs for rehabilitation and care. Consequently, they are also known to be visited by many people with a disability who not only use it to inform themselves about new aids, but also to meet up and network with their peers.

For us, this meant extremely high numbers of potential (primary and secondary) users, which do not need to explicitly travel only to join our studies, but were basically already on-location. For people who were previously associated with our projects (e.g. due to the participatory design), we were able to offer discounts on the entry. These conditions allow for a way tighter definition of the study subjects, simply because of the associated shift in the local distribution: Instead of rough estimates of potentially care-receiving individuals, substantial sample sizes can be reached with tighter and more fitting inclusion criteria, e.g. acquiring only wheelchair users with limited mobility in their upper extremities.

However, this special environment also greatly influences the objectives that can reasonably be evaluated: We have to assume our participants to be less focused, both due to more external interference, as well as individual agendas as trade fair guests. The latter also strictly limits the available time per user. As a result, it is reasonable to focus on qualitative objectives, relying more on interviews and personal user remarks, rather than interpreting too much into individual trials. For us, this meant selecting objectives that focus on user feedback, acceptance and preferences.

3.1 Experimental Design

As the main reason to select this type of study is a shortage of suitable participants, the most valuable resource whilst conducting the

[2] *REHAB* trade fair. https://www.rehab-karlsruhe.com, last retrieved August 20, 2024
[3] *REHACARE* trade fair. https://www.rehacare.de, last retrieved August 20, 2024.

study is the users' time and willingness to contribute. We therefore aimed to optimize their experience as far as possible, by isolating temporal bottlenecks and widening them by conducting the study partially in parallel using multiple researchers: One person recruits the next user and give introductory information, while another one performs the robot interaction with the current user, and a third person debriefs the previous user and runs an final interview.

The physical setup of our studies at trade fairs was relatively minimalistic: a robot arm is mounted to a table in a standard booth with sufficient room around it, such that both wheelchair users as well as researchers can easily access the system. A simple sketch of an exemplary booth setup is shown in Figure 2: Our main experimentation area is shown on the left, with the right side representing our project partner's booth, which we could partially use for the interviews. For the participants, we assured a minimal distance to the robot as a safety measure to the semi-chaotic nature of the environment. We purposefully avoided completely enclosing structures on the booth to preserve the visibility of the study as an advertising element for recruitment. Finally, cameras and microphones were set up around the experimentation area to allow a subsequent analysis of user feedback and remarks in addition to notes taken by hand.

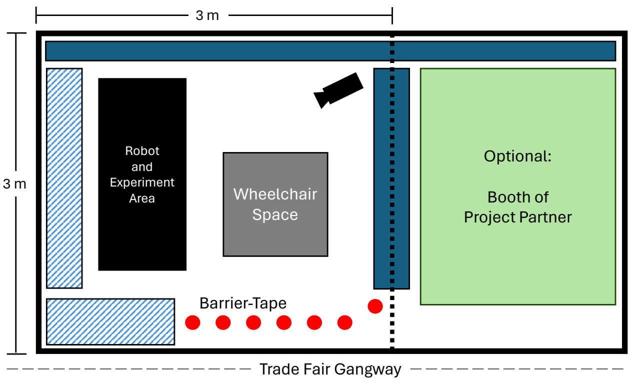


Fig. 2. Sketch of exemplary booth setup

In addition to technical and logistical preparations, a sufficiently thorough ethical evaluations was necessary. This included limiting the trial time to not cause exhaustion for our participants, which we attempted by structuring the trials such that the duration for every participant would not be longer than 60 minutes all together.

Another topic of ethical review was privacy and data security during the collection of personal data in such a semi-public space. As previously mentioned, completely enclosing the booth for this did not seem a realistically implementable scenario, as it would result in higher costs (e.g. building materials, extra booth space) and a less attractive booth, thus making recruiting more difficult. Instead, we used natural obstructions (e.g. semi-transparent shelves, striped blue in Figure 2) and planed to positioned the setup in such a way that the participants were oriented with their back towards the aisle and most other visitors passing by. Personal data was collected with a questionnaire to be filled out directly by the participants or their companions. Ethic approvals were obtained for both studies.

4 EXAMPLE STUDIES

We executed two very distinct studies on trade fairs at very different stages of projects: Study A was conducted at the *REHAB* trade fair with a minimalistic physical setup and served mostly for an evaluation of the state of the art, harvesting of subjective user requirements, and recruitment of primary users of assistive robotics for the ongoing participatory research. Close to the very end of our project, Study B was conducted at the *REHACARE* trade fair in Düsseldorf (Germany) and served more as the final stage of evaluation: We used this chance and our experience from study A to assess our solution with subjects sampled solely from the target population.

This section provides a brief overview of the studies, focusing on their comparative user requirements, meta results and resonance.

4.1 Study A: Explorative User Tests

Having had only very limited experience of studies during a trade fair, we cautiously designed this first study to be mainly explorative, focus on qualitative user feedback, and potentially recruit people to join us for our upcoming participatory development. In this initial study, we aimed to compare existing options of manual control for an assistive robot arm by gathering feedback from people associated with care. This included care recipients, as well as informal and professional caregivers. We intentionally set these relatively loose conditions, as we were looking for diverse perspectives on the matter and were uncertain of the actual user distribution at the trade fair.

In total, 26 participants joined our experiment, 10 of which were care-receiving wheelchair users. As we conducted the study in a booth during the 3-day *REHAB* trade fair and each subject stayed with us for 30 to 45 minutes, this sample size came close to our full capacity and exceeded our previous expectations.

The results showed a preference for a direct control mapping combined with a minimization of mode switches, as well as a willingness to be confronted with more complex input devices and train with them. In addition, we gathered invaluable insights into requirements and situations of our target population; By basically sampling from the wild, we were able to include user profiles that might have otherwise be forgotten, overlooked, or incorrectly excluded for our studies (e.g. users with spastics).

All together, the participants were all very interested and reported enthusiasm in joining our study. They shared previous experiences with similar systems, as well as contextual anecdotes, with the professional care-givers often providing technical expert clarifications.

4.2 Study B: Final User Evaluation

Based on the positive resonance and high number of participants of the previous study, we decided to also conduct the final evaluation of our project-developed shared control approach at a trade fair. In this study, we selected the larger *REHACARE* and defined the inclusion criteria to be more specific: wheelchair-users with limited mobility of their upper limbs. Impressions from the study can be seen in Figure 1.

Compared to the previous study, this evaluation was more structured and less explorative in order to allow us a more substantiated analysis of our shared control. While the concept was previously shown to be functional with able-bodied users [10, 28], the verification with the target group was still lacking. Bridging this gap was the main goal of the study presented here: Can the users learn the control sufficiently quick; and do they perceive it well and accessible? By adjusting to people's needs, we also analyzed the generalizability of the control to different input devices. In contrast to Section 4.1, the tasks were defined with measurable brief goals such that trials could be recorded and (partially quantitatively) compared afterwards.

Due to the more complex nature of the study and including interviews and preparation, participants spend roughly 60 minutes with us. In total, we managed to gather data from 24 people of the target population (wheelchair users with limited mobility in their upper limbs) during the 4-day long trade fair. As the physical capabilities of users varied vastly and were not known by us in advance, spontaneous adjustments to the mechanical setup were often necessary.

Nevertheless, participants were again very enthusiastic and even glad to be included in the research process. The results showed the capabilities of the proposed control, as well as an increase in understanding and acceptance of the control during the trials.

5 LESSONS LEARNED

Trade fairs are messy: There is a lot of noise, it is at times very crowded, and there are various distractions. For example, spotlights disrupt visual interfaces, while the huge number of transmitting technical devices interfere with wireless connections. But with the challenges do come opportunities and unexpected results that would not occur in more controlled setting.

Running Trials with an Audience: During both studies, seeing the robot in action was very attractive for other guests, who stopped to watch, often asking questions or giving comments, however also partially generating unwanted performance pressure for study participants. Consequently in Study B some participants remarked on the audience or other distractions during the interviews, implying that a calmer setting could have led to a better performance. One participant stated "*But at the trade fair, there are people, time pressure. (...) and that's a bit more strenuous than at home.*"[4] Another one said: "*Oh, if you try that a few times and there's no audience there, (...) you'll become more confident.*" Other participants did not mind being watched during testing, with some even asking their companions take pictures or record them.

Acquisition of Participants: At both trade fairs, experience showed that guests are keen to test new technologies and are generally open to new ideas. We especially found our target audience of wheelchair users with limited mobility in their upper limbs to be very curious and open to us, which greatly simplified and accelerated acquisition. Many participants expressed their joy in testing our robot control: "*I just think it's really great that this option exists. And it was nice to be able to test it out.*" Confirming our assumptions, participants told us that visiting these events is a regular (mostly annual) activity for them: They use it to stay on track with technology, find new assistive devices and gather with their peers.

[4]This and all further direct quotes are translated from German.

Recruitment for the study was therefore relatively unstructured: by basically sampling in-the-wild, we approached potential participants and had short condense introductory talks to get them interested in the study. As both the participants and our time on the trade fair was limited, the pre-study briefing was held minimalistic, especially when compared to recruitment talks in a lab-based study with travel time. While this interchange was at all possible due to the reduction of hurdles and consequently quick launches into participation, unexpected difficulties also arose due to diverse and previously unknown user situations. This included spastics which made holding controllers difficult, head rests or vision impairments that prevented the use of smart glasses, breathing aids or speech impediments that restricted communication, and neurological impairments. For most users, we were able to find spontaneous workarounds (e.g. repositioning and propping up controllers to lessen spastics, or setting up an external screen for vision impairments), but some trials had be aborted. In addition to blocking valuable time, this often left participants and researchers unsatisfied.

Specific to the trade fairs, we observed that appointments for trials tended not to work. The agenda of the guests changed too rapidly, as to allow them to return to our booth at a predefined time. In addition, it has to be considered that people participating in the afternoon after several hours at the trade fair tended to be more exhausted.

Interviewing on a Trade Fair: The interviews of study B were conducted without an extra booth or fixed place. They mainly took place in a corner of our partner's booth and in some cases even at the edge of the aisle next to our booth. Not interviewing in our booth had the advantage that the next person could start their trial while the previous one was still in their closing interview. For the researcher, it was inconvenient to constantly search for a place to conduct the interview, however, it provided the participants with a moment to clear their mind before the questions started. Further challenges that arose due to the conditions of the fair included the noise level, crowded space, as well as trials or demonstrations of our project partners in clear view. This led to a lot of distractions and interruptions during the interviews and made it hard for both parties to stay focused. Sometimes, caused by the background noises, the interviewers missed parts of an answer, thus loosing the opportunity for follow-up questions. This noise level also affected the transcription of the audio recording, resulting in 5 interviews, where part of the answers could not be transcribed. In 6 further interviews, it was challenging to distinguish the speaker. Although an AI-based software was used for transcription, the transcripts needed to be manually corrected more thoroughly than usual. To find a practical solution to conduct interviews with less noise on a trade fair is not that easy. Still, a bigger booth with a fixed place for the interviews might have led to fewer distractions and a better interviewing quality.

6 DISCUSSION

Typically, similar studies are conducted in laboratories of research institutes, i.e. subjects are recruited and invited in advance and the environment is known and completely controllable by the researchers. This comes with various advantages, accumulating to generally more predictable procedures: (1) There are none to very few external influences in a lab, resulting in less distractions and consequently cleaner trials, (2) the participants are known and associated possible complications can be surmised in advance, and (3) it is possible to schedule participants, as they will arrive independently and no study-related equipment needs to be transported.

However, this requires a pre-existing cohort of subjects to sample from (which is non-trivial for sufficiently tight requirements) and loads the burden of traveling to the lab onto the participants. Especially for people with physical impairments, this can be a major task, which, among other things, involves the availability of accompanying persons, options of transport, and space in one's own timetable, which is often enough stuffed with therapies.

In contrast, the presented trade fair-based approach inverts the situation. Both the environment as well as the participants are unknown variables; however, the latter have already traveled to the study location, such that they are not additionally burdened. This requires, of course, an existing trade fair that specifically addresses the target group.

Other imaginable alternatives are purely virtual off-site studies (e.g. [27]), purely ethnographic studies where researchers travel to users' homes without equipment (e.g. [26]), or expensive and complicated evaluations where researchers visit users' homes and bring along equipment such as robots (e.g. [4]). However, each of these variations come with extensive downsides.

Consistent throughout these alternative methodologies is the requirement of known predetermined users. In contrast, the presented studies recruited participants on-the-fly from a cohort of "free-roaming" trade fair guests. This resulted in a much larger diversity of participants, both in terms of physical capabilities, as well as previous technical experience and acceptance, thus improving the scientific significance. In particular, this includes less tech-savvy users who might be more critical towards such systems and would therefore regularly not get involved with technical research [34, 37].

On the downside, this came with unforeseen challenges (e.g. partial blindness on one eye) which required spontaneous adjustments, not all of which could be met on-site. Still, providing a larger group of potential future users with a chance to evaluate the technology meant that, in the spirit of action research, as many people from our target group were involved in our final evaluation as possible: They shared their knowledge as experts in their own field, including former experiences with assistive technologies, life situation and the resulting requirements; thus supplementing their perspective and influencing further research [18, 40].

In Section 3.1, we discussed the situation regarding privacy and data security of personal data and its impact on the design of the booth. Arriving at the trade fair (for study B) and seeing the situation on site for the first time, it soon became clear that the set up could not be implemented as planned. Instead of placing the demonstrator in a way that participants would face the back wall of the booth, they would now be positioned lateral and be more visible for other passing visitors. That meant less privacy and more distractions during the trial. However, it is quite common on such trade fairs for visitors to try out new aids (e.g. wheelchairs or robotic eating devices) in public and without any consideration to privacy. Doing so, visitors of the trade fair are well aware that others stop to watch or even take pictures. Still, whenever we noticed someone taking pictures, we asked them to only depict the robot and ensure the participants to not be recognizable, e.g. by taking the picture from behind. We therefore conclude that the reduction of privacy at a trade fair is a reasonable circumstance for the participants.

Another challenge was a secure way of collecting of personal data. Even though some participants were able to manually fill out the questionnaire (sometimes with assistance from their companions), they often requested a researcher to collaborate. This was done as discreetly as the circumstances allowed. The trials itself took longer than anticipated: Instead of planned 45 minutes, most participants stayed with us for about 60 minutes. Luckily, this showed not to be a problem at all, as all participants were happily ready to invest this time. Some expressed excitement and joy during or after their trail.

Study Result Validity: In our investigation of inclusive and assistive human-robot interaction, we conducted two distinct studies, each offering unique insights into the effectiveness and acceptance of our approach.

The first study, Study A, took place at the REHAB trade fair in Karlsruhe, Germany. With a focus on qualitative feedback and exploration, we engaged 26 participants over the course of three days. Building upon the success of Study A, we proceeded to Study B, a final evaluation conducted at the REHACARE trade fair in Düsseldorf. Here, we targeted wheelchair users with limited upper limb mobility, totaling 24 participants over four days. Adopting a more structured methodology, including interviews and defined task trials, we sought to validate the effectiveness and acceptance of our shared control approach.

Together, these studies provide a robust foundation – involving 50 participants – for assessing the viability of assistive robotics in inclusive environments. With a diverse participant pool and a combination of qualitative and structured methodologies, we have garnered valuable insights into user preferences, requirements, and acceptance, paving the way for future advancements in inclusive human-robot interaction. In general, participants demonstrated enthusiasm and willingness to engage with the research process, reinforcing the validity of our findings.

7 CONCLUSION

We presented an in-depth analysis of the capabilities that arise from running robotic studies at trade fairs, with a special focus on assistive technologies designed for care-receiving individuals. For this, we showed a generalized methodology, provided brief summaries of the approaches and results of two different studies we performed on trade fairs, and discussed our experiences with this unconventional study setting.

As discussed, the study conditions on a trade fair differ vastly from those of a typical lab-based evaluation. In short, one exchanges a bit of predictability and general control in the lab with a way better adjusted localized target population (i.e. easier recruitment of appropriate subjects) and a more realistic in-the-wild environment. As shown, the setting of a trade fair has its own, partially chaotic, dynamics. Therefore a thorough planing is needed, including situation-aware preparations, but also researchers that are willing to react spontaneously and be ready to improvise.

This definitely does not expose trade fairs as the general go-to option for user evaluations, but instead highlights them as a valid alternative to a lab-based study. This is especially the case for studies that are mostly qualitative and focus on a user group that is either limited in size, has difficulties traveling to research laboratories, or difficult to contact in the first place.

While this requires a mobile study setup, custom preparations in advance, and capabilities for spontaneous problem solving at the trade fair itself, we can highly recommend examining this option. For us, it provided larger numbers of more appropriate participants and extremely valuable insights from exchanges with the real target group. Besides, quite a lot of our participants stated, that they had fun during the trails and enjoyed being able to join the study and partake in our research.

ACKNOWLEDGMENTS

This research is supported by the *German Federal Ministry of Education and Research* (BMBF, FKZ: 16SV8563, 16SV8564, and 16SV8565). Our studies were approved by the Ethics Committee of the *Faculty of Business Administration and Economics of the University of Duisburg-Essen* and the Ethics Committee of the Carl von Ossietzky University of Oldenburg.

REFERENCES

[1] Wayne L. Anderson and Joshua M. Wiener. 2013. The Impact of Assistive Technologies on Formal and Informal Home Care. *The Gerontologist* 55, 3 (Dec. 2013), 422–433. https://doi.org/10.1093/geront/gnt165

[2] Richard L. Baskerville and A. Trevor Wood-Harper. 1996. A critical perspective on action research as a method for information systems research. *Journal of Information Technology* 11, 3 (01 Sep 1996), 235–246. https://doi.org/10.1080/026839696345289

[3] Fran Baum, Colin MacDougall, and Danielle Smith. 2006. Participatory action research. *Journal of Epidemiology & Community Health* 60, 10 (2006), 854–857. https://doi.org/10.1136/jech.2004.028662 arXiv:https://jech.bmj.com/content/60/10/854.full.pdf

[4] Jonathan Behrens, Jonas Schulz, Felix Goldau, Christian Kowalski, Daniel Dorniok, Sandra Helmers, Stephanie Raudies, Rebekka Stellfeldt, Yashaswini Shivashankar, Niels Will, Serge Autexier, Cletus Brauer, Thomas Breisig, Andreas Hein, and Manfred Hülsken-Giesler. 2024. Interdisciplinary Development and Reflection of a Robotic System for Nursing. *International Journal of Social Robotics* (2024). Issue Special Issue on Robotic Systems for Nursing Care. Submitted.

[5] Karol Connors, Lisa Mahony, and Prue Morgan. 2019. Variation in assistive technology use in Motor Neuron Disease according to clinical phenotypes and ALS Functional Rating Scale – Revised Score: A prospective observational study. *NeuroRehabilitation* 44, 2 (4 2019), 303–313. https://doi.org/10.3233/nre-182511

[6] Katharina Dalko, Bernhard Kraft, Patrick Jahn, Jan Schildmann, and Sebastian Hofstetter. 2023. Cocreation of Assistive Technologies for Patients With Long COVID: Qualitative Analysis of a Literature Review on the Challenges of Patient Involvement in Health and Nursing Sciences. *J Med Internet Res* 25 (15 Aug 2023), e46297. https://doi.org/10.2196/46297

[7] Louise Demers, W. Ben Mortenson, Marcus J. Fuhrer, Jeffrey W. Jutai, Michelle Plante, Jasmine Mah, and Frank DeRuyter. 2016. Effect of a tailored assistive technology intervention on older adults and their family caregiver: a pragmatic study protocol. *BMC Geriatrics* 16, 1 (May 2016), 1–11. https://doi.org/10.1186/s12877-016-0269-3

[8] Jochen Fahrenberg. 2016. *Labor-Feld-Problem.* Dorsch - Lexikon der Psychologie, Bern. https://dorsch.hogrefe.com/stichwort/labor-feld-problem last visited 11.04.2024.

[9] Susanne Frennert and Britt Östlund. 2014. Review: Seven Matters of Concern of Social Robots and Older People. *International Journal of Social Robotics* 6, 2 (Jan. 2014), 299–310. https://doi.org/10.1007/s12369-013-0225-8

[10] Felix Goldau and Udo Frese. 2024. Probabilistic Combination of Heuristic Behaviors for Shared Assistive Robot Control. In *The PErvasive Technologies Related to Assistive Environments (PETRA) conference (PETRA '24)* (Crete, Greece). Association for Computing Machinery, New York, NY, USA, 465–467. https://doi.org/10.1145/3652037.3652071

[11] Deepak Gopinath, Siddarth Jain, and Brenna D. Argall. 2017. Human-in-the-Loop Optimization of Shared Autonomy in Assistive Robotics. *IEEE Robotics and Automation Letters* 2, 1 (2017), 247–254. https://doi.org/10.1109/LRA.2016.2593928

[12] Murray Goulden, Christian Greiffenhagen, Jon Crowcroft, Derek McAuley, Richard Mortier, Milena Radenkovic, and Arjuna Sathiaseelan. 2017. Wild interdisciplinarity: ethnography and computer science. *International Journal of Social Research Methodology* 20, 2 (2017), 137–150. https://doi.org/10.1080/13645579.2016.1152022 arXiv:https://doi.org/10.1080/13645579.2016.1152022

[13] Gillian R. Hayes. 2011. The relationship of action research to human-computer interaction. *ACM Trans. Comput.-Hum. Interact.* 18, 3, Article 15 (aug 2011), 20 pages. https://doi.org/10.1145/1993060.1993065

[14] Laura V. Herlant, Rachel M. Holladay, and Siddhartha S. Srinivasa. 2016. Assistive Teleoperation of Robot Arms via Automatic Time-Optimal Mode Switching. In *The Eleventh ACM/IEEE International Conference on Human Robot Interaction* (Christchurch, New Zealand) *(HRI '16)*. IEEE Press, Piscataway, NJ, 35–42.

[15] Siddarth Jain, Ali Farshchiansadegh, Alexander Broad, Farnaz Abdollahi, Ferdinando Mussa-Ivaldi, and Brenna Argall. 2015. Assistive robotic manipulation through shared autonomy and a Body-Machine Interface. In *2015 IEEE International Conference on Rehabilitation Robotics (ICORR)*. IEEE, Piscataway, NJ, 526–531. https://doi.org/10.1109/ICORR.2015.7281253

[16] Barbara Klein. 2020. *Hilfsmittel, Assistive Technologien und Robotik: Selbstständigkeit und Lebensqualität im Alter erhalten.* Kohlhammer Verlag, Stuttgart, Germany.

[17] Leah Kulp and Aleksandra Sarcevic. 2018. Design In The "Medical" Wild: Challenges Of Technology Deployment. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems* (<conf-loc>, <city>Montreal QC</city>, <country>Canada</country>, </conf-loc>) *(CHI EA '18)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3170427.3188571

[18] Siegfried Lamnek and Claudia Krell. 2016. *Qualitative Sozialforschung: mit Online-Material* (6. ed.). Beltz, Weinheim. https://content-select.com/de/portal/media/view/56cc0a3b-a3c0-4460-bf58-5eeeb0dd2d03

[19] Bente Martinsen, Ingegerd Harder, and Fin Biering-Sorensen. 2008. The meaning of assisted feeding for people living with spinal cord injury: a phenomenological study. *Journal of Advanced Nursing* 62, 5 (May 2008), 533–540. https://doi.org/10.1111/j.1365-2648.2008.04637.x arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-2648.2008.04637.x

[20] Colleen McMillan, Joseph Lee, James Milligan, Loretta M. Hillier, and Craig Bauman. 2015. Physician perspectives on care of individuals with severe mobility impairments in primary care in Southwestern Ontario, Canada. *Health & Social Care in the Community* 24, 4 (March 2015), 463–472. https://doi.org/10.1111/hsc.12228

[21] Josephine McMurray, Gillian Strudwick, Cheryl Forchuk, Adam Morse, Jessica Lachance, Arani Baskaran, Lauren Allison, and Richard Booth. 2017. The Importance of Trust in the Adoption and Use of Intelligent Assistive Technology by Older Adults to Support Aging in Place: Scoping Review Protocol. *JMIR Res Protoc* 6, 11 (02 11 2017), e218. https://doi.org/10.2196/resprot.8772

[22] Sebastian Merkel and Alexander Kucharski. 2018. Participatory Design in Gerontechnology: A Systematic Literature Review. *The Gerontologist* 59, 1 (5 2018), e16–e25. https://doi.org/10.1093/geront/gny034

[23] Kyle Montague, André Rodrigues, Hugo Nicolau, and Tiago Guerreiro. 2015. TinyBlackBox: Supporting Mobile In-The-Wild Studies. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility* (Lisbon, Portugal) *(ASSETS '15)*. Association for Computing Machinery, New York, NY, USA, 379–380. https://doi.org/10.1145/2700648.2811379

[24] P. Muthu, Yongqi Tan, S. Latha, Samiappan Dhanalakshmi, Khin Wee Lai, and Xiang Wu. 2023. Discernment on assistive technology for the care and support requirements of older adults and differently-abled individuals. *Frontiers in Public Health* 10 (1 2023), 19 pages. https://doi.org/10.3389/fpubh.2022.1030656

[25] Daehyung Park, Yuuna Hoshi, Harshal P. Mahajan, Ho Keun Kim, Zackory Erickson, Wendy A. Rogers, and Charles C. Kemp. 2020. Active robot-assisted feeding with a general-purpose mobile manipulator: Design, evaluation, and lessons learned. *Robotics and Autonomous Systems* 124 (2 2020), 103344. https://doi.org/10.1016/j.robot.2019.103344

[26] Max Pascher, Annalies Baumeister, Stefan Schneegass, Barbara Klein, and Jens Gerken. 2021. Recommendations for the Development of a Robotic Drinking and Eating Aid - An Ethnographic Study. In *Human-Computer Interaction – INTERACT 2021* (2021-09-01), Carmelo Ardito, Rosa Lanzilotti, Alessio Malizia, Helen Petrie, Antonio Piccinno, Giuseppe Desolda, and Kori Inkpen (Eds.). Springer, Cham, Bari, Italy, 331–351. https://doi.org/10.1007/978-3-030-85623-6_21

[27] Max Pascher, Felix Ferdinand Goldau, Kirill Kronhardt, Udo Frese, and Jens Gerken. 2024. AdaptiX – A Transitional XR Framework for Development and Evaluation of Shared Control Applications in Assistive Robotics. *Proc. ACM Hum.-Comput. Interact.* 8, EICS (2024), 28 pages. https://doi.org/10.1145/3660243 Preprint on arXiv: https://arxiv.org/abs/2310.15887.

[28] Max Pascher, Kevin Zinta, and Jens Gerken. 2024. Exploring of Discrete and Continuous Input Control for AI-enhanced Assistive Robotic Arms. In *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction* (Boulder, CO, USA) *(HRI '24)*. Association for Computing Machinery, New York, NY, USA, 828–832. https://doi.org/10.1145/3610978.3640626

[29] Albert B. Schultz. 1992. Mobility impairment in the elderly: Challenges for biomechanics research. *Journal of Biomechanics* 25, 5 (May 1992), 519–528. https://doi.org/10.1016/0021-9290(92)90092-f

[30] Tobias Sonne, Carsten Obel, and Kaj Grønbæk. 2015. Designing Real Time Assistive Technologies: A Study of Children with ADHD. In *Proceedings of the Annual Meeting of the Australian Special Interest Group for Computer Human Interaction (OzCHI '15)*. ACM, New York, NY, USA, 34–38. https://doi.org/10.1145/2838739.2838815

[31] Statistisches Bundesamt (Destatis). 2022. Disability Facts and Figures – Brief Report 2021. Published online: https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Gesundheit/Behinderte-Menschen/Publikationen/Downloads-Behinderte-Menschen/sozial-schwerbehinderte-kb-5227101219004.html (last visited on August 20, 2024).

[32] Constantine Stephanidis and Gavriel Salvendy. 1998. Toward an Information Society for All: An International Research and Development Agenda. *International Journal of Human-Computer Interaction* 10, 2 (June 1998), 107–134. https://doi.org/10.1207/s15327590ijhc1002_2

[33] Constantine Stephanidis, Gavriel Salvendy, Margherita Antona, Jessie Y. C. Chen, Jianming Dong, Vincent G. Duffy, Xiaowen Fang, Cali Fidopiastis, Gino Fragomeni, Limin Paul Fu, Yinni Guo, Don Harris, Andri Ioannou, Kyeong-ah (Kate) Jeong, Shin'ichi Konomi, Heidi Krömker, Masaaki Kurosu, James R. Lewis, Aaron Marcus, Gabriele Meiselwitz, Abbas Moallem, Hirohiko Mori, Fiona Fui-Hoon Nah, Stavroula Ntoa, Pei-Luen Patrick Rau, Dylan Schmorrow, Keng Siau, Norbert Streitz, Wentao Wang, Sakae Yamamoto, Panayiotis Zaphiris, and Jia Zhou. 2019. Seven HCI Grand Challenges. *International Journal of Human–Computer Interaction* 35, 14 (7 2019), 1229–1269. https://doi.org/10.1080/10447318.2019.1619259

[34] Rahul Thakurta. 2017. Identifying the Motives for User Participation in Information System Projects. *Pacific Asia Journal of the Association for Information Systems* 9 (2017), 67–96. Issue 3. https://doi.org/10.17705/1pais.09304

[35] John Vines, Rachel Clarke, Peter Wright, John McCarthy, and Patrick Olivier. 2013. Configuring Participation: On How We Involve People in Design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 429–438. https://doi.org/10.1145/2470654.2470716

[36] Hella von Unger, Martina Block, and Michael T. Wright. 2007. *Aktionsforschung im deutschsprachigen Raum: Zur Geschichte und Aktualität eines kontroversen Ansatzes aus Public Health Sicht.* Technical Report SP I 2007-303. WZB Discussion Paper, Berlin. https://hdl.handle.net/10419/47408

[37] Pontus Wallgren, Maral Babapour, and Siw Eriksson. 2021. HOW TO MAKE ADVANCED ONLINE USER STUDIES MEANINGFUL. *Proceedings of the Design Society* 1 (2021), 1787–1796. https://doi.org/10.1017/pds.2021.440

[38] Andree Woodcock, Louise Moody, Deana McDonagh, Ajita Jain, and Lakhmi C Jain. 2020. *Design of Assistive Technology for Ageing Populations.* Springer International Publishing, Cham, Switzerland. https://doi.org/10.1007/978-3-030-26292-1

[39] World Health Organization (WHO). 2023. Disability & Health Report. Published online: https://www.who.int/news-room/fact-sheets/detail/disability-and-health (last visited on August 20, 2024).

[40] Michael T. Wright, Martina Block, Holger Kilian, and Karl Lemmen. 2013. Förderung von Qualitätsentwicklung durch Partizipative Gesundheitsforschung. *Prävention und Gesundheitsförderung* 8, 3 (2013), 147–154. https://doi.org/10.1007/s11553-013-0396-z